DAGSTUHL REPORTS

Volume 14, Issue 6, June 2024

 ISSN 2192-5283

ISSN 2192-5283

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at
https://www.dagstuhl.de/dagpub/2192-5283

Publication date
December, 2024

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at `https://dnb.d-nb.de`.

License
This work is licensed under a Creative Commons Attribution 4.0 International license (CC BY 4.0).
In brief, this license authorizes each and everybody to share (to copy, distribute and transmit) the work under the following conditions, without impairing or restricting the authors' moral rights:

- Attribution: The work must be attributed to its authors.

The copyright is retained by the corresponding authors.

Digital Object Identifier: 10.4230/DagRep.14.6.i

Aims and Scope
The periodical *Dagstuhl Reports* documents the program and the results of Dagstuhl Seminars and Dagstuhl Perspectives Workshops.
In principal, for each Dagstuhl Seminar or Dagstuhl Perspectives Workshop a report is published that contains the following:

- an executive summary of the seminar program and the fundamental results,
- an overview of the talks given during the seminar (summarized as talk abstracts), and
- summaries from working groups (if applicable).

This basic framework can be extended by suitable contributions that are related to the program of the seminar, e. g. summaries from panel discussions or open problem sessions.

Editorial Board

- Elisabeth André
- Franz Baader
- Goetz Graefe
- Reiner Hähnle
- Barbara Hammer
- Lynda Hardman
- Steve Kremer
- Rupak Majumdar
- Heiko Mantel
- Lennart Martens
- Albrecht Schmidt
- Wolfgang Schröder-Preikschat
- Raimund Seidel (*Editor-in-Chief*)
- Heike Wehrheim
- Verena Wolf
- Martina Zitterbart

Editorial Office
Michael Wagner *(Managing Editor)*
Michael Didas *(Managing Editor)*
Jutka Gasiorowski *(Editorial Assistance)*
Dagmar Glaser *(Editorial Assistance)*
Thomas Schillo *(Technical Assistance)*

Contact
Schloss Dagstuhl – Leibniz-Zentrum für Informatik
Dagstuhl Reports, Editorial Office
Oktavie-Allee, 66687 Wadern, Germany
reports@dagstuhl.de
https://www.dagstuhl.de/dagrep

Report from Dagstuhl Seminar 24231

Stochastic Games

Nathanaël Fijalkow*[1], Jan Kretinsky*[2], Ann Nowé*[3], and Gabriel Bathie†[4]

1 **CNRS – Talence, FR.** nathanael.fijalkow@gmail.com
2 **Masaryk University – Brno, CZ.** jan.kretinsky@gmail.com
3 **Free University of Brussels, BE.** ann.nowe@vub.ac.be
4 **University of Bordeaux, FR.** gabriel.bathie@gmail.com

Abstract

The Dagstuhl Seminar on Stochastic Games brought together leading researchers and practitioners in the field to discuss recent advances, challenges, and future directions. The seminar featured a series of tutorials, invited talks, and contributed talks, which provided a comprehensive overview of the latest developments in Markov decision processes, reinforcement learning, and stochastic game theory. The seminar fostered lively discussions during open problem sessions and working groups, culminating in a collaborative exploration of open questions and potential research directions.

Seminar June 2–7, 2024 – https://www.dagstuhl.de/24231

2012 ACM Subject Classification Theory of computation → Algorithmic game theory and mechanism design; Theory of computation → Convergence and learning in games; Theory of computation → Probabilistic computation; Theory of computation → Representations of games and their complexity

Keywords and phrases Algorithmic Game Theory, Algorithms, Optimisation, Reinforcement Learning, Stochastic Games

Digital Object Identifier 10.4230/DagRep.14.6.1

1 Executive Summary

Nathanaël Fijalkow
Jan Kretinsky
Ann Nowé

License Creative Commons BY 4.0 International license
© Nathanaël Fijalkow, Jan Kretinsky, and Ann Nowé

The Dagstuhl Seminar on Stochastic Games brought together leading researchers and practitioners in the field to discuss recent advances, challenges, and future directions. The seminar featured a series of tutorials, invited talks, and contributed talks, which provided a comprehensive overview of the latest developments in Markov decision processes, reinforcement learning, and stochastic game theory. Key results from the seminar include novel insights into branching stochastic games, the development of new algorithms for solving concurrent and population games, and advancements in the theoretical understanding of efficient solutions for Markov decision processes. The seminar fostered lively discussions during open problem sessions and working groups, culminating in a collaborative exploration of open questions and potential research directions.

* Editor / Organizer
† Editorial Assistant / Collector

Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license
Stochastic Games, *Dagstuhl Reports*, Vol. 14, Issue 6, pp. 1–18
Editors: Nathanaël Fijalkow, Jan Kretinsky, and Ann Nowé
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Overview of the Invited Talks and Tutorials:

- Yinyu Ye: "Progresses and Open Questions on the Markov Decision / Game Process" – Discussed recent advancements and open problems in the field of Markov decision processes and games.
- Dave Parker: "PRISM-games" – Provided a tutorial on PRISM-games, a tool for modeling and analyzing probabilistic systems.
- Kousha Etessami: "Branching MDPs, Branching Stochastic Games, and Generalizations of Newton's Method" – Explored branching Markov decision processes and games, and introduced generalizations of Newton's method for these models.
- Aaron Sidford: "Theoretical Advances in Efficiently Solving Markov Decision Processes" – Highlighted recent theoretical progress in solving Markov decision processes more efficiently.
- Sven Schewe: "Automata for Profit and Pleasure" – Discussed the applications of automata theory in both practical and theoretical contexts.

Overall, the seminar was a highly productive event, advancing the collective understanding of stochastic games and fostering future research collaborations.

2 Table of Contents

3 Overview of Talks

3.1 Solving irreducible stochastic mean-payoff games and entropy games by relative Krasnoselskii-Mann iteration

Marianne Akian (Inria & CMAP, Ecole polytechnique – Palaiseau, FR)

License Creative Commons BY 4.0 International license
© Marianne Akian
Joint work of Marianne Akian, Stéphane Gaubert, Ulysse Naepels, Basile Terver
Main reference Marianne Akian, Stéphane Gaubert, Ulysse Naepels, Basile Terver: "Solving Irreducible Stochastic Mean-Payoff Games and Entropy Games by Relative Krasnoselskii-Mann Iteration", in Proc. of the 48th International Symposium on Mathematical Foundations of Computer Science, MFCS 2023, August 28 to September 1, 2023, Bordeaux, France, LIPIcs, Vol. 272, pp. 10:1–10:15, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023.
URL https://doi.org/10.4230/LIPICS.MFCS.2023.10

We analyse an algorithm solving stochastic mean-payoff games, combining the ideas of relative value iteration and of Krasnoselskii–Mann damping. We derive parameterized complexity bounds for several classes of games including turn-based or concurent games satisfying irreducibility or ergodicity conditions. These bounds improve the ones of Chatterjee and Ibsen-Jensen (ICALP 2014), under the same ergodicity condition, and the one of Allamigeon, Gaubert, Katz and Skomra (ICALP 2022) in the particular case of turn-based games. We also establish parameterized complexity bounds for entropy games, a class of matrix multiplication games introduced by Asarin, Cervelle, Degorre, Dima, Horn and Kozyakin (2016). We derive all these results by methods of variational analysis, establishing contraction properties of the relative Krasnoselskii–Mann iteration with respect to Hilbert's semi-norm.

3.2 Improved bounds for strategy improvement algorithms for energy games

Dani Dorfman (Tel Aviv University, IL)

License Creative Commons BY 4.0 International license
© Dani Dorfman
Joint work of Dani Dorfman, Haim Kaplan, Uri Zwick

Strategy improvement is a natural and well-studied family of algorithms for solving various classes of stochastic and deterministic games. We present an improved upper bound of $O(n2^n)$ on the number of iterations performed by the most natural, and most greedy, variant of the algorithm when applied to n-vertex *Energy Games*. We also obtain a similar upper bound on the expected number of iterations performed by *Random-Edge*, one of the most natural randomized variants of the algorithm.

3.3 Activating Formal Verification of Deep Reinforcement Learning Policies by Model Checking Bisimilar Latent Space Models

Florent Delgrange (Free University of Brussels, BE)

License Creative Commons BY 4.0 International license
© Florent Delgrange
Joint work of Florent Delgrange, Ann Nowé, Guillermo A. Pérez
Main reference Florent Delgrange, Ann Nowé, Guillermo A. Pérez: "Distillation of RL Policies with Formal Guarantees via Variational Abstraction of Markov Decision Processes", in Proc. of the Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 – March 1, 2022, pp. 6497–6505, AAAI Press, 2022.
URL https://doi.org/10.1609/AAAI.V36I6.20602

Intelligent agents are computational entities that autonomously interact with an environment to achieve their design objectives. On the one hand, reinforcement learning (RL) encompasses machine learning techniques that allow agents to learn by trial and error a control policy, prescribing how to behave in the environment. Although RL is proven to converge to an optimal policy under some assumptions, the guarantees vanish with the introduction of advanced techniques, such as deep RL, to deal with high-dimensional state and action spaces. This prevents them from being widely adopted in real-world safety-critical scenarios.

On the other hand, formal methods are mathematical techniques that provide guarantees about the correctness of systems. In particular, model checking allows formally verifying the agent's behaviors in the environment. However, this typically relies on a formal description of the interaction, as well as conducting an exhaustive exploration of the state space. This poses significant challenges because the environment is seldom explicitly accessible. Even when it is, model checking suffers from the curse of dimensionality and struggles to scale to high-dimensional state and action spaces, which are common in deep RL.

We propose to tackle this challenge by leveraging the strengths of deep RL to handle realistic scenarios while integrating formal methods to provide guarantees on the agent's behaviors. Specifically, we enable formal verification of deep RL policies by learning a latent model of the environment, over which we distill the deep RL policy. The outcome is amenable for model checking and is endowed with bisimulation guarantees, which allows to lift the verification results to the original environment.

3.4 Branching MDPs, branching stochastic games, and generlizations of Newton's method

Kousha Etessami (University of Edinburgh, GB)

License Creative Commons BY 4.0 International license
© Kousha Etessami

Over the last 20 years there has been a large body of research in theoretical computer science and verification on algorithms and complexity of analyzing and model checking infinite-state, but finitely presented, Markov chains, Markov decision processes (MDPs), and stochastic games. Many of these models add probabilistic/control/game behavior to some classic automata-theoretic models, like context-free grammars, pushdown automata, one-counter automata, etc. These models, it turns out, are also intimately related to some classic stochastic processes.

In this talk I will give a flavor of one piece of this research. I will focus on a series of results I have been involved with on algorithms and complexity of analyzing Multi-type Branching processes, Branching MDPs, and Branching stochastic games. A key aspect of these results is new algorithms and complexity bounds based on (generalizations of) Newton's method for computing the least fixed point solution for systems of monotone/probabilistic (min/max)-polynomial equations. Such equations arise, e.g., as Bellman optimality equations for Branching MDPs.

(Based on a series of joint works (2005-2020) with Alistair Stewart and Mihalis Yannakakis.)

3.5 Solving tropical polynomial systems using parametric mean-payoff games

Stéphane Gaubert (INRIA & CMAP, Ecole polytechnique – Palaiseau, FR)

License Creative Commons BY 4.0 International license
© Stéphane Gaubert
Joint work of Stéphane Gaubert, Marianne Akian, Antoine Béreau
Main reference Marianne Akian, Antoine Béreau, Stéphane Gaubert: "The Tropical Nullstellensatz and Positivstellensatz for Sparse Polynomial Systems", in Proc. of the 2023 International Symposium on Symbolic and Algebraic Computation, ISSAC 2023, Tromsø, Norway, July 24-27, 2023, pp. 43–52, ACM, 2023.
URL https://doi.org/10.1145/3597066.3597089

Grigoriev and Podolskii (2018) have established a tropical analogue of the effective Nullstellensatz, showing that a system of tropical polynomial equations is solvable if and only if a linearized system obtained from a truncated Macaulay matrix is solvable. They provided an upper bound of the minimal admissible truncation degree, as a function of the degrees of the tropical polynomials. We establish a tropical Nullstellensatz adapted to *sparse* tropical polynomial systems. Our approach is inspired by a construction of Canny-Emiris (1993), refined by Sturmfels (1994). This leads to an improved bound of the truncation degree, which coincides with the classical Macaulay degree in the case of $n+1$ equations in n unknowns. We also establish a tropical Positivstellensatz, allowing one to decide the inclusion of tropical basic semialgebraic sets. This allows one to reduce decision problems for tropical semi-algebraic sets to the solution of systems of tropical linear equalities and inequalities. Finally, we shall discuss the recent development of a tropical analogue of the eigenvalue method for polynomial system solving: we show how to compute solutions of systems of tropical polynomial (in)equalities using parametric mean payoff games.

3.6 Similarities between ARRIVAL and Simple Stochastic Games

Sebastian Haslebacher (ETH Zürich, CH)

License Creative Commons BY 4.0 International license
© Sebastian Haslebacher

The two problems ARRIVAL and Simple Stochastic Games (SSG) share the same complexity status: Both are known to lie in NP and coNP, but not known to lie in P. Besides this, it turns out that recent algorithms for both problems use similar ideas. Concretely, I talked about similarities underlying three results in the area:

- A polynomial-time algorithm for SSG on almost acyclic graphs due to Auger, Coucheney, Strozecki [1],
- the subexponential-time algorithm for ARRIVAL due to Gärtner, Haslebacher, Hoang [3],
- and the quasi-polynomial-time algorithm for SSG on graphs of bounded treewidth due to Chatterjee, Meggendorfer, Saona, Svoboda [2].

The main question is whether the common framework of these three results can be further exploited to give better algorithms for either problem.

References

1 David Auger, Pierre Coucheney, Yann Strozecki. *Finding Optimal Strategies of Almost Acyclic Simple Stochastic Games*. Theory and Applications of Models of Computation, 8402:67–85, 2014.
2 Krishnendu Chatterjee, Tobias Meggendorfer, Raimundo Saona, Jakub Svoboda. *Faster Algorithm for Turn-based Stochastic Games with Bounded Treewidth*. In Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 4590–4605, 2023.
3 Bernd Gärtner, Sebastian Haslebacher, Hung P. Hoang. *A Subexponential Algorithm for ARRIVAL*. In 48th international colloquium on automata, languages, and programming (ICALP), pages 69:1–69:14, 2021.

3.7 Open Problems in Parametric Markov Decision Processes

Sebastian Junges (Radboud University Nijmegen, NL)

License Creative Commons BY 4.0 International license
© Sebastian Junges
Joint work of Sebastian Junges, Joost-Pieter Katoen, Guillermo A. Pérez, Tobias Winkler
Main reference Sebastian Junges, Joost-Pieter Katoen, Guillermo A. Pérez, Tobias Winkler: "The complexity of reachability in parametric Markov decision processes", J. Comput. Syst. Sci., Vol. 119, pp. 183–210, 2021.
URL https://doi.org/10.1016/J.JCSS.2021.02.006

This talk considers parametric Markov decision processes (pMDPs), an extension to Markov decision processes (MDPs) where transitions probabilities are described by polynomials over a finite set of parameters rather than precise values. Parametric MDPs have been studied in the context of verifying robust randomizing algorithms and to study decision making in partially observable environments. We first review results regarding the complexity of finding values for these parameters such that the induced MDP satisfies some maximal or minimal reachability probability constraints and then discuss three intriguing open problems.

3.8 Complexity and Representations of Controllers in Reactive Synthesis

James C. A. Main (F.R.S.-FNRS & UMONS – University of Mons, BE)

License Creative Commons BY 4.0 International license
© James C. A. Main
Joint work of James C. A. Main, Mickael Randour

We consider controller synthesis via game theory. Controllers are derived from strategies of players in games whenever suitable strategies can be computed. Traditionally, strategies are represented by Mealy machines, i.e., automata with outputs along their transitions. This model has been extensively studied and is particularly well-suited for automata-based game solving techniques.

The Mealy machine standard for strategy representation, in spite of its established usefulness, may not be the most relevant in certain contexts. For instance, some strategies require exponential-size Mealy machines despite admitting a small program-like representation based on counters that is thus better suited for controller implementation. In a certain sense, Mealy machines can somewhat obfuscate the structure of strategies and their complexity (e.g., with respect to implementation). These concerns have recently led to a surge of alternative models to represent strategies such as decision trees, neural networks and programs.

We motivate a multifaceted vision of strategy complexity. For instance, strategy representations provide natural measures of complexity (e.g., their size). These measures are not necessarily directly comparable to one another. We are also interested in the applicability and relevance of strategy representations, and their relationships. In addition to model-based notions of complexity, measures that are independent of the chosen representation are also relevant, such as the inclusion of randomness in decision making.

This talk is based on joint work with Mickael Randour.

3.9 Strategy shapes for population games

Corto Mascle (University of Bordeaux, FR)

License Creative Commons BY 4.0 International license
© Corto Mascle
Joint work of Corto Mascle, Hugo Gimbert, Patrick Totzke

Population games are played on a non-deterministic finite automaton. Some number of tokens are placed on an initial state. At each round, one player picks a letter x, and the other answers by moving each token along an x-transition. The first player wins if she can make all tokens reach a final state, for any initial number of tokens.

In this talk we will survey the results obtained so far on this model, with a focus on the model where the second player picks transitions at random. We will sketch the latest algorithm to compute a winning strategy for the first player in that case, obtained with Hugo Gimbert and Patrick Totzke. We will also present and motivate several open problems concerning this model.

3.10 Trustworthy Reinforcement Learning, challenges and opportunities

Ann Nowé (Free University of Brussels, BE)

License Creative Commons BY 4.0 International license
© Ann Nowé
Main reference References are mentioned on the slides

Abstract Reinforcement Learning (RL) has long outgrown the traditional representations that guaranteed policy convergence but severely limited its application to complex domains. Modern Deep RL enables far richer and complex behaviour, yet at the cost of transparency and explainability, as well as convergence guarantees. The same has been observed in Multi-agent RL settings, where in the past it has been shown that independent learning agents, i.e. not having access to the states nor actions of the other agents, could converge to interesting solution concepts, especially. When equipped with an additional protocol, fair solutions could even be obtained. Today the state-of-the art is Centralised learning, decentralized execution, and while it can handle larger state spaces, it gives in on distributivity. After discussing some convergence proofs which hold for tabular settings, some recent developments for policy distillation were discussed and for providing formal guarantees (see also talk by Florent Delgrange).

3.11 PRISM games: Model Checking for Stochastic Games

David Parker (University of Oxford, GB)

License Creative Commons BY 4.0 International license
© David Parker

This tutorial presents an overview of PRISM-games, which is a tool for probabilistic model checking of stochastic games. This technique provides formal modelling and analysis of multi-agent systems with probabilistic behaviour. I will describe ways to model such systems using turn-based or concurrent stochastic games, and to formally specify desired properties of the agents' strategies using probabilistic temporal logic. I will discuss the techniques that PRISM-games implements for model checking these logics, covering both zero-sum and equilibria-based properties, and present some illustrative case studies.

3.12 Synthesizing "more probabilistic" systems

Jakob Piribauer (TU Dresden, DE)

License Creative Commons BY 4.0 International license
© Jakob Piribauer

In order to provide guarantees on the behavior of a system, probabilism and non-determinism pose different challenges. Non-determinism typically necessitates a worst-case analysis, while probabilism allows for more nuanced assurances. This raises the question whether systems in which uncertainty is mainly subject to probabilism are more desirable. This talk presents an idea on how to measure the influence of non-determinism and probabilism, respectively, on the uncertainty of some quantitative aspect, e.g., the runtime, of a system. The talk aims to open a discussion on whether "more probabilistic" systems are indeed desirable and in which situations using such an objective (potentially in conjunction with further objectives) might be useful.

3.13 Automata for Profit and Pleasure

Sven Schewe (University of Liverpool, GB)

License Creative Commons BY 4.0 International license
© Sven Schewe
Joint work of Ernst Moritz Hahn, Mateo Perez, Sven Schewe, Fabio Somenzi, Qiyi Tang, Ashutosh Trivedi, Dominik Wojtczak, Tansholpan Zhanabekova
Main reference Ernst Moritz Hahn, Mateo Perez, Sven Schewe, Fabio Somenzi, Ashutosh Trivedi, Dominik Wojtczak: "Good-for-MDPs Automata for Probabilistic Analysis and Reinforcement Learning", in Proc. of the Tools and Algorithms for the Construction and Analysis of Systems – 26th International Conference, TACAS 2020, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2020, Dublin, Ireland, April 25-30, 2020, Proceedings, Part I, Lecture Notes in Computer Science, Vol. 12078, pp. 306–323, Springer, 2020.
URL https://doi.org/10.1007/978-3-030-45190-5_17
Main reference Sven Schewe, Qiyi Tang, Tansholpan Zhanabekova: "Deciding What Is Good-For-MDPs", in Proc. of the 34th International Conference on Concurrency Theory, CONCUR 2023, September 18-23, 2023, Antwerp, Belgium, LIPIcs, Vol. 279, pp. 35:1–35:16, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023.
URL https://doi.org/10.4230/LIPICS.CONCUR.2023.35

What could be greater fun than toying around with formal structures? One particularly beautiful structure to play with are automata over infinite words, and there is really no need to give any supporting argument for the *pleasure* part in the title. But how about profit? When using ω-regular languages as target languages for practical applications like Markov chain model checking, MDP model checking and reinforcement learning, reactive synthesis, or as a target for an infinite word played out in a two player game, the classic approach has been to first produce a deterministic automaton D that recognises that language. This deterministic automaton is quite useful: we can essentially play on the syntactic product of the structure and use the acceptance mechanism it inherits from the automaton as target. This is beautiful and moves all the heavy lifting to the required automata transformations. But when we want even more profit in addition to the pleasure, the question arises whether deterministic automata are the best we can do. They are clearly good enough: determinism is as restrictive as it gets, and easily guarantees that one can just work on the product. But what we really want is the reverse: we want an automaton, so that we can work on the product, and determinism is just maximally restrictive, and therefore good enough for everything. At Dagstuhl, all will know that we can lift quite a few restrictions and instead turn to the gains we can make when we focus on the real needs of being able to work on the product. For Markov chains, this could be unambiguous automata, for MDPs this could be good-for-MDP automata, and for synthesis and games, this could be good-for-games automata. We will shed a light to a few nooks and corners of the vast room available open questions and answers, with a bias towards MDPs analysis in general and reinforcement learning in particular.

3.14 Theoretical Advances in Efficiently Solving Markov Decision Processes

Aaron Sidford (Stanford University, US)

License Creative Commons BY 4.0 International license
© Aaron Sidford
Joint work of Yujia Jin, Ishani Karmarkar, Aaron Sidford, Jiayi Wang
Main reference Yujia Jin, Ishani Karmarkar, Aaron Sidford, Jiayi Wang: "Truncated Variance Reduced Value Iteration", CoRR, Vol. abs/2405.12952, 2024.
URL https://doi.org/10.48550/ARXIV.2405.12952

Markov Decision Processes (MDPs) are a fundamental mathematical model for reasoning about uncertainty and have a foundational role in the theory of reinforcement learning. Over the past decade, there have been substantial advances in the design and analysis of algorithms for provably computing approximately optimal policies for MDPs in a variety of settings. In this talk, I will survey these advances touching upon optimization tools of potential broader utility. Additionally, I will discuss recent joint work with Ishani Karmarkar, Jiayi Wang, and Yujia Jin on solving MDPS (arXiv:2405.12952).

3.15 Solving concurrent games in PSPACE

Patrick Totzke (University of Liverpool, GB)

License Creative Commons BY 4.0 International license
© Patrick Totzke
Joint work of Sougata Bose, Rasmus Ibsen-Jensen, Patrick Totzke
Main reference Sougata Bose, Rasmus Ibsen-Jensen, Patrick Totzke: "Bounded-Memory Strategies in Partial-Information Games", in Proc. of the 39th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2024, Tallinn, Estonia, July 8-11, 2024, pp. 17:1–17:14, ACM, 2024.
URL https://doi.org/10.1145/3661814.3662096

Very recently, progress has been made on new nondeterministic upper bounds for solving concurrent Mean-Payoff and related games, resulting in approximating algorithms at level 2 of the polynomial hierarchy, FNP[NP], for approximating values and equilibria. In this talk I give a brief overview of the technique, which is based on Frederiksen and Miltersen's (ISAAC 2013) work that uses floating point representations to approximate doubly-exponentially small values. We have extended this to mean-payoff objectives under partial info, and for any fixed number of players and shown that this implies improved upper bounds for various well-known types of games. Directions for discussions:

- Is there hope to get it down to plain FNP?
- Ultimately, I would like to use SAT/SMT solvers to implement these procedures. I would welcome any advice or discussions about implementation details: Which libraries etc to use to deal with (doubly exponentially small) probabilities?
- I would be interested in hearing suggestions about more applications, theoretical and practical.

References

1 Søren Kristoffer Stiil Frederiksen and Peter Bro Miltersen. *Approximating the Value of a Concurrent Reachability Game in the Polynomial Time Hierarchy* In International Symposium on Algorithms and Computation (ISAAC 2013). 457–467.

3.16 Decidability of Omega-Regular Objectives for POMDPs with Revelations

Pierre Vandenhove (University of Bordeaux, FR)

License Creative Commons BY 4.0 International license
© Pierre Vandenhove
Joint work of Pierre Vandenhove, Fijalkow Nathanaël, Hugo Gimbert, Guillermo A. Pérez

We consider partially observable Markov decision processes (POMDPs) with parity objectives. We study the qualitative problem of deciding whether there exists an almost surely winning strategy. Such a problem is undecidable in general, already for coBüchi objectives.

We introduce two decidable properties requiring that, almost surely and infinitely often, the exact state can be deduced (which is called a revelation). Assuming the first property, we show that coBüchi is decidable, while parity with three priorities is still undecidable. Assuming the second property, we show that parity is decidable. Technically, the decidable cases all reduce to the analysis of the finite belief support MDP. We also consider partially observable zero-sum games and show that coBüchi is still undecidable under the revealing properties.

3.17 General-sum stochastic games: Turn-based v.s. simultaneous play

Muthukumar Vidya (Georgia Institute of Technology – Atlanta, US)

License Creative Commons BY 4.0 International license
© Muthukumar Vidya
Joint work of Muthukumar Vidya, Aaron Sidford
Main reference Yujia Jin, Vidya Muthukumar, Aaron Sidford: "The Complexity of Infinite-Horizon General-Sum Stochastic Games", in Proc. of the 14th Innovations in Theoretical Computer Science Conference, ITCS 2023, January 10-13, 2023, MIT, Cambridge, Massachusetts, USA, LIPIcs, Vol. 251, pp. 76:1–76:20, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023.
URL https://doi.org/10.4230/LIPICS.ITCS.2023.76

We study the complexity of computing stationary Nash equilibrium (NE) in n-player infinite-horizon general-sum stochastic games. We focus on the problem of computing NE in such stochastic games when each player is restricted to choosing a stationary policy and rewards are discounted. First, we prove that computing such NE is in PPAD (in addition to clearly being PPAD-hard). Second, we consider turn-based specializations of such games where at each state there is at most a single player that can take actions and show that these (seemingly simpler) games remain PPAD-hard. Third, we show that under further structural assumptions on the rewards computing NE in such turn-based games is possible in polynomial time. Towards achieving these results we establish structural facts about stochastic games of broader utility, including monotonicity of utilities under single-state single-action changes and reductions to settings where each player controls a single state.

3.18 An Overview of Stochastic Games Benchmarks

Maximilian Weininger (IST Austria – Klosterneuburg, AT)

License Creative Commons BY 4.0 International license
© Maximilian Weininger

Evaluating, validating and comparing algorithms for solving stochastic games requires a set of benchmarks. This set should be "realistic", easy-to-use, and structurally diverse. I provide an overview of the state-of-the-art benchmark sets, pointing out their shortcomings and sketching how we can improve them.

My insights on this topic are based on my continued work on algorithms for stochastic games, in particular [1, 2, 3, 4, 5], and the discussions we had about the MDP benchmark set at the previous Dagstuhl Seminar 24134.

References

1. Roman Andriushchenko, Alexander Bork, Carlos E. Budde, Milan Ceska, Kush Grover, Ernst Moritz Hahn, Arnd Hartmanns, Bryant Israelsen, Nils Jansen, Joshua Jeppson, Sebastian Junges, Maximilian A. Köhl, Bettina Könighofer, Jan Kretínský, Tobias Meggendorfer, David Parker, Stefan Pranger, Tim Quatmann, Enno Ruijters, Landon Taylor, Matthias Volk, Maximilian Weininger, and Zhen Zhang. Tools at the frontiers of quantitative verification. *CoRR*, abs/2405.13583, 2024.
2. Muqsit Azeem, Alexandros Evangelidis, Jan Kretínský, Alexander Slivinskiy, and Maximilian Weininger. Optimistic and topological value iteration for simple stochastic games. In *ATVA*, volume 13505 of *Lecture Notes in Computer Science*, pages 285–302. Springer, 2022.
3. Arnd Hartmanns, Sebastian Junges, Tim Quatmann, and Maximilian Weininger. A practitioner's guide to MDP model checking algorithms. In *TACAS (1)*, volume 13993 of *Lecture Notes in Computer Science*, pages 469–488. Springer, 2023.
4. Jan Kretínský, Emanuel Ramneantu, Alexander Slivinskiy, and Maximilian Weininger. Comparison of algorithms for simple stochastic games. *Inf. Comput.*, 289(Part):104885, 2022.
5. Tobias Meggendorfer and Maximilian Weininger. Playing games with your PET: extending the partial exploration tool to stochastic games. *CoRR*, abs/2405.03885, 2024.

4 Working groups

4.1 Auction-Based Scheduling

Guy Avni (University of Haifa, IL)

License Creative Commons BY 4.0 International license
© Guy Avni

Joint work of Guy Avni, Kaushik Mallik, Suman Sadhukhan
Main reference Guy Avni, Kaushik Mallik, Suman Sadhukhan: "Auction-Based Scheduling", in Proc. of the Tools and Algorithms for the Construction and Analysis of Systems, pp. 153–172, Springer Nature Switzerland, 2024.
URL https://doi.org/10.1007/978-3-031-57256-2_8

We discussed the details of the auction-based scheduling framework and its required background from bidding games. We discussed possible relations with history deterministic automata. We discussed extensions of the framework in other domains, specifically in MDPs, which we agreed might be an interesting future direction for study.

4.2 Solving Markov Decision Processes by adding a discount

Nathanaël Fijalkow (CNRS – Talence, FR)

License Creative Commons BY 4.0 International license
© Nathanaël Fijalkow

In this working group we discussed how we could use the solutions for solving discounted Markov Decision Processes to the undiscounted case. This was an excuse to revisit the existing methods for the discounted case. Some interesting directions were discussed.

4.3 Finding Tarski Fixed Points

Sebastian Haslebacher (ETH Zürich, CH)

License Creative Commons BY 4.0 International license
© Sebastian Haslebacher

The complexity of solving Condon's and (approximately solving) Shapley's stochastic games are longstanding open questions. Etessami et al. [1] proved that both problems reduce to the problem of finding a fixed point of a monotone function on an integer grid. Concretely, such a function maps $\{1, \ldots, N\}^d$ into itself and is monotone with respect to the coordinate-wise partial order $\leq_c$. A well-known theorem of Tarski guarantees that these conditions guarantee the existence of a fixed point. Hence, the computational problem of finding it is often called TARSKI.

The best algorithms to date solve TARSKI using at most $O(\log^{\lceil \frac{2d}{3} \rceil} N)$ queries [2], or $O(\log^{\lceil \frac{d+1}{2} \rceil} N)$ queries [3] if d is considered to be a constant, respectively. All of these algorithms are time efficient, i.e. the next query can be determined in time polynomial in d and $\log N$. Conversely, the best query lower bound is $\Omega(\log^2(N)$ [1]. This means that there is a big gap in terms of query complexity, and closing this gap would be a major achievement.

Recently, Chen et al. [4] gave a query-efficient algorithm for the problem of finding an approximate fixed point of a contraction map (with respect to the infinity norm). Unfortunately, their algorithm is not time-efficient (if it were, this would have tremendous implications e.g. for stochastic games). But naturally, the question arises whether similar techniques could be used to obtain better query upper bounds for Tarski.

The discussion in the working session revolved around these recent results. Concretely, the above mentioned literature was discussed, and we tried to understand whether the techniques from [4] might indeed be useful for Tarski as well.

References

1 Kousha Etessami, Christos Papadimitriou, Aviad Rubinstein, and Mihalis Yannakakis. *Tarski's Theorem, Supermodular Games, and the Complexity of Equilibria*. In 11th Innovations in Theoretical Computer Science Conference (ITCS), pages 18:1–18:19, 2020.

2 John Fearnley, Dömötör Pálvölgyi, and Rahul Savani. *A Faster Algorithm for Finding Tarski Fixed Points*. ACM Transactions on Algorithms, 18(3):1–23, July 2022.

3 Xi Chen and Yuhao Li. *Improved Upper Bounds for Finding Tarski Fixed Points*. In Proceedings of the 23rd ACM Conference on Economics and Computation, EC '22, pages 1108–1118, July 2022.

4 Xi Chen, Yuhao Li, and Mihalis Yannakakis. *Computing a Fixed Point of Contraction Maps in Polynomial Queries*. In Proceedings of the 56th Annual ACM Symposium on Theory of Computing, STOC 2024, pages 1364–1373, June 2024.

4.4 Discussion on Parametric Markov Chains and Partially Observable MDPs

Sebastian Junges (Radboud University Nijmegen, NL)

License Creative Commons BY 4.0 International license
© Sebastian Junges
Main reference Sebastian Junges, Nils Jansen, Ralf Wimmer, Tim Quatmann, Leonore Winterer, Joost-Pieter Katoen, Bernd Becker: "Finite-State Controllers of POMDPs using Parameter Synthesis", in Proc. of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence, UAI 2018, Monterey, California, USA, August 6-10, 2018, pp. 519–529, AUAI Press, 2018.
URL https://auai.org/uai2018/proceedings/papers/195.pdf

This blackboard session studied the connection between decision making in Partially Observable Markov Decision Processes (POMDPs) and parametric Markov chains (pMCs), which were discussed at length in an earlier talk. After reviewing results from the main reference given below a discussion revolved mostly about the possibilities to extend these results in different directions, e.g., by partitioning the set of parameters in controllable and uncontrollable parameters, by connecting Partially Observable Stochastic Games (or Concurrent Stochastic Games) with pMDPs. This lead to a discussion of the important features which make parameter synthesis in MDPs hard.

4.5 Measures of benchmarks difficulty and strategy structure

Jan Kretinsky (Masaryk University – Brno, CZ)

License Creative Commons BY 4.0 International license
© Jan Kretinsky
Main reference Jan Kretínský, Emanuel Ramneantu, Alexander Slivinskiy, Maximilian Weininger: "Comparison of algorithms for simple stochastic games", Inf. Comput., Vol. 289(Part), p. 104885, 2022.
URL https://doi.org/10.1016/J.IC.2022.104885

This break-out session was concerned with two questions. Firstly, how to measure the difficulty and structure of benchmarks so that we can pinpoint what makes real ones easy (or hard) and what to focus one in order to design practical algorithms [1]. Second, how to describe the fine structure of strategies and utilize it in order to represent them (i) explainably and (ii) so that they are easier to compute [2].

References

1 Jan Kretinsky, Emanuel Ramneantu, Alexander Slivinskiy, Maximilian Weininger. Comparison of Algorithms for Simple Stochastic Games. In GandALF, pp. 131-148. EPTCS, 2020.
2 Pranav Ashok, Mathias Jackermeier, Jan Kretinsky, Christoph Weinhuber, Maximilian Weininger, Mayank Yadav. dtControl 2.0: Explainable Strategy Representation via Decision Tree Learning Steered by Experts. In TACAS (2), pp. 326-345. Springer, 2021.

4.6 Speed of victory in population games

Corto Mascle (University of Bordeaux, FR)

License Creative Commons BY 4.0 International license
© Corto Mascle
Joint work of Corto Mascle, Nathanaël Fijalkow

In 2017, Bertrand, Dewaskar, Genest and Gimbert introduced a population games, inspired by questions from biology on the control of yeast populations [1]. A population game is described by an NFA with one initial and one final state. Some number of tokens are placed on an initial state of an NFA. Two players then play alternately: Controller chooses a letter a, then Environment moves each token along an a-labelled transitions chosen at random. Controller wins if all tokens end up on the same final state eventually. The question is then: does Controller have a winning strategy against any number of tokens?

A randomized version of those games was later presented by Colcombet, Fijalkow and Ohlmann [2]: this is the model studied in this working group. In this framework, the Environment picks the next transition of each token uniformly at random. The new question is: Can Controller win with probability 1 against any number of tokens? That first paper established decidablity of this problem. However, the expected time needed by Controller to win may in general be exponential in the number of tokens.

This working group was dedicated to the search for decidable characterizations of games where Controller can win in polynomial and polylogarithmic time. We made progress in several directions, Most notably, we built a clear plan to characterize the polylogarithmic case, leaving reducing the initial problem to a couple of lemmas that we believe we can prove with a little more work.

References

1 Nathalie Bertrand and Miheer Dewaskar and Blaise Genest and Hugo Gimbert. *Controlling a population*. CONCUR 2017., https://doi.org/10.4230/LIPIcs.CONCUR.2017.12, 2017

2 Thomas Colcombet and Nathanaël Fijalkow and Pierre Ohlmann. *Controlling a random population*. Log. Methods Comput. Sci., https://doi.org/10.46298/lmcs-17(4:12)2021, 2021

Participants

- Marianne Akian
Inria & CMAP, Ecole polytechnique – Palaiseau, FR
- Xavier Allamigeon
INRIA & Ecole Polytechnique – Palaiseau, FR
- Guy Avni
University of Haifa, IL
- Christel Baier
TU Dresden, DE
- Gabriel Bathie
University of Bordeaux, FR
- Eline Bovy
Radboud University Nijmegen, NL
- Antonio Casares
University of Bordeaux, FR
- Florent Delgrange
Free University of Brussels, BE
- Dani Dorfman
Tel Aviv University, IL
- Kousha Etessami
University of Edinburgh, GB
- Nathanaël Fijalkow
CNRS – Talence, FR
- Stéphane Gaubert
INRIA & CMAP, Ecole polytechnique – Palaiseau, FR
- Kristoffer Arnsfelt Hansen
Aarhus University, DK
- Sebastian Haslebacher
ETH Zürich, CH
- Sebastian Junges
Radboud University Nijmegen, NL
- Marcin Jurdzinski
University of Warwick – Coventry, GB
- Jan Kretinsky
Masaryk University – Brno, CZ
- Antonin Kucera
Masaryk University – Brno, CZ
- Karoliina Lehtinen
Aix-Marseille University, FR
- Georg Loho
FU Berlin, DE & University of Twente – Enschede, NL
- James C. A. Main
F.R.S.-FNRS & UMONS – University of Mons, BE
- Corto Mascle
University of Bordeaux, FR
- Théo Matricon
University of Bordeaux, FR
- Petr Novotný
Masaryk University – Brno, CZ
- Ann Nowé
Free University of Brussels, BE
- David Parker
University of Oxford, GB
- Guillermo A. Pérez
University of Antwerp, BE
- Jakob Piribauer
TU Dresden, DE
- Roxana Radulescu
VU – Brussels, BE
- Sven Schewe
University of Liverpool, GB
- Guruprerana Shabadi
IMDEA Software Institute – Madrid, ES
- Aaron Sidford
Stanford University, US
- Mateusz Skomra
LAAS – Toulouse, FR
- Yann Strozecki
University of Versailles, FR
- Marnix Suilen
Radboud University Nijmegen, NL
- K. S. Thejaswini
IST Austria – Klosterneuburg, AT
- Patrick Totzke
University of Liverpool, GB
- Pierre Vandenhove
University of Bordeaux, FR
- Muthukumar Vidya
Georgia Institute of Technology – Atlanta, US
- Maximilian Weininger
IST Austria – Klosterneuburg, AT
- Yinyu Ye
Stanford University, US & Shanghai JiaoTong University, CN

Report from Dagstuhl Seminar 24232

Designing Computers' Control Over Our Bodies

Nadia Bianchi-Berthouze*[1], Mar Gonzalez-Franco*[2], Florian 'Floyd' Mueller*[3], Misha Sra*[4], and Rakesh Patibanda†[5]

1 **University College London, GB.** n.berthouze@ucl.ac.uk
2 **Google – Seattle, US.** margonzalezfranco@gmail.com
3 **Exertion Games Lab, Human-Centred Computing, Monash University – Melbourne, AU.** floyd@exertiongameslab.org
4 **University of California – Santa Barbara, US.** sra@cs.ucsb.edu
5 **Exertion Games Lab, Human-Centred Computing, Monash University – Melbourne, AU.** rakesh@exertiongameslab.org

Abstract

The classical human-computer interaction (HCI) model characterised interactions as predominantly user-directed, with the computer acting as a responder to human commands. This paradigm had been foundational, yet contemporary research pivoted towards more reciprocal roles, where the machine not only responded but also asserted control. This emerging domain, characterised by technologies such as electrical muscle stimulation, galvanic vestibular stimulation, and exoskeletons, introduced a new dynamic – computational control over the human body. Such technologies offered significant benefits, like enhanced safety in autonomous vehicles and increased mobility through autonomous exoskeletons. However, these advancements also ushered in ethical, psychological, and physical concerns, paralleling earlier fears associated with technologies that aimed to control human psychology.

The absence of a structured theoretical framework to articulate and evaluate the experiences of being controlled by a machine was evident, as was a comprehensive understanding of how to design such interactions responsibly. This Dagstuhl Seminar sought to bridge these gaps by convening experts from academia and industry. The seminar explored multidisciplinary approaches to the design challenges and societal implications of computational control over the human body. Through collaborative discourse, the event aimed to foster a deeper understanding of this complex interaction paradigm, contributing towards a more humane technological future by integrating diverse insights and expertise.

Seminar June 2–7, 2024 – https://www.dagstuhl.de/24232

2012 ACM Subject Classification Human-centered computing → Human computer interaction (HCI); Human-centered computing → Interaction paradigms

Keywords and phrases Body, Control, Human-Computer Integration, Human-Computer Interaction, HCI, Computational Control, Electrical Muscle Stimulation, Galvanic Vestibular Stimulation, Exoskeletons, Ethics, Design Framework

Digital Object Identifier 10.4230/DagRep.14.6.19

* Editor / Organizer
† Editorial Assistant / Collector

Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Designing Computers' Control Over Our Bodies, *Dagstuhl Reports*, Vol. 14, Issue 6, pp. 19–51
Editors: Nadia Bianchi-Berthouze, Mar Gonzalez-Franco, Florian 'Floyd' Mueller, Misha Sra, and Rakesh Patibanda
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Nadia Bianchi-Berthouze
Mar Gonzalez-Franco
Florian 'Floyd' Mueller
Misha Sra

License Creative Commons BY 4.0 International license
© Nadia Bianchi-Berthouze, Mar Gonzalez-Franco, Florian 'Floyd' Mueller, and Misha Sra

The Dagstuhl Seminar 24232, titled *"Designing Computers' Control Over Our Bodies",* was held from June 2 to June 7, 2024, at Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany. This seminar brought together leading experts from diverse fields to explore a new paradigm in human-computer interaction (HCI) where technologies such as electrical muscle stimulation, galvanic vestibular stimulation, and exoskeletons enable computers to exert direct control over the human body [1, 2, 3, 4, 5]. The seminar addressed both the opportunities and challenges of this emerging domain, which extends beyond the traditional model of user-directed interaction to consider reciprocal roles where machines also assert control [3, 6, 7].

Over the course of five days, participants engaged in a variety of activities designed to foster interdisciplinary collaboration and develop a comprehensive understanding of this complex interaction paradigm. The seminar's objectives included the following:

Identifying Theoretical Gaps. Participants explored the absence of a structured theoretical framework for articulating and evaluating experiences of being controlled by a machine. Discussions focused on such interactions' ethical, psychological, and physical dimensions, highlighting the need for a multidisciplinary approach to understand and design for these new dynamics [8, 9, 10].

Design Challenges and Opportunities. The seminar investigated the design challenges associated with computational control over the human body, considering both the benefits, such as enhanced safety in autonomous systems and increased mobility, and the concerns, including issues of user agency, consent, and trust [11]. Hands-on sessions enabled participants to prototype and critique design concepts that address these complexities.

Societal Implications. The discussions extended to the societal impact of these technologies, examining potential applications and unintended consequences. Ethical considerations, particularly around the autonomy of users and the transparency of machine actions, were central to these debates [12]. The seminar aimed to outline guidelines for the responsible development and deployment of such technologies in various contexts, from healthcare to entertainment.

Envisioning Future Directions. A key focus of the seminar was envisioning the future landscape of HCI where computational control technologies become more prevalent. Speculative design sessions invited participants to imagine both utopian and dystopian scenarios for the year 2050, facilitating discussions on the potential trajectories of these technologies with- and without-AI, and their integration into daily life.

Developing a Research Agenda. The seminar concluded with the development of a research agenda to guide future work in this field. This agenda included key questions to address, potential methodologies for study, and proposals for new theoretical frameworks. Working groups were established to continue collaborative efforts beyond the seminar.

The seminar's key outcomes included identifying several grand challenges in this emerging field, formulating initial design frameworks, and establishing a network of researchers and practitioners committed to advancing knowledge and practice in computational control technologies. The seminar highlighted the importance of interdisciplinary approaches to understanding and designing for these new interaction paradigms, which blend the digital and physical realms where the computer can control the human body.

The seminar's findings will contribute to the ongoing discourse on the future of HCI, particularly in areas where control dynamics between humans and machines are increasingly blurred. Moving forward, the insights gained from this event will help shape a more humane technological future, ensuring that innovations in computational control are aligned with ethical and societal values.

References

1 Rakesh Patibanda, Chris Hill, Aryan Saini, Xiang Li, Yuzheng Chen, Andrii Matviienko, Jarrod Knibbe, Elise Van Den Hoven, and Florian 'Floyd' Mueller. Auto-paizo games: Towards understanding the design of games that aim to unify a player's physical body and the virtual world. *Proceedings of the ACM on Human-Computer Interaction*, 7(CHI PLAY):893–918, 2023.

2 Pedro Lopes, Sijing You, Lung-Pan Cheng, Sebastian Marwecki, and Patrick Baudisch. Providing haptics to walls & heavy objects in virtual reality by means of electrical muscle stimulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1471–1482, 2017.

3 Florian Floyd Mueller, Pedro Lopes, Paul Strohmeier, Wendy Ju, Caitlyn Seim, Martin Weigel, Suranga Nanayakkara, Marianna Obrist, Zhuying Li, Joseph Delfa, Jun Nishida, Elizabeth M. Gerber, Dag Svanaes, Jonathan Grudin, Stefan Greuter, Kai Kunze, Thomas Erickson, Steven Greenspan, Masahiko Inami, Joe Marshall, Harald Reiterer, Katrin Wolf, Jochen Meyer, Thecla Schiphorst, Dakuo Wang, and Pattie Maes. Next steps for human-computer integration. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–15, New York, NY, USA, 2020. Association for Computing Machinery.

4 Romain Nith, Yun Ho, and Pedro Lopes. Splitbody: Reducing mental workload while multitasking via muscle stimulation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.

5 Rakesh Patibanda, Elise Van Den Hoven, and Florian 'Floyd' Mueller. Towards understanding the design of body-actuated play. In *Extended Abstracts of the 2022 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '22, page 388–391, New York, NY, USA, November 2022. Association for Computing Machinery.

6 Eun-Soo Jung, Suh-Yeon Dong, and Soo-Young Lee. Neural correlates of variations in human trust in human-like machines during non-reciprocal interactions. *Scientific Reports*, 9, 2019.

7 Steve Benford, Richard Ramchurn, Joe Marshall, Max L. Wilson, Matthew Pike, Sarah Martindale, Adrian Hazzard, Chris Greenhalgh, Maria Kallionpää, Paul Tennent, and Brendan Walker. Contesting control: journeys through surrender, self-awareness and looseness of control in embodied interaction. *Human–Computer Interaction*, 36(5-6):361–389, 2021.

8 Pedro Lopes, Lewis L. Chuang, and Pattie Maes. Physiological i/o. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21, New York, NY, USA, 2021. Association for Computing Machinery.

9 Rakesh Patibanda, Aryan Saini, Nathalie Overdevest, Maria F. Montoya, Xiang Li, Yuzheng Chen, Shreyas Nisal, Josh Andres, Jarrod Knibbe, Elise van den Hoven, and Florian 'Floyd' Mueller. Fused spectatorship: Designing bodily experiences where spectators become players. *Proc. ACM Hum.-Comput. Interact.*, 7(CHI PLAY), oct 2023.

10 Rakesh Patibanda, Nathalie Overdevest, Shreyas Nisal, Aryan Saini, Don Samitha Elvitigala, Jarrod Knibbe, Elise Van Den Hoven, and Florian 'Floyd' Mueller. Shared bodily fusion: Leveraging inter-body electrical muscle stimulation for social play. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, DIS '24, page 2088–2106, New York, NY, USA, July 2024. Association for Computing Machinery.

11 Florian 'Floyd' Mueller, Nathan Semertzidis, Josh Andres, Joe Marshall, Steve Benford, Xiang Li, Louise Matjeka, and Yash Mehta. Toward understanding the design of intertwined human–computer integrations. *ACM Trans. Comput.-Hum. Interact.*, 30(5), sep 2023.

12 Saul Greenberg, Sebastian Boring, Jo Vermeulen, and Jakub Dostal. Dark patterns in proxemic interactions: a critical perspective. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, DIS '14, page 523–532, New York, NY, USA, 2014. Association for Computing Machinery.

2 Table of Contents

3 Introduction

The Dagstuhl Seminar, "Designing Computers' Control Over Our Bodies", gathered leading experts from academia and industry to explore the emerging paradigm of human-computer interaction (HCI) characterised by computational control over the human body. The participant pool included 6 PhD candidates, 1 postdoctoral researcher, 3 industry professionals, 4 assistant professors, 2 associate professors, and 8 full professors (Figure 1). This seminar marked a departure from the classical HCI model, which traditionally positioned the human as the sole director and the computer as a passive responder. Instead, the discussions focused on a more dynamic, reciprocal interaction model where technologies such as electrical muscle stimulation, galvanic vestibular stimulation, and exoskeletons allow computers to assert direct control over human actions and movements.

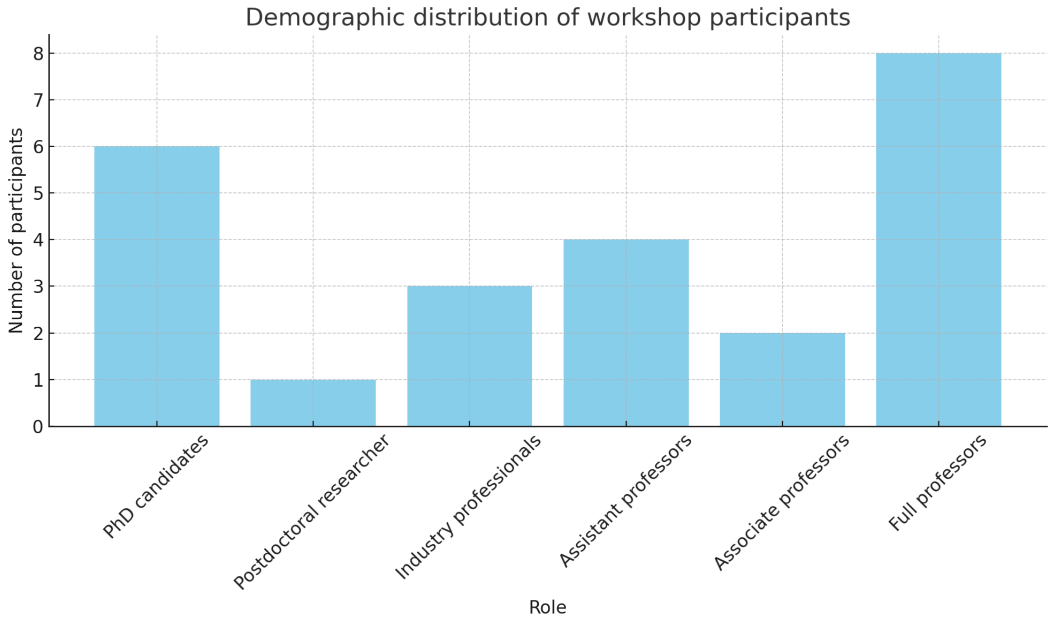


Figure 1 Demographics of participants.

The seminar was structured over five intensive days, each with a distinct thematic focus, ranging from theoretical foundations to practical design implications and future visions. The participants engaged in a mix of presentations, interactive sessions, group discussions, and collaborative design activities to foster a comprehensive understanding of the new dynamics in HCI.

3.1 Day 1: Theory Day

The seminar began with an introductory session led by the organisers, where participants were welcomed and introduced to the seminar's goals and structure. The day continued with a series of "PechaKucha" presentations, where each participant had five minutes to present their work and perspectives on computational control technologies. This fast-paced format helped set the stage for a deeper exploration of the grand challenges and opportunities that these emerging technologies present. The day ended with a group activity where participants identified and categorised key challenges, which were then organised into broader themes of grand challenges to be addressed throughout the seminar.

3.2 Day 2: Design Day

Building on the theoretical discussions from Day 1, Day 2 was dedicated to exploring the design challenges associated with computational control over the human body. Participants were invited to bring prototypes or share design concepts during hands-on sessions. These sessions facilitated a deeper understanding of how to create responsible and ethical design frameworks that accommodate the complexities of machine control over human bodies. The day also included further PechaKucha presentations, followed by group activities to synthesise insights from both the theoretical and practical design perspectives.

3.3 Day 3: Theory Meets Design

Day 3 bridged the gap between theory and practice, focusing on the integration of theoretical frameworks with practical design approaches. Participants worked in breakout groups to align their individual seminar goals with the continued process of identifying grand challenges. This day featured extensive discussions on how to merge theoretical insights with design principles to create holistic approaches to computational control technologies. The morning session included collaborative group work, and the afternoon focused on presenting and critiquing the developed ideas.

3.4 Day 4: Day of Envisioning

On Day 4, the seminar shifted towards envisioning future directions for research and development in this field. Participants engaged in speculative design sessions to imagine scenarios where computational control technologies (controlled by AI or otherwise) are fully integrated into daily life by 2050. These speculative exercises aimed to identify utopian and dystopian possibilities, encouraging participants to consider such technologies' ethical, social, and psychological implications. The day concluded with a plenary session where groups reported on their envisioned futures and debated their desirability and feasibility.

3.5 Day 5: What's Next?

The final day was dedicated to synthesising the seminar's outcomes and identifying tangible next steps, particularly directed at writing an article on the grand challenges that surfaced in this area throughout this seminar. Participants worked on drafting concrete outputs, including research papers, frameworks, and guidelines that could serve as foundational documents for this emerging field. The day also included reflective sessions where participants shared their experiences, insights, and ideas for future collaboration. The seminar ended with a discussion on the formation of working groups to continue the momentum generated during the week.

This structured and intensive programme allowed participants to explore the complexities of computational control technologies from multiple perspectives, paving the way for future research and collaboration in designing humane and ethical interactions where computers assert control over the human body.

4 PechaKutchas

The seminar featured a series of dynamic PechaKucha sessions, a presentation format where each speaker had five minutes to present their slides and three minutes for questions from the audience. This format allowed for fast-paced and engaging presentations that succinctly conveyed key ideas. These sessions provided a platform for participants to share their latest research, insights, and reflections on the seminar's central theme: designing computers' control over the human body. They played a crucial role in setting the stage for broader discussions and stimulating dialogue around foundational concepts and emerging ideas.

Participants covered a diverse range of topics that spanned the theoretical, practical, and experiential dimensions of computational control technologies. Some presentations focused on novel technological approaches, such as "TickleFoot" [4], which explored how unconventional stimuli could create playful and positive bodily experiences under computer control. This work highlighted the potential of playful interactions to ease into the complex topic of shared control between humans and machines.

Other sessions explored the more conceptual and philosophical aspects of control. For example, the presentation on "If control is here, how is it?" [20] investigated the nuanced and often ambiguous nature of control in systems where humans and machines function symbiotically [2, 26]. Meanwhile, "Losing Control" examined what it means to relinquish control to a machine, not just from a functional perspective but as a transformative playful experience that challenges existing notions of autonomy and agency in HCI and game design [12, 2].

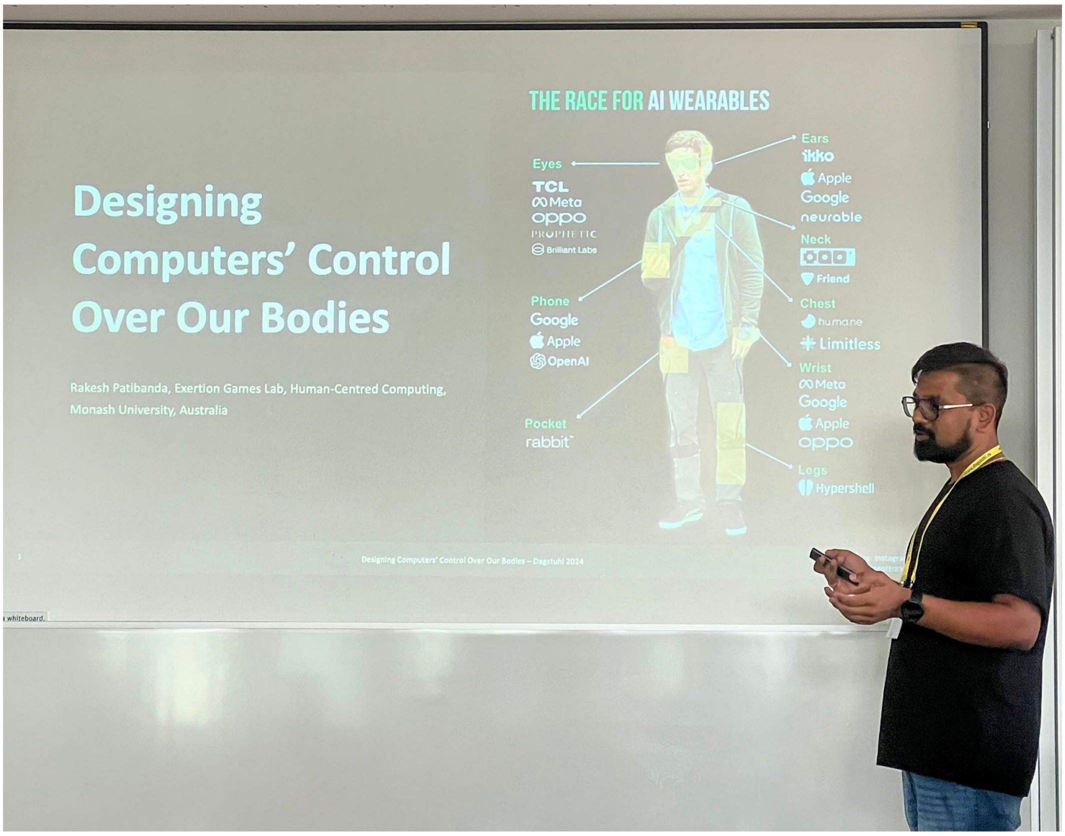


Figure 2 Rakesh Patibanda, participant, presenting his research.

Presentations like "Charting the Landscape of Body Control Techniques" provided an overview of current research in body control technologies, mapping out various methods, from closed-loop to open-loop systems, and identifying areas for future exploration [1]. Talks such as "Control in HCI" and "Closed-loop and Open-loop Control" offered insights into how different control mechanisms impact user experiences and highlighted these approaches' ethical and practical implications [8, 7, 38, 5].

The theme of playfulness was a recurrent topic, as seen in the presentation "Playing by Sharing Bodily Control with Computers." This talk explored how shared control dynamics could create playful and engaging interactions, emphasising the potential benefits of designing experiences where computers and users co-navigate control [28] (Figure 2). By treating the body as both a controller and a site of play, this approach illustrated how playful designs could create novel forms of engagement and enjoyment in contrast to traditional HCI paradigms [21, 26, 24, 23, 22, 27].

Figure 3 Florian 'Floyd' Mueller, organiser, introducing the seminar theme "Designing Computers' Control Over Our Bodies".

Similarly, sessions like "Sharing Bodily Control with Computers" and "Designing Computers' Control over the Body that centres around Bodily Play" (Figure 3) emphasised the value of integrating play into design frameworks. These presentations argued that allowing users to share control with computers opens up new possibilities for creativity, spontaneity, and user agency, ultimately enriching the user experience [16, 18, 15, 13].

A number of talks, such as "Negotiating Augmented Bodies" [29, 30] and "Designing Bodily Extensions," [32, 31] examined the interface between humans and technology, focusing on how digital and physical extensions of the body can be designed to respect bodily autonomy while allowing for innovative forms of interaction. These discussions were complemented by presentations like "Haptic Experiences and Bodily Control," [33] which explored the sensory aspects of control technologies, demonstrating how haptic feedback can be used to create more immersive and nuanced control experiences.

These technologies' social and embodied dimensions were also a focal point, with sessions like "Social and Embodied Perspectives" [6, 14] exploring how body control technologies intersect with social norms and behaviours. These presentations emphasised the importance of designing technically feasible, socially and ethically responsible technologies, fostering acceptance and trust among users.

Some participants were happy to share their presentations for everyone to view, while others chose to keep their content exclusive to the seminar, reflecting the sensitive and pioneering nature of some of the presented research.

Overall, the PechaKucha sessions set the tone for the seminar by providing a plethora of perspectives and ideas, ranging from the playful and provocative to the deeply theoretical and critical. They accelerated discussions that would continue throughout the seminar, laying a foundation for developing new frameworks, methodologies, and collaborative efforts in the field of designing computers' control over the human body.

4.1 Emergent themes of discussion from Day 1

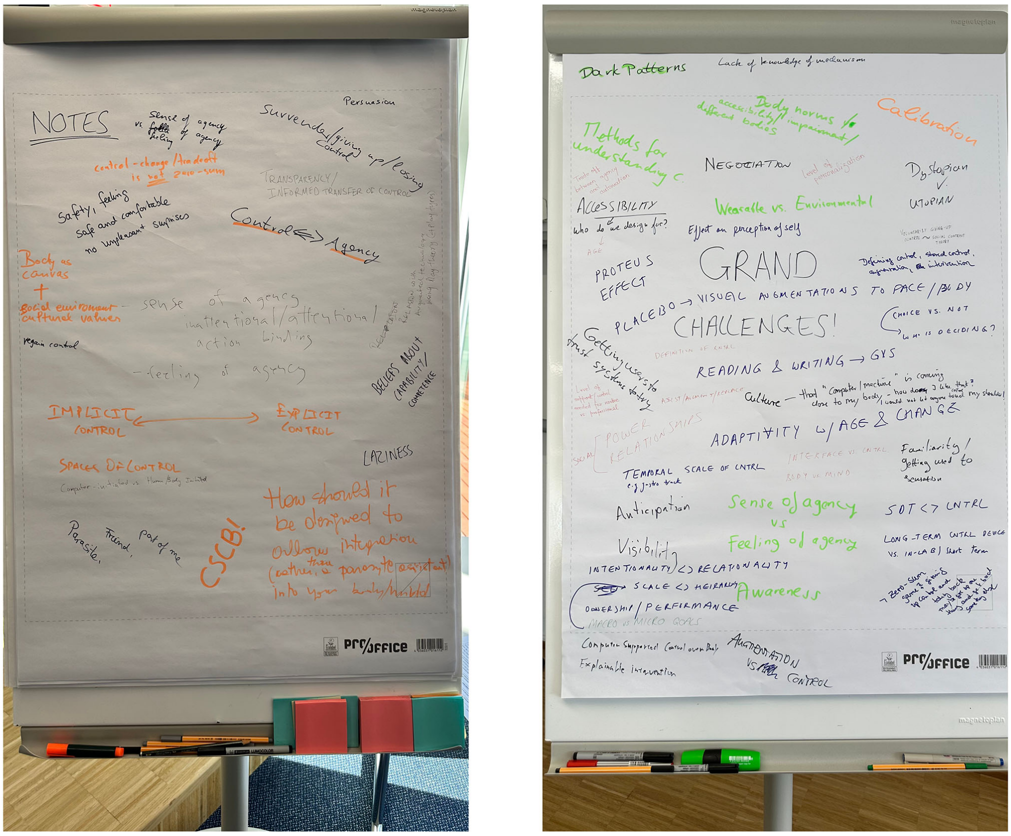


Figure 4 Results from Day 1: Notes on the grand challenges participants took during and after the presentation sessions.

From Day 1, several key themes emerged as participants discussed the grand challenges related to computational control over the human body (Figure 4). These themes were documented in the notes taken during and after the presentation sessions:

1. **Control and agency:**
   - A significant theme centred around the relationship between control and agency. Participants debated how computational systems could either enhance or diminish a user's sense of agency. Discussions included both *implicit control* (where control is seamlessly integrated and less noticeable to the user) and *explicit control* (where the user is fully aware of and potentially directing the control dynamics).
   - Questions emerged regarding how control should be designed to balance these dynamics, ensuring transparency and maintaining a user's feeling of autonomy.
2. **Implicit vs. explicit control:**
   - The balance between implicit and explicit control was identified as a critical design consideration. Participants explored how different forms of control might affect the user's experience, comfort, and acceptance of such technologies. For example, while explicit control allows for more transparency and user awareness, implicit control might lead to smoother, less intrusive experiences but could also raise concerns about trust and manipulation.
3. **Designing for sensory and perceptual experiences:**
   - Discussions highlighted the importance of designing technologies that consider sensory and perceptual aspects, such as visual augmentations to the body or face, haptic feedback, and other multisensory experiences. Themes like the *Proteus Effect* (how a user's self-perception changes through interaction with technology) and the placebo effect were discussed, pointing to the need for careful consideration of how sensory augmentations might alter user experiences and perceptions.
4. **Adaptivity and personalisation:**
   - Another theme focused on the need for adaptive systems that can adjust to individual user differences, such as age, physical ability, and cognitive load. Participants raised questions about how these systems could dynamically modify their behaviours or outputs to better suit the user's evolving needs, thereby enhancing usability and user satisfaction.
5. **Negotiation of control:**
   - The notion of negotiating control between the human and the machine was a recurring topic. This included discussions on *collaboration vs. automation* and how users and systems can share or shift control in various contexts. The concept of "negotiation" in this space addresses how to design interactions that are flexible, adaptable, and responsive to changes in user intent or environmental conditions.
6. **Ethical considerations and dark patterns:**
   - Ethical issues such as transparency, informed consent, and avoiding manipulative "dark patterns" were heavily discussed. The theme of *responsible design* emerged, focusing on creating technologies that respect user autonomy and ensure ethical standards are upheld. This theme underscored the importance of designing systems that can communicate their intentions clearly to users and provide meaningful choices regarding control dynamics.

7. **Social and embodied perspectives:**
   - Participants explored the social implications of these technologies, examining how control over the body might affect social norms, behaviours, and interactions. Themes related to the social environment, the effect of such technologies on shared experiences, and how they could either enhance or disrupt existing social practices were identified.
8. **Awareness and performance:**
   - The role of awareness, both in terms of a user's self-awareness and their understanding of the system's behaviour, was another key theme. Discussions also touched on performance metrics, questioning how these technologies could optimise human performance without compromising ethical standards or user comfort.
9. **Future design directions:**
   - Participants speculated on future directions for design, asking how to create systems that balance autonomy, adaptability, and user comfort. They considered what frameworks could guide the ethical development of technologies that exert control over the human body and how to align these technologies with human values and needs.

These themes reflect a broad and deep engagement with the challenges and opportunities presented by emerging control technologies, highlighting the complexities involved in designing systems that can control the human body.

5 Group Activities on Day 2 and Day 3

During Days 2 and 3 of the seminar, participants engaged in several group activities designed to deepen their understanding of computational control over the human body. The activities were structured around five distinct groups, each focused on a specific aspect of the seminar's theme. These groups collaborated to explore definitions, perspectives, challenges, and implications associated with control, synthesising ideas that emerged from the PechaKucha sessions and other discussions.

1. **Defining control:**
   - One group concentrated on developing a comprehensive definition of "control" within the context of human-computer interaction. They debated the nuances of control, such as its various forms (implicit versus explicit), its dynamic nature, and how control is negotiated between humans and machines (Figure 5). Their discussions aimed to articulate a shared understanding that could serve as a foundation for further research and design frameworks.
2. **Exploring perspectives on control:**
   - Another group sought to map the diverse perspectives of control, scoping the topic across different dimensions, including psychological, social, physiological, and environmental contexts. This group discussed how control can be experienced differently by individuals and how these differences could be accounted for in designing systems that assert control over the human body. Their goal was to establish a multidimensional framework that reflects the complex, layered nature of control in human-computer interactions.

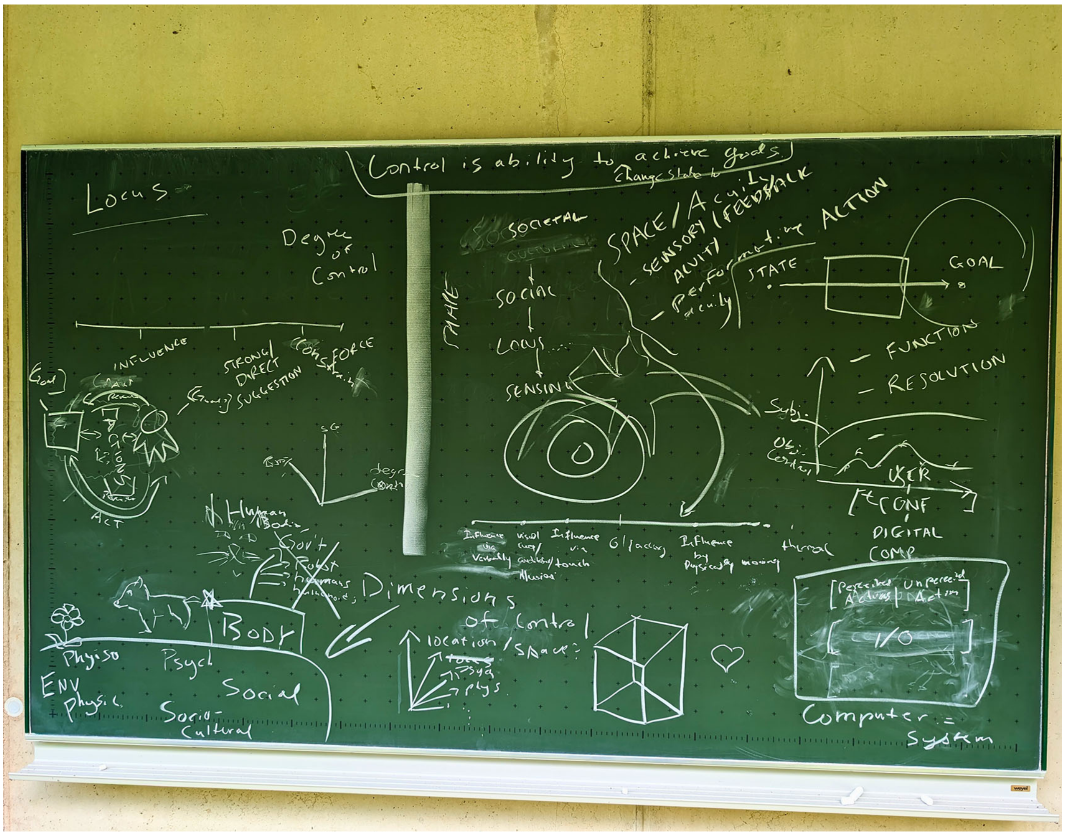


Figure 5 Group discussion on defining "control" in human-computer interaction, illustrating key concepts such as influence, sensory feedback, and decision-making processes.

3. **Social and ethical implications:**
   - A third group focused on control technologies' social and ethical implications. They examined how these technologies might affect social norms, interpersonal relationships, and individual autonomy (Figure 6). Topics such as user consent, data privacy, and the avoidance of dark patterns were central to their discussions. The group aimed to identify ethical guidelines and best practices for designing and deploying control technologies in ways that respect user rights and societal values.
4. **Synthesising challenges:**
   - This group aimed to synthesise the grand challenges identified during the PechaKucha sessions and the first two days of discussions. They reviewed the notes and flipcharts created throughout the sessions, looking for common themes, gaps, and emerging questions. Their objective was to produce a coherent set of challenges that encapsulate the key issues facing researchers and practitioners in the field, providing a roadmap for future research and development.
5. **Engineering perspectives on control:**
   - The fifth group explored the engineering side of control technologies. They focused on technical challenges such as calibration, measurement, and benchmarking methods, and the need for robust, adaptable algorithms that can dynamically adjust to user needs and contexts. This group also examined how to design for user confidence and

trust, ensuring that control mechanisms are reliable, transparent, and predictable. They looked into how engineering solutions could support both explicit and implicit control forms, aligning with broader ethical and social considerations.

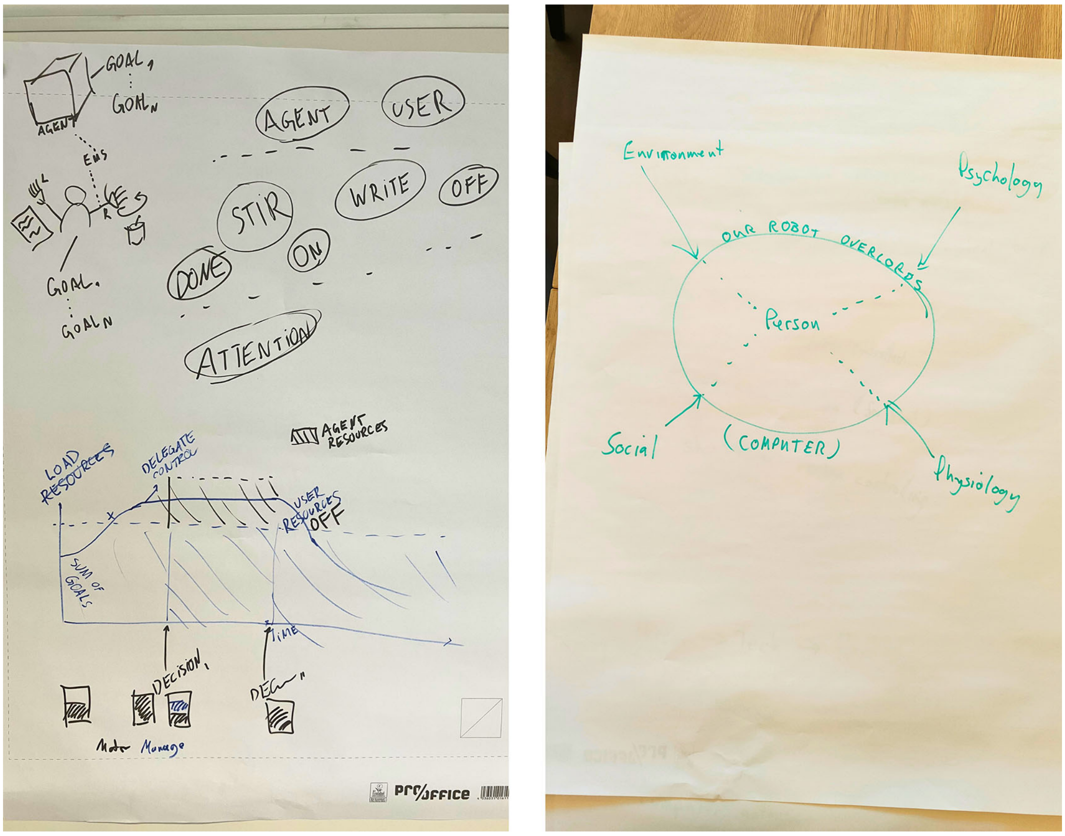


Figure 6 Exploration of social, ethical, and diverse perspectives of control, highlighting the implications of control technologies on human autonomy, privacy, and societal norms.

5.1 Key Outcomes from Group Activities

- **Shared understanding:** By the end of Day 3, there was a clearer shared understanding of the core concept of control and its complexities within human-computer interactions. This was marked by a convergence of ideas on the need for new frameworks and models that account for the diverse experiences and expectations of control.
- **Framework for future research:** The synthesis of challenges led to the creation of a preliminary framework outlining the key areas for future research. This included ethical guidelines, design principles, and technical requirements that would guide the responsible development of control technologies.
- **Interdisciplinary collaboration:** The discussions highlighted the importance of interdisciplinary collaboration, bringing together perspectives from psychology, engineering, ethics, and design to address the multifaceted nature of control technologies. Participants identified potential collaboration opportunities and established working groups to continue exploring these themes beyond the seminar.

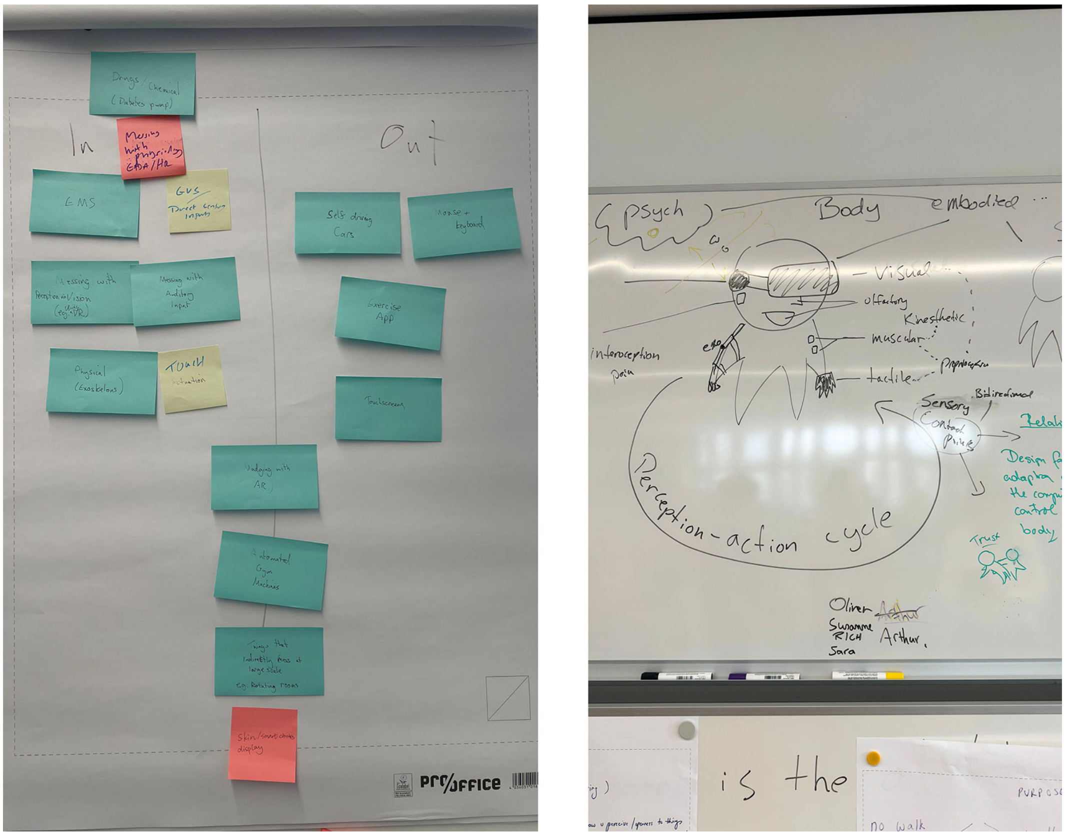


Figure 7 Scoping and perspectives session on control, examining different viewpoints from physiological, psychological, environmental, and social dimensions to understand the full impact of control technologies.

- **Engineering challenges:** The group focusing on engineering challenges provided insights into the technical constraints and possibilities of implementing control technologies. They emphasised the importance of rigorous testing, user feedback loops, and iterative design processes to ensure these technologies are both effective and aligned with user needs.

These group activities set the stage for subsequent discussions and sessions, creating a collaborative environment where participants could further refine their ideas and contribute to the development of a comprehensive understanding of computational control over the human body.

6 Hiking Activity at Saar-Hunsrück-Steig

On the afternoon of Day 3, participants took a break from the intense seminar discussions and embarked on a hiking tour along the scenic Saar-Hunsrück-Steig trail (Figure 8). The hike provided a unique opportunity for informal networking and reflection in a more relaxed and natural setting. Scheduled to start at 1:30 PM, the activity was designed for leisure and as a platform for meaningful career development and peer exchange.

Figure 8 Group photo of seminar participants during the hiking tour at Saar-Hunsrück-Steig, nurturing informal networking, career discussions, and reflection.

6.1 Career Advice in Motion

The hiking tour incorporated a dedicated career advice session where participants could seek guidance and share insights about professional growth and development. Senior academics and experienced researchers walked alongside early-career participants, facilitating open and candid conversations about navigating the complexities of academic and research careers.

Participants were encouraged to switch groupings periodically, allowing them to interact with a variety of peers and mentors. This dynamic format fostered an environment where advice could be exchanged freely, and participants could gain diverse perspectives on topics such as:

- Building a research profile and securing funding.
- Balancing academic responsibilities with personal life.
- Navigating interdisciplinary collaborations.
- Strategies for publishing in high-impact journals.
- Leveraging professional networks for career advancement.

Many participants found this informal setting particularly conducive to discussing challenges and aspirations they might not typically share in more formal seminar settings.

6.2 Ending the Hike with a Group Dinner

After the hike, the participants reconvened for a group dinner, where they continued their conversations and reflected on the day's experiences. The dinner provided an additional networking opportunity, solidifying new connections made during the hike and sharing reflections on the seminar's themes.

The combination of physical activity, career-focused discussions, and social interaction helped deepen the sense of community among participants, enhancing the collaborative spirit developing throughout the seminar.

Overall, the hiking activity was a refreshing and productive break that enriched the seminar experience, offering personal and professional growth opportunities.

7 Thinking Through Writing

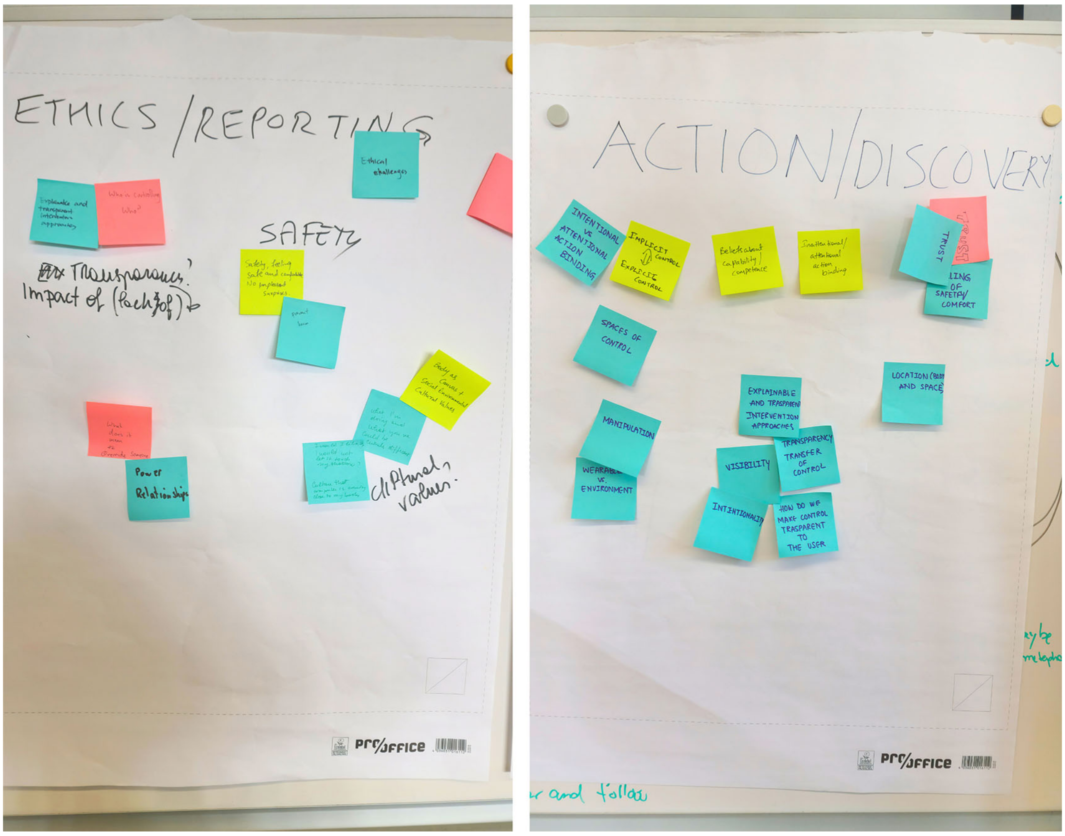


Figure 9 Thematic grouping exercise: Participants organised key discussion points into themes, such as "Ethics/Reporting" and "Action/Discovery," which became the high-level grand challenges.

On Day 4, participants engaged in a group activity titled "Thinking Through Writing," where they collaboratively organised all discussion points gathered over the first three days of the seminar. The process began by thematically categorising the diverse ideas, questions, and insights that were documented on sticky notes and flipcharts throughout the seminar. These discussions ranged across multiple topics, including control mechanisms, ethical considerations, user agency, social implications, and technical challenges.

Participants worked in small groups to cluster these points into higher-level themes that could encapsulate the broad spectrum of ideas that had surfaced. These groupings allowed for a deeper examination of the emergent themes and helped identify connections between seemingly disparate points. For example, discussions on "Calibration" and "Measurement/Benchmarking" were grouped under technical challenges, while topics like "Ethics/Reporting," "Social Transparency," and "Power Dynamics" were clustered under social and ethical considerations (Figure 9).

This thematic organisation was dynamic, involving active debate and negotiation among participants to reach a consensus on which points should be grouped together and what overarching themes they represented. The result was a refined set of grand challenges that captured the core issues and opportunities in designing computers' control over human bodies.

These themes were then visually represented on new flipcharts, providing a more coherent framework for understanding the various dimensions of the seminar's central topic. This framework became a reference point for further discussion and reflection.

8 Writing the Grand Challenges

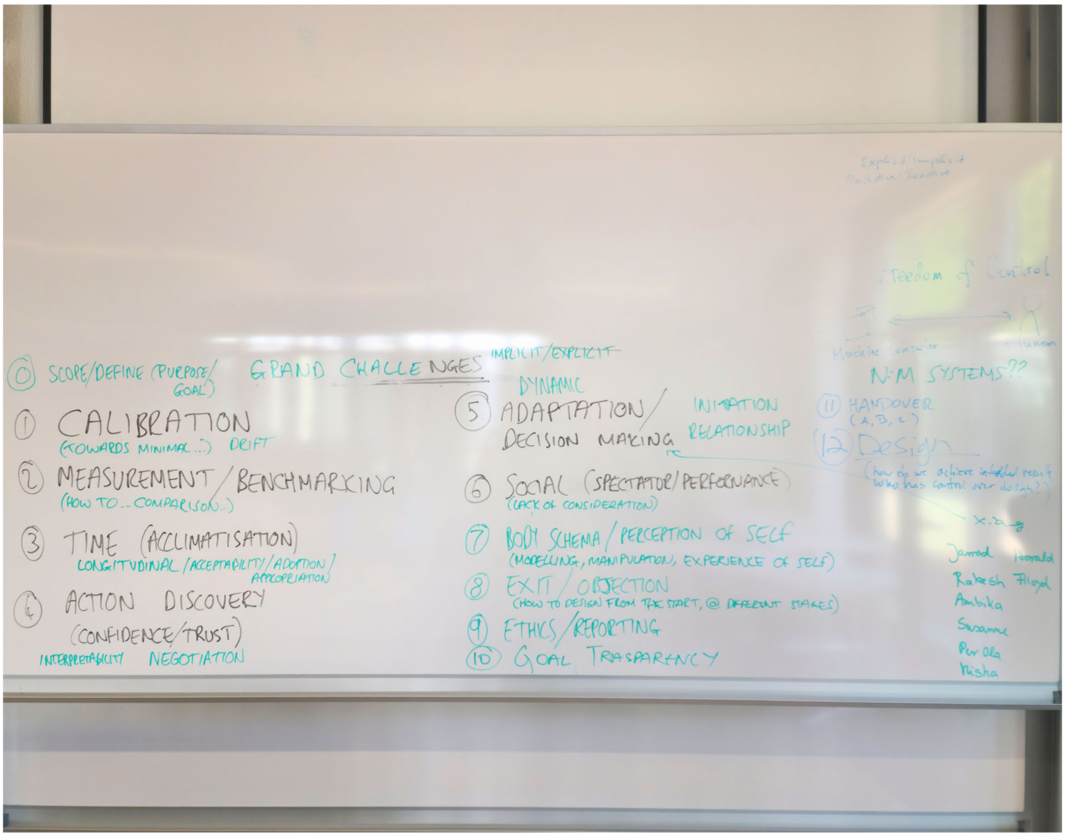


Figure 10 Final list of grand challenges identified during the seminar, covering themes such as calibration, measurement, adaptation, social implications, ethical reporting, and transparency in designing technologies that control human bodies.

On Day 5, the seminar shifted focus towards drafting a collective article that synthesised the grand challenges identified through the previous day's thematic grouping activity (Figure 10). Using the flipcharts as a guide, participants worked in their respective groups to flesh out the key elements of each challenge, drawing on the rich discussions and insights generated throughout the seminar.

The writing process was collaborative and iterative, with participants contributing their expertise to different sections of the article. Each group focused on elaborating their assigned themes, providing detailed explanations, relevant examples, and proposing potential research directions or solutions. For instance, the group handling the "Action Discovery" theme expanded on how confidence and trust play a crucial role in user acceptance of control technologies, while those working on "Ethics/Reporting" outlined guidelines for ethical transparency and user safety.

The aim was to produce a comprehensive document that not only summarised the seminar's key findings but also provided a roadmap for future research and practice in the field of human-computer interaction involving bodily control. Throughout the day, participants engaged in discussions to refine their sections, ensuring that the article was coherent and aligned with the seminar's goals.

By the end of Day 5, the draft article was completed, capturing the collective knowledge and perspectives of all participants. It represented a significant outcome of the seminar, encapsulating the core challenges and opportunities identified over the five days. This article is intended to be a starting point for ongoing dialogue and collaboration in the field, fostering a shared understanding of how to design for and navigate the complex interplay between human autonomy and computational control.

9 Overview of Talks

9.1 Digital Vertigo Games

Richard James Byrne (Made Tech – Bristol, GB)

License Creative Commons BY 4.0 International license
© Richard James Byrne

Many people enjoy "vertigo" sensations caused by intense, playful bodily activities such as riding fairground rides. Game scholar Caillois calls such experiences "vertigo play", elucidating that these enjoyable activities result from confusion between sensory channels.

In HCI, designers are often cautious to avoid deliberately causing sensory confusion in players, but we believe there is an opportunity to transition and extend Caillois' thinking to the digital realm, allowing designers to create novel and intriguing digital bodily experiences inspired by traditional vertigo play activities, through exploring methods of allowing computers varying degrees of control over one's body.

The Digital Vertigo Experience framework is derived from four case studies and the development of three different digital vertigo experiences. This framework aims to bring the excitement of traditional vertigo play experiences to the digital world, allowing designers to create novel body-based games and providing players with more possibilities to enjoy exciting play experiences [3].

9.2 Charting the Landscape of Body Control Techniques

Arthur Caetano (University of California – Santa Barbara, US)

License Ⓒ Creative Commons BY 4.0 International license
© Arthur Caetano

Body control techniques, such as Electric Muscle Stimulation (EMS), Galvanic Vestibular Stimulation (GVS), and Exoskeletons, enable computational agents to act on the user's body. These techniques can offload control of voluntary movements to assist in physical tasks, stimulate body functions for rehabilitation, and redirect walking and reaching movements in virtual reality applications. Despite promising applications, more work is needed to outline the design space of body control techniques. Mapping out the available methods from the perspective of body entry points and targets for control signals can help researchers identify opportunities for enabling more comprehensive body control. What signal entry points could be used to control a specific body part? Which body parts can be controlled through the same entry point? Which control modalities can be combined with which sensing modalities without interference? What are the modalities most resilient to the user's resistant attitude? Synthesizing this landscape will support design decisions in body control applications.

9.3 TickleFoot: Design, Development and Evaluation of a Novel Foot-Tickling Mechanism That Can Evoke Laughter

Don Samitha Elvitigala (Monash University – Clayton, AU)

License Ⓒ Creative Commons BY 4.0 International license
© Don Samitha Elvitigala
Joint work of Don Samitha Elvitigala, Roger Boldu, Suranga Nanayakkara, Denys J. C. Matthies

Tickling is a type of sensation that is associated with laughter, smiling or other similar reactions. Psychology research has shown that tickling and laughter can significantly relieve stress. Although several tickling artifacts have been suggested in prior work, limited knowledge is available if those artifacts could evoke laughter. In this article, we aim at filling this gap by designing and developing a novel foot-tickling mechanism that can evoke laughter. We first developed an actuator that can create tickling sensations along the sole of the foot utilising magnet-driven brushes. Then, we conducted two studies to identify the most ticklish locations of the foot's sole and stimulation patterns that can evoke laughter. In a follow-up study with a new set of participants, we confirmed that the identified stimuli could evoke laughter. From the participants' feedback, we derived several applications that such a simulation could be useful. Finally, we embedded our actuators into a flexible insole, demonstrating the potential of a wearable tickling insole [4].

9.4 Control in HCI

Jarrod Knibbe (University of Queensland – Brisbane, AU)

License Creative Commons BY 4.0 International license
© Jarrod Knibbe
Joint work of Jarrod Knibbe, Rachel Freire, Marion Koelle, Paul Strohmeier

Control in HCI is poorly understood. We lack a systematic approach to modelling, predicting, and benchmarking control. At the same time, however, control is arguably the central challenge of interaction design: controlling a system towards a goal. Across its diverse history, HCI has adopted two primary levels for studying control, implicit control and explicit control. Implicit control concerns itself with the low-level study of control perception. To study this, we have adopted neuroscience methods, primarily relying on intentional binding and the Libet clock, but the demands of these methods make them hard to apply for interaction design, and the results appear to have limited links with our subjective feeling of control (our feelings of explicit control). Explicit control methods are primarily "qualitative", effectively asking "do you feel in control?" I posit that this question is broadly unanswerable and the data it yields almost arbitrary. We need better methods [5].

9.5 Closed-loop and open-loop control

Per Ola Kristensson (University of Cambridge, GB)

License Creative Commons BY 4.0 International license
© Per Ola Kristensson

Bestowing computer systems with control can amplify people's ability to input information into systems. Gesture keyboard systems enable users to write using touchscreen and other input media by sliding a finger across a keyboard. Serial single-finger typing is a closed-loop control task with performance limited by visual attention. Using a gesture keyboard, prolonged use results in motor memory consolidation of frequently used gestures for words. This allows users to transition to open-loop direct recall from motor memory, which is faster. However, for this to be possible the system has to interpret imprecise open-loop articulation using a decoder. In effect, the user gains efficiency by transitioning from precise closed-loop to imprecise open-loop and this transition is only possible because a decoder interprets the imprecise input. Effectively, control is shifted towards the computer system in exchange for efficiency. The same principle can be exploited in other applications, such as dwell-free eye typing, which removes the time-consuming and fatiguing dwell timeouts of regular eye typing by interpreting eye gaze data using a decoder. Other examples include typing in thin air and automatic activation of virtual tools in virtual reality by the user simply articulating the tool motion to trigger the tool, removing the need for the user to first select the tool. However, bestowing a computer system with this additional control may also come with a cost. It is possible to empirically measure a user's sense of agency and we can demonstrate that, in at least one instance, at a certain point increased computer assistance results in a dramatic reduction in the user's sense of agency.

9.6 Swarm Bodily Interfaces in Virtual Environments

Xiang Li (University of Cambridge, GB)

License Creative Commons BY 4.0 International license
© Xiang Li

Our bodies, serving as intermediaries between physical and virtual realities, play a crucial role in transmitting our perceptions and presenting our behaviors. Previous research primarily focused on creating one-to-one replicas of the real world as avatars in virtual environments or utilizing real-world on-body interactions to enhance immersion. This approach, however, has not fully exploited the potential of bodily interfaces. For example, if our bodies were composed of swarm particles with a fluid and liquid form in virtual reality, how would our perception adapt to such a dynamic state? Would our viewpoint fluctuate with the liquid's movements, and how could we exert control over such a form? Future research should aim to design diverse bodily interfaces to investigate the implications for agency, ownership, perception, affordance, and interaction. This exploration is essential to comprehensively understand and enhance our presence and interaction within virtual environments [9].

9.7 Sharing Bodily Control with Computers

Zhuying Li (Southeast University – Nanjing, CN)

License Creative Commons BY 4.0 International license
© Zhuying Li

In this talk, I first introduced several examples of my work related to sharing control with computers, exploring the interplay between human autonomy and technological influence. The first part focuses on ingestible play, a game based on sensors. In these games, participants swallow a digital sensor that monitors bodily information to serve as a game controller. However, control over the sensor is limited as participants cannot direct its movement within the body. In some cases, players can influence it indirectly, for example, by consuming coffee or bananas to affect digestion speed. Also, it would be hard to control the data the technology captured, as we don't have direct control over our interior body. These kinds of play make people aware that we have limited control over our body as the invisible part of it. The second work I shared in this talk involves sharing bodily control with an intelligent exoskeleton in body games. This exoskeleton is designed to predict and actuate optimal movements during gameplay. We developed three modes for this system. In the first one, the exoskeleton autonomously moves the player's body to compete against a computer-controlled opponent. In the second mode, one arm controlled by the exoskeleton competes against the player's other free arm. With the last one, the exoskeleton controls one arm while the player decides whether to move the other arm to align with the AI's predictions to play. Both projects aim to augment human capabilities – whether to enhance gaming performance or to increase self-awareness. However, these augmentations also raise important questions about how much we should allow machines to influence or determine human actions and whether we should let machines decide what is good for humans [10, 11, 17].

9.8 Losing Control

Joe Marshall (University of Nottingham, GB)

License Creative Commons BY 4.0 International license
© Joe Marshall
Joint work of Steve Benford, Christopher Greenhalgh, Gabriella Giannachi, Brendan Walker, Joe Marshall, Tom Rodden

Our work on uncomfortable interactions explored technology interactions in which people deliberately expose themselves to discomfort. Experiences such as rollercoasters, theatre shows, and horror films have in common the requirement that people gradually surrender control to technology over the experience. Two examples: In Bronocomatic, a bucking bronco ride is controlled by the rider's breathing. Initially, this means the rider is fully in control of the ride movement, but as the movement becomes more extreme, riders are no longer able to control their breathing, and control shifts to the machine. Teslasuit uses electrical muscle stimulation to create a range of sensations in the wearer. At the lowest stimulation, it is like a gentle touch, and users stay fully in control. As the stimulation gets more intense, the wearer becomes less in control of their body until, at full strength, the intensity is such that wearers fall to the ground and cannot control their muscles.

9.9 Designing computers' control over the body that centres around bodily play

Louise Petersen Matjeka (Copenhagen, DK)

License Creative Commons BY 4.0 International license
© Louise Petersen Matjeka
Joint work of Louise Petersen Matjeka, Dag Svanæs, Alf Inge Wang

My focus for designing computers' control over the body centers around bodily play and to answer the following question: How can we design for bodily playfulness? A key element in my work with bodily play is movement – and making (game) moves. I have developed two games, each of which manipulates the player to move and not move in specific ways. The first game, Space Agent, used an interactive binaural soundscape to drive the players to twist and turn. The other game, Move Maker, explored different ways of restraining the body as it had to move around playing with the game's paraphernalia as the surrounding conditions for movement. My recommended reading is about the latter game. My work merges my background as a PhD in game and interaction design (Norwegian University of Science and Technology) and a Music and Movement Teacher (Rhythmic Music Conservatory, Copenhagen).

9.10 Sharing control over our bodies with computers

Florian 'Floyd' Mueller (Monash University – Clayton, AU)

License Creative Commons BY 4.0 International license
© Florian 'Floyd' Mueller

Sharing control over our bodies with computers benefits from the integration of the human body and interactive technology. We demonstrate this through a series of research design works around sharing control experiences, including eBike integrations, Wifi integration and emotion-EMS integration. The results of these works suggest interesting ways forward for control in HCI research, in particular, how the design of integrated systems can consider experiential aspects, facilitating playful control experiences. Ultimately, with our work, we want to enhance our knowledge around the design of integrated control experiences to help people understand who they are, who they want to become, and how to get there [19].

9.11 If control is here, how is it? Engendering subtleties of "control" observable in symbiotic systems through an enactive approach

Minna Nygren (University College London, GB)

License Creative Commons BY 4.0 International license
© Minna Nygren

Since the emergence of the notion of body-machine couplings dating back to the 1950s (e.g., Clynes and Kline, 1960), interface design and engineering has increasingly explored developing systems with varying levels of human control (e.g., full, partial) and machine control (e.g., autonomous, partial) (e.g., Casado, 2013). Enactivist perspective to cognition (e.g., Thompson and Varela, 2001) can open a window into exploring the notion of control in subtle processes of situated interaction, perception and action (e.g., Rowlands, 2010). Methodologically, this approach may engender microinteractional dimensions of "computers" control over our bodies' observable, and develop our understanding of the details about where, and how, the control is. This may offer insights about how "control" is perceived and acted upon in care contexts applying tactile robotics (e.g., Parviainen et al., 2019), or mirroring digital environments (e.g., Nygren, 2023).

9.12 Playing by Sharing Bodily Control with Computers

Rakesh Patibanda (Monash University – Clayton, AU)

License Creative Commons BY 4.0 International license
© Rakesh Patibanda

Traditional digital games predominantly treat the human body as an input mechanism, overlooking its potential as an output medium. This oversight limits the depth of interaction and experiential richness possible in human-computer interaction (HCI). Body-Actuating Play (BAP) [21, 26], discussed in this seminar, is a concept where bodily control is shared between the player and the computer, enabling a novel form of play where the body is both an input and output medium. This novel concept transforms the user's role from a mere operator to a collaborator with the computer, shifting their role to that of both influencer and responder.

As body-actuating technologies like electrical muscle stimulation and exoskeletons begin to control and augment human actions, we find ourselves at a pivotal moment in redefining the relationship between humans and technology. The potential for shared bodily control in BAP challenges our traditional notions of autonomy and control in digital play, provoking deeper reflection on the nature of the control we have over our bodies and the ethics of machine intervention in our physical selves [25].

Crucially, the exploration of BAP compels us to consider the psychological impact of sharing bodily control with such technologies. What does it mean for our sense of self when a machine can respond to our movements and direct them? How do we reconcile the benefits of enhanced physical capabilities with the potential loss of personal control over our bodies? While BAP is an exciting concept to explore, the ethical landscape of these innovations must be navigated with care, ensuring that the design of these technologies prioritises user consent, safety, and well-being.

Ultimately, BAP invites us to envision a future where humans and machines are intertwined in a seamless, symbiotic relationship, pushing the boundaries of what is possible in digital play experiences. This paradigm shift calls for a re-evaluation of traditional design frameworks within the HCI community, embracing the body as a central component of digital play. By designing future BAP experiences, we can create technologies that entertain and enrich and expand the human experience in previously unimaginable ways.

9.13 Negotiating Augmented Bodies

Henning Pohl (Aalborg University, DK)

License 🅭 Creative Commons BY 4.0 International license
© Henning Pohl

Our appearance and our interactions with the world around us are closely intertwined. How we dress, how we do our hair, and what makeup we put on signals something about who we are to others. Currently, more and more facets of our appearance are becoming adaptive and changeable through computational processes. Physically, this includes shape- and colour-changing clothing and accessories, as well as electronic tattoos, beauty products, and jewellery. Virtually, many more changes are possible, such as beautification filters, digital fashion, or blurring out of tattoos. Control over such body augmentations right now is mostly in the hands of the person themselves. However, such control might well transfer to others or to computational systems. For example, we can envision a future cocktail party setting where I put virtual name tags on people (i.e., I have control over their appearance), wear glasses that can turn heart-shaped on their own (i.e., I delegated control to a system and asked it to accentuate my flirting), or everybody relinquishes control over their looks (i.e., a central party AI takes charge to maybe ensure the party is eventful). Thus, as our appearance becomes interactively malleable, we also have to negotiate how much control is retained or delegated [29, 30].

9.14 Social and Embodied Perspectives

Sara Price (University College London, GB)

License 🅭 Creative Commons BY 4.0 International license
© Sara Price

Our bodies are intimate and personal: social, embodied, and physiological. This demands considerations for design – beyond the flesh – of computer "control" on our bodily engagement, raising 3 key questions: what do we mean by "bodies", what do we mean by "control", and what methods can foster important social design considerations? Bodies are not just fleshy entities; they are complex sensorial systems embedded in societal, community, and individual lived experiences, cultures and contexts, emotions, histories, etiquettes, and practices. Computers can exert control on our bodies explicitly (e.g., exoskeletons), but also implicitly, in less recognised ways. Research shows how a haptic baby sensing monitor changes parent-baby bodily interaction/communication practices: is this control? Socially oriented methodologies, such as socio-technical imaginaries/speculations, draw attention to these broader aspects of bodily engagement, to the body's socially shaped and interpreted sensorial experiences, foregrounding dystopian/utopian discourses, changing practices, power relationships, privacy, and ethical implications.

9.15 "Best Buddy": AI & XR-Enhanced Human

Harald Reiterer (Universität Konstanz, DE)

License 🅭 Creative Commons BY 4.0 International license
© Harald Reiterer

My research focus involves developing interaction techniques for Extended Reality (XR) Applications. Building on my previous research on cross-reality interaction, we combine different input devices (e.g., tabletops, tablets, smartphones) with Extended Reality devices (e.g., Head-Mounted Displays), harnessing their complementary strengths. We refer to these as Complementary Interfaces (Zagermann et al., 2022) and utilise them for various Visual Computing scenarios (e.g., Immersive Analytics, Situated Analytics). In my talk, I introduced the vision of a "Best" Buddy, an AI- and XR-enhanced Human. "Best Buddy" consists of an ecosystem of devices and technologies that, while promising, also raise significant ethical concerns. The ecosystem comprises in-ear headphones or hearing aids, Augmented Reality (AR) contact lenses or AR glasses, and a smartphone. All devices are connected to a cloud service that uses AI technology (e.g., LLM like GPT-4o), leading to intertwined human-computer integration. In real-time, the LLM gets input from the lenses and in-ear headphones, empowering the human user with relevant information and giving hints on how to behave in specific situations. But who is in control? What are this vision's potential Dark or Deceptive Patterns of an intertwined human-computer integration? How is "Best Buddy" positioned in the design space of intertwined systems (Butler, Angel, Influencer, Adversary)? [16]

9.16 Designing Bodily Extensions

Aryan Saini (Monash University – Clayton, AU)

License 🅭 Creative Commons BY 4.0 International license
© Aryan Saini

HCI researchers have increasingly investigated bodily augmentations due to the associated benefits, such as accessibility, offering novel input space, and sensing environment. Most of these investigations have adopted a utilitarian perspective, providing impairment, mobility, and rehabilitation support. However, more recently, a subset called bodily extensions has emerged that appears to also embrace experiential aspects. These bodily extensions extend the human body beyond traditional sensing and accessibility perspectives. Initial designs were mostly rigid, focusing on short-term use, missing out on the opportunities for these bodily extensions to be integrated into everyday life. While the adoption of the same will become more prevalent owing to technological advances in the near future, there is limited knowledge on how to cater to controlling these bodily extensions with respect to the human body. In my work, I build soft pneumatic-based bodily extensions that users can incorporate into their everyday lives. I detail the design and rationale behind these bodily extensions, along with the novel scenarios they enable. Finally, I share my insights from creating these systems and the associated user experiences from our field study, hoping to better understand control over bodily extensions physically, physiologically, and psychologically [32, 31].

9.17 Haptic Experiences and Bodily Control

Oliver Schneider (University of Waterloo, CA)

License Creative Commons BY 4.0 International license
© Oliver Schneider

Haptics and physical interaction let us engage with digital systems in a variety of ways. Typically, this is structured as a system we use as a tool. However, increasing understanding of what haptic experiences are, new types of haptic devices, and capabilities for participatory design may influence our decisions about who controls what with our bodies. I will discuss how we can conceptualize haptic experiences, employ interaction modalities like implicit interaction in autonomous vehicles and intermittent interaction with fabrication devices, and then report on efforts for participatory design that give creators control over their haptic experiences and threats to bodily autonomy from touch.

9.18 Technology Acceptance of Electrical Muscle Stimulation

Ambika Shahu (TU Wien, AT)

License Creative Commons BY 4.0 International license
© Ambika Shahu
Joint work of Ambika Shah, Sonja Dorfbauer, Philipp Wintersberger, Florian Michahelles

Electrical muscle stimulation (EMS) has many potential applications in many fields, but the parameters necessary for EMS acceptance are poorly understood. We investigated a wide range of characteristics to gain a better understanding of EMS user acceptance. We created four scenarios on the basis of a literature review, assessed the technology acceptance of participants in an online survey with the help of the TAM model, and conducted in-depth interviews. Our results suggest that the use of EMS to enhance a VR experience is supported by a good attitude towards it and the intention to use it. We recommend that researchers and designers focus their efforts on keeping users in control of their activities rather than forcing them through EMS. We advise researchers to design systems that circumvent the involuntary nature of the system and allow the user to feel controlled. We found that the use of EMS to speed up learning was not widely accepted and that the desire to maintain control outweighed the desire to learn faster [34, 36, 35].

9.19 Exploring Control: Galvanic Vestibular Stimulation and Bodily Autonomy in Human-Computer Interaction

Misha Sra (University of California – Santa Barbara, US)

License Creative Commons BY 4.0 International license
© Misha Sra

Galvanic Vestibular Stimulation (GVS) has been a known technique for over a century, yet its potential in Human-Computer Interaction (HCI) remains largely untapped and poorly understood. Despite its long history, there is a notable absence of systematic approaches to modeling, predicting, and calibrating GVS-based systems designed for end-user applications. This gap in methodology has limited the widespread adoption of GVS in HCI applications, despite its potential for reducing cybersickness in virtual reality (VR), enhancing balance rehabilitation, or developing novel experiences.

In parallel to GVS control of the body, I am also interested in the questions of control in situations where individuals experience a loss or lack of control over their physical self. These situations range from medical conditions like paralysis, neuromuscular diseases, or amputation, to the temporary disorientation or diminished bodily sensation experienced when transitioning from VR back to physical reality. This prompts a fundamental question: When there's no sense of body, what does bodily control mean?

Understanding and addressing control in these contexts could help develop more accessible, inclusive, and empowering experiences. Both GVS applications and control-related challenges in HCI necessitate interdisciplinary research, combining insights from neuroscience, engineering, design, and HCI [37].

References

1 Alejandro Aponte, Arthur Caetano, Yunhao Luo, and Misha Sra. Grav: Grasp volume data for the design of one-handed XR interfaces. In Anna Vallgårda, Li Jönsson, Jonas Fritsch, Sarah Fdili Alaoui, and Christopher A. Le Dantec, editors, *Designing Interactive Systems Conference, DIS 2024, IT University of Copenhagen, Denmark, July 1-5, 2024*. ACM, 2024.

2 Steve Benford, Richard Ramchurn, Joe Marshall, Max L. Wilson, Matthew Pike, Sarah Martindale, Adrian Hazzard, Chris Greenhalgh, Maria Kallionpää, Paul Tennent, and Brendan Walker. Contesting control: journeys through surrender, self-awareness and looseness of control in embodied interaction. *Human–Computer Interaction*, 36(5-6):361–389, 2021.

3 Richard Byrne, Joe Marshall, and Florian 'Floyd' Mueller. Designing digital vertigo experiences. *ACM Trans. Comput.-Hum. Interact.*, 27(3), may 2020.

4 Don Samitha Elvitigala, Roger Boldu, Suranga Nanayakkara, and Denys J. C. Matthies. Ticklefoot: Design, development and evaluation of a novel foot-tickling mechanism that can evoke laughter. *ACM Trans. Comput.-Hum. Interact.*, 29(3), jan 2022.

5 Graeme Gange and Jarrod Knibbe. Optimising automatic calibration of electric muscle stimulation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(5):3759–3767, May 2021.

6 Carey Jewitt, Sara Price, Kerstin Leder Mackley, Nikoleta Yiannoutsou, Douglas Atkinson, Carey Jewitt, Sara Price, Kerstin Leder Mackley, Nikoleta Yiannoutsou, and Douglas Atkinson. Social norms of touch. *Interdisciplinary insights for digital touch communication*, pages 57–72, 2020.

7 Jarrod Knibbe, Rachel Freire, Marion Koelle, and Paul Strohmeier. Skill-sleeves: Designing electrode garments for wearability. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '21, New York, NY, USA, 2021. Association for Computing Machinery.

8 Jarrod Knibbe, Paul Strohmeier, Sebastian Boring, and Kasper Hornbæk. Automatic calibration of high density electric muscle stimulation. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3), sep 2017.

9 Xiang Li, Jin-Du Wang, John J. Dudley, and Per Ola Kristensson. Swarm manipulation in virtual reality. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction*, SUI '23, New York, NY, USA, 2023. Association for Computing Machinery.

10 Zhuying Li, Tianze Huang, Rakesh Patibanda, and Florian Mueller. Ai in the shell: Towards an understanding of integrated embodiment. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '23, page 1–7, New York, NY, USA, April 2023. Association for Computing Machinery.

11 Zhuying Li, Rakesh Patibanda, Felix Brandmueller, Wei Wang, Kyle Berean, Stefan Greuter, and Florian 'Floyd' Mueller. The guts game: Towards designing ingestible games. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '18, page 271–283, New York, NY, USA, October 2018. Association for Computing Machinery.

12 Joe Marshall, Duncan Rowland, Stefan Rennick Egglestone, Steve Benford, Brendan Walker, and Derek McAuley. Breath control of amusement rides. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 73–82, New York, NY, USA, 2011. Association for Computing Machinery.

13 Louise Petersen Matjeka, Dag Svanæs, and Alf Inge Wang. Turning eight family homes into interactive, pervasive playgrounds during the COVID-19 pandemic lockdown. *Frontiers Comput. Sci.*, 4:822337, 2022.

14 Nygren Minna, Thomas Rhiannon, and Sara Price. Designing for embodied proximity. *Visitor Studies*, 2024.

15 Florian Mueller, Marianne Graves Petersen, and Zhuying Li. Technology futures: Towards understanding how to design awe-inspiring futures. *International Journal of Human-Computer Studies*, 170:102961, 2023.

16 Florian Floyd Mueller, Pedro Lopes, Paul Strohmeier, Wendy Ju, Caitlyn Seim, Martin Weigel, Suranga Nanayakkara, Marianna Obrist, Zhuying Li, Joseph Delfa, Jun Nishida, Elizabeth M. Gerber, Dag Svanaes, Jonathan Grudin, Stefan Greuter, Kai Kunze, Thomas Erickson, Steven Greenspan, Masahiko Inami, Joe Marshall, Harald Reiterer, Katrin Wolf, Jochen Meyer, Thecla Schiphorst, Dakuo Wang, and Pattie Maes. Next steps for human-computer integration. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–15, New York, NY, USA, 2020. Association for Computing Machinery.

17 Florian 'Floyd' Mueller, Pedro Lopes, Josh Andres, Richard Byrne, Nathan Semertzidis, Zhuying Li, Jarrod Knibbe, and Stefan Greuter. Towards understanding the design of bodily integration. *International Journal of Human-Computer Studies*, 152:102643, 2021.

18 Florian 'Floyd' Mueller, Nathan Semertzidis, Josh Andres, Joe Marshall, Steve Benford, Xiang Li, Louise Matjeka, and Yash Mehta. Toward understanding the design of intertwined human–computer integrations. *ACM Trans. Comput.-Hum. Interact.*, 30(5), sep 2023.

19 Florian 'Floyd' Mueller, Nathan Semertzidis, Josh Andres, Martin Weigel, Suranga Nanayakkara, Rakesh Patibanda, Zhuying Li, Paul Strohmeier, Jarrod Knibbe, Stefan Greuter, Marianna Obrist, Pattie Maes, Dakuo Wang, Katrin Wolf, Liz Gerber, Joe Marshall, Kai Kunze, Jonathan Grudin, Harald Reiterer, and Richard Byrne. Human–computer integration: Towards integrating the human body with the computational machine. *Foundations and Trends® in Human-Computer Interaction*, 16(1):1–64, October 2022.

20 Minna O. Nygren, Sara Price, and Rhiannon Thomas Jha. The role of embodied scaffolding in revealing "enactive potentialities" in intergenerational science exploration. *Science Education*, 108(2):495–523, 2024.

21 Rakesh Patibanda, Chris Hill, Aryan Saini, Xiang Li, Yuzheng Chen, Andrii Matviienko, Jarrod Knibbe, Elise Van Den Hoven, and Florian 'Floyd' Mueller. Auto-paizo games: Towards understanding the design of games that aim to unify a player's physical body and the virtual world. *Proceedings of the ACM on Human-Computer Interaction*, 7(CHI PLAY):893–918, 2023.

22 Rakesh Patibanda, Chris Hill, Aryan Saini, Xiang Li, Yuzheng Chen, Andrii Matviienko, Jarrod Knibbe, Elise van den Hoven, and Florian 'Floyd' Mueller. Auto-paizo games: Towards understanding the design of games that aim to unify a player's physical body and the virtual world. *Proceedings of the ACM on Human-Computer Interaction*, 7(CHI PLAY):408:893–408:918, October 2023.

23 Rakesh Patibanda, Xiang Li, Yuzheng Chen, Aryan Saini, Christian N Hill, Elise van den Hoven, and Florian Floyd Mueller. Actuating myself: Designing hand-games incorporating electrical muscle stimulation. In *Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '21, page 228–235, New York, NY, USA, October 2021. Association for Computing Machinery.

24 Rakesh Patibanda, Nathalie Overdevest, Shreyas Nisal, Aryan Saini, Don Samitha Elvitigala, Jarrod Knibbe, Elise Van Den Hoven, and Florian 'Floyd' Mueller. Shared bodily fusion: Leveraging inter-body electrical muscle stimulation for social play. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, DIS '24, page 2088–2106, New York, NY, USA, July 2024. Association for Computing Machinery.

25 Rakesh Patibanda, Nathalie Overdevest, Aryan Saini, Zhuying Li, Josh Andres, Jarrod Knibbe, Elise van den Hoven, and Florian 'Floyd' Mueller. Exploring shared bodily control: Designing augmented human systems for intra- and inter-corporeality. In *Proceedings of the Augmented Humans International Conference 2024*, AHs '24, page 318–323, New York, NY, USA, 2024. Association for Computing Machinery.

26 Rakesh Patibanda, Aryan Saini, Nathalie Overdevest, Maria F. Montoya, Xiang Li, Yuzheng Chen, Shreyas Nisal, Josh Andres, Jarrod Knibbe, Elise van den Hoven, and Florian 'Floyd' Mueller. Fused spectatorship: Designing bodily experiences where spectators become players. *Proc. ACM Hum.-Comput. Interact.*, 7(CHI PLAY), oct 2023.

27 Rakesh Patibanda, Nathan Arthur Semertzidis, Michaela Scary, Joseph Nathan La Delfa, Josh Andres, Mehmet Aydin Baytaş, Anna Lisa Martin-Niedecken, Paul Strohmeier, Bruno Fruchard, Sang-won Leigh, Elisa D. Mekler, Suranga Nanayakkara, Josef Wiemeyer, Nadia Berthouze, Kai Kunze, Thanassis Rikakis, Aisling Kelliher, Kevin Warwick, Elise van den Hoven, Florian Floyd Mueller, and Steve Mann. Motor memory in hci. page 1–8, April 2020.

28 Rakesh Patibanda, Elise Van Den Hoven, and Florian 'Floyd' Mueller. Towards understanding the design of body-actuated play. In *Extended Abstracts of the 2022 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '22, page 388–391, New York, NY, USA, November 2022. Association for Computing Machinery.

29 Henning Pohl, Tor-Salve Dalsgaard, Vesa Krasniqi, and Kasper Hornbæk. Body layars: A toolkit for body-based augmented reality. In *26th ACM Symposium on Virtual Reality Software and Technology*, VRST '20, New York, NY, USA, 2020. Association for Computing Machinery.

30 Henning Pohl, Kasper Hornbæk, and Jarrod Knibbe. Wanding through space: Interactive calibration for electric muscle stimulation. In *Proceedings of the 9th Augmented Human International Conference*, AH '18, New York, NY, USA, 2018. Association for Computing Machinery.

31 Aryan Saini, Haotian Huang, Rakesh Patibanda, Nathalie Overdevest, Elise Van Den Hoven, and Florian Floyd Mueller. Somaflatables: Supporting embodied cognition through pneumatic bladders. In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22 Adjunct, page 1–4, New York, NY, USA, October 2022. Association for Computing Machinery.

32 Aryan Saini, Rakesh Patibanda, Nathalie Overdevest, Elise Van Den Hoven, and Florian 'Floyd' Mueller. Pneuma: Designing pneumatic bodily extensions for supporting movement in everyday life. In *Proceedings of the Eighteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '24, New York, NY, USA, 2024. Association for Computing Machinery.

33 Oliver Schneider, Karon MacLean, Colin Swindells, and Kellogg Booth. Haptic experience design: What hapticians do and where they need help. *International Journal of Human-Computer Studies*, 107:5–21, 2017. Citation Key: SCHNEIDER20175.

34 Ambika Shahu, Sonja Dorfbauer, Philipp Wintersberger, and Florian Michahelles. Skillab – a multimodal augmented reality environment for learning manual tasks. In José Abdelnour Nocera, Marta Kristín Lárusdóttir, Helen Petrie, Antonio Piccinno, and Marco Winckler, editors, *Human-Computer Interaction – INTERACT 2023*, pages 588–607, Cham, 2023. Springer Nature Switzerland.

35 Ambika Shahu, Kahled Kassem, Leonhard Zehetgruber, Philipp Winterberger, and Florian Michahelles. Electrifying obstacle avoidance: Enhancing teleoperation of robots with ems-assisted obstacle avoidance. In *HHAI 2023: Augmenting Human Intellect*, pages 224–233. IOS Press, 2023.

36 Ambika Shahu, Philipp Wintersberger, and Florian Michahelles. Would users accept electric muscle stimulation controlling their body? insights from a scenario-based investigation. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, New York, NY, USA, 2022. Association for Computing Machinery.

37 Misha Sra, Abhinandan Jain, and Pattie Maes. Adding proprioceptive feedback to virtual reality experiences using galvanic vestibular stimulation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–14, New York, NY, USA, 2019. Association for Computing Machinery.

38 Dennis Wolf, John J. Dudley, and Per Ola Kristensson. Performance envelopes of in-air direct and smartwatch indirect control for head-mounted augmented reality. In Kiyoshi Kiyokawa, Frank Steinicke, Bruce H. Thomas, and Greg Welch, editors, *2018 IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2018, Tuebingen/Reutlingen, Germany, 18-22 March 2018*, pages 347–354. IEEE Computer Society, 2018.

Participants

- Nadia Bianchi-Berthouze
University College London, GB
- Susanne Boll
Universität Oldenburg, DE
- Richard James Byrne
Made Tech – Bristol, GB
- Arthur Caetano
University of California – Santa Barbara, US
- Don Samitha Elvitigala
Monash University – Clayton, AU
- Masahiko Inami
University of Tokyo, JP
- Jarrod Knibbe
University of Queensland – Brisbane, AU
- Per Ola Kristensson
University of Cambridge, GB
- Xiang Li
University of Cambridge, GB
- Zhuying Li
Southeast University – Nanjing, CN
- Joe Marshall
University of Nottingham, GB
- Louise Petersen Matjeka
Copenhagen, DK
- Florian 'Floyd' Mueller
Monash University – Clayton, AU
- Minna Nygren
University College London, GB
- Rakesh Patibanda
Monash University – Clayton, AU
- Henning Pohl
Aalborg University, DK
- Sara Price
University College London, GB
- Harald Reiterer
Universität Konstanz, DE
- Aryan Saini
Monash University – Clayton, AU
- Oliver Schneider
University of Waterloo, CA
- Ambika Shahu
TU Wien, AT
- Misha Sra
University of California – Santa Barbara, US
- Jürgen Steimle
Universität des Saarlandes, DE

Remote Participants

- Mar Gonzalez-Franco
Google – Seattle, US

Figure 11 Group photo of seminar participants at Schloss Dagstuhl, celebrating a week of collaborative exploration and discussion on designing computers' control over our bodies.

Report from Dagstuhl Seminar 24241

Geometric modeling: Challenges for Additive Manufacturing, Design and Analysis

Tor Dokken*[1], Xiaohong Jia*[2], Géraldine Morin*[3], Elissa Ross*[4], and Zoë Marschner†[5]

1 **SINTEF – Oslo, NO.** tor.dokken@sintef.no
2 **Chinese Academy of Sciences, CN.** xhjia@amss.ac.cn
3 **IRIT – University of Toulouse, FR.** geraldine.morin@irit.fr
4 **Metafold 3D – Toronto, CA.** elissa@metafold3d.com
5 **Carnegie Mellon University – Pittsburgh, US.** zoem@cmu.edu

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 24241 "Geometric modeling: Challenges for Additive Manufacturing, Design and Analysis". The seminar has returned to the on-site participants only format. On-site participation is essential for establishing a good dialogue between participants. The industry participation was higher than in previous events. Many of the participants were newcomers that brought new ideas and life into the discussions. This report summarizes the seminar communications by providing the abstracts of the talks which present recent results in geometric modeling and its applications. Organized scientific exchanges were structured into three working groups that each provided a report included in this document. The working group reports highlight new and future challenges within Geometric Modeling in general, and its use within Additive Manufacturing, Isogeometric Analysis, and Design Optimization.

Seminar June 9–14, 2024 – https://www.dagstuhl.de/24241

2012 ACM Subject Classification Applied computing → Physical sciences and engineering; Computing methodologies → Artificial intelligence; Computing methodologies → Computer graphics; Mathematics of computing → Mathematical software; Mathematics of computing → Numerical analysis

Keywords and phrases Additive Manufacturing, Computer Graphics, Computer-Aided Design, Design Optimization, Geometric Modeling, Geometry, Geometry Processing, Isogeometric Analysis, Shape Design

Digital Object Identifier 10.4230/DagRep.14.6.52

1 Executive Summary

Tor Dokken (SINTEF - Oslo, NO)
Xiaohong Jia (Chinese Academy of Sciences, CN)
Géraldine Morin (IRIT - University of Toulouse, FR)
Elissa Ross (Metafold 3D - Toronto, CA)

License Creative Commons BY 4.0 International license
© Tor Dokken, Xiaohong Jia, Géraldine Morin, and Elissa Ross

The Dagstuhl Seminar 24241 "Geometric Modeling: Challenges for Additive Manufacturing, Design and Analysis" took place in the week of June 9–14, 2024. This year, the seminar returned to having on-site participants only. In the previous seminar (2021, during the COVID

* Editor / Organizer
† Editorial Assistant / Collector

Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

pandemic), a hybrid format was used. In 2021, close to two thirds of participants joined remotely from eleven different time zones. This year, with all participants on site, all could participate in discussions during breaks, meals, and in the evenings. The popular hike on the afternoon of the third day of the seminar was well-attended, combining scientific discussion with a walk through the beautiful surroundings of Dagstuhl. With most participants joining all five days, we benefited from high attendance during all sessions.

One of the challenges when planning a new Dagstuhl Seminar on geometric modeling was to achieve the right balance between renewal and continuation. This was particularly challenging for this iteration of the seminar, as the targeted number of participants was around 40, in comparison to the more than 50 who attended the 2021 seminar (including both on-site and online participation). Consequently, when making the initial list for invitations, some names that have contributed over many seminars were not included. This was necessary to allow new scientists to be invited in the interest of renewal. As always, there were some late cancellations that allowed us to also invite many of those on the reserve list. Industry participation was higher than in previous events. Many of the participants were newcomers who brought new ideas and life into the discussions. For the next seminar, the organizers will once again have to balance renewal and continuation.

As with previous seminars, geometric modeling remained the core topic of the seminar. However, in recent seminars, the focus has shifted from representation of shape for computer aided design to the challenges posed by a wide use of these technologies in industry and society. In particular, the use of the geometric model is considered within a thorough and complete process in order to design, optimize, and create manufactured 3D content. As the title of the seminar suggests, applications in additive manufacturing, design, and analysis were central. During the seminar, challenges of geometry representation and processing for architecture also arose as a major topic.

The abstracts of the 35 talks presented at the seminar are included in this report as well as the conclusions of the three working groups that addressed challenges in:

- geometric modeling for additive manufacturing,
- geometric modeling for design optimization, and
- computer aided geometric design and isogeometric analysis.

The topics discussed in these working groups ranged from theoretical challenges in spline technology, to the need for improved digital technology in geometric modeling, to manufacturing of novel shape concepts in architecture. An emerging theme was the topic of geometric data, which has become more diverse in shape and nature. At the same time, geometric models are also strongly linked and considered in a context broader than the representation of shape, which includes their physical properties and the capabilities to manufacture these models in a sustainable manner. The reflection, work, and opportunities for geometric design and analysis are now more open, and Dagstuhl meetings and the diversity of participants created an opportunity to further this wide vision of the research field.

As always, both the organizers and participants of the Dagstuhl Seminar appreciated the smooth execution of the event, due to the great support and organization from the research center, the great food and lunch meetings, and the opportunity to discuss scientific challenges in a friendly atmosphere at a beautiful location and great venue.

2 Table of Contents

Working groups

3 Overview of Talks

3.1 Numerical characterization and experimental validation of 3D printed lattice structures

Massimo Carraturo (University of Pavia, IT)

License Creative Commons BY 4.0 International license
© Massimo Carraturo

Numerical characterization of 3D printed lattice structure mechanical behavior is a challenging task due to the inhomogeneous material micro-structure and the complex geometry of these components. In this contribution an empirical procedure suitable to define a lattice material model is presented and validated. Starting from micro-indentation measurements the yield stress and the Young modulus of the material are obtained at nodal and truss locations on two different planes and an exponential plastic law is used to define four different isotropic material models with J2 plasticity. Experimental tensile tests were conducted using Digital Image Correlation (DIC) technique showing that the actual mechanical behavior of a lattice tensile specimen lies between the numerical curves. The geometrical issues of lattice components is also considered. In fact, it is well known form the literature that the elastic behavior of lattice structures is dramatically underestimated when computed on the as-designed (CAD) geometry. Therefore, the actual as-built geometry as acquired for instance by Computed Tomography (CT) scan has been used for the analysis. However, such a geometry can be very challenging to mesh, thus an efficient immersed boundary method, namely the Finite Cell Method, has been investigated to perform accurate numerical simulations of 3D printed lattice components.

3.2 Topology Optimization of Self-supporting Structures for Additive Manufacturing via Implicit B-spline Representations

Falai Chen (Univ. of Science & Technology of China – Anhui, CN)

License Creative Commons BY 4.0 International license
© Falai Chen
Joint work of Nan Zheng, Xiaoya Zhai, Jingchao Jiang, Falai Chen
Main reference Nan Zheng, Xiaoya Zhai, Jingchao Jiang, Falai Chen: "Topology Optimization of Self-supporting Structures for Additive Manufacturing via Implicit B-spline Representations", Comput. Aided Des., Vol. 175, p. 103745, 2024.
URL https://doi.org/10.1016/J.CAD.2024.103745

Owing to the rapid development in additive manufacturing, the potential to fabricate intricate structures has become a reality, emphasizing the importance of designing structures conducive to additive manufacturing processes. One crucial consideration is the ability to design structures requiring no additional support during manufacturing. In this talk, I will introduce an optimization framework for self-supporting structure design via implicit B-spline representations. The method employs the control coefficients of the implicit tensor product B-spline as the design variables and integrates a topology optimization model with self-supporting constraints analytically derived from the implicit B-spline representation. Compared with the traditional voxel-based methods, it effectively expedites the optimization process by reducing the number of design variables. Furthermore, several acceleration techniques are introduced to significantly enhance the efficiency.

References

1 X. Qian, Undercut and overhang angle control in topology optimization: a density gradient based integral approach, International Journal for Numerical Methods in Engineering 111 (3) , 247–272, 2017.

2 G. Allaire, C. Dapogny, R. Estevez, A. Faure, G. Michailidis, Structural optimization under overhang constraints imposed by additive manufacturing technologies, Journal of Computational Physics 351, 295–328, 2017.

3 C. Wang, W. Zhang, L. Zhou, T. Gao, J. Zhu, Topology optimization of self-supporting structures for additive manufacturing with b-spline parameterization, Computer Methods in Applied Mechanics and Engineering 374, 113599, 2021.

4 N. Zheng, X. Zhai, F. Chen, Topology optimization of self-supporting porous structures based on triply periodic minimal surfaces, Computer-Aided Design 161, 103542, 2023.

3.3 Locally Refined B-splines over hierarchical meshes

Tor Dokken (SINTEF – Oslo, NO)

License 🅭 Creative Commons BY 4.0 International license
© Tor Dokken
Main reference Tor Dokken, Tom Lyche, Kjell Fredrik Pettersen: "Polynomial splines over locally refined box-partitions", Comput. Aided Geom. Des., Vol. 30(3), pp. 331–356, 2013.
URL https://doi.org/10.1016/J.CAGD.2012.12.005

Locally Refined B-splines (LRB) represents a multivariate generalization of knot insertion for univariate spline spaces. For LRB refinement is through sequential insertion of axes parallel mesh line segments. For each refinement step the mesh-line segment must divide into two disjoint regions the support of at least one tensor product (TP) B-spline from the collection of TP B-splines spanning the spline space. In this presentation we propose a new subclass of LRB. It performs better than Truncated Hierarchical B-splines when both methods are defined over the same mesh of knotlines: 1. The spline space defined by the knotlines are always filled which is not always the case for THB; 2. The LRB basis functions have better scaling factors than for THB, resulting in better condition numbers of both the mass and stiffness matrices. The minimal refinements regions of this approach will be slightly longer or wider than the minimal refinement for THB to ensure that the support of at least one TP B-splines is split. The resulting TP B-splines for this subclass are guaranteed to be linearly independent.

References

1 Gael Kermarrec, Vibeke Skytt, Tor Dokken, Optimal Surface Fitting of Point Clouds Using Local Refinement. Springer Vibeke Skytt, Gael Kermarrec 2023 (ISBN 978-3-031-16954-0) 112 s. Springer Briefs in Earth System Sciences

2 Vibeke Skytt, Gaël Kermarrec, Tor Dokken: LR B-splines to approximate bathymetry datasets: An improved statistical criterion to judge the goodness of fit. Int. J. Appl. Earth Obs. Geoinformation 112: 102894 (2022)

3 Vibeke Skytt, Tor Dokken, Scattered Data Approximation by LR B-Spline Surfaces: A Study on Refinement Strategies for Efficient Approximation, in Geometric Challenges in Isogeometric Analysis, Springer 2022

4 Stangeby, I; Dokken, T, Properties of Spline Spaces Over Structured Hierarchical Box Partitions, in Isogeometric Analysis and Applications 2018, Springer 2021, pp 177-207

5 Francesco Patrizi, Tor Dokken: Linear dependence of bivariate Minimal Support and Locally Refined B-splines over LR-meshes. Comput. Aided Geom. Des. 77: 101803 (2020)

6 Tor Dokken, Vibeke Skytt, Oliver J. D. Barrowclough: Trivariate spline representations for computer aided design and additive manufacturing. Comput. Math. Appl. 78(7): 2168-2182 (2019)
7 Vibeke Skytt, Quillon K. Harpham, Tor Dokken, Heidi E. I. Dahl: Deconfliction and Surface Generation from Bathymetry Data Using LR B-splines. CoRR abs/1610.09992 (2016)
8 T. Dokken, V. Skytt, O. Barrowclough, Locally refined splines representation for geospatial big data. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 2015 ;Volum XL-3/W3. 565-570
9 Vibeke Skytt, Oliver Joseph David Barrowclough, Tor Dokken: Locally refined spline surfaces for representation of terrain data. Comput. Graph. 49: 58-68 (2015)
10 K.A. Johannsesen, T. Kvamsdal and T. Dokken, Isogeometric analysis using LR B-splines, Computer Methods in Applied Mechanics and Engineering, Computer Methods in Applied Mechanics and Engineering, Volume 269, 1 February 2014, 471–514
11 Tor Dokken, Tom Lyche, Kjell Fredrik Pettersen: Polynomial splines over locally refined box-partitions. Comput. Aided Geom. Des. 30(3): 331-356 (2013)

3.4 Volumetric Representations (V-reps): the Geometric Modeling of the Next Generation

Gershon Elber (Technion – Haifa, IL)

License Creative Commons BY 4.0 International license
© Gershon Elber
Joint work of many, including Gershon Elber, Ben Ezair, Fady Massarwi, Boris Van Sosin, Jinesh Machchhar, Ramy Masalha, Q Youn Hong, Emiliano Cirillo, Sumita Dahiya, Pablo Antolin, Massimiliano Martinelli, Annalisa Buffa, Giancarlo Sangalli, Stefanie Elgeti, Robert Haimes

The needs of modern (additive) manufacturing (AM) technologies can be satisfied no longer by boundary representations (B-reps), as AM enables the manipulation and fabrication of interior (graded) materials as well as porosity. Further, while the need for a tight coupling between design and analysis has been recognized as crucial almost since geometric modeling (GM) was conceived, contemporary GM systems only offer a loose link between the two, if at all. For about half a century, since the 70's, (trimmed) Non Uniform Rational B-spline (NURBs) surfaces have been the B-rep of choice for virtually all of the GM industry. Fundamentally, B-rep GM has evolved little during this period and is no longer able to fulfill the needs of modern (additive) manufacturing, namely heterogeneity and lattice/porosity support. In this talk, we seek to examine an extended (trimmed) NURBs volumetric representation (V-rep) that successfully confronts the existing and anticipated design, analysis, and manufacturing foreseen challenges. We extend all fundamental B-rep GM operations, such as primitive and surface constructors and Boolean operations, to trimmed trivariate V-reps. This enables the much-needed tight link to (Isogeometric) analysis (IGA) on one hand and the full support of (porous, heterogeneous, and anisotropic) modern/additive manufacturing needs on the other.

3.5 Stackable surface rationalization for freeform architectural design

Konstantinos Gavriil (SINTEF – Oslo, NO)

License Creative Commons BY 4.0 International license
© Konstantinos Gavriil

Surface stackability is the measure of how well a freeform surface can be decomposed into stackable components. Under this view, a freeform surface is treated as one modality of a single bimodal geometric object which adheres to two configurations: the deployed surface state and the stacked volume state. This dual configuration introduces a geometric link between freeform surfaces and volume foliations.

Stackability has applications in freeform architecture and digital fabrication methods, such as hot blade cutting and conformal 3D printing, and allows for efficiency in material use, packing, storage and transportation.

I presented work-in-progress results and discussed possible future research questions.

3.6 A data-driven approach to adaptive THB-spline fitting

Carlotta Giannelli (University of Firenze, IT)

License Creative Commons BY 4.0 International license
© Carlotta Giannelli
Joint work of Carlotta Giannelli, Sofia Imperatore, Angelos Mantzaflaris, Felix Scholz
Main reference Carlotta Giannelli, Sofia Imperatore, Angelos Mantzaflaris, Felix Scholz: "BIDGCN: boundary-informed dynamic graph convolutional network for adaptive spline fitting of scattered data," *Neural Computing and Applications*, pp. 1–24, 2024.
URL https://doi.org/10.1007/s00521-024-09997-0

In this talk, we combine computer aided geometric design methods with deep learning technologies. The final objective is to develop and apply adaptive data fitting schemes for the design of data-driven free-form spline geometries. Depending on the acquisition process, the nature of the data can strongly vary, from uniformly distributed to scattered and affected by noise; yet the reconstructed geometric models are required to be compact, highly accurate, and smooth, while simultaneously capturing key geometric features. We propose data-driven parameterization methods based on (geometric) deep learning, considering either structured or unstructured point cloud configurations. On the basis of this learning model, we introduce THB-spline fitting schemes with moving parameterization and present a selection of numerical examples.

3.7 Algebras from Algebras

Ron Goldman (Rice University – Houston, US)

License Creative Commons BY 4.0 International license
© Ron Goldman

An algebra is a vector space A together with a rule for multiplication $A \times A \rightarrow A$. The real numbers, complex numbers, quaternions, dual quaternions, and conformal quaternions are examples of algebras with potential applications in geometric modeling. The goal of this talk is to show how to generate new algebras from known algebras doubling the size of the algebra by adjoining one new element along with new multiplication rules characterizing how this new element interacts with preexisting elements of the algebra. Examples are provided to illustrate the method. The connection to even dimensional subalgebras of Clifford algebras is also explained. We will close with some open questions for future research.

3.8 Simulating and Visualizing the Functionality of Surfaces

Hans Hagen (RPTU Kaiserslautern-Landau, DE)

License Creative Commons BY 4.0 International license
© Hans Hagen

The functionality and the quality of surfaces are important topics in engineering.How can we visualize these aspects in a preprocessing step during simulations? A surface is under pressure and deforming. Is it bending without or with deforming the surface-metric? Mathematical concepts to deal with these kind of problems are differential geometry and infinitesimal bendings.Shadow-curves are an intuitive visualization tool. We show in this talk that as long as the shadow-lines stay stationary during the deformation the surface is infinitesimal rigid.

3.9 Curvature continuous corner cutting

Kai Hormann (University of Lugano, CH)

License Creative Commons BY 4.0 International license
© Kai Hormann
Joint work of Kai Hormann, Claudio Mancinelli

Subdivision schemes are used to generate smooth curves by iteratively refining an initial control polygon. The simplest such schemes are corner-cutting schemes, which specify two distinct points on each edge of the current polygon and connect them to get the refined polygon, thus cutting off the corners of the current polygon. While de Boor [1] showed that this process always converges to a Lipschitz continuous limit curve, no matter how the points on each edge are chosen, Gregory and Qu [2] discovered that the limit curve is differentiable under certain constraints. We extend this result and show that the limit curve can even be curvature continuous for a specific sequence of cut ratios.

References

1 Carl de Boor. Cutting corners always works. *Computer Aided Geometric Design*, 4(1–2):125–131, July 1987.
2 John A. Gregory and Ruibin Qu. Nonuniform corner cutting. *Computer Aided Geometric Design*, 13(8):763–772, November 1996.

3.10 Theory and Applications of Moving Curves and Moving Surfaces

Xiaohong Jia (Chinese Academy of Sciences, CN)

License Creative Commons BY 4.0 International license
© Xiaohong Jia
Joint work of Xiaohong Jia, Falai Chen, Ron Goldman
Main reference Xiaohong Jia, Xiaoran Shi, Falai Chen: "Survey on the theory and applications of μ-bases for rational curves and surfaces", J. Comput. Appl. Math., Vol. 329, pp. 2–23, 2018.
URL https://doi.org/10.1016/J.CAM.2017.07.023
Main reference Jieyin Yang, Xiaohong Jia, Dong-Ming Yan: "Topology Guaranteed B-Spline Surface/Surface Intersection", ACM Trans. Graph., Vol. 42(6), pp. 211:1–211:16, 2023.
URL https://doi.org/10.1145/3618349
Main reference Xiaohong Jia, Falai Chen, Shanshan Yao: "Singularity Computation for Rational Parametric Surfaces Using Moving Planes", ACM Trans. Graph., Vol. 42(1), pp. 12:1–12:14, 2023.
URL https://doi.org/10.1145/3551387

Moving curves that follow a rational curve and moving surfaces that follow a rational surface serve as a bridge between the parametric forms and implicit forms. Their algebraic counterparts are special syzygies of the parametric equations of rational curves or surfaces. Over the past thirty years, the technique of moving curves and moving surfaces have been proven to be significant in solving many important problems in geometric modeling, such as fast implicitization, intersection computation, singularity computation, reparametrization as well as providing easy inversion formulas for points. We review the state-of-the-art results in µ-bases theory and applications for rational curves and surfaces, and raise unsolved problems for future research.

3.11 A point-normal interpolatory subdivision scheme preserving conics

Jiri Kosinka (University of Groningen, NL)

License Creative Commons BY 4.0 International license
© Jiri Kosinka
Joint work of Niels Bügel, Lucia Romani, Jirí Kosinka
Main reference Niels Bügel, Lucia Romani, Jirí Kosinka: "A point-normal interpolatory subdivision scheme preserving conics", Comput. Aided Geom. Des., Vol. 111, p. 102347, 2024.
URL https://doi.org/10.1016/J.CAGD.2024.102347

The use of subdivision schemes in applied and real-world contexts requires the development of conceptually simple algorithms that can be converted into fast and efficient implementation procedures. In the domain of interpolatory subdivision schemes, there is a demand for developing an algorithm capable of (i) reproducing all types of conic sections whenever the input data (in our case point-normal pairs) are arbitrarily sampled from them, (ii) generating a visually pleasing limit curve without creating unwanted oscillations, and (iii) having the potential to be naturally and easily extended to the bivariate case. In this paper we focus on the construction of an interpolatory subdivision scheme that meets all these conditions simultaneously. At the centre of our construction lies a conic fitting algorithm that requires as few as four point-normal pairs for finding new edge points (and associated normals) in a subdivision step. Several numerical results are included to showcase the validity of our algorithm.

3.12 Quadratic-surface-preserving parameterization of point clouds

Bert Jüttler (Johannes Kepler Universität Linz, AT)

License Creative Commons BY 4.0 International license
© Bert Jüttler
Joint work of Dany Rios, Felix Scholz, Bert Jüttler
Main reference Dany Rios, Felix Scholz, Bert Jüttler: "Quadratic surface preserving parameterization of unorganized point data", Comput. Aided Geom. Des., Vol. 110, p. 102287, 2024.
URL https://doi.org/10.1016/J.CAGD.2024.102287

Finding parameterizations of spatial point data is a fundamental step for surface reconstruction in Computer Aided Geometric Design. Especially the case of unstructured point clouds is challenging and not widely studied. In this work, we show how to parameterize a point cloud by using barycentric coordinates in the parameter domain, with the aim of reproducing the parameterizations provided by quadratic triangular Bézier surfaces. To this end, we train an artificial neural network that predicts suitable barycentric parameters for a fixed number of data points. In a subsequent step we improve the parameterization using non-linear optimization methods. We then use a number of local parameterizations to obtain a global parameterization using a new overdetermined barycentric parameterization approach. We study the behavior of our method numerically in the zero-residual case (i.e., data sampled from quadratic polynomial surfaces) and in the non-zero residual case and observe an improvement of the accuracy in comparison to standard methods.

3.13 Physics-Informed Geometric Operators to Support Generative Models for Shape Optimisation

Shahroz Khan (BAR Technologies – Portsmouth, GB)

License Creative Commons BY 4.0 International license
© Shahroz Khan
Joint work of Shahroz Khan, Panagiotis Kaklis

In engineering, the applications of machine learning and deep learning models have predominantly focused on reducing the computational costs of simulations by creating low-fidelity surrogate models that predict performance nearly instantly. Until recently, the capability of these models to generate innovative solutions was limited. The advent of modern generative models (GMs) and their application in engineering have significantly altered the landscape by enabling the creation of innovative shapes in addition to performance prediction.

However, as with these models, both the input and output streams consist of low-level 3D shape representations, they fail to capture structural and shape characteristics that are essential for performance analysis. Consequently, common issues in the generated designs include lack of surface smoothness and a large number of invalid designs, such as non-watertight or self-intersecting designs. This is particularly critical in engineering analysis where maintaining surface smoothness and validity is vital, as even minor local variations in surface quality can significantly impact performance. Additionally, due to their unsupervised nature, these models fail to incorporate any notion of physics.

In this work, we propose the use of physics-informed geometric operators (GOs) to enrich the geometric data provided to the employed GMs. We claim that this addition enables the extraction of useful high-level shape characteristics, even when using simple model architectures or low-level data representations like design parameters. GOs leverage the shape's differential and/or integral properties – retrieved via Fourier analysis, curvature, geometric moments, and their invariants – thereby introducing high-level intrinsic geometric information and physics into the resulting shape descriptors. These operators capture both global and local shape features by explicitly encoding the relevant shape information. This not only augments the training dataset with a compact geometric representation of free-form shapes but also embeds physical information. The latter is achieved through the selection of shape characteristics that correlate with the design's performance metrics (such as wave-making resistance, lift, and drag), thereby making the model physics-informed. Our experiments further demonstrate that GO-augmented shape descriptors result in measurable improvements in modelling accuracy and enhancements in the model's generalisation capability. Various finite combinations of GO-induced values can capture different sets of underlying geometric information, which are studied here for their efficacy in acting as sufficient and compact signatures of the corresponding shape surfaces.

3.14 Robust Pose Graph Optimization with Loop Closure Outliers

Tae-wan Kim (Seoul National University, KR)

License Creative Commons BY 4.0 International license
© Tae-wan Kim
Joint work of Tae-wan Kim, Camille Kesseler
Main reference Luca Carlone, Giuseppe Carlo Calafiore: "Convex Relaxations for Pose Graph Optimization With Outliers", IEEE Robotics Autom. Lett., Vol. 3(2), pp. 1160–1167, 2018.
URL https://doi.org/10.1109/LRA.2018.2793352

With the increasing need of autonomous robots for complicated environment situation such as underwater application, more robust algorithm is needed. In simultaneous localization and mapping algorithm, one of the core parts is the back end. The noisy measurement and robot trajectory are process to correct the drifting error using loop closure measurement (recognition of previously visited place). The process of optimizing the robot poses with respect to the sensor measurements is called pose graph optimization (PGO). Solving a PGO problem is equivalent to solve a maximum likelihood estimation problem where the objective function is the error between the measurement and the poses. The classical framework is to use a least-square formulation. However, this formulation has several drawbacks: The sensitivity to poses initialization first can lead to a local minima solution as it is a nonconvex problem. Then the presence of wrong measurement with large error, also called outliers, can lead to arbitrary wrong solution. In this research, we aim at studying a method for PGO which leverages the problem of initialization and is robust to outliers' presence. The initialization sensitivity problem comes from the nonconvexity of the minimization problem as it introduces multiple local minima. The proposed solution is to relax the problem into a convex one with a single global minimum. The solution of the relaxed problem can be reprojected on the initial nonconvex problem feasible set. Additionally, using this method we have a contract on the certifiability of our solution, i.e we can ensure that the solution is the global minima, or we detect the failure. For the sensitivity to outliers, it mainly comes from the fact that the ii formulation is quadratic in the error terms so if one measurement contains a wrong large error it will dominate the objective function. The proposed approach here is the use of M-estimator. A M-estimator is adding a loss function around the error term to mitigate its impact if it is too large. This thesis aims at comparing different loss function that can be used on the chosen convex relaxation approach. Additionally, we suppose that only edge which are loop closure can be outliers. After deriving the formulation corresponding to our choice, we test on 3 synthetic datasets the different loss function and compare them. Our results show that the convex loss function, i.e L^1, L^2, identity tested here do well for highly connected pose graph but failed to stay robust for low connectivity pose graph.

3.15 Reconstruction of Geometries using Sensor Data

Stefan Kollmannsberger (Bauhaus-Universität Weimar, DE)

License 🅭 Creative Commons BY 4.0 International license
© Stefan Kollmannsberger
Joint work of Tim Bürchner, Philipp Kopp, Stefan Kollmannsberger, Ernst Rank
Main reference Tim Bürchner, Philipp Kopp, Stefan Kollmannsberger, Ernst Rank: "Immersed boundary parametrizations for full waveform inversion", Computer Methods in Applied Mechanics and Engineering, Vol. 406, p. 115893, 2023.
URL https://doi.org/10.1016/j.cma.2023.115893

Sensor Data is diverse. Examples are pictures recorded by CMOS sensors common in any cell-phone, point clouds recorded by laser scanners, computed tomography scans as in medical imaging, recordings of vibrations by accelerometers, or pressure waves recorded by piezoelectric elements.

The presentation will give an insight on how to use such data for the mechanical analysis of structures [1, 2]. It finishes by presenting a particularly interesting combination of point clouds and wave signals to reconstruct the geometry of internal defects of as-built structures in concrete additive manufacturing.

References

1 Kudela, László, Stefan Kollmannsberger, Umut Almac, und Ernst Rank. "Direct structural analysis of domains defined by point clouds." *Computer Methods in Applied Mechanics and Engineering*, 2019. https://doi.org/DOI: 10.1016/j.cma.2019.112581.

2 Bürchner, Tim, Philipp Kopp, Stefan Kollmannsberger, und Ernst Rank. "Immersed Boundary Parametrizations for Full Waveform Inversion." *Computer Methods in Applied Mechanics and Engineering* 406 (März 2023): 115893. https://doi.org/10.1016/j.cma.2023.115893.

3.16 A few issues in intersections beyond triangle meshes

Zoë Marschner (Carnegie Mellon University – Pittsburgh, US)

License 🅭 Creative Commons BY 4.0 International license
© Zoë Marschner
Joint work of Zoë Marschner, Silvia Sellán, Hsueh-Ti Derek Liu, Alec Jacobson, Paul Zhang, David Palmer, Justin Solomon

Choosing a geometric representation is often about trade-offs – representations beyond the ubiquitous triangle mesh can make certain tasks fundamentally easier, but come with new challenges. In this talk, I will give two examples of these trade-offs I have encountered in my research by discussing some issues that arise when computing geometric intersections on higher-order patches and neural SDFs. I'll then discuss some of my work that builds towards solving these issues: detecting intersections between higher-order patches using an optimization technique called sum-of-squares relaxation and computing constructive solid geometry operations on neural SDFs.

References

1 Z. Marschner, S. Sellán, H. D. Liu, A. Jacobson, "Constructive solid geometry on neural signed distance fields," SIGGRAPH Asia 2023 Conference Papers, 1-12, 2023.
2 Z. Marschner, P. Zhang, D. Palmer, and J. Solomon. "Sum-of-squares geometry processing," ACM Transactions on Graphics (TOG) 40, no. 6, 1-13, 2021.

3.17 Alternating and joint surface approximation with THB-splines: results and industrial perspective

Dominik Mokriš (MTU Aero Engines – München, DE)

License Creative Commons BY 4.0 International license
© Dominik Mokriš
Joint work of Dominik Mokriš, Carlotta Giannelli, Sofia Imperatore, Angelos Mantzaflaris
Main reference Carlotta Giannelli, Sofia Imperatore, Angelos Mantzaflaris, Dominik Mokriš: "Leveraging Moving Parameterization and Adaptive THB-Splines for CAD Surface Reconstruction of Aircraft Engine Components", in Proc. of the Smart Tools and Applications in Graphics – Eurographics Italian Chapter Conference, The Eurographics Association, 2023.
URL https://doi.org/10.2312/stag.20231301

We will have a closer look at three surface approximation techniques: the *alternating point distance minimisation* (A-PDM), an alternating method that is a *hybrid between the PDM and the tangent-distance minimization* (A-HDM) and the *joint optimization method.* A common feature of these three methods is that they approximate a parametrised point cloud by not only searching for suitable control points but also by adapting the parameters of the data points. While the literature provides considerable experience with these methods in the curve case, relatively little is known about their performance when constructing surfaces. We have successfully combined the three methods with the truncated hierarchical B-splines (THB-splines) and our examples demonstrate a significant reduction of degrees of freedom necessary to construct a satisfying approximation.

The second focus of the talk will be the industrial viewpoint. Introducing MTU and its activities in developing, manufacturing and repairing aircraft engines will enable putting the methods into an application context. We will review our team's past experience in industrialising geometric methods; based on this, we will discuss the "boring" and technical part of what the next steps and hurdles towards productive use would be. While the problematics of the geometric modeling at MTU is fairly specific, the principles discussed could be interesting for anyone intent on seeing their methods applied in practice.

3.18 A clean, robust 3D medial axis

Géraldine Morin (IRIT – University of Toulouse, FR)

License Creative Commons BY 4.0 International license
© Géraldine Morin
Joint work of Géraldine Morin, Bastien Durix, Kathryn Leonard

Computing the medial axis of a 3D surface mesh is challenging. Points on the discrete medial axis can be defined as interior Voronoï vertices of the surface mesh, but the resulting medial structure rarely has clean connectivity and consistent geometry. In this work, we propose a medial axis computation of an arbitrary surface mesh based on the Voronoï diagram able to generate manifold medial sheets with coherent topology in terms of homothopy and orientability, that is, generating consistent geometric structures similar to those in the continuous setting. Because of the correspondences between the surface mesh and resulting medial mesh, we also provide an efficient method for separating the shape into coherent regions associated to medial structures. This correspondence allows for a medial-axis-based filtration of surface structures to generate a built-in Hausdorff ϵ-approximation of the surface points based on a simplified medial axis, thereby providing a robust medial representation with guaranteed surface approximation.

3.19 New developments in machine learning for geometry reconstruction

Georg Muntingh (SINTEF – Oslo, NO)

License 🅭 Creative Commons BY 4.0 International license
© Georg Muntingh
Joint work of Georg Muntingh, Sverre Briseid

Recent developments in machine learning from sensor data have enabled new possibilities in geometry reconstruction. We will discuss some of these new methods, such as neural implict representations and autoregressive models, and investigate some preliminary results.

3.20 Design without Boundaries

Suraj R. Musuvathy (nTopology – New York, US)

License 🅭 Creative Commons BY 4.0 International license
© Suraj R. Musuvathy
URL www.ntop.com

Advances in manufacturing processes and design exploration technology have enabled creation of objects with unprecedented complexity and customizability. Design exploration methods such as multi-disciplinary optimization, topology optimization, and AI have enabled rapid exploration of large design spaces. Additive manufacturing enables production of shapes with complex topology and geometric features varying across many length scales. The choice of a geometric representation, and the core modeling algorithms it enables, are crucial in order for a design system to enable engineers to leverage the benefits of these advances and create better mechanical products. Most, if not all, major CAD software systems today are built on Boundary Representations (B-Reps). In our view B-Reps have reached their limits on addressing the design opportunities available today. At nTop, we are building a design system using hybrid implicit modeling that presents advantages in terms of geometric scalability, modeling operation reliability that enables automation, and performance over B-Reps. Our system also provides gradients that are beneficial for design optimization. Using a different geometric representation introduces challenges with interoperability with existing B-Rep based systems. We will present recent advances in this area in creating B-Reps from implicits as well as in direct interop with implicits with an API for our kernel.

3.21 Shape preserving interpolation on surfaces via variable-degree splines

Panagiotis Kaklis (The University of Strathclyde – Glasgow, GB)

License Creative Commons BY 4.0 International license
© Panagiotis Kaklis
Joint work of Panagiotis D. Kaklis, S. Stamatelopoulos, Alexandros I. Ginnis
Main reference Panagiotis D. Kaklis, S. Stamatelopoulos, Alexandros I. Ginnis: "Shape-preserving interpolation on surfaces via variable-degree splines", Comput. Aided Geom. Des., Vol. 109, p. 102276, 2024.
URL https://doi.org/10.1016/J.CAGD.2024.102276

This work proposes two, geodesic-curvature based, criteria for shape-preserving interpolation on smooth surfaces, the first criterion being of non-local nature, while the second criterion is a local (weaker) version of the first one. These criteria are tested against a family of on-surface C^2 splines obtained by composing the parametric representation of the supporting surface with variabledegree (≥ 3) splines amended with the preimages of the shortest-path geodesic arcs connecting each pair of consecutive interpolation points. After securing that the interpolation problem is well posed, we proceed to investigate the asymptotic behaviour of the proposed on-surface splines as degrees increase. Firstly, it is shown that the local-convexity sub-criterion of the local criterion is satisfied. Second, moving to non-local asymptotics, we prove that, as degrees increase, the interpolant tends uniformly to the spline curve consisting of the shortest-path geodesic arcs. Then, focusing on isometrically parametrized developable surfaces, sufficient conditions are derived, which secure that all criteria of the first (strong) criterion for shape-preserving interpolation are met. Finally, it is proved that, for adequately large degrees, the aforementioned sufficient conditions are satisfied. This permits to build an algorithm that, after a finite number of iterations, provides a C^2 shape-preserving interpolant for a given data set on a developable surface.

3.22 Robust Geometry Processing for Differentiable Physical Simulation

Daniele Panozzo (New York University, US)

License Creative Commons BY 4.0 International license
© Daniele Panozzo
Main reference Zizhou Huang, Davi Colli Tozoni, Arvi Gjoka, Zachary Ferguson, Teseo Schneider, Daniele Panozzo, Denis Zorin: "Differentiable solver for time-dependent deformation problems with contact", ACM Trans. Graph., Vol. 43(3), pp. 31:1–31:30, 2024.
URL https://doi.org/10.1145/3657648

The numerical solution of partial differential equations (PDE) is ubiquitously used for physical simulation in scientific computing, computer graphics, and engineering. Ideally, a PDE solver should be opaque: the user provides as input the domain boundary, boundary conditions, and the governing equations, and the code returns an evaluator that can compute the value of the solution at any point of the input domain. This is surprisingly far from being the case for all existing open-source or commercial software, despite the research efforts in this direction and the large academic and industrial interest. To a large extent, this is due to lack of robustness and generality in the geometry processing algorithms used to convert raw geometrical data into a format suitable for a PDE solver.

I will discuss the limitations of the current state of the art, and present a proposal for an integrated pipeline, considering data acquisition, meshing, basis design, and numerical optimization as a single challenge, where tradeoffs can be made between different phases to increase automation and efficiency. I will demonstrate that this integrated approach offers many advantages, while opening exciting new geometry processing challenges, and that a fully opaque meshing and analysis solution is already possible for heat transfer and elasticity problems with contact. I will present a set of applications enabled by this approach in reinforcement learning for robotics, force measurements in biology, shape design in mechanical engineering, stress estimation in biomechanics, and simulation of deformable objects in graphics.

3.23 Polyhedral-net Splines (PnS)

Jorg Peters (University of Florida – Gainesville, US)

License Creative Commons BY 4.0 International license
© Jörg Peters
Joint work of Jörg Peters, Kestutis Karciauskas, Kyle Lo, Bhaskar Mishra
Main reference Jörg Peters, Kyle Shih-Huang Lo, Kestutis Karciauskas: "Algorithm 1032: Bi-cubic Splines for Polyhedral Control Nets", ACM Trans. Math. Softw., Vol. 49(1), pp. 7:1–7:12, 2023.
URL https://doi.org/10.1145/3570158
Main reference Bhaskar Mishra, Jörg Peters: "Polyhedral control-net splines for analysis", Comput. Math. Appl., Vol. 151, pp. 215–221, 2023.
URL https://doi.org/10.1016/J.CAMWA.2023.09.041

Polyhedral-net splines generalize tensor-product splines by allowing additional control net configurations: isotropic patterns, like n quads surrounding a vertex, an n-gon surrounded by quads, and preferred direction patterns, that adjust parameter line density, such as T-junctions. PnS2 generalize C1 bi-2 splines. PnS3 generalize C2 bi-3 splines.

There are two instances of PnS2 in the public domain: a Blender add-on and a ToMS distribution. A web interface now offers solving elliptic PDEs on PnS2 surfaces using PnS2 finite elements. The output for PnS2 is tensor-product polynomial pieces in BB-form (Bezier form).

3.24 Optimization in Aerospace Engineering

Jeff Poskin (The Boeing Company – Seattle, US)

License Creative Commons BY 4.0 International license
© Jeff Poskin

Mixed integer optimization (MIO) is leveraged regularly across aerospace engineering, addressing problems in the design, manufacturing, production, and in-service life of commercial aircraft. This presentation will start with a brief overview of how Boeing's Applied Math Group applies MIO to problems at Boeing. This presentation will then focus on a mixed integer nonlinear programming (MINLP) formulation for stringer centerline design in a commercial vehicle. Fuselage stringers are load-bearing components that run lengthwise along an aircraft and transfer aerodynamic loads acting on the skin into the stringers and frames. The design of stringer centerlines in commercial airplanes has traditionally been performed manually by structural engineers since stringer configurations are subject to a wide variety of design and integration requirements. These requirements include minimum and maximum spacing between pairs of centerlines, maximum area constraints on frame bays, and a host of integration requirements with additional features in the fuselage. The MINLP formulation uses discrete variables to assign stringers to drop out of the design at specific frame stations and continuous variables to control the path stringers take on the fuselage surface. We will efficiently enumerate this design space by checking the feasibility of several linear programs representing relaxations of the stringer spacing constraints. This enumeration is incorporated into a branch and bound algorithm that solves the centerline design problem to global optimality. This algorithm is compared to a piecewise linear modeling approach to solve the MINLP.

3.25 Current Challenges in Additive Manufacturing: An Industry Perspective

Elissa Ross (Metafold 3D – Toronto, CA)

License Creative Commons BY 4.0 International license
© Elissa Ross

Additive manufacturing holds enormous potential as a fabrication methodology, offering "free complexity" and flexible, sustainable, local manufacturing. Examples of lightweight parts, high-efficiency heat exchangers, mass customized medical devices and the explosion of new results in metamaterials illustrate the opportunities. Yet, the growth of the AM industry has been slower than projected. Estimates of industry growth from 2020 showed a 25-30% year over year growth, but were recently adjusted downward to a new projected growth of just 13.9% in 2024 (source: AM Power). Some of the barriers to growth include cost (which remains high), a labour shortage, and a general conservatism/change resistance from the manufacturing industry. Key technical barriers to adoption include repeatability/consistency, material characterization, part qualification, availability and education on DfAM tools, and interoperability. Three observations from the perspective of my company Metafold are that:
1. manufacturing is increasingly defined by software,
2. iteration speed is under pressure for commercial manufacturing, putting pressure on simulation-informed design, and
3. appetite for trusted simulation is enormous.

Underpinning these themes is the need for a robust digital foundation. Choosing an appropriate geometry representation is critical to ensuring a seamless workflow between design, simulation, production, and validation, yet available software is divided in approach between meshes, BReps, and implicits, making interoperability a challenge.

3.26 Natural control for multi-sided surfaces

Péter Salvi (Budapest University of Technology and Economics, HU)

License Creative Commons BY 4.0 International license
© Péter Salvi
Joint work of Márton Vaitkus, Péter Salvi, Tamás Várady
Main reference Márton Vaitkus, Péter Salvi, Tamás Várady: "Interior control structure for Generalized Bézier patches over curved domains", Comput. Graph., Vol. 121, p. 103952, 2024.
URL https://doi.org/10.1016/J.CAG.2024.103952

The past decade has seen a large variety of new multi-sided surface representations. These works focused primarily on the patch equation, while the input – boundary constraints or control points – was assumed to be given. In our recent publications we have investigated methods for the automatic creation of natural control structures, and we have also proposed different tools for the intuitive modification of multi-sided, multi-connected surfaces.

For the placement of control points we recommend an algorithm based on the (curved) domain, which is in turn generated from the boundary curves. A refinement routine resembling degree elevation provides a means for adding further details to the patch, and also ensures a reasonable distribution of controls.

In the context of editing, control points can be divided into two types: (1) boundary control points (BCPs), responsible for the continuous connection to adjacent surfaces, and (2) interior control points (ICPs), governing the middle of the patch. BCPs have associated algebraic or numerical constraints, rendering direct modification infeasible. Here we propose an indirect approach for modification via control vectors. On the other hand, ICPs are unconstrained, but there are generally too many of them to handle manually. We show how hierarchical and proportional editing schemes can be useful in this situation.

References

1 T. Várady, P. Salvi, M. Vaitkus. *Genuine multi-sided parametric surface patches – A survey*. Computer Aided Geometric Design, Vol. 110, #102286, 2024.
2 P. Salvi, M. Vaitkus, T. Várady. *Constrained modeling of multi-sided patches*. Computers and Graphics, Vol. 114, pp. 86-95, 2023.
3 P. Salvi. *Intuitive interior control for multi-sided patches with arbitrary boundaries*. Computer-Aided Design and Applications, Vol. 21(1), pp. 143-154, 2024.
4 M. Vaitkus, P. Salvi, T. Várady. *Interior control structure for Generalized Bézier patches over curved domains*. Computers and Graphics, Vol. 121, #103952, 2024.

3.27 C^1 simplex splines on a triangulation and numerical simulations

Maria Lucia Sampoli (University of Siena, IT)

License Creative Commons BY 4.0 International license
© Maria Lucia Sampoli
Joint work of Maria Lucia Sampoli, Francesca Pelosi, Hendrik Speleers, Jean-Louis Merrien

Splines over triangulations or tetrahedral partitions are useful in many applications, such as finite element analysis (FEA), computer aided design (CAD), and other engineering problems. For several of these applications, C^0 piecewise linear polynomials do not suffice. In some cases, one needs smoother elements for modeling, or higher polynomial degrees to increase the approximation order. The smoothness over a triangular partition is obtained either by raising the degree of polynomials or keeping the degrees low and splitting the triangles into subtriangles. In the context of Isogeometric Analysis (IgA), the key concept is the development of a new isoparametric paradigm for FEA, where the same basis functions used for geometry representations in CAD systems are adopted for the approximation of field variables. In its original formulation, IgA is based on tensor-product B-splines and their rational version NURBS. Unfortunately, the tensor-product structure has some drawbacks, especially when the adaptivity of the mesh is required. This motivates the interest in alternative structures for IgA, including T-splines, hierarchical B-splines, LR-splines, THB-splines. However, due to their (local) tensor-product structures, there are still restrictions on the refinement. To overcome the tensor-product restriction, many advances have been made in using non-tensor product splines, such as triangular/tetrahedral Bezier patches, Powell-Sabin splines, Box-splines, and so forth. An alternative to the above so-called macro-elements are spline spaces spanned by compactly supported, smooth functions. A well-understood example of such functions is the so-called simplex spline. Recently in[1] it was proposed a new simplex-spline basis for spline spaces built on Clough-Tocher splits with C^1 smoothness defined on a general triangulation. In this talk we present the use of spline spaces based on this particular class of simplex splines as a tool in numerical simulation.

References

1 T. Lyche and J-L Merrien, Simplex-splines on the Clough-Tocher element, Comput. Aided Geom. Design 65 (2018), 76-92.

3.28 Robust mass lumping and outlier removal strategies in isogeometric analysis

Espen Sande (EPFL – Lausanne, CH)

License Creative Commons BY 4.0 International license
© Espen Sande
Joint work of Yannis Voet, Espen Sande, Annalisa Buffa
Main reference Yannis Voet, Espen Sande, Annalisa Buffa: "Robust mass lumping and outlier removal strategies in isogeometric analysis", CoRR, Vol. abs/2402.14956, 2024.
URL https://doi.org/10.48550/ARXIV.2402.14956

Mass lumping techniques are commonly employed in explicit time integration schemes for problems in structural dynamics to both avoid solving costly linear systems with the consistent mass matrix and increase the critical time step. In isogeometric analysis, the critical time step is constrained by so-called "outlier" frequencies, representing the inaccurate high frequency part of the spectrum. Removing or dampening these high frequencies is paramount for fast explicit solution techniques. In this work, we propose robust mass lumping and outlier removal techniques for nontrivial geometries, including multipatch and trimmed geometries. Our lumping strategies provably do not deteriorate (and often improve) the CFL condition of the original problem and are combined with deflation techniques to remove persistent outlier frequencies. Numerical experiments reveal the advantages of the method, especially for simulations covering large time spans where they may halve the number of iterations with little or no effect on the numerical solution.

3.29 Star-Shaped Elements

Scott Schaefer (Texas A&M University – College Station, US)

License Creative Commons BY 4.0 International license
© Scott Schaefer
Joint work of Anshul Mendiratta, Scott Schaefer, Wenping Wang

This talk considers a generalization of the Clough-Tocher interpolant to star-shaped polygons. We show how to automatically tessellate the shape into triangle elements. These elements are quartic functions that have cubic boundaries and quadratic cross-boundary derivatives. We then show how to satisfy all of the smoothness constraints including an exact count of the numbers of degrees of freedom. We remove those degrees of freedom through a fairing functional that reproduces cubic functions and show pictures of several basis functions created via this approach.

3.30 Adaptive optimization of isogeometric multi-patch discretizations using artificial neural networks

Felix Scholz (Johannes Kepler Universität Linz, AT)

License Creative Commons BY 4.0 International license
© Felix Scholz
Joint work of Dany Rios, Felix Scholz, Thomas Takacs
Main reference Dany Rios, Felix Scholz, Thomas Takacs: "Adaptive optimization of isogeometric multi-patch discretizations using artificial neural networks", CoRR, Vol. abs/2403.19286, 2024.
URL https://doi.org/10.48550/ARXIV.2403.19286

In isogeometric analysis, isogeometric function spaces are used for accurately representing the solution to a partial differential equation (PDE) on a parameterized domain. They are generated from a tensor-product spline space by composing its basis functions with the inverse of a parameterization of the physical domain. Depending on the geometry of the domain and on the data of the PDE, the solution might not have maximum Sobolev regularity close to every point of the domain, leading to a reduced convergence rate. In this case, it is necessary to reduce the local mesh size close to the singularities. Based on the concept of r-adaptivity we can find a suitable isogeometric function space for a given PDE without sacrificing the tensor-product structure. In particular, we use the fact that different reparameterizations of the same computational domain lead to different isogeometric function spaces while preserving the geometry. Starting from a multi-patch domain consisting of bilinearly parameterized patches, we aim to find the biquadratic multi-patch parameterization of the domain that leads to the smallest approximation error of the solution. In order to estimate the location of the optimal control points, we use a trained residual neural network that predicts optimal parameters for point sets sampled from the graph surface of an approximate solution to the PDE. An iterative procedure leads to a reparameterization of the computational domain that is adapted to the solution of the given PDE and leads to vastly improved approximation errors.

3.31 Parametric Geometric Modeling Techniques for Additive Manufacturing: Commercial Applications and Technological Insights

Gunnar Schulze (ttrinckle 3D – Berlin, DE)

License Creative Commons BY 4.0 International license
© Gunnar Schulze
URL https://www.trinckle.com/

In the evolving landscape of additive manufacturing (AM), the integration of parametric geometric modeling with cloud-based platforms is revolutionizing single-lot production and mass customization. This talk delves into the commercial and technological advancements enabled by our parametric modeling techniques, which are pivotal in harnessing the full potential of AM. We present compelling industry and medical use cases that demonstrate significant enhancements in design efficiency and customization, achieved through our innovative algorithms. Our discussion includes a detailed examination of a novel mesh morphing strategy, which innovatively employs field lines of an (1/r) potential to seamlessly enfold one mesh onto another. This method streamlines the design process and ensures precision and adaptability in complex geometries.

3.32 Splits and Simplex Splines

Tom Lyche (University of Oslo, NO)

License Creative Commons BY 4.0 International license
© Tom Lyche

Smooth splines of low degree on triangular or simplicial partitions are considered. These kind of spaces have applications in computer aided design and isogeometric analysis. One way to obtain both smoothness and low degree is to split each element in the space into sub-pieces. Examples we consider are splits named after Clough-Tocher, Alfeld, Powell-Sabin and Wang-She.

On a split we want to construct a basis of multivariate B-splines known as simplex splines and then use Bernstein-Bézier techniques to obtain a global representation.

3.33 Isogeometric Analysis for higher order problems – Challenges and Prospects

Ulrich Reif (TU Darmstadt, DE)

License Creative Commons BY 4.0 International license
© Ulrich Reif

The construction of spline spaces for the isogeometric analysis of higher order PDEs is partially understood for bivariate problems, but offers great challenges for trivariate problems. In this talk, we discuss the current situation and identify tasks to be addressed in the future. In particular, we consider volumetric subdivision as a possible candidate for the construction of function spaces with sufficient Sobolev regularity. While many analytic questions still remain unsolved, we can report on progress concerning the construction of algorithms with favorable properties.

3.34 Generative Manufacturing: AI + IGA, Digital Twins and Reduced Order Modeling for Applications in Additive Manufacturing

Yongjie Jessica Zhang (Carnegie Mellon University – Pittsburgh, US)

License Creative Commons BY 4.0 International license
© Yongjie Jessica Zhang

Generative manufacturing applies the power of artificial intelligence (AI) to generate and execute optimal solutions given customer-defined constraints and parameters, such as functional specifications, cost, and lead time, by exploring vast combinations of design and production alternatives based on material and process availability. In this talk, we present our latest research on combining AI with isogeometric analysis (IGA) for applications in additive manufacturing (AM). It includes a machine learning (ML) framework for inverse design and manufacturing of self-assembling fiber-reinforced composites in 4D printing [1], IGA-based topology optimization for AM of heat exchangers [2, 3], as well as data-driven residual deformation prediction to enhance metal component printability and lattice support structure design in the laser powder bed fusion (LPBF) AM process [4]. By speeding up geometry distortion predictions from several hours to mere seconds with uncertainty quantification, our model can be deployed to prevent generation of infeasible designs. Our on-going efforts also include developing digital twins to enable prediction and control of process parameters for minimal melt pool variability in LPBF manufacturing, where reduced order modeling [5, 6] is one key technique to efficiently simulate underlying physics such as the melt pool dynamics.

References

1 Y. Yu, K. Qian, H. Yang, L. Yao, Y. J. Zhang. *Hybrid IGA-FEA of Fiber Reinforced Thermoplastic Composites for Forward Design of AI-Enabled 4D Printing*. Journal of Materials Processing Technology, 302:117497, 2022
2 X. Liang, A. Li, A. D. Rollett, Y. J. Zhang. *An Isogeometric Analysis-Based Topology Optimization Framework for 2D Cross-Flow Heat Exchangers with Manufacturability Constraints*. Engineering with Computers, 38(6):4829-4852, 2022
3 X. Liang, L. White, J. Cagan, A. D. Rollett, Y. J. Zhang. *Unit-Based Design of Cross-Flow Heat Exchangers for LPBF Additive Manufacturing*. ASME Journal of Mechanical Design, 145:012002, 2023
4 L. White, G. Zhang, J. Seo, N. Lamprinakos, A. Rollett, J. Cagan, Y. J. Zhang. *A Multi-Sized Unit Cell Method to Design Lattice Support Structures for Complex Geometries in LPBF*. ASME International Design Engineering Technical Conference & Computers and Information in Engineering Conference (IDETC/CIE). Washington, DC. Aug 25-28, 2024
5 A. Prakash, Y. J. Zhang. *Projection-Based Reduced Order Modeling and Data-Driven Artificial Viscosity Closures for Incompressible Fluid Flows*. Computer Methods in Applied Mechanics and Engineering, 425:116930, 2024
6 A. Prakash, Y. J. Zhang. *Data-Driven Identification of Stable Sparse Differential Operators Using Constrained Regression*. Computer Methods in Applied Mechanics and Engineering, 429:117149, 2024

3.35 Designing Multi-Material Distributions in 3D Parts for Desired Deformation Behavior

Jianmin Zheng (Nanyang TU – Singapore, SG)

License Creative Commons BY 4.0 International license
© Jianmin Zheng
Joint work of Jianmin Zheng, Haoxiang Li

Objects made from different materials have demonstrated their potential to offer distinct functional properties across regions. They can be customized to meet specific application requirements by properly allocating materials to areas in need, showcasing remarkable design flexibility. However, multi-material 3D printing technology for fabricating such objects is usually limited by the number of printable base materials. This talk begins with a framework for design, optimization and fabrication of deformable 3D objects and then introduces an approach for efficiently designing the distribution of available base materials within an object to achieve the desired deformation behavior. The approach takes displacements and forces at a set of mesh vertices as input and uses FEM to compute the material distribution. Two formulations are proposed. The first one is a discrete formulation based on L0-minimization, achieving the computation of sparse material distribution in one step, which is beneficial for additive manufacturing with multi-material printers. The second one is a continuous formulation through mathematical relaxation, which facilitates the numerical process. The work is partially supported by MOE AcRF Tier 1 Grant of Singapore (RG12/22).

4 Working groups

4.1 Additive Manufacturing

Massimo Carraturo (University of Pavia, IT), Gershon Elber (Technion – Haifa, IL), Stefan Kollmannsberger (Bauhaus-Universität Weimar, DE), Elissa Ross (Metafold 3D – Toronto, CA), and Gunnar Schulze (ttrinckle 3D – Berlin, DE)

License Creative Commons BY 4.0 International license
© Massimo Carraturo, Gershon Elber, Stefan Kollmannsberger, Elissa Ross, and Gunnar Schulze

Additive Manufacturing (AM) is an array of different technologies that are revolutionizing production systems worldwide. All AM technologies are sharing a common feature: the final part is produced depositing material in a layer-by-layer fashion. Contrary to subtractive technologies, the material is not incrementally removed starting from a bulk part anymore and the final component is instead generated by sequentially depositing new material. Such an approach allows the generation of close-to-freeform components since most of the traditional manufacturing constraints are removed. AM opens the possibility to produce very complex geometries that are not possible with traditional subtractive technologies, leading to more efficient usage of resources in terms of both material and energy consumption. Moreover, AM is a native digital technology and the overall workflow can be entirely digitalized involving minimal human intervention; therefore, it can be seen as a key enabling technology toward a green and digital transition of manufacturing ecosystems. Yet, at the industrial level, AM technologies are limited to a few sectors (e.g., aerospace, biomedical implants, jewelry). In many other industries, its broader adoption is still hindered by several factors. The multi-disciplinary working group in Dagstuhl focused on current challenges and limitations that still burden AM, limiting its widespread adoption in many manufacturing industries.

4.2 Design Optimization

Konstantinos Gavriil (SINTEF – Oslo, NO), Falai Chen (Univ. of Science & Technology of China – Anhui, CN), Panagiotis Kaklis (The University of Strathclyde – Glasgow, GB), Shahroz Khan (BAR Technologies – Portsmouth, GB), Dominik Mokriš (MTU Aero Engines – München, DE), Géraldine Morin (IRIT – University of Toulouse, FR), Caitlin Mueller (MIT – Cambridge, US), Georg Muntingh (SINTEF – Oslo, NO), Suraj R. Musuvathy (nTopology – New York, US), Baldwin Nsonga (Universität Leipzig, DE), Jorg Peters (University of Florida – Gainesville, US), Jeff Poskin (The Boeing Company – Seattle, US), David Reeves (Metafold 3D – Toronto, CA), and Eric Zimmermann (FU Berlin, DE)

License Creative Commons BY 4.0 International license
© Konstantinos Gavriil, Falai Chen, Panagiotis Kaklis, Shahroz Khan, Dominik Mokriš, Géraldine Morin, Caitlin Mueller, Georg Muntingh, Suraj R. Musuvathy, Baldwin Nsonga, Jorg Peters, Jeff Poskin, David Reeves, and Eric Zimmermann

We identify four challenges in Design Optimization: providing an adapted geometric representation, building a robust and efficient design system, providing the right degree of automation in order to secure a manufacturing pipeline that satisfies the criteria of accuracy and smoothness implied by the application area, and appropriate model formulation for manufacturing. We elaborate below.

4.2.1 Geometric representation

The first challenge is to find a geometric representation able to support efficient and robust Design Optimization. This geometric model should not be conceived as a singular choice but rather represents a system that is flexible enough to address several industrial challenges. More precisely, a geometric modeling system should provide the following features:

- **Scalability** to support arbitrarily complex geometries and large datasets;
- **Precise geometry** representations with guaranteed, quantified geometric fidelity up to a chosen threshold;
- Reliable and automated **interoperability** between geometric modeling, simulation, and optimization systems;
- **Volume friendliness** for securing seamless integration with manufacturing, especially AM, processes.

In addition, the system should provide the following tools/assets:

- **Reliable modeling operators** that succeed in any configuration thus avoiding expert intervention and solving;
- Geometric **analysis algorithms** that
  - leverage parallel/GPU processing when possible, ensuring efficiency and scalability;
  - compute analytic gradients of shapes with respect to design parameters to leverage gradient based optimization techniques.

Different representations have been proposed and used: B-reps are the standard today but severely limit automated design optimization due to insufficient support or lack of the above features. Alternative representations are desired, such as point set representations, and volume representations. Two such alternative attractive volumetric representations are implicit representations and V-reps, which come with their own challenges. Further work is required to bring them to professional industrial standards. Point data representing geometrical objects lack connectivity, prompting the development of methods to resemble point neighborhoods. These methods are sensitive to parameters, and as point sets grow, simplified structures like simplicial surfaces or skeletons become useful. These structures reduce complexity, enable interaction, and preserve topological properties absent in the original point sets. Overall, richer geometric representations are desired and further research is needed to mature them for professional design engineering.

4.2.2 Characteristics and properties of the system

As discussed above, the first challenge for geometric design is to provide a representation that offers features and tools to support the design process. Along the design process, analysis and optimization is performed, thus requiring assets of the underlying model, but also good properties of its parameterization.

Controlled approximation and numerical stability. That is, numerical methods may fail on noisy data or a badly conditioned model. Thus, robustness should be ensured when handling geometric data; visual datasets generated by computer vision methods, like digital twins, are known to be subject to noise and sometimes outliers. Also, approximation is sometimes necessary but for precise engineering processes, quantization of the approximation should be possible to ensure a model within a chosen threshold, and stable optimization methods are also required in order to further maintain a controlled approximation rate. The stability and measure of the approximation is particularly relevant when resorting to a surrogate model, or when reducing the dimensionality of a problem.

Differentiability and local vs global. Modeling the design process requires paying attention to the choice of the design parameters to both ensure a valid and viable solution, but also to allow a diverse and original set of solutions. Another desired property to ensure good performance from an optimization problem is to include differential operators whenever possible. But learning from existing or real world design, generative models may be of great use for design optimization. Differentiability relative to the chosen model parameters is completely necessary in this context. Future, explainable AI may also help us meet the need for controlled solutions.

4.2.3 Balance between automation and human interaction

Another challenge is achieving the optimal **balance between automation and human interaction in design processes**, one that reduces the user burden while allowing for meaningful and informed human input. Deciding which stages of the design optimization process can be automated and how the automation module itself should be "open" to exploration and interaction is a challenging task.

One example is the dimensionality reduction of parametric models to the important and relevant design parameters, and the inclusion of parameters where human input is most appropriate. This reduction can be a cumbersome process that should be preferably automated while remaining adequately accurate. Design decisions that are informed by hard-to-model human factors, such as aesthetic preferences or industry-specific expert knowledge, should be exposed for human input.

Another important issue is that the inclusion of human interaction during design exploration can lead to results affected by **user bias**. Such bias can work both positively and negatively. In industries that are sensitive to abrupt design changes, a designer can control the gradual introduction of modifications to the design. Conversely, user bias can negatively hinder progress in change-averse fields. For example, historical data can act as "attractors" in naval ship design which favors traditional designs, and as "repellers" in fields that explore novelty such as architectural design. We highlight the importance of **bias awareness** during human interaction with design systems.

Informed human input in design optimization requires appropriate interaction with data, e.g. by means of **appropriate visualization techniques**. The field of visual analytics is concerned with the question of designing visualization tools to interactively explore abstract data in a human-computer-environment system. Here the data and the information one wants to obtain from them give directions for suitable visualization techniques. These design principles and the derived analysis could be of potential interest on the way towards design optimization, in areas such as latent space exploration, understanding "black box" techniques better, and the choice of important parameters in sensitivity analysis. Furthermore, the insights gained from investigating visual analytics systems for design optimization can be leveraged to develop tools for decision support. These visualization systems prioritize the most relevant information and interactions for a specific task, which can positively impact decision-making based on data.

We also highlight the importance of communication between industry and academia, and between different fields. Seminars such as the Dagstuhl Seminar which generated the current report, can be greatly beneficial by guiding academic research towards real world industrial problems and questions. Moreover, transfer of knowledge between different disciplines and applications can improve practices and encourage progress. In that sense, systems that allow input from different fields or users lead to more open or creative designs.This aspect should be also taken into account by educators responsible for updating the curricula of designers and engineers at universities.

4.2.4 Model formulation for manufacturing: through objectives, constraints, or design space formulation

Finally, we highlight the importance of **appropriate model formulation in manufacturing**. The model should not only aim to provide a solution to the design problem at hand, but should also include aspects from the whole life-cycle of the product, such as **sustainable practices**, while avoiding common issues, such as bias related to specific choices in the model formulation.

Sustainability in AM is intrinsically linked to the **material choice**, the fabrication method employed, and the strategies implemented for **waste reduction**. These are all factors that can be incorporated to the formulation of the design optimization model. Focusing on the manufacturing approach for example, we can see that it can act as both a constraint in the design process, or as a variable that is determined by the design optimization result.

We already discussed the importance of geometric representation in a previous section. Here, we want to highlight that these geometric models should meet the requirements of a given AM technology, both constraints and capabilities. The model formulation should ensure that the modeled design is feasible, and that factors such as material properties, simulation of the manufacturing process, limitations and tolerances, are all **integrated into the early stages of design**, which has a heavy impact on the construction and operation cost.

Advanced algorithms for simulating material properties and the manufacturing process, supported by machine learning and computational geometry, can help with the automatic identification of both the geometric and the manufacturing constraints present in AM processes. This way, expert knowledge and experience are incorporated into the design stage, avoiding an otherwise lengthy and costly experimentation stage that requires trial and error.

4.3 Computer Aided Geometric Design and Isogeometric Analysis

Yongjie Jessica Zhang (Carnegie Mellon University – Pittsburgh, US), Tor Dokken (SINTEF – Oslo, NO), Ron Goldman (Rice University – Houston, US), Hans Hagen (RPTU Kaiserslautern-Landau, DE), Kai Hormann (University of Lugano, CH), Xiaohong Jia (Chinese Academy of Sciences, CN), Bert Jüttler (Johannes Kepler Universität Linz, AT), Myung-Soo Kim (Seoul National University, KR), Tae-wan Kim (Seoul National University, KR), Jiri Kosinka (University of Groningen, NL), Zoë Marschner (Carnegie Mellon University – Pittsburgh, US), Dominik Mokriš (MTU Aero Engines – München, DE), Géraldine Morin (IRIT – University of Toulouse, FR), Georg Muntingh (SINTEF – Oslo, NO), Daniele Panozzo (New York University, US), Helmut Pottmann (TU Wien, AT), David Reeves (Metafold 3D – Toronto, CA), Péter Salvi (Budapest University of Technology and Economics, HU), Maria Lucia Sampoli (University of Siena, IT), Scott Schaefer (Texas A&M University – College Station, US), Felix Scholz (Johannes Kepler Universität Linz, AT), Gunnar Schulze (ttrinckle 3D – Berlin, DE), and Jianmin Zheng (Nanyang TU –Singapore, SG)

License Creative Commons BY 4.0 International license
© Yongjie Jessica Zhang, Tor Dokken, Ron Goldman, Hans Hagen, Kai Hormann, Xiaohong Jia, Bert Jüttler, Myung-Soo Kim, Tae-wan Kim, Jiri Kosinka, Zoë Marschner, Dominik Mokriš, Géraldine Morin, Georg Muntingh, Daniele Panozzo, Helmut Pottmann, David Reeves, Péter Salvi, Maria Lucia Sampoli, Scott Schaefer, Felix Scholz, Gunnar Schulze, and Jianmin Zheng

During the group discussion session on the current status of Computer Aided Geometric Design (CAGD), isogeometric analysis (IGA) and the interaction between them, we identified future challenges and main areas of interest. This report briefly summarizes this discussion in four topics, including (4.3.1) theoretical foundation of CAGD, (4.3.2) what about Artificial Intelligence (AI)? (4.3.3) IGA, and (4.3.4) practical issues and datasets.

4.3.1 Theoretical foundations of CAGD

CAGD is concerned with the design, computation, and representation of curved objects on a computer. It has strong ties to approximation theory (approximation by polynomial and piecewise polynomial functions), differential geometry (parametric surfaces), algebraic geometry (algebraic surfaces), functional analysis and differential equations (surface design by minimizing functionals), and numerical analysis. CAGD has broad applications in many engineering fields, such as computer-aided design and manufacturing (CAD/CAM), engineering simulation, biomedical imaging, robotics, and computer vision. However, there are some theoretical issues that still need to be addressed. For example,

- Higher order geometric modeling, local refinement (LR), and modeling with tolerances
- What approximation theory do we need? Simplex splines and splines over triangulation, manifolds
- Splines on triangulations: dimension of spline spaces using approximation theory, topology and algebraic geometry methods, finding good bases
- Interpolating orthogonal polynomial basis for simplicial elements, C-infinity surfaces from triangle meshes (approximating)
- Generalized Barycentric coordinates combined with blossoming, n-D Barycentric rational interpolation Subdivision volumes and analysis, tools for analyzing nonlinear subdivisions
- B-reps is the technology of the 1980s standardized in the 1990s addressing the material shape. We need solutions for the micro and nano scales, as well as a new foundation for 2030
- V-reps to support Additive Manufacturing and Isogeometric Analysis

- Automatic mesh generation for complex domains
- Algebraic geometry, surface intersection, collision, singularity, contact surfaces
- Implicit models and operations on implicit models, contact problems, downstream finite element analysis, and dynamic remeshing
- Signed/gap distance function, distance measuring functions with precision, good data structure
- CAGD optimization, shape and topology optimization
- Integration of sensor data and simulation

4.3.2 What about artificial intelligence (AI)?

Traditional CAD software requires users to manually design every aspect of a project. With AI-powered tools, designers can now leverage algorithms to automate repetitive tasks and generate designs based on specific parameters and user preferences. Much progress has been made on deep neural networks for both explicit and implicit representations. Machine learning has helped improve the interoperability between geometric modeling, simulation and manufacturing. We summarize a set of new challenges and potential opportunities where machine learning could be used to improve CAGD, such as

- Overview of current CAGD experiments using AI, experiences and results
- Neural network combined with traditional algorithms; implicit neural representations, e.g. neural radiance fields, as an intermediate representation
- Autoregressive shape modelling, e.g. MeshGPT; learn from large datasets
- Splines as activation functions in neural networks, the local properties of B-splines for neural networks; Kolmogorov-Arnold Networks (KANs) replace the usual weights at each edge with a spline function, whose knot sequences can be adapted during training, yielding a nonparametric regression approach to machine learning of geometries.
- AI for esthetic design and art, AI for class A-surfaces, AI for controlling geometric properties such as curvature
- Parametrization, shape grammar and semantics, challenge of design patents
- Data-driven approaches to generate geometry from sensor data
- Multi-scale modeling, statistical methods, materials science
- Stability issues of AI/ML, Uncertainty Quantification, preconditioners
- Dimension reduction, reduced order modeling, smooth and better representation, digital twins

4.3.3 Isogeometric Analysis (IGA)

The root idea of IGA is to use the same basis suitable for both geometry and analysis, integrating design with analysis seamlessly. IGA has been a well-established technology and was successfully researched in academia since 2005. However, it is facing some difficult challenges in technology transfer to industry. Here we summarize some theoretical and practical issues related to software development and industrial inroads, including

- Spline element in FEM-software, for example, PolyFEM has spline elements
- Parameterization of irregular meshes in the object space, extraordinary nodes
- Analysis suitability properties: local refinement, linear independence, partition of unity (good for geometry, but not necessary for analysis), continuity, and so on
- Mesh generation is needed, what features are suitable for geometry which might not be ideal for analysis

- Benchmark problems comparing IGA vs FEM, need to investigate solver technologies (engage more analysis researchers in this seminar in the future)
- Data structure, software, sparse matrix solvers, libraries
- IGA in industry: LS-Dyna, find real applications such as Additive Manufacturing (AM)

4.3.4 Practical issues and datasets

Finally, the group also discussed practical issues and datasets used in CAGD and IGA. We summarize our findings as follows

- Medical applications
- Additive manufacturing applications
- Architecture
- Industrial needs
- How to get funding from Agencies
- How to attract young people to join the CAGD field
- Examples that must use higher order, where lower order does not work
- Benchmarks such as shell structure
- Education tools, sketch-based methods combining with AI, arts, spline-related linear algebra
- Sensor data of micro and nano structures: images, thermal, eddy current, acoustics
- Points cloud
- Lacking commercially available local refinement based spline libraries targeting IGA and AM

Participants

- Massimo Carraturo
University of Pavia, IT
- Falai Chen
Univ. of Science & Technology of China – Anhui, CN
- Tor Dokken
SINTEF – Oslo, NO
- Gershon Elber
Technion – Haifa, IL
- Konstantinos Gavriil
SINTEF – Oslo, NO
- Carlotta Giannelli
University of Firenze, IT
- Ron Goldman
Rice University – Houston, US
- Hans Hagen
RPTU Kaiserslautern-Landau, DE
- Stefanie Hahmann
INRIA Grenoble Rhône-Alpes, FR
- Kai Hormann
University of Lugano, CH
- Xiaohong Jia
Chinese Academy of Sciences – Beijing, CN
- Bert Jüttler
Johannes Kepler Universität Linz, AT
- Panagiotis Kaklis
The University of Strathclyde – Glasgow, GB
- Shahroz Khan
BAR Technologies – Portsmouth, GB
- Myung-Soo Kim
Seoul National University, KR
- Tae-wan Kim
Seoul National University, KR
- Stefan Kollmannsberger
Bauhaus-Universität Weimar, DE
- Jiri Kosinka
University of Groningen, NL
- Tom Lyche
University of Oslo, NO
- Zoë Marschner
Carnegie Mellon University – Pittsburgh, US
- Dominik Mokriš
MTU Aero Engines – München, DE
- Géraldine Morin
IRIT – University of Toulouse, FR
- Caitlin Mueller
MIT – Cambridge, US
- Georg Muntingh
SINTEF – Oslo, NO
- Suraj R. Musuvathy
nTopology – New York, US
- Baldwin Nsonga
Universität Leipzig, DE
- Daniele Panozzo
New York University, US
- Jorg Peters
University of Florida – Gainesville, US
- Jeff Poskin
The Boeing Company – Seattle, US
- Helmut Pottmann
TU Wien, AT
- David Reeves
Metafold 3D – Toronto, CA
- Ulrich Reif
TU Darmstadt, DE
- Elissa Ross
Metafold 3D – Toronto, CA
- Péter Salvi
Budapest University of Technology and Economics, HU
- Maria Lucia Sampoli
University of Siena, IT
- Espen Sande
EPFL – Lausanne, CH
- Scott Schaefer
Texas A&M University – College Station, US
- Felix Scholz
Johannes Kepler Universität Linz, AT
- Gunnar Schulze
ttrinckle 3D – Berlin, DE
- Yongjie Jessica Zhang
Carnegie Mellon University – Pittsburgh, US
- Jianmin Zheng
Nanyang TU – Singapore, SG
- Eric Zimmermann
FU Berlin, DE

Report from Dagstuhl Seminar 24242

Computational Analysis and Simulation of the Human Voice

Sten Ternström[*1], **Nathalie Henrich Bernardoni**[*2], **Peter Birkholz**[*3], **Oriol Guasch**[*4], **and Amelia Gully**[†5]

1 **KTH Royal Institute of Technology – Stockholm, SE.** stern@kth.se
2 **Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, FR.** nathalie.henrich@gipsa-lab.fr
3 **TU Dresden, DE.** peter.birkholz@tu-dresden.de
4 **Ramon Llull University – Barcelona, ES.** oriol.guasch@salle.url.edu
5 **University of York, GB.** amelia.gully@york.ac.uk

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 24242 "Computational Analysis and Simulation of the Human Voice", which was held from the 9th to the 14th of June, 2024. The seminar addressed key issues for a better understanding of the human voice by focusing on four main areas: voice analysis, visualisation techniques, simulation methods, and data analysis with machine learning. There has been enormous progress in recent years in all these fields. The seminar brought together a number of experts from fields as diverse as computer science, logopedics and phoniatrics, clinicians, acoustics and audio engineering, electronics, musicology, speech and hearing sciences, physics and mathematics. The schedule was quite flexible, including inspirational talks in the main areas, interactive working groups, sharing of conclusions and discussions, presentation of successes and failures to learn from, and a large number of free talks that emerged throughout the days. The variety of topics and participants created a highly enriching environment from which novel proposals for future research and collaboration emerged, as well as the collective writing of a paper on the state of the art and future perspectives in human voice research.

Seminar June 9–14, 2024 – https://www.dagstuhl.de/24242

2012 ACM Subject Classification Human-centered computing → Sound-based input / output; Human-centered computing → Auditory feedback; Human-centered computing → Visualization theory, concepts and paradigms; Computing methodologies → Speech recognition; Applied computing → Molecular structural biology; Applied computing → Health informatics; Applied computing → Performing arts; Applied computing → Sound and music computing; Applied computing → Physics

Keywords and phrases voice science, voice analysis, voice simulation, visualization, big data, machine learning, clinical voice treatment

Digital Object Identifier 10.4230/DagRep.14.6.84

* Editor / Organizer
† Editorial Assistant / Collector

Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

1 Executive Summary

Sten Ternström
Nathalie Henrich Bernardoni
Oriol Guasch
Peter Birkholz

License Creative Commons BY 4.0 International license
© Sten Ternström, Nathalie Henrich Bernardoni, Oriol Guasch, and Peter Birkholz

The human voice is able to produce a very rich set of different sounds, making it the single most important channel for communication human-to-human, and also potentially for human-computer interaction. Spoken communication can be thought of as a stack of layered transport protocols that includes language, speech, voice, and sound. This Dagstuhl Seminar was concerned with the voice and its function as a transducer from neurally encoded speech patterns to sound. This very complex mechanism remains insufficiently explained both in terms of analysing voice sounds, as for example in medical assessment of vocal function, and of simulating them from first principles, as in talking or singing machines. There were four main themes to the seminar:

Voice Analysis. Measures derived from voice recordings are clinically attractive, being non-invasive and relatively inexpensive. For clinical voice assessment, however, quantitative objective measures of vocal status have been researched for some seven decades, yet perceptual assessment by listening is still the dominating method. Isolating the properties of a voice (the machine) from those of its owner's speech or singing (the process) is far from trivial. Computational approaches are expected to facilitate a functional decomposition that can advance beyond conventional cut-off values of metrics and indices.

Voice Visualization. Trained listeners can deduce some of what is going on in the larynx and the vocal tract, but we cannot easily see it or document it. The multidimensionality of the voice poses interesting challenges to the making of effective visualizations. Most current visualizations are textbook transforms of the acoustic signal, but they are not as clinically or pedagogically relevant as they could be. Can functionally or perceptually informed visualizations improve on this situation?

Voice Simulation. Balancing low- and high-order models. A "complete" physics-based computational model of the voice organ would have to account for bidirectional energy exchange between fluids and moving structures at high temporal and spatial resolutions, in 3D. Computational brute force is still not an option to represents voice production in all its complexity, and a proper balance between high and low order approaches has to be found. We discussed strategies for choosing effective partitionings or hybrids of the simulation tasks that could be suitable for specific sub-problems.

Data science and voice research. With today's machine learning and deep neural network methods, end-to-end systems for both text-to-speech and speech recognition have become remarkably successful, but they remain quite ignorant of the basics of vocal function. Yet machine learning and big data science approaches should be very useful for helping us deal with and account for the variability in voices. Rather than seeking for automated discrimination between normal and pathological voice, clinicians wish for objective assessments of the progress of an intervention, while researchers wish for ways to distil succinct models of voice production from multi-modal big-data observations. We have explored how techniques such as domain-specific feature selection and auto-encoding can make progress toward these goals.

This seminar has resulted in (1) shared knowledge and data about the science of voice from the perspectives of scientists in fields as diverse as computer science, voice pathology and therapy, clinicians, acoustics and audio engineering, electronics, musicology, speech and hearing sciences, physics and mathematics, (2) identifying areas of common interest where significant progress is being made and needs to be made, such as individual voice variability, physical replicas for modelling and validation, synthesis and computational modelling, motor control, and availability of data and resources, (3) sharing and discussing failures to learn lessons and ideas for future developments, and (4) envisioning the future of progress in human voice analysis and simulation in the medium to long term: what is needed to make a big leap forward in this field? These ideas will be captured by the publication of a collaborative article in a leading voice journal.

2 Table of Contents

3 Inspirational Talks

For each of the four main topics of the seminar, a participant was invited to prepare an inspirational talk of about 30 minutes.

3.1 Inspirational Talk on Analysis: Voice analysis: looking at the subsystems of the vocal apparatus

Filipa M. B. Lã (UNED – Madrid, ES)

License Creative Commons BY 4.0 International license
© Filipa M. B. Lã

This presentation shows examples of voice analysis that are currently available in the voice studio to assist the teaching of singers. It also sets the ground for discussing possibilities to simplify current recording setups and voice analysis, especially when the aim is to describe and understand not only acoustical but also physiological and aerodynamic aspects of voice production that are relevant in voice education.

Discussion. Discussion took place across five groups. Group 1 emphasized that averages are not reliable for representing data like vibrato, and that SPL significantly impacts measurements. Group 2 stressed the need for organizing and simplifying tools for studio and clinical use, noting the overwhelming nature of existing tools. Group 3 discussed the potential of simulations to isolate and understand specific signals, suggesting a forum for sharing tools and methods. Group 4 pointed out the challenges in translating measurements into actionable information, the need for benchmark data, and standardization. Group 5 explored the potential of AI and visualization to enhance feedback and teaching methods, suggesting that additional or combined metrics might improve current practices. Overall, the discussions underscored the importance of better data representation, tool simplification, standardization, and the integration of advanced technologies in acoustic analysis of the voice.

3.2 Inspirational Talk on Visualisation: Collaboration between Domain Scientists and Visualization Experts

Tino Weinkauf (KTH Royal Institute of Technology – Stockholm, SE)

License Creative Commons BY 4.0 International license
© Tino Weinkauf

Visualization and data analysis have become increasingly integral components of research workflows in all academic disciplines. Data is sourced from experiments, sensors, modeling, simulations, data repositories, and other means, making it a ubiquitous presence in the scientific landscape. Scientific progress heavily relies on comprehending this data. Across all domains, scientists face growing challenges in data analysis, such as dealing with massive datasets, high dimensionality, noisy data, and intricate complexity. Addressing these challenges often extends beyond the expertise of the respective scientists, and rather requires specialized knowledge in data analysis and visualization. Hence, it is necessary to develop a joint language between visualization experts and domain scientists with the goal of clearly expressing the visualization tasks and the properties of the data. This talk investigates the opportunities and potential pitfalls in these regards, and gives practical advice on how to facilitate the interdisciplinary work.

Discussion. The discussion took place in 4 groups. Group 1 discussed how spectrograms were a breakthrough and that modern visualisation techniques offer a similar paradigm shift. They also noted the need for context-specific visualization tools. Group 2 emphasized the importance of domain knowledge for effective visualization, the potential role of AI, and the need for a repository of visualization tools. Group 3 focused on the value of visualization in distinguishing noise from meaningful data, understanding AI findings, and exploring data. Group 4 considered how visualization can reveal unknowns, facilitate discussions, and present data in new narrative forms, stressing the importance of multi-modal understanding. Across the groups, there was a call for developing a domain-specific visualization toolkit and addressing issues like data reduction and meaningful representation.

3.3 Inspirational Talk on Simulation: Articulatory Speech Synthesis: Modelling and Simulation – or – Can we make a voice instrument?

Sidney Fels (University of British Columbia – Vancouver, CA)

License 🅭 Creative Commons BY 4.0 International license
© Sidney Fels

In this inspirational talk, I make the case that the goal of creating a human controlled voice instrument embodies many of the major challenges in speech modeling and synthesis research. Starting with a inspiration drawn from the successes and shortcomings of various voice instruments of the past, like Bell's skull-based actuator, the Vodor and my own Glove-TalkII instrument, I discuss the driving aspects of my own research on biomechanical modelling of the human vocal tract, forays into making a real-time speech synthesis engine and human control of complex systems. The first shortcoming in Glove-TalkII was that the control space used formant which led to a difficult mapping between the relatively slow hand gesture movements and the faster dynamics of the formant space. The thinking was, if we could build an articulatory speech synthesizer that uses synthetic muscle activations, rather than kinematics, the mapping between hand gestures to simulated muscles would be easier to learn for the person and neural networks. This led to, first, creating a computational, muscle activated, hybrid rigid-body/FEM biomechanical model of the human vocal tract and the necessary modelling and simulation infrastructure, Artisynth.com, suitable for simulating the complexity of the human vocal tract dynamics in real time. Second, the geometry of the simulated 3D vocal tract models the dynamic 3D geometry of the human vocal tract, thus, like a number of researchers in our field, we are working on generating the voice acoustics from the 3D geometry of the vocal tract. We require real-time performance, hence, 3D FEM approach are not suitable. Plus, we need the full range of vocal sounds, including frication and stops, leading to our 2.5D approach rather than modification of the 1D wave equation. Finally, we have been investigating how the human brain represents and performs speech motor control to discover ways to reduce the information requirements needed for control. Our thinking is that the human motor control system takes advantage of constraints in the physical world to reduce the needed degrees of freedom to make takes have a reduce index of difficulty, thus, reducing the information load for control. Thus, the real-time articulatory speech synthesis engine embodies the physics of the world that can be used by a machine learned adaptive mapping and human motor control system to make the voice instrument easier to play, whether by brain signals or gesture. During the talk, I cover how much progress has been done by us and others and how much is left to do. Finally, I conclude with the assertion that creating a new voice instrument for human vocal expression pushes limits of knowledge for modeling, simulation and control of voice and speech.

Discussion. The discussion took place across five groups. Group 1 considered the challenges in defining constraints, and wondered how much knowledge of soft tissue biomechanics is required for appropriate emergent behaviour. The interaction between different systems with different levels of detail was considered a very positive step. Group 2 explored what makes a musical instrument expressive, emphasizing effort, perceiver presence, control, and the creation and perception of time patterns, linking pleasure in performance to flow and transcendence. Group 3 discussed the balance between passion projects and career sustainability, the need to contribute to a larger goal, and the implications of the model presented for training, healthcare, and physiology. Group 4 highlighted the prediction of surgical outcomes, the role of auditory feedback, and the interplay between speech production and perception, including the importance of visual aspects. Group 5 focused on the reasons for pursuing expressive instruments, clinical applications, patient-specific simulations, and the historical significance of voice models, noting the need for storytelling to secure support and funding.

3.4 Inspirational Talk on Data Science and ML: Data Science and Machine Learning on Clinical Applicability of Voice Studies

Pedro Gómez-Vilda (NeuSpeLab – Las Rozas de Madrid, ES)

License Creative Commons BY 4.0 International license
© Pedro Gómez-Vilda

Machine Learning (ML), Deep Neural Network (DNN) architectures, and End-to-End Systems (EES) for both text-to-speech and speech recognition have become remarkably successful, but they remain quite agnostic of the basics of neurological foundations of vocal and speech function. Nevertheless, ML, DNNs and EES are very useful for dealing with pathological speech characterization, either organic, functional, or neurogenic. Rather than seeking for automated discrimination between normal and pathological voice, clinicians wish for objective assessments of the progress of an intervention. AI researchers, on the other hand, look for ways to distil predictive scores about voice pathology from multi-modal big-data observations. A comprehensive search on domain-specific shallow architectures having an impact toward these goals is nowadays the Holy Grail of ML.

Discussion. The discussions took place in four groups. All groups reported some variation of the idea that AI should assist, rather than replace, human decision-making, since AI tools are adept at distinguishing patterns but should lack human expertise and intuition. Group 1 questioned the clinical relevance of binary (pathological / non-pathological) classification, and noted the use of AI tools for learning something about features of relevance. Group 2 discussed an analogy with cars, suggesting something akin to a driving licence for using AI – we don't need to know exactly how cars work, but we do need to follow certain rules to ensure everyone's safety. Group 3 pointed out that the humans using the AI are the major concern, and the risks of AI amplifying errors due to human misuse, also suggesting for certifications for AI users. Group 4 discussed the task-sensitive nature of AI, its role in simplifying clinical decisions, and the varying levels of trust in AI among clinicians and patients. An additional point was raised about the concentration of power and resources required for big data, highlighting the need for ethical considerations.

4 Free papers

Participants were invited to volunteer presentations without constraints, which resulted in ten free talks of 10-15 minutes each. The first three free talks were given on day 2 and the remaining ones on day 4.

4.1 Revisiting Laver's voice qualities

Philip Aichinger (Medizinische Universität Wien, AT)

License Creative Commons BY 4.0 International license
© Philip Aichinger

Voice quality is a relevant feature in healthy as well as pathological speech. In particular, the use of different voice qualities in connected speech may be impeded in pathological speakers. Laver had defined phonatory settings relating to different voice qualities in terms of longitudinal tension, medial compression, and adductive tension [1]. This enabled modal, whispery, breathy, creaky, and harsh voice, as well as falsetto. A laryngeal high-speed video synthesizer [2] is extended here to visualize both the cartilaginous and the muscular parts of the glottis. A few example videos showing different voice qualities are presented.

References

1 J. Laver, *The Phonetic Description of Voice Quality*. Cambridge University Press, 2009.
2 P. Aichinger; S. Kumar; S. Lehoux; J. Švec, *Simulated Laryngeal High-Speed Videos for the Study of Normal and Dysphonic Vocal Fold Vibration*, J. Speech, Lang. Hear. Res, 65:7, 2431–2445, 2022.

4.2 Voice conversion approaches to synthetic substitution voices

Philip Aichinger (Medizinische Universität Wien, AT)

License Creative Commons BY 4.0 International license
© Philip Aichinger

Communication disorders, including speech pathologies, have a 12-month incidence of approximately 10%, varying in severity. In severe, persistent cases, these disorders can significantly impact quality of life, potentially leading to social isolation. Assistive devices such as speech synthesizers can restore communication during and after rehabilitation. However, their use is often limited by difficulties in controlling them and dissatisfaction with the voice sound. Voice conversion using deep learning offers a new approach by emulating a target speaker's identity based on a reference microphone recording. Essentially, the user speaks normally, and the voice converter outputs improved speech. This study aims to assess the quality of speech output by voice converters, focusing on perceived voice quality improvement while monitoring potential adverse effects on intelligibility.

Test material includes speech audio of pathological speakers and a healthy speaker using an electrolarynx. A few examples of speech converted to normal are presented. In conclusion, current advancements in voice conversion technology promise improvements in output quality and reduced latency, making voice conversion technology a potential game-changer in the field of speech-assisting devices, especially in telecom applications.

4.3 Sampling rate bias of vocal cycle perturbations

Jean Schoentgen (Free University of Brussels, BE)

License 🅭 Creative Commons BY 4.0 International license
© Jean Schoentgen

Acoustic features that report vocal cycle length or vocal cycle amplitude perturbations are biased when they involve length or amplitude data that are sampled at the pace of the vocal cycles. The reason of the bias is that the features involve high-pass filtering of the cycle lengths or amplitudes to separate fast (i.e. jitter or shimmer) from slow perturbations followed by averaging the magnitude of the filtered data. Indeed, the cut-off frequency of a digital filter depends on the sampling frequency that is equal to the vocal frequency (f_0) when the speech signal is sampled once every cycle. The cutoff frequency above which the vocal perturbations are reported therefore depends on the vocal frequency. As a consequence, the jitter or shimmer bandwidth is token-dependent loose from any physiological or anatomical causes. For the same reason, the vocal perturbation bandwidth evolves within a same token when the intonation is not flat. This difficulty concerns most of the known features that report vocal jitter or shimmer. Typically, these features are obtained by means of popular analysis software such as MDVP or PRAAT.

We have compared five biased features that are reported by PRAAT and that describe the jitter of cycle lengths sampled at the rate of fo to one unbiased feature describing the jitter of cycle lengths sampled at a fixed rate. The purpose of the comparison has been to examine whether the dispersion of the feature values within a corpus of speech tokens differs for fixed and variable rate features as well as whether variable-rate features may be substituted for fixed-rate features. A second topic has been the correlation of the magnitude and frequency of vocal jitter with vocal frequency f_0.

4.4 Inverse filtering

Johan Sundberg (KTH Royal Institute of Technology – Stockholm, SE)

License 🅭 Creative Commons BY 4.0 International license
© Johan Sundberg

This talk presented the finding that after inverse filtering, harmonic partials located at or close to $F1$ are reduced in amplitude, likening the effect to the partial "doing the limbo" under the formant. A discussion then ensued about the cause of this observation. In the discussion it was noted that the inertic part of the impedance rises with frequency towards the peak of a resonance, and then drops rapidly, becoming compliant at the resonance frequency itself. The fact that this occurs over a wider frequency range than just at the resonance frequency is due to the wider bandwidths associated with real vocal tracts. This suggestion is borne out by the observation that the harmonic amplitude increases slightly with the inertance in the frequency region below the resonance.

4.5 Model based speech research

Brad Story (University of Arizona – Tucson, US)

License Creative Commons BY 4.0 International license
© Brad Story

During the production of speech, a talker-specific, baseline configuration of the airway is modulated almost continuously by the movements of the tongue, jaw, lips, velum, and larynx. From an acoustic perspective, the airway can be considered a non-uniform conduit whose shape at a given instant of time supports a specific pattern of acoustic resonances that transmit information related to both the intended message and the identity of the talker. This presentation begins with recollection of attempts to simulate connected speech about 30 years, and then summarizes the recent development of a model in which individual speech segments that comprise a word, phrase, or sentence are specified as relative deflections of the resonance frequencies of the baseline vocal tract configuration, and then transformed to time-dependent modulations of the airway. The output of the model is artificial speech that can be presented to listeners. Examples will demonstrate the construction of speech with the model, results from a recent perceptual experiment, effects that may occur with constraints imposed on the vocal tract, and engage the audience in an interactive listening experience. [Research supported in part by NIH 5R01DC017998 and Galileo Circle Fellows grant from the University of Arizona.]

4.6 The Laryngonaut

Scott Reid Moisik (Nanyang TU – Singapore, SG)

License Creative Commons BY 4.0 International license
© Scott Reid Moisik

Laryngeal simulation and visualization overwhelmingly emphasize the vocal folds. The structures of the larynx above the level of the vocal folds, known collectively as the epilarynx (which includes the ventricular folds, aryepiglottic folds, and epiglottis), are often removed to provide a better view of the vocal folds below, despite the important contribution of epilaryngeal structures to overall laryngeal behaviour in speech, singing, and life-supporting functions. In this work, I discuss a broadly targeted phonetics-oriented laryngeal simulation platform that provides interactive 3D visualization and sound synthesis. I will discuss how the work draws on a range of numerical methods for fast simulation of the physics of rigid and deformable bodies, representing the laryngeal structures, and a simplified 1D aeroacoustic model coupled to Titze's (1973) dual particle-chain model for vocal fold vibration, all with the express purpose of real-time responsiveness to user interaction. The entire larynx is represented, including the often-neglected structures of the epilarynx, which, along with the vocal folds, undergo natural-looking deformations in response to changes of muscle activity and virtual manipulation of the structures. With this system, it is possible to explore a range of laryngeal states which map onto important phonetic categories (including various phonation types, such as creaky voice, and articulations, such as epiglottal stop) while simultaneously generating an appropriate approximation of the sound output expected for the current laryngeal state. While considerable progress has been made, much still remains, and some future plans will be discussed, such as extending the system to provide a simulation of the various types of epilaryngeal vibration (such as aryepiglottic trilling).

4.7 Pressure in, SPL out (where is the flow?)

Peter Pabon (Utrecht, NL)

License Creative Commons BY 4.0 International license
© Peter Pabon

Variation seen in acoustic voice quality measurement is often seen as an inevitable random factor happening along the time line. However, bringing in the SPL- and fo-based localisation principle of voice mapping refutes the notion that a voice can be characterized with some single representative quality value with some inevitable and uncontrollable variation. Rather, metrics exibit much less intra-subject variation when mapped against SPL and fo, while inter-subject differences prove to be larger than expected. The localising principle rests on (1) a calibrated SPL scale (dual mic headset, constant microphone distance checking and calibration), (2) pitch-period synced information processing and metrics sampled in sync, (3) clean, unbiased, preferable salient metrics on log scales (no hidden contraints by hard links to axes) – "where does this voice hit its own constraints?" – limited control space; and (4) by not allowing disinformation, that is, include information (including noise) from the voice only.

References

1 Pabon, P., and Ternström, S. (2020). "Feature Maps of the Acoustic Spectrum of the Voice," J. Voice, 34, 161.e1-161.e26. doi:10.1016/j.jvoice.2018.08.014

4.8 First steps to a pacemaker for phonation?

Oriol Guasch (Ramon Llul University – Barcelona, ES)

License Creative Commons BY 4.0 International license
© Oriol Guasch

We present first ideas about the possibility of building a pacemaker for phonation. Vocal fold mass models show that small changes in muscle restoring forces, such as excessive stiffness typical of Parkinson's patients, or too much subglottal pressure, can cause chaotic oscillations of the VF, resulting in abnormal glottal volume flow. Regular oscillations could be restored by a pacemaker made of a smart material with adjustable damping to control chaos [1]. To evaluate the effectiveness of the pacemaker on voiced sounds, the glottal volume velocity for normal, chaotic, and controlled vocal fold oscillations can be calculated and convolved with MRI-derived vocal tract impulse responses for the vowels /a/, /i/, and /u/ [2]. Audiovisual files in spectral and temporal analysis show that chaotic oscillations significantly distort vowel sounds, but the control strategy restores them to normal.

References

1 O Guasch; A. Van Hirtum; A.I Fernández; M. Arnela, *Controlling chaotic oscillations in a symmetric two-mass model of the vocal folds*, Chaos Solit. Fractals, 159, 112188, 2022.
2 O Guasch; M. Freixes: M. Arnela; A. Van Hirtum, *Controlling chaotic vocal fold oscillations in the numerical production of vowel sounds*, Chaos Solit. Fractals, 182, 114740, 2024.

4.9 Formant tuning of 3D vocal tracts for FEM vowel synthesis

Marc Arnela (Ramon Llul University, ES)

License Creative Commons BY 4.0 International license
© Marc Arnela

This talk introduces a methodology for tuning formants in 3D vocal tract geometries obtained via MRI to produce specific voice or singing effects. The process involves converting 3D MRI-based geometries into 1D area functions, iteratively adjusting these functions with an algorithm based in sensitivity functions, and reconstructing the 3D vocal tract with modified cross-sections. This method enables the tuning of vowel formants while maintaining the high energy spectrum of 3D models at low computational cost. Examples include shifting the first formant ($F1$) and generating formant clusters, such as those found in singing.

References

1. O Guasch; M. Arnela; A. Pont, *Resonance tuning in vocal tract acoustics from modal perturbation analysis instead of nonlinear radiation pressure*, J. Sound Vib., 493, 115826, 2021.
2. M. Arnela; O Guasch, *Formant frequency tuning of three-dimensional MRI-based vocal tracts for the finite element synthesis of vowels*, IEEE/ACM Trans. Audio Speech Lang. Process, 32, 2790–2799, 2024.

4.10 Development of an AI assisted data assimilation framework for voice research and clinics

Qian Xue (Rochester Institute of Technology, US)

License Creative Commons BY 4.0 International license
© Qian Xue

Capturing real-time high-resolution 3D images of tissue dynamics remains a challenge due to several inherent factors including organ accessibility and low temporal/ spatial resolution of imaging and image reconstruction. This study introduces a novel hybrid physics-informed neural network (PINN) differentiable learning algorithm that integrates a recurrent neural network model of 3D continuum soft tissue with a differentiable fluid solver to infer 3D flow-induced tissue dynamics and other physical quantities from sparse 2D images. To enhance the scalability and convergence, the designed algorithm leverages the prior knowledge in solid mechanics by projecting the governing equation onto the numerical eigenmode space, reducing the infinite dimensions of the continuous solution space to a finite dimension of discrete search space. The dimensions of the problem are further reduced by only using truncated eigenmodes, which can effectively represent the whole dynamics with negligible errors. To better capture the temporal dependence of flow-structure interaction (FSI) dynamics and enhance the predictive accuracy, a Long short-term memory (LSTM)-based recurrent encoder-decoder connected with a fully connected neural network (FCNN) is designed to learn the time history of modal coefficients, which, combined with eigenmodes, enable the spatiotemporal predictions of tissue dynamics. The effectiveness and merit of the proposed algorithm is demonstrated in subject-specific models by using synthetic data from a canine vocal fold model and experimental data from four excised pigeon syringes. The results showed that, by only using sparse 2D vibration profiles, the algorithm was able to accurately reconstruct the full 3D tissue dynamics as well as other high-dimensional physical quantities, such as aerodynamics and acoustics quantities, which are otherwise very difficult/impossible to measure. The algorithm can advance disease diagnosis beyond the current morphological and 2D dynamics criterion. It also allows a significant expansion of the measurable quantities in both experimental/clinical research, which could broadly enhance biomedical research capabilities.

5 Working Groups

5.1 Spontaneous workshops

On the first day of the seminar, a whiteboard poll was conducted for identifying the topics of greatest interest to the participants. Working groups were then formed around these topics:

- Individual variability in the voice
- Physical replicas for modelling and validation
- Synthesis and computational modelling
- Motor control
- Data and resources

In a breakout session, the groups discussed these topics for about 30 minutes and then returned for a plenary discussion.

5.1.1 Individual variability in the voice

The first two groups focused on individual variability, considering sources of variation, and research questions around voice pathologies and how to compare voices. There was also discussion of how voices vary from moment to moment, leading to questions about robustness and the validation of variability measurements. Longitudinal data (comparing an individual to themselves) was preferred over normative values, which have little value at the individual level.

5.1.2 Physical replicas for modelling and validation

The discussion of physical replicas considered their purpose (primarily for validating computational models, but also to understand voice production when existing computational models break down, such as during fricatives or with large deformations). The current issues and needs in this area were identified as being the choice of suitable materials (e.g. active materials), reproducibility, and a question of how much detail is necessary.

5.1.3 Synthesis and computational modelling

The discussion of synthesis focused on how much detail is "enough" in simulation, concluding that there are several very different cases where simulation is used, for which the modelling cost varies greatly. A spectrum exists, from "drastically simplified but real-time", via "perceptually sufficient but physically lacking", to "detailed reconstructions of the physics taking many hours to compute". There are considerable issues around validation, but the data necessary for this is lacking, and depends to a large extent how detailed the simulation method is. Nevertheless it was agreed that some benchmark data could be agreed – if available – to further this goal.

5.1.4 Motor control

Finally, the discussion of motor control noted how – including here! – simulation considerations are often decoupled from the idea of motor control, and suggested a future where the two could be considered together. It remains unclear what control variables the system is actually working with, which is a priority for future research, and may not relate to the variables currently used for simulation. The group also highlighted how perception should be included in models of voice motor control, as the message passes from one brain to another by way of muscle movement.

6 Workshops and talks on dreams and failures

The idea of these workshops and talks was to explore the journey of innovation in voice science through three key questions, each addressed in a sub-section. *What do you need to make a leap?* focuses on the essential elements and conditions required for significant advancements in the field. By understanding these factors, we can better navigate the complexities of research and development. *My favorite failure* reflects on setbacks that provided invaluable lessons and insights for future developments. Failures often pave the way for breakthroughs, making them crucial for sustained innovation. Finally, *My dream come true* envisions the future of progress in human voice analysis and simulation, setting ambitious goals for the years ahead.

The workshops were organised around spontaneous presentations followed by discussions, with the aim of encouraging participants to dream big and push the boundaries of what is currently possible in voice science.

6.1 What would you need to make a leap?

This was a one-hour break-out session with group discussions, summarized *in plenum*.

Summary of Discussions. The five discussion groups presented various ideas and desires to advance voice science and technology. Group 1 proposed creating a 100M€ voice center, developing non-invasive physiological signal collection systems, and employing a "voice influencer" to fund research. They also envisioned a robot head with soft tissue activation and advanced imaging systems for vocal tract shapes. Group 2 focused on AI-driven data analysis, distance voice teaching with machine learning, and portable devices for comprehensive physiological monitoring. Group 3 aimed for a digital twin of the human body, systems visualizing interconnected data, predictive tools, and general models that can be personalized. Group 4 sought real-time imaging of muscle activation, non-invasive measurement of subglottal pressure, and pacemakers for vocal folds. Group 5 emphasized the need for funding, advanced measurement techniques, comprehensive computational models, and collaborative efforts. They stressed the importance of storytelling in securing funding, the need for standardized data-sharing protocols, and a community-driven approach to set priorities and justify research efforts.

6.2 My favourite failure: Learning from failures and outliers: "Oh my... that was embarrassing..."

Eric J Hunter (University of Iowa, US)

License Creative Commons BY 4.0 International license
© Eric J Hunter

Isaac Asimov said, "The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' but rather 'hmm... that's funny...'". In its simplest form, the practice of science can be thought of as learning from our mistakes and noticing those things that don't quite fit theory. While we work to approach research rationally, avoiding dogmatic opinions, and remaining open to opposing views, we really can never claim absolute truth yet always in pursuit of it. Applying the scientific method involves objective analysis of issues, recognizing that induction plays a role, but the true method for advancing knowledge is often filled outliers, oddities, and even some immense and embarrassing failures. Yet, learning from these is integral to both scientific progress and personal growth. This presentation includes a few of my stories.

6.3 My favourite failure: can subglottal pressure be estimated from intra-oral pressure in speech and singing?

Nathalie Henrich Bernardoni (Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, FR)

License Creative Commons BY 4.0 International license
© Nathalie Henrich Bernardoni

The direct measurement of subglottal pressure is challenging, because it requires a very invasive approach. It consists in placing a pressure transducer below the glottis by tracheal puncture between the cricoid cartilage and the trachea first ring. Other methods have been proposed, which estimate the subglottal pressure using less invasive approaches. The most common one is to estimate subglottal pressure from intra-oral pressure measured during the closed phase of voiceless consonants (Smitheran and Hixon, 1981; Hertegard et al., 1995). This approach has been validated in the case of normal speaking voice. However, few studies have explored its validity for soft or loud voice, for whisper or pressed voice, and in the case of singing. This study explores the possibilities and limitations of estimating subglottal pressure from intra-oral pressure in speech and singing. Two subjects (a speaker and a trained singer) were recorded while uttering CV segments (plosive consonant followed by a vowel) with different voice qualities (normal, soft, loud, whisper, pressed). The singer sung sentences at several pitches covering his comfortable tessitura in the two main laryngeal mechanisms. Two recording sessions were conducted with a one-year time interval in between. Several methods for estimating subglottal pressure from intra-oral pressure signal are compared. A good agreement between estimates and direct measures is found in many cases. Bad agreement may be found in the case of soft phonation, for productions in laryngeal mechanism M2, and in the case of pressed speech. We thought a lot about the reasons for these discrepancies, and found that they may be due to the absence of a nose clip during measurements. The conclusion of this favourite failure is that it is of much importance to be concerned about the nose in speech and singing production, even in cases where it should not interfere . . .

References

1 Smitheran JR, Hixon TJ. (1981) A clinical method for estimating laryngeal airway resistance during vowel production. J Speech Hear Disord. 46(2):138-46.

2 Hertegård S, Gauffin J, Lindestad PA. (1995) A comparison of subglottal and intraoral pressure measurements during phonation. J Voice 9(2):149-55.

6.4 My favourite failure: Spontaneous confessions

Sten Ternström (KTH Royal Institute of Technology – Stockholm, SE)

License 🅭 Creative Commons BY 4.0 International license
© Sten Ternström

In addition to the above accounts, there were a few spontaneous and amusing participant “confessions” to failures.

Jean Schoentgen related how he had once been trying to synthesise rough voices, with some success – they sounded rough and reasonably natural. Then he had the “good bad idea” to apply the general knowledge to include soft f_o contours that decrease 10-30% over 1 s. But then the roughness disappeared! He thought it was an error, but... information in the frequency contour is in the spectral side-bands – and so is the roughness. Somehow one information masked the other. If this is correct / relevant – then if you are looking for voice quality in connected speech, don't look for roughness in there.

Sten Ternström described how for the Singing Synthesis Competition at Interspeech 2007 in Antwerpen he had prepared a source-filter synthesis, working with headphones with a soprano voice in the left ear only. But on playback over loudspeakers at the venue, the soprano was too shrill, resulting in a low rating in the competition. Sten had forgotten to account for an imbalance in his own hearing, which has a dip at 3-5 kHz in the left ear. If he had tried reversing the headphones, or first played it to someone else, the competing entry might have done better!

Peter Pabon recounted how he once had been on the brink of a very high-profile failure, on a live television show in the Netherlands. He had been tasked to break a huge crystal glass, one meter high, by catastrophic resonance, which involves matching a strong loudspeaker tone very precisely to the previously measured eigenfrequency of the glass. The glass was vibrating vigorously, but refused to break. Then Peter noticed that the spectrum analyzer of the audio was showing a slightly different frequency – the tone had been shifted by the studio's public-address system, as a counter-feedback measure. He tweaked the frequency just that little, and the glass broke, spectacularly. And the shooter with an air-gun, waiting in the wings, did not have to fake it after all.

6.5 My dream come true: the story of Pinocchio

Nathalie Henrich Bernardoni Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, FR)

License 🅭 Creative Commons BY 4.0 International license
© Nathalie Henrich Bernardoni

Speech production results from a fine neuro-controlled coordination between breathing, phonatory and articulatory movements. In physical terms, these movements induce fluid-structure-acoustic interactions within the vocal tract. In the framework of source-filter theory (Fant 1960), the aerodynamic phase of speech is often overlooked, yet being of much importance for voiced and unvoiced speech sound production (Catford 1977). In the case of voiced speech sound, it is taken into account by myoelastic-aerodynamic theory of phonation (Jan G. Švec et al. 2021), in which glottal constriction and vocal folds vibration generate aeroacoustic sources. Our aim is to advance our understanding of speech sound production by developing a biomimetic in vitro test-bed. Over the last thirty years, vocal-fold testbeds have evolved in complexity and biomimicry. However, most testbeds explore the physics of phonation on geometrically-fixed replicas capable of self-sustained oscillations in fluid-structure interaction, but unable to produce intonative variations. Most of the time, the testbeds are not coupled to vocal tract, and whenever they are, the resonant cavities are static 3D-printed tracts. Reproducing in vitro the dynamic movement of speech articulators, such as jaw, tongue, velum and larynx, together with phonation, remains a challenge. We present here the first steps in the design of a biomimetic mechatronic testbed that would integrate all phonatory and articulatory aspects important for voice and speech production.

6.6 My dream come true: Design of self-oscillating biomimetic vocal folds

Lucie Bailly (Université Grenoble Alpes – Saint Martin d'Hères, FR)

License 🅭 Creative Commons BY 4.0 International license
© Lucie Bailly

This work aims to contribute to an in-depth understanding of the link between the histo-mechanical properties of the vocal folds and their remarkable vibratory performances, which still remains elusive. In vitro simulators of the phonatory system have been developed for decades. Yet, most of them are made of materials with structural and mechanical properties still far from those of the native tissues. Thus, this work proposes to design new in vitro self-oscillating biomimetic vocal folds with tailored structural and mechanical properties, and study the impact of material properties on the fluid/structure/acoustics interactions driving phonation. To this end, three complementary families of materials were studied: gelatin-based hydrogels cross-linked with glutaraldehyde, biocompatible polyethylene glycol-based hydrogels and silicone elastomers used in voice research. Formulations were optimised to best fit the mechanics of vocal tissues under physiological multiaxial loadings in tension, compression and shear. Fibre-reinforced composites were then produced to mimic the microscale collagen structure of tissues, and their macroscale non-linear and anisotropic behaviour. Finally, the ability of the tailored materials to vibrate under realistic morphological and aerodynamic conditions was tested using an articulated larynx test-bed. Promising vibratory patterns were obtained for the optimised hydrogel-based replicas, able to self-oscillate with subglottal pressures close to physiological reality. The data have evidenced the link between the material properties and the aeroacoustic performances of several synthetic oscillators, paving the way for future investigation of the impact of the fibrous architecture of vocal tissue on voice quality.

6.7 My dream come true: Discussion on particle-based simulations

Peter Birkholz (TU Dresden, DE)

License Creative Commons BY 4.0 International license
© Peter Birkholz

A short discussion followed on the possibility of particle-based simulation of the aeroacoustics of the vocal system. Some outline calculations suggested that the computational complexity was feasible at a suggested density of 1000 particles/mm^3. Discussion indicated the principle was the same as the lattice Boltzmann method and highlighted the necessity for more complex simulations with more particles for accurate modelling. It was suggested that high-performance computing resources would be suitable for this kind of task.

7 Workshop on the elusive Inversion Problem

An age-old problem with studying the voice organ is that is difficult to access experimentally. Much effort continues to be directed to the inversion problem, that is, to inferring what the voice organ is doing with access only to external signals. This is closely related to acquisition technologies for biophysical data, which are in rapid development. There were two invited talks (Laprie, Birkholz) and one contributed paper (Zhang).

7.1 Acoustic-articulatory inversion with rtMRI

Yves Laprie (LORIA – Nancy, FR)

License Creative Commons BY 4.0 International license
© Yves Laprie

The use of real-time MRI (rt-MRI) data is dramatically changing the way acoustic-to-articulatory inversion can be addressed. In general, inversion is based on electro-magnetographic (EMA) data, but they provide a very limited information in the form of just a few points. rt-MRI data have the advantage of covering the entire vocal tract from glottis to lips, and of enable larger databases for each speaker. On the disadvantage side, rt-MRI data need to be pre-processed to track articulator contours, and the speech signal needs to be denoised. We have developed effective solutions to both these problems. It is also necessary to be able to move from the speech signal acquired in a low-noise environment in the sitting or standing position to the MRI signal produced in a strong noise in the supine position. The inversion itself can be carried out using deep learning techniques, inspired by those already used for EMA-based work.

7.2 Two examples using VocalTractLab and neural networks

Peter Birkholz (TU Dresden, DE)

License 🅭 Creative Commons BY 4.0 International license
© Peter Birkholz

The inverse problem tries to estimate parameters of the speech production process from the speech audio signal. In this presentation I show how the articulatory speech synthesizer VocalTractLab (www.vocaltractlab.de) was used to generate synthetic data to train artificial neural networks (ANN) to solve two kinds of inversion problems: the estimation of the glottal flow and the estimation of the articulatory movements from the speech audio signal. In both cases, the ANNs trained on purely synthetically generated data showed a promising performance when used on human speech signals. Directions for future research in this direction are discussed.

7.3 Inversion in voice production

Zhaoyan Zhang (UCLA, US)

License 🅭 Creative Commons BY 4.0 International license
© Zhaoyan Zhang

In this talk, I will present a simulation-based machine learning model toward solving the inverse problem in voice production, i.e., to estimate vocal fold geometry, stiffness, position, and subglottal pressure from the produced voice acoustics and aerodynamics, toward clinical and voice technology applications. Unlike previous voice inversion research that often uses lumped-element models of phonation, this study explores the feasibility of voice inversion using data generated from a three-dimensional voice production model. Neural networks are trained to estimate vocal fold properties and subglottal pressure from voice features extracted from the simulation data. Results show reasonably good estimation accuracy, particularly for vocal fold properties with a consistent global effect on voice production, and reasonable agreement with excised human larynx experiment [1]. Human subject studies further showed that the neural network was able to monitor the subglottal pressure with reasonable accuracy and predict the alternating vocal fold adduction and abduction pattern during consonant-vowel- consonant transitions [2]. All subjects simultaneously increased the subglottal pressure and vocal fold approximation when producing a louder voice, although the degrees of laryngeal and respiratory adjustments were speaker-specific. These results demonstrate the potential of this neural network toward monitoring and identifying potentially unhealthy vocal behaviors outside the clinic.

References

1 Z. Zhang, *Estimation of vocal fold physiology from voice acoustics using machine learning*, J. Acoust. Soc. Am., 147, EL264–EL270, 2021.
2 Z. Zhang, *Estimating subglottal pressure and vocal fold adduction from the produced voice in a single subject study*, J. Acoust. Soc. Am., 151, 1337–1340, 2022.

8 Challenges in voice analysis and simulation

As a researcher, it is easy to fall into "tunnel vision" and lose the perspective of the larger questions. Therefore, two participants were invited to take a step back and present their view of some over-arching issues in voice analysis and voice simulation.

8.1 What do clinicians need?

Meike Brockmann-Bauser (Universitätsspital Zürich, CH)

License Creative Commons BY 4.0 International license
© Meike Brockmann-Bauser

Instrumental acoustic measurements of quantitative and qualitative human voice characteristics have enormous potential to objectively describe vocal pathology and, thereby, to assist clinical treatment decisions. Despite an increasing application and availability of these techniques, recent research has highlighted a lack of understanding of physiologic and speech related influencing factors. This contribution critically reviews the state of the art in the clinical application of currently recommended instrumental acoustic voice measures and points out future directions. Recent research in vocally healthy and voice disordered adults has shown, that the most widely recommended voice quality measures jitter (%), shimmer (%) harmonics-to-noise ratio (HNR) and Cepstral Peak Prominence (CPP) are affect by natural variations in speaking voice intensity and pitch, vowel and voice task type. Moreover, age, voice training and gender related differences have not been comprehensively described for qualitative, but also quantitative measurements including Voice Range Profiles (SPL and F0 plots). In summary, main limitations include a lack of (a) normative data for known physiologic covariables, such as age, training, speaking voice sound pressure level (SPL) and fundamental frequency (f0) (b) standardization and reporting of analysis procedures (including voice tasks) and techniques (c) understanding of the relation between audible dysphonia, vocal dysfunction, and instrumental acoustic voice features. Future directions include the exploration of Voice Range Profiles complemented with a third dimension of voice quality or electroglottographic measures as clinical tools for pre-post comparisons of voice functionality, related to specific tasks and pathologies. This calls for further research to transfer currently available techniques into clinical applications.

Details are given in [1].

References

1 Brockmann-Bauser, M. and M.F. de Paula Soares, Do We Get What We Need from Clinical Acoustic Voice Measurements? Applied Sciences, 2023. 13(2): p. 941.

8.2 Numerical simulations

Michael Döllinger (Universitätsk-Klinikum Erlangen, DE)

License Creative Commons BY 4.0 International license
© Michael Döllinger

This talk introduced the different types of modelling approach used for numerical simulation of phonatory processes. Models were presented on a spectrum ranging from highly simplified (but fast to run) to highly detailed (but computationally extremely expensive). Lumped mass models offer the lowest-cost approach, and can simulate basic phonatory principles. Their low cost makes them suitable for optimisation or inversion studies, but they constitute a rough approximation of real physiology, it is difficult to incorporate pathologies, and models are difficult to compare to one another due to a lack of standardisation. Increasing in complexity, computational fluid dynamics (CFD) approaches simulate the 3D flow field in high resolution, but only in relation to proscribed vocal fold movements. Although this is only the second approach on the list, computational cost is already high enough that HPC resources are typically required here, along with detailed expert knowledge about simulation parameters and conditions. Computational aeroacoustics (CAA) methods are based on CFD approaches and also include acoustics, allowing use of the models to investigate sound generation processes, again with a higher computational cost. Increasing in detail further, fluid-structure interaction (FSI) models permit the simulation of flow-induced vocal fold oscillations, provided an appropriate biomechanical tissue model is available for the vocal folds. This method is able to capture the vocal fold-airflow interactions, again at the expense of increased computational cost over CFD/CAA models. Finally, the most detailed models currently in use are fluid-structure-acoustic interaction (FSAI) models, which couple FSI models with acoustics, providing a comprehensive account of the whole phonation process, but with requirements for deep expert knowledge of the systems and parameters involved, and a computational cost higher than all of the above approaches. In all cases, validation is a critical issue. Validation should be performed in order to ensure the models are correct, but there is a critical lack of data to validate against. In summary, there are a range of modelling approaches available, each with pros and cons, and the choice of model should be based upon the research question under study.

9 Concluding Plenary

9.1 Towards the future

Amelia Gully (University of York, GB)

License Creative Commons BY 4.0 International license
© Amelia Gully

This talk is both a summary of many of the key points raised throughout the week, and a perspective from an early-career researcher looking towards the future of the field. I make three main arguments: 1) we are an interdisciplinary community, and we need to understand our differences so that we can communicate in a shared language; 2) we need to share high-quality resources to avoid re-inventing the wheel; 3) a concerted, international effort to promote voice science is needed in the face of public indifference and AI voices created without any understanding of the underlying systems. I also share some of my experiences developing interdisciplinary communities at a national level. This Dagstuhl Seminar has gone a long way towards building the necessary trust between researchers, and if we can build upon this to promote excellent international voice research, we can make a real difference in the world.

9.2 Final discussions

Participants expressed various perspectives and challenges in their fields. Several highlighted the difficulty of funding and recruiting for research, emphasizing the importance of sustained financial support and collaborative efforts. Issues such as data sharing, standardization, and the need for interdisciplinary collaboration were recurrent themes. There was enthusiasm for initiatives like World Voice Day to raise public awareness, and the suggestion of potential avenues for broader outreach, through media like documentaries or social media platforms. Overall, all participants agreed on the value of personal interactions and ongoing dialogue among researchers, aiming to foster continued collaboration and innovation beyond the event itself. A commitment was made to produce a joint position article on the state of the art and future perspectives in research on the human voice.

Participants

- Philipp Aichinger
Medizinische Universität Wien, AT
- Marc Arnela
Ramon Llull University – Barceolona, ES
- Lucie Bailly
Univ. Grenoble Alpes, CNRS, Grenoble INP, 3SR, Grenoble, FR
- Peter Birkholz
TU Dresden, DE
- Meike Brockmann-Bauser
Universitätsspital Zürich, CH
- Helena Daffern
University of York, GB
- Michael Döllinger
Universitäts-Klinikum Erlangen, DE
- Mennatallah El-Assady
ETH Zürich, CH
- Sidney Fels
University of British Columbia, Vancouver, CA
- Mario Fleischer
Charité – Berlin, DE
- Andrés Goméz-Rodellar
NeuSpeLab, Las Rozas de Madrid, ES
- Pedro Gomez-Vilda
NeuSpeLab, Las Rozas de Madrid, ES
- Oriol Guasch
Ramon Llull University – Barcelona, ES
- Amelia Gully
University of York, GB
- Nathalie Henrich Bernardoni
Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, FR
- Eric Hunter
University of Iowa, US
- Filipa M.B. Lã
UNED Madrid, ES
- Yves Laprie
LORIA, Nancy, FR
- Sarah Lehoux
UCLA, US
- Matthias Miller
ETH Zürich, CH
- Scott Reid Moisik
Nanyang TU, Singapore, SG
- Peter Pabon
Utrecht, NL
- Jean Schoentgen
Free University of Brussels, BE
- Brad Story
University of Arizona – Tucson, US
- Johan Sundberg
KTH Royal Institute of Technology – Stockholm, SE
- Sten Ternström
KTH Royal Institute of Technolog – Stockhom, SE
- Tino Weinkauf
KTH Royal Institute of Technology – Stockholm, SE
- Qian Xue
Rochester Institute of Technology, US
- Zhaoyan Zhang
UCLA, US

Report from Dagstuhl Seminar 24251

Teaching Support Systems for Formal Foundations of Computer Science

Tiffany Barnes*[1], Jan Vahrenhold*[2], Thomas Zeume*[3], and Florian Schmalstieg†[4]

1 **North Carolina State University – Raleigh, US.** tiffany.barnes@gmail.com
2 **Universität Münster, DE.** jan.vahrenhold@uni-muenster.de
3 **Ruhr-Universität Bochum, DE.** thomas.zeume@rub.de
4 **Ruhr-Universität Bochum, DE.** florian.schmalstieg@rub.de

Abstract

Introductory courses on formal foundations of computer science – including basic courses on theoretical computer science (regular and context-free languages, computability theory, and complexity theory) as well as on logic in computer science (propositional and first-order logic, modeling, and algorithms for evaluation and satisfaction of formulas) – are a cornerstone of computer science curricula, yet many students struggle with their often theoretical contents. The recent influx of students in computer science, as well as the shift towards the inclusion of more online-based teaching ask for advanced teaching support systems that aid both students and instructors.

This Dagstuhl Seminar focussed on fostering discussion between researchers in computing education, builders of systems for teaching formal foundations, as well as instructors of these foundations in order to facilitate more robust research and development of systems to support teaching and learning of the formal foundations of computer science.

Seminar June 16–21, 2024 – https://www.dagstuhl.de/24251
2012 ACM Subject Classification Social and professional topics → Computing education; Theory of computation; Applied computing → Education
Keywords and phrases artificial intelligence in education, computing education research, educational data mining, formal foundations of computer science, intelligent tutoring systems, user modeling and adaptive personalization, user studies
Digital Object Identifier 10.4230/DagRep.14.6.108

1 Executive Summary

Thomas Zeume (Ruhr-Universität Bochum, DE)
Tiffany Barnes (North Carolina State University – Raleigh, US)
Jan Vahrenhold (Universität Münster, DE)

License Creative Commons BY 4.0 International license
© Thomas Zeume, Tiffany Barnes, and Jan Vahrenhold

The primary goal of this Dagstuhl Seminar was to determine how to enable communication between between researchers in computing education, builders of systems for teaching formal foundations, as well as instructors of these foundations. While these groups have very similar interests, they also have very different notions, foci, and methods. In particular, participants from the "formal foundations" community talk about the "hardness" of a problem in terms of

* Editor / Organizer
† Editorial Assistant / Collector

Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

its computational complexity, participants from the "intelligent tutoring systems" community are concerned with whether or not a system can scale or how to best provide feedback to the learner, and participants from the "computing education research" community study the effectiveness of teaching methods for learning, e.g., the cognitive load, student learning, etc.

Within the first one-and-a-half days of the seminar, tutorials on Formal Foundations of CS, CS Education Research, and Intelligent Tutoring Systems given by experts of the respective domains set the stage for the rest of the seminar. The tutorial on CS Education Research was interspersed with breakout sessions for applying the theoretical content of the tutorial to projects of seminar participants, leading to intense discussions across the different communities and therefore being very effective also in bridging barriers between the communities. On the afternoon of the first day, tools and tutoring systems in the formal foundation domain were presented in teaser and poster sessions as well.

The rest of the seminar was centered around breakout sessions, whose research and discussion topics were proposed and voted on by participants. There were a few contributed research talks and occasional ad-hoc tutorial-like sessions as they became relevant for the breakout sessions.

Participants noted that the seminar had a very open atmosphere and that the different research communities were eager to learn from each other. This welcoming spirit was also reflected by a music event on one of the evenings where three of the participants gave a concert and a fare-well magician's show by one of the participants as part of the closing session.

In summary, it was a very fruitful seminar – both with respect to research collaborations and personal interactions. The goal of bringing together the communities and bridging the gaps between them was fully achieved. Several collaborative research projects were initiated during the seminar and are currently being followed-up on.

2 Table of Contents

Working groups

3 Overview of Talks

3.1 Intelligent Tutoring Systems – An Introduction

Johan Jeuring (Utrecht University, NL)

License Creative Commons BY 4.0 International license
© Johan Jeuring

I presented a brief introduction to Intelligent Tutoring Systems. In this abstract I will review the themes I discussed, and include pointers for further reading.

A good introduction to Tutoring Systems is VanLehn's the behavior of tutoring systems [17]. He distinguishes two components:

- the inner loop, in which a student works on a task, takes steps towards solving the task, and a tutoring system provides support in the form of feedback and hints;
- the outer loop, which helps a student with finding a path in the learning material, for example by suggesting next tasks to work on.

Through the years, many approaches to supporting a student when solving an exercise (the inner loop) have been developed:

- cognitive tutors, built upon theories such as ACT-R [7];
- constraint-based tutors [6];
- domain reasoners [5, 4, 3];
- data-driven tutors [2];
- LLM-based tutors [15];
- and more.

Some approaches use a student model to keep track of the learning progress of a student, and to adapt the kind of feedback given to a student. There exist many approaches to student modelling [1]; some well known examples are:

- ELO ratings
- Overlay models
- Knowledge space theory
- Constraint-based modelling
- Bayesian modelling
- Model tracing

Student models are also used to support the outer loop. The outer loop typically presents tasks to a student. The sequence in which tasks are offered is often fixed, sometimes determined by the student, and sometimes supported by a giving suggestions for a next task to work on. Recommendations can be based on a learner model and task attributes, on behavior from other students, on ratings on earlier items, and sometimes other components [8].

Many experiments have been performed to study the effectiveness of Tutoring Systems. VanLehn has shown that if you compare the effects of a tutor that supports stepwise exercises with the help of teaching assistants, there is little difference [16]. Quite a few other meta-reviews on the effectiveness of tutoring systems have been performed, but it is hard to use these reviews to make general statements: they regularly compare apples and pears [9, 10].

There is quite a lot of recent work on teaching-support systems for formal foundations of computer science [11, 12, 13, 14]. A complete overview would require a more systematic approach.

References

1 Konstantina Chrysafiadi and Maria Virvou: Student modeling approaches: A literature review for the last decade. Expert Syst. Appl. 40(11): 4715-4729 (2013)

2 Behrooz Mostafavi and Tiffany Barnes: Evolution of an Intelligent Deductive Logic Tutor Using Data-Driven Elements. Int. J. Artif. Intell. Educ. 27(1): 5-36 (2017)

3 Bastiaan Heeren and Johan Jeuring: Feedback services for stepwise exercises. Sci. Comput. Program. 88: 110-129 (2014)

4 Bastiaan Heeren, Johan Jeuring and Alex Gerdes: Specifying Rewrite Strategies for Interactive Exercises. Math. Comput. Sci. 3(3): 349-370 (2010)

5 Josje Lodder, Bastiaan Heeren and Johan Jeuring: A Domain Reasoner for Propositional Logic. J. Univers. Comput. Sci. 22(8): 1097-1122 (2016)

6 Antonija Mitrovic, Michael Mayo, Pramuditha Suraweera and Brent Martin: Constraint-Based Tutors: A Success Story. IEA/AIE 2001: 931-940

7 J.R. Anderson, A. T. Corbett, K.R. Koedinger and Ray Pelletier. (1995). Cognitive Tutors: Lessons Learned. Journal of the Learning Sciences, 4(2), 167–207. `https://doi.org/10.1207/s15327809jls0402_2`

8 M. Deschênes. Recommender systems to support learners' Agency in a Learning Context: a systematic review. Int J Educ Technol High Educ 17, 50 (2020). `https://doi.org/10.1186/s41239-020-00219-w`

9 J. A. Kulik and J. D. Fletcher (2016). Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. Review of Educational Research, 86(1), 42-78. `https://doi.org/10.3102/0034654315581420`

10 Alan C.K. Cheung and Robert E. Slavin. The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis, Educational Research Review, Volume 9, 2013, Pages 88-113, ISSN 1747-938X, `https://doi.org/10.1016/j.edurev.2013.01.001`.

11 Marko Schmellenkamp, Alexandra Latys, Thomas Zeume: Discovering and Quantifying Misconceptions in Formal Methods Using Intelligent Tutoring Systems. SIGCSE (1) 2023: 465-471

12 Seth Poulsen, Mahesh Viswanathan, Geoffrey L. Herman and Matthew West: Proof Blocks: Autogradable Scaffolding Activities for Learning to Write Proofs. ITiCSE (1) 2022: 428-434

13 Matthew Farrugia-Roberts, Bryn Jeffries, Harald Søndergaard: Programming to Learn: Logic and Computation from a Programming Perspective. ITiCSE (1) 2022: 311-317

14 Preya Shabrina, Behrooz Mostafavi, Mark Abdelshiheed, Min Chi and Tiffany Barnes. Investigating the impact of backward strategy learning in a logic tutor: Aiding subgoal learning towards improved problem solving. *International Journal of Artificial Intelligence in Education*, pages 1–37, 2023.

15 Angelo Sifaleras. *Generative Intelligence and Intelligent Tutoring Systems: 20th International Conference, ITS 2024, Thessaloniki, Greece, June 10–13, 2024, Proceedings, Part I.* Springer Nature, 2024.

16 Kurt VanLehn. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4):197–221, 2011.

17 Kurt VanLehn. The behavior of tutoring systems. *International journal of artificial intelligence in education*, 16(3):227–265, 2006.

3.2 Cognitive Science Concepts You Can Use

Shriram Krishnamurthi (Brown University – Providence, US), Rodrigo Duran (Federal Institute of Mato Grosso do Sul, BR), and R. Benjamin Shapiro (University of Washington – Seattle, US)

License Creative Commons BY 4.0 International license
© Shriram Krishnamurthi, Rodrigo Duran, and R. Benjamin Shapiro

A tour of some of the central theories of cognitive science that are directly applicable in tools for teaching computing foundations (and also in generic pedagogy).

3.3 A Formal-Language-Based Framework for Computing Feedback Information Generically

Martin Lange (Universität Kassel, DE)

License Creative Commons BY 4.0 International license
© Martin Lange
Joint work of Florian Bruse, Martin Lange
Main reference Florian Bruse, Martin Lange: "Computing All Minimal Ways to Reach a Context-Free Language", in Proc. of the Reachability Problems – 18th International Conference, RP 2024, Vienna, Austria, September 25-27, 2024, Proceedings, Lecture Notes in Computer Science, Vol. 15050, pp. 38–53, Springer, 2024.
URL https://doi.org/10.1007/978-3-031-72621-7_4

We present a theory of rewriting a word into a given target language. We show that the natural notion of equivalence between corrections as sequences of edit operations can be captured syntactically by means of a rather simple rewrite system. Completeness relies on a normal form for corrections that is then also used to develop a notion of minimality for corrections. This is not based on edit distance between words and languages but on a subsequence order on corrections, capturing the intuitive notion of doing a minimal number of rewriting steps. We show that the number of minimal corrections is always finite, and that they are computable for context-free languages.

The motivation for this theory and the more intricate notion of minimality is drawn from the study of digital classroom environments where language learning is required. Minimal corrections can be used to give individually targeted feedback and thus guide the learning process automatically.

3.4 Intelligent Tutoring Systems – Tools (What makes them intelligent?)

Martin Lange (Universität Kassel, DE), Tiffany Barnes (North Carolina State University – Raleigh, US), Felix Freiberger (Universität des Saarlandes – Saarbrücken, DE), Michael Goedicke (Universität Duisburg – Essen, DE), Norbert Hundeshagen (Universität Kassel, DE), Johan Jeuring (Utrecht University, NL), Alexandra Mendes (University of Porto, PT & INESC TEC – Porto, PT), Seth Poulsen (Utah State University, US), and Francois Schwarzentruber (IRISA – ENS Rennes, FR)

License Creative Commons BY 4.0 International license
© Martin Lange, Tiffany Barnes, Felix Freiberger, Michael Goedicke, Norbert Hundeshagen, Johan Jeuring, Alexandra Mendes, Seth Poulsen, and Francois Schwarzentruber

The original motivation for this breakout session was given by a simple question: how can intelligent tutoring systems (ITS) be made intelligent? Is the power of LLMs for example sufficient to guarantee a level of machine intelligence that is sufficient for generating feedback in ITS for guiding students through particular learning processes? But then, ITS have been designed and in use before LLMs came up, and therefore other techniques have been used to create specific forms of intelligence in these tools.

The aim of this breakout session was then to analyse and perhaps categorise and quantify such forms of intelligence in order to answer the question above in a way that ideally would tell the developers of ITS what technology to implement in their tools in order to achieve certain forms of intelligent feedback.

Not surprisingly, the concept of intelligence – even in the restricted setting of tutoring systems for formal foundations of computer science – is not easily categorised and quantified. So a large amount of time in this breakout session was initially spent on personal reports on what is used in particular tools. Examples of such methods include the following general methods.

- Comparing the way that a student constructs a solution to successful paths taken from previous attempt (of other students).
- Comparing a student's solution – either the final result or, in interactive tools, the construction path – to some master solution in the form of distance measures or, more generally, as inputs to some abstract problems, for instance comparing actions traces.
- Provide a set of rules that are allowed to be applied in order to construct a correct solution.

This has also sparked off a brief discussion on what should be judged as a correct solution: just the final answer, or the entire construction path. The latter perhaps needs a higher level of intelligence in an ITS. Another very much related question that has been discussed is: what technology can be used to intelligently create good exercises automatically?

The breakout session indentified some general technologies that can be used to create some form of intelligence or other, either in marking, feedback generation or creation of examples and exercise:

- LLMs
- SAT solvers and, more generally, SMT/CSP solvers,
- enumeration algorithms researched primarily in combinatorics and discrete math.

3.5 Automated Proof by Induction Feedback

Seth Poulsen (Utah State University – Logan, US)

License Creative Commons BY 4.0 International license
© Seth Poulsen
Joint work of Seth Poulsen, Chenyan Zhao, Yael Gertner, Benjamin Cosman, Shubhang Kulkarni, Mahesh Viswanathan, Hongxuan Chen, Geoffrey Herman, Matthew West
Main reference Chenyan Zhao, Mariana Silva, Seth Poulsen: "Autograding Mathematical Induction Proofs with Natural Language Processing", CoRR, Vol. abs/2406.10268, 2024.
URL https://doi.org/10.48550/ARXIV.2406.10268

This talk is about software tools that help students learn to write mathematical proofs. Research has shown that timely feedback can be very helpful to students learning new skills. First I introduce Proof Blocks, a tool which enables students to construct mathematical proofs by dragging and dropping prewritten proof lines into the correct order instead of needing to write them from scratch. The instructor specifies the dependency graph of the lines of the proof, so that any correct arrangement of the lines can receive full credit. We develop a novel algorithm which enables assigning students' partial credit on Proof Blocks problems based on the number of edits that their submission is from a correct solution.

For assessment, we provide statistical evidence that Proof Blocks are easier than written proofs, which are typically very difficult. We also show that Proof Blocks problems provide about as much information about student knowledge as written proofs. Survey results show that students believe that the Proof Blocks user interface is easy to use, and that the questions accurately represent their ability to write proofs.

Next, I present a set of training methods and models capable of autograding freeform mathematical proofs by leveraging existing large language models and other machine learning techniques. We recruit human graders to grade the same proofs as the training data, and find that the best grading model is also more accurate than most human graders.

With the development of these grading models, we create and deploy an autograder for proof by induction problems and perform a user study with students. Results from the study shows that students are able to make significant improvements to their proofs using the feedback from the autograder, but students still do not trust the AI autograders as much as they trust human graders. Future work can improve on the autograder feedback and figure out ways to help students trust AI autograders.

3.6 Theory & Methods in Computing Education Research

R. Benjamin Shapiro (University of Washington – Seattle, US) and Shriram Krishnamurthi (Brown University – Providence, US)

License Creative Commons BY 4.0 International license
© R. Benjamin Shapiro and Shriram Krishnamurthi

A tour through some of the foundational ideas in education research applicable in computing. This was followed by a description of good and poor research questions. Finally, we presented three "grammars" for structuring a study.

3.7 Formal Foundations of Computer Science: A personal perspective

Thomas Zeume (Ruhr-Universität Bochum, DE)

License Creative Commons BY 4.0 International license
© Thomas Zeume

In this tutorial, I will outline typical topics and methods included in introductory courses on formal foundations of computer science. Material from introductory courses in logic and theoretical computer science at the Ruhr-Universität Bochum will be used as examples.

4 Demos

4.1 pseuCo Book

Felix Freiberger (Universität des Saarlandes – Saarbrücken, DE)

License Creative Commons BY 4.0 International license
© Felix Freiberger
Main reference Felix Freiberger: "pseuCo Book: An Interactive Learning Experience", in Proc. of the ITiCSE 2022: Innovation and Technology in Computer Science Education, Dublin, Ireland, July 8 – 13, 2022, Volume 1, pp. 414–420, ACM, 2022.
URL https://doi.org/10.1145/3502718.3524801

In this demo, we present pseuCo Book, a truly interactive textbook experience designed to help teachers and students alike. In pseuCo Book, interactive demonstrations and exercises are interwoven with traditional textual elements. Its technical foundation, the Hybrid Document Framework, is a toolset that makes authoring interactive textbooks as easy as possible. PseuCo Book contains three chapters: The first one, covering Milner's Calculus of Communicating systems, is built around an interactive editor for CCS semantics derivations. The second chapter, teaching notions of equality for concurrent processes, features custom-built exercises covering proofs and algorithms around trace equality, bisimilarity, and observation congruence. The third chapter, which covers practical concurrent programming in a minimal, academic programming language called pseuCo, is built around a set of verification technologies that allow deep inspection of the concurrency-related features of pseuCo programs, enabling fast autograding of user-submitted solutions to programming tasks. PseuCo Book has been used extensively as part of the Concurrent Programming lecture at Saarland University. A comprehensive user study, run as part of the course, demonstrates that pseuCo Book is both well-received by students and has a measurable, positive impact on student performance.

4.2 FLACI – Formal Languages, Automata, Compilers and Interpreters

Michael Hielscher (Pädagogische Hochschule Schwyz, CH)

License Creative Commons BY 4.0 International license
© Michael Hielscher
Main reference Michael Hielscher, Christian Wagenknecht: "FLACI – Eine Lernumgebung für theoretische Informatik, Informatik für alle, 18. GI-Fachtagung Informatik und Schule", INFOS 2019: Page 211-220
URL https://flaci.com
Main reference Christian Wagenknecht, Michael Hielscher: "Formale Sprachen, abstrakte Automaten und Compiler: Lehr- und Arbeitsbuch mit FLACI für Grundstudium und Fortbildung", Springer Nature, 2022.
URL https://doi.org/10.1007/978-3-658-36853-1

I gave a brief demo session on FLACI, a web-based system designed for working with formal languages, context-free grammars, and automata. FLACI simplifies the application of these theories to compiler construction, making it accessible even at the high school level. The system allows users to visually construct and simulate automata, derivations, and compiler processes. The goal is to help students learn formal foundations while they work towards translating their own simple language into visual or acoustic output using a compiler they generate from a formal definition.

4.3 TeachingBook

Norbert Hundeshagen (Universität Kassel, DE)

License Creative Commons BY 4.0 International license
© Norbert Hundeshagen
Joint work of Norbert Hundeshagen, Maurice Herwig, John Hundhausen

The TeachingBook (TB) is a prototype of a web-based platform to create interactive learning materials in a cell-based Jupyter-Notebook-like environment. Its main feature is the flexibility in creating content for lectures such as interactive scripts or exercise sheets by simply arranging cells of different types (see attached screenshot). Currently, several cell-types are supported. Besides markdown cells to provide LateX content and quizzes, also cells are available to foster the learning of topics in the realm of formal foundations in computer science (regular languages, reductions, ...). Furthermore, TB is designed to ease the integration of existing tools either as iFrames or natively as front-end components. Several extensions of the TeachingBook are currently under development in student projects and a preliminary usability study has been conducted in a lecture on computability theory. It is planned that a publication will be ready for next year.

4.4 DiMo

Martin Lange (Universität Kassel, DE) and Norbert Hundeshagen (Universität Kassel, DE)

License Creative Commons BY 4.0 International license
© Martin Lange and Norbert Hundeshagen
Joint work of Maurice Herwig, Norbert Hundeshagen, John Hundhausen, Stefan Kablowski, Martin Lange, Georg Siebert
Main reference Maurice Herwig, Norbert Hundeshagen, John Hundhausen, Stefan Kablowski, Martin Lange: "Problem-Specific Visual Feedback in Discrete Modelling", in Proc. of the DELFI 2024 – Die 22. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V., DELFI 2024, Fulda, Germany, September 9-11, 2024, LNI, Vol. P-356, Gesellschaft für Informatik e.V., 2024.
URL https://doi.org/10.18420/DELFI2024_08

DiMo, short for *Discrete Modelling*, is a tool that supports learning of skills to use propositional logic as a backbone for general problem solving. Typical exercises in this area ask for the construction of propositional formulas depending on problem instance, A good example is: write a propositional formula Φ_n for $n \geq 1$ that is satisfiable iff the n-queens problem has a solution, i.e. it is possible to place n queens on a chessboard with no two of them sharing a row, a column or a diagonal line. Another example is: write a formula Φ_G for any undirected graph G that is satisfiable iff G is 3-colourable.

DiMo provides a language that is reminiscent of simple imperative programming languages in order to specify formulas. For example, $\bigvee_{i=0}^{n-1} D(i,0) \wedge \bigwedge_{\substack{j=0 \\ j\neq i}}^{n-1} \neg D(j,0)$ would be written as

```
FORSOME i: {0,..,n-1}. D(i,0) & FORALL j: {0,..,n-1} \ {i}. -D(j,0)
```

DiMo translates formulas in this formal language automatically into the mathematical form above so that students see the connection to the way formulas are presented in lectures etc.

DiMo's crown feature is a programming language with for-loops and propositions as data types. It is supposed to be used by teachers in order to write programs that turn propositional evaluations into any graphical form, for instance in HTML. DiMO then executes these programs after satisfiability checks on the students' formulas. This way, it is possible to check correctness of the formulas graphically, for instance by depicting the placement of queens on a chessboard.

DiMo is publically available to try out via a webinterface, located at `https://dumbarton.tifm.cs.uni-kassel.de/`. Further and more detailed information can be found in two papers on the technical aspects of DiMo [2] and on the graphical feedback interface [1].

References

1 M. Herwig, N. Hundeshagen, J. Hundhausen, S. Kablowski, and M. Lange. Problem-specific visual feedback in discrete modelling. In *Proc. 21. Fachtagung Bildungstechnologien der GI e.V., Delfi'24*, LNI, 2024. To appear.

2 N. Hundeshagen, M. Lange, and G. Siebert. Dimo – discrete modelling using propositional logic. In *Proc. 24th Int. Conf. on Theory and Applications of Satisfiability Testing, SAT'21*, number 12831 in LNCS, pages 242–250. Springer, 2021.

4.5 Proof Blocks

Seth Poulsen (Utah State University, US)

License Creative Commons BY 4.0 International license
© Seth Poulsen
Joint work of Mahesh Viswanathan, Geoffrey L. Herman, Matthew West
Main reference Seth Poulsen, Mahesh Viswanathan, Geoffrey L. Herman, Matthew West: "Proof Blocks: Autogradable Scaffolding Activities for Learning to Write Proofs", in Proc. of the ITiCSE 2022: Innovation and Technology in Computer Science Education, Dublin, Ireland, July 8 – 13, 2022, Volume 1, pp. 428–434, ACM, 2022.
URL https://doi.org/10.1145/3502718.3524774

In this software tool paper we present Proof Blocks, a tool which enables students to construct mathematical proofs by dragging and dropping prewritten proof lines into the correct order. We present both implementation details of the tool, as well as a rich reflection on our experiences using the tool in courses with hundreds of students. Proof Blocks problems can be graded completely automatically, enabling students to receive rapid feedback. When writing a problem, the instructor specifies the dependency graph of the lines of the proof, so that any correct arrangement of the lines can receive full credit. This innovation can improve assessment tools by increasing the types of questions we can ask students about proofs, and can give greater access to proof knowledge by increasing the amount that students can learn on their own with the help of a computer.

4.6 JFLAP (Java Formal Language and Automata Package)

Susan Rodger (Duke University – Durham, US)

License Creative Commons BY 4.0 International license
© Susan Rodger
Main reference Susan H. Rodger, Thomas W. Finley: "JFLAP: An Interactive Formal Languages and Automata Package". Jones and Bartlett Publishers, Inc., Sudbury, Massachussetts, USA, 2006.
URL https://www.jflap.org

We have been designing the JFLAP tool now for about thirty years. JFLAP allows one to experiment with finite state automata, pushdown automata, Turing machines, all kinds of grammars, and L-systems. In addition one can experiment with algorithms and proofs such as converting a nondeterministic finite automaton (NFA) to a deterministic finite automaton (DFA), a DFA to a regular expression, a context-free grammar (CFG) to an nondeterministic pushdown automaton (NPDA), or explore examples with the Pumping Lemma. We presented a demo on the JFLAP tool to show how to take a CFG to build an LR(1) parse table, and then parse a string from that CFG using the table. First, we loaded the following CFG: S -> aSb, S -> aBb, B -> cB, B -> b. We then showed how to convert that CFG to an equivalent NPDA that corresponds to the LR(1) parsing algorithm. We then traced the string "aacbbb" showing how the NPDA is nondeterministic and that the string is accepted. Next, we started with the same grammar and showed how to use JFLAP to construct the LR(1) parse table from the CFG. First we calculated FIRST and FOLLOW sets. Then we built a DFA that models how the LR(1) parsing stack works. Each state in the DFA has marked rules associated with it, indicating how much of the rule has been processed. Using the DFA, we constructed the equivalent LR(1) parse table. The table showed no conflicts. Finally, we parsed the same string "aacbbb", seeing which entry is being executed in the table and which symbols are on the stack, showing the string is accepted.

We have included two figures from this example. One of the figures shows the grammar on the left and on the right shows the corresponding FIRST set, FOLLOW set, DFA with marked rules associated with each state, and the corresponding LR(1) parse table. The other figure shows one step in the parsing of the string "aacbbb$" (we add $ to the right end, an end of string marker). The top left shows the LR(1) parse table, highlighting row 0 (state 0 in the DFA) and column S with an entry of 1, meaning the S and 1 were just pushed onto the parsing stack on top of the 0. The top right shows the input remaining, only "$", and the current stack contents of 1S0 (with 1 the top of the stack). The grammar is shown in the bottom left, highlighting the rule that is currently being reduced. The parse tree being built is shown in the bottom right, and is now complete.

We like to show applications with the theory! `www.jflap.org`.

References

1 Susan H. Rodger and Thomas W. Finley. *JFLAP: An Interactive Formal Languages and Automata Package*. Jones and Bartlett Publishers, Inc., Sudbury, Massachussetts, USA, 2006.

2 Susan H. Rodger, Eric N. Wiebe, Kyung Min Lee, Chris Morgan, Kareem Omar, Jonathan Su, *Increasing engagement in automata theory with JFLAP*, Proceedings of the 40th SIGCSE Technical Symposium on Computer Science Education, SIGCSE 2009, Chattanooga, TN, USA, pages 403-407, March 4-7, 2009.

4.7 Teaching Formal Foundations of Computer Science with Iltis

Marko Schmellenkamp (Ruhr-Universität Bochum, DE)

License Creative Commons BY 4.0 International license
© Marko Schmellenkamp
Joint work of Marko Schmellenkamp, Fabian Vehlken, Thomas Zeume
Main reference Marko Schmellenkamp, Fabian Vehlken, Thomas Zeume: "Teaching formal foundations of computer science with Iltis". Educational Column of the Bulletin of EATCS 14, 2024.
URL http://bulletin.eatcs.org/index.php/beatcs/article/download/797/842

Iltis is a web-based educational support system for the formal foundations of computer science. In the field of logic, Iltis offers exercises for many typical reasoning workflows for propositional, modal, and first-order logic. This includes exercises for modelling a scenario with formulas, transforming these formulas into appropriate normal forms, and testing these formulas for satisfiability using different methods. In the field of formal languages, Iltis includes exercises on regular expressions, finite automata, context-free grammars, and push-down automata. In the field of computational and complexity theory, Iltis supports students with exercises for working with graph problems and designing graph-based reductions. Core objectives in the development of Iltis were to facilitate the straightforward incorporation of new educational tasks, the sequencing of these individual tasks into multi-step exercises, and the cascading of sophisticated feedback mechanisms. Iltis is regularly used in courses with more than 300 students. We welcome all readers to try Iltis at `https://iltis.cs.tu-dortmund.de`.

4.8 Automata Tutor

Maximilian Weininger (IST Austria – Klosterneuburg, AT)

License Creative Commons BY 4.0 International license
© Maximilian Weininger
Joint work of Loris D'Antoni, Martin Helfrich, Jan Kretinsky, Emanuel Ramneantu, Maximilian Weininger
Main reference Loris D'Antoni, Martin Helfrich, Jan Kretinsky, Emanuel Ramneantu, and Maximilian Weininger. Automata tutor v3. In CAV (2), volume 12225 of Lecture Notes in Computer Science, pages 3–14. Springer, 2020.
URL 10.1007/978-3-030-53291-8_1

I shortly demonstrate the teaching-support system Automata Tutor [1], showing the ability to generate feedback quickly, in the form of counter-examples. The three core messages, relating to the communities present at the seminar, are:

1. As TCS-teacher (working in foundations), you might want to use Automata Tutor.
2. As an educations researcher, you might be interested in the data Automata Tutor generates.
3. As a tool-developer, you might be interested in exchanging experiences on technical challenges.

References

1 Loris D'Antoni, Martin Helfrich, Jan Kretínský, Emanuel Ramneantu, and Maximilian Weininger. Automata tutor v3. In *CAV (2)*, volume 12225 of *Lecture Notes in Computer Science*, pages 3–14. Springer, 2020.

4.9 Karp

Chenhao Zhang (Northwestern University – Evanston, US)

License Creative Commons BY 4.0 International license
© Chenhao Zhang

In CS theory courses, NP reductions are a notorious source of pain for students and instructors alike. Invariably, students use pen and paper to write down reductions that work in many but not all cases. When instructors observe that a student's reduction deviates from the expected one, they have to manually compute a counterexample that exposes the mistake. We introduce Karp, a language for programming and testing NP reductions.

5 Working groups

5.1 Formal foundations in schools

Erik Barendsen (Radboud University – Nijmegen, NL & Open University – Heerlen, NL), Rodrigo Duran (Federal Institute of Mato Grosso do Sul, BR), Judith Gal-Ezer (The Open University of Israel – Ra'anana, IL), Sandra Kiefer (University of Oxford, GB), Dennis Komm (ETH Zürich, CH), Tilman Michaeli (TU München, DE), Liat Peterfreund (The Hebrew University of Jerusalem, IL), Ramaswamy Ramanujam (Azim Premji University – Bengaluru, IN), Susan Rodger (Duke University – Durham, US), Florian Schmalstieg (Ruhr-Universität Bochum, DE), R. Benjamin Shapiro (University of Washington – Seattle, US), and John Slaney (Australian National University – Canberra, AU)

License Creative Commons BY 4.0 International license
© Erik Barendsen, Rodrigo Duran, Judith Gal-Ezer, Sandra Kiefer, Dennis Komm, Tilman Michaeli, Liat Peterfreund, Ramaswamy Ramanujam, Susan Rodger, Florian Schmalstieg, R. Benjamin Shapiro, and John Slaney

The members exchanged the current state of affairs concerning the role of formal methods in K-12 curricula and teacher education in Switzerland, Israel, The Netherlands, Germany, Brazil, the United States, India, and Australia.

Formal methods include algorithmic reasoning, automata, formal languages, and logic. The group discussed possible advantages and hindrances related to teaching aspects of formal methods in K-12. It was useful in distinguishing societal needs, CS as a discipline, curricular content, and pedagogies.

Potential themes for follow-up activities of the group are:

- Why are formal foundations important in K-12? What should be in a dedicated subject and what is important for "everyone" (or just some group) and why? Which elements are critical for compulsory education? (The situation for CS seems to be different from, eg, physics.) An opinion article (column) could be a result.
- A research study on teachers' perspectives on the "formal foundations": content knowledge, teacher beliefs, PCK, and a cross-cultural inventory of teaching practices w.r.t. formal methods.
- A research study on students' point of view on Computing in society and which elements are based on formal foundations of CS.
- Identifying and utilizing the opportunities w.r.t. 'reasoning' at primary and upper primary levels and increasing the logic content across secondary school curricula.
- Cross-cultural studies on specific aspects of formal methods, taking the respective cultural contexts into account.

5.2 Using LLMs in the process of learning how to perform reductions – An idea

Michael Hielscher (Pädagogische Hochschule Schwyz, CH), Norbert Hundeshagen (Universität Kassel, DE), Johan Jeuring (Utrecht University, NL), Martin Lange (Universität Kassel, DE), and Tilman Michaeli (TU München, DE)

License Creative Commons BY 4.0 International license
© Michael Hielscher, Norbert Hundeshagen, Johan Jeuring, Martin Lange, and Tilman Michaeli

Reductions are a key method in computability and complexity theory to identify unsolvable or intractable problems. Therefore, learning how to reduce one problem onto another is part of every standard curriculum in such courses. We here report on ideas discussed in a working group with the goal of designing an intervention to help students solving exercises on reductions and thus, foster a better understanding of this rather difficult topic. Moreover, our focus is on tool-supported learning by using generative AI. From a learners perspective, a reduction task between two problems A and B essentially can be seen as a programming exercise, where students need to write an algorithm that converts instances of A into instances of B, such that a solution for B can be used to solve A. Our idea of an intervention aims at this programming part of reductions. More specifically, an educational tool should be designed to help students plan the solution to a given reduction task. That is, before implementing a reduction between two given problems, students use our intended tool to describe in natural language how their algorithm should convert problem instances. This description is then used as a prompt for an LLM (e.g. ChatGPT 4.0) with the task of actually producing an implementation, e.g. in Python. The latter allows feedback to be computed in two ways. First, the AI-generated algorithm can be tested for correctness of the solution (and the plan), by running it on positive and negative examples of problem A and providing its answer to the students. In the case of incorrect solution attempts, i.e. positive instances of A are mapped to negative instances of B, or vice versa, further feedback on how to improve the plan could also be provided by the LLM. We suspect that a carefully designed prompt that includes problem-specific information and meta-information on planning can be used for the latter. As the approach described above is intended to be a first idea on the subject, it is clear that a number of issues need to be addressed before the intervention is actually implemented. Among others, it is unclear how planning of reductions in natural language is perceived by students, and how the quality of a plan can be measured. Furthermore, the quality of LLM-output needs to be investigated, especially, if reduction tasks are complex and/or the plans of students are incomplete or ambiguous.

5.3 Brainstorm on Recording Teachers' Observations of Errors for Formal Foundations of CS

Daphne Miedema (University of Amsterdam, NL), Norbert Hundeshagen (Universität Kassel, DE), Martin Lange (Universität Kassel, DE), Alexandra Mendes (University of Porto, PT & INESC TEC – Porto, PT), Sophie Pinchinat (University of Rennes, FR), Anne Remke (Universität Münster, DE), Vaishnavi Sundararajan (Indian Institute of Technology – New Delhi, IN), Maximilian Weininger (IST Austria – Klosterneuburg, AT), and Thomas Zeume (Ruhr-Universität Bochum, DE)

License Ⓒ Creative Commons BY 4.0 International license
© Daphne Miedema, Norbert Hundeshagen, Martin Lange, Alexandra Mendes, Sophie Pinchinat, Anne Remke, Vaishnavi Sundararajan, Maximilian Weininger, and Thomas Zeume

There is not much existing research on misconceptions for FF and TCS. Therefore, we propose to build a base to build such work on. The format will be to gather teachers' observations on students' struggles in courses such as Theory of Computer Science. We do this by creating a form asking for descriptions and example questions that these mistakes occur on. These are collected and organized in a document, which could be written up for a discussion paper. In a later stage, we could follow-up on this research by studying error prevalence and identifying underlying misconceptions.

Additionally, the document could be of use for those building Teaching-Support Systems, as they can gain insight into problems teachers or students typically run into. They could translate this into additional exercises or hints within the TSS.

5.4 Automated Proof Feedback in the Wild

Seth Poulsen (Utah State University, US), Erik Barendsen (Radboud University – Nijmegen, NL & Open University – Heerlen, NL), Felix Freiberger (Universität des Saarlandes – Saarbrücken, DE), Dennis Komm (ETH Zürich, CH), and Thomas Zeume (Ruhr-Universität Bochum, DE)

License Ⓒ Creative Commons BY 4.0 International license
© Seth Poulsen, Erik Barendsen, Felix Freiberger, Dennis Komm, and Thomas Zeume

We discussed the feasibility of doing a large scale study of using teaching support systems to help students learn how to write reduction proofs as part of a computing theory course. Such a study would involve helping 2nd year computer science students learn reduction proofs by using Proof Blocks, Faded Proof Blocks, and feedback from large language models fine-tuned on mathematical proof data. Student proofs would be analyzed through the lens of the following rubric: (1) Overall proof structure is there, (2) Reduction Function is clearly defined, (3) Proving the computability/complexity, (4) Proving the Reduction Property (both directions).

5.5 What makes translating formal languages in set notation to context free grammar difficult?

Florian Schmalstieg (Ruhr-Universität Bochum, DE), Rodrigo Duran (Federal Institute of Mato Grosso do Sul, BR), Liat Peterfreund (The Hebrew University of Jerusalem, IL), Jakob Schwerter (TU Dortmund, DE), John Slaney (Australian National University – Canberra, AU), and Maximilian Weininger (IST Austria – Klosterneuburg, AT)

License Creative Commons BY 4.0 International license
© Florian Schmalstieg, Rodrigo Duran, Liat Peterfreund, Jakob Schwerter, John Slaney, and Maximilian Weininger

The conversion of formal languages (e.g. set notation to context free grammar or PDA) is a difficult problem for students to tackle. But some tasks are more difficult than others. This leads to the question: What properties make a conversion task difficult? We call such properties *difficulty-generating factors.* They can help to understand why a specific task might be difficult and to adapt tasks for different needs.

But in which ways can such factors be found systematically? And how can hypothesized factors be verified?

To this end we propose a mixture of qualitative and quantitative research designs. First we want to assess possible difficulty-generating factors. For this we will be doing expert and novice interviews to find out, which aspects of such a task they look at. The tasks for these interviews are informed by existing student performance data.

We will then try to verify the found factors in a randomized control trial study by systematically manipulating the aspects and using Rasch-analysis to compare the result with the expected ranking.

5.6 Can teaching support systems affect students self-regulated learning behavior?

Jakob Schwerter (TU Dortmund, DE), Michael Hielscher (Pädagogische Hochschule Schwyz, CH), Daphne Miedema (University of Amsterdam, NL), Marko Schmellenkamp (Ruhr-Universität Bochum, DE), Jan Vahrenhold (Universität Münster, DE), and Thomas Zeume (Ruhr-Universität Bochum, DE)

License Creative Commons BY 4.0 International license
© Jakob Schwerter, Michael Hielscher, Daphne Miedema, Marko Schmellenkamp, Jan Vahrenhold, and Thomas Zeume

Self-regulated learning (SLR) has been shown to be positively correlated to academic succes (Zimmerman et al. 1992, Greene et al. 2021). In this breakout group, we considered the question of whether – and if so how – data from digital learning environments (DLEs) can be used to gain insights into the state of a self-regulated learning process the learner is currently in. For the sake of conciseness, we decided to focus on the ILTIS DLE while reminding ourselves of the fact that there is a broad spectrum of such systems. Building on Zimmerman's conceptualization (2008), we used the SRL cycle consisting of "Forethought Phase" – "Performance Phase" – "Self-Reflection Phase". Going through each of the phases of this cycle, we started with the question which SRL-related variables can be generated or derived by the system and what gaps currently exist. We hypothesize that using these variables, we can create an individual learner profile which then in turn can ultimately be used to suggest tailored interventions to help learners improve their study and learning behavior.

References

1 Greene, J. A., Plumley, R. D., Urban, C. J., Bernacki, M. L., Gates, K. M., Hogan, K. A., Demetriou, C., & Panter, A. T. (2021). Modeling temporal self-regulatory processing in a higher education biology course. Learning and Instruction, 72, 101201. https://doi.org/10.1016/j.learninstruc.2019.04.002

2 Zimmerman, B. J. (2008). Investigating Self-Regulation and Motivation: Historical Background, Methodological Developments, and Future Prospects. American Educational Research Journal, 45(1), 166-183. doi.org/10.3102/0002831207312909

3 Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-Motivation for Academic Attainment: The Role of Self-Efficacy Beliefs and Personal Goal Setting. American Educational Research Journal, 29(3), 663-676. doi.org/10.3102/00028312029003663

5.7 Tool building: Experience exchange

Maximilian Weininger (IST Austria – Klosterneuburg, AT), Tiffany Barnes (North Carolina State University – Raleigh, US), Felix Freiberger (Universität des Saarlandes – Saarbrücken, DE), Michael Goedicke (Universität Duisburg – Essen, DE), Michael Hielscher (Pädagogische Hochschule Schwyz, CH), Dennis Komm (ETH Zürich, CH), Daphne Miedema (University of Amsterdam, NL), Seth Poulsen (Utah State University, US), Susan Rodger (Duke University – Durham, US), Florian Schmalstieg (Ruhr-Universität Bochum, DE), Marko Schmellenkamp (Ruhr-Universität Bochum, DE), Vaishnavi Sundararajan (Indian Institute of Technology – New Delhi, IN), and Thomas Zeume (Ruhr-Universität Bochum, DE)

License Ⓒ Creative Commons BY 4.0 International license
© Maximilian Weininger, Tiffany Barnes, Felix Freiberger, Michael Goedicke, Michael Hielscher, Dennis Komm, Daphne Miedema, Seth Poulsen, Susan Rodger, Florian Schmalstieg, Marko Schmellenkamp, Vaishnavi Sundararajan, and Thomas Zeume

In this working group, we discussed best practices for building and maintaining teaching support systems. Below, we summarize the key takeaways:

- Maintainability: This is a crucial problem for teaching-support tools, and often a cause for them to not stay available after some years. A key problem is that they are usually developed by undergraduate or PhD-students and thus not maintained after some time. We identified several ways to mitigate this: i) avoid using the "fanciest" new technology, ii) limit the inclusion of other libraries and dependencies, iii) host a web app and do not require local installation for users, iv) modularize solutions, v) have a plan for how the tool will be maintained (which requires permanent staff to be involved).
  We also discussed the idea of building on top of existing systems (like ILTIS or moodle) which can take care of processes like user management, course management and task management. On the one hand, this simplifies the code of the actual teaching-support system; on the other hand, it introduces a dependency to the other system and thus a potential maintainability problem in the future (keeping versions up to date, what if the existing system is discountinued).
- Scalability: Several tools run on a cluster of virtual machines, which has only limited scalability. Alternative solutions include using systems like Apache Kafka, distributed caching or client-side processing. The latter effectively eliminates scalability problems, however at a cost: Firstly, collecting performance data requires is less immediate, but this can be fixed as has been done for PseuCo. Secondly, client-side grading introduces a vulnerability, as students can reverse-engineer the grader and obtain the sample solution.

- Funding: Acquiring the funding to host a teaching-support system is a key problem, in particular as maintenance and scaling are expensive, but often only development of new features is funded. The following possible solutions were suggested: using teaching improvement funds or asking universities using the tool to pay for a student assistant (since they anyway pay students to grade assignments, they are saving money this way).
- Software-development best practices: Proper practices like writing requirements, performing code-reviews and unit tests should be applied. However, this requires knowledge of software engineering as well as the funding and time to adhere to these practices.
- Data-collection: For education purposes, it can be very useful to not only have the final result of a task, but also a trace of all the actions the student performed when solving the task. Moreover, for analysis, it is very useful if the tool offers a “replay” feature. This allows to “execute” the trace and for every action of the student see the state of the system like the student did.

Participants

- Efthimia Aivaloglou
TU Delft, NL
- Erik Barendsen
Radboud University – Nijmegen, NL & Open University – Heerlen, NL
- Tiffany Barnes
North Carolina State University – Raleigh, US
- Rodrigo Duran
Federal Institute of Mato Grosso do Sul, BR
- Felix Freiberger
Universität des Saarlandes – Saarbrücken, DE
- Judith Gal-Ezer
The Open University of Israel – Ra'anana, IL
- Michael Goedicke
Universität Duisburg – Essen, DE
- Michael Hielscher
Pädagogische Hochschule Schwyz, CH
- Norbert Hundeshagen
Universität Kassel, DE
- Johan Jeuring
Utrecht University, NL
- Sandra Kiefer
University of Oxford, GB
- Dennis Komm
ETH Zürich, CH
- Shriram Krishnamurthi
Brown University – Providence, US
- Martin Lange
Universität Kassel, DE
- Alexandra Mendes
University of Porto, PT & INESC TEC – Porto, PT
- Tilman Michaeli
TU München, DE
- Daphne Miedema
University of Amsterdam, NL
- Liat Peterfreund
The Hebrew University of Jerusalem, IL
- Sophie Pinchinat
University of Rennes, FR
- Seth Poulsen
Utah State University, US
- Ramaswamy Ramanujam
Azim Premji University – Bengaluru, IN
- Anne Remke
Universität Münster, DE
- Susan Rodger
Duke University – Durham, US
- Florian Schmalstieg
Ruhr-Universität Bochum, DE
- Marko Schmellenkamp
Ruhr-Universität Bochum, DE
- Francois Schwarzentruber
IRISA – ENS Rennes, FR
- Thomas Schwentick
TU Dortmund, DE
- Jakob Schwerter
TU Dortmund, DE
- R. Benjamin Shapiro
University of Washington – Seattle, US
- John Slaney
Australian National University – Canberra, AU
- Vaishnavi Sundararajan
Indian Institute of Technology – New Delhi, IN
- Jan Vahrenhold
Universität Münster, DE
- Maximilian Weininger
IST Austria – Klosterneuburg, AT
- Thomas Zeume
Ruhr-Universität Bochum, DE
- Chenhao Zhang
Northwestern University – Evanston, US

Report from Dagstuhl Seminar 24261

Computational Creativity for Game Development

Duygu Cakmak*[1], Setareh Maghsudi*[2], Diego Perez Liebana*[3], and Pieter Spronck*[4]

1 **Creative Assembly – Horsham, GB.** duygucakmak@gmail.com
2 **Ruhr-Universität Bochum, DE.** setareh.maghsudi@rub.de
3 **Queen Mary University of London, GB.** diego.perez@qmul.ac.uk
4 **Tilburg University, NL.** p.spronck@gmail.com

Abstract

Developments in artificial intelligence are currently dominated by deep neural networks, trained on large data sets, which excel at pattern recognition. Variants of the "classic" deep neural networks have the ability to generate new data with statistical properties similar to the training set. Despite the impressive products of such creative artificial intelligence, the results are usually lacking in meaning. They contain mistakes that humans would avoid, and often produce content which is not functional. While the product of creative artificial intelligence can be used as a strong basis for humans to build upon, human intelligence and human creativity are almost always a necessary ingredient of the creative process. Moreover, the more relevant the meaning, purpose, and functionality of the product are, the less the creative process benefits from the involvement of artificial intelligence.

Game design and implementation are tasks which require a high amount of creativity, and which must lead to products which require a high amount of fine-tuned functionality. For example, a game "level" should not only look appealing, it should also be playable and it should be interesting to play. These are not features which can be acquired simply by "training on big data," which is what most developments in modern artificial intelligence are based on.

This report on Dagstuhl Seminar 24261 discusses to what extent modern artificial intelligence techniques can produce meaningful and functional game content.

Seminar June 23–28, 2024 – https://www.dagstuhl.de/24261
2012 ACM Subject Classification Computing methodologies → Artificial intelligence; Human-centered computing → Human computer interaction (HCI); Information systems → Multimedia databases
Keywords and phrases artificial intelligence, computational creativity, computational intelligence, game design, game development
Digital Object Identifier 10.4230/DagRep.14.6.130

1 Executive Summary

Pieter Spronck (Tilburg University, NL)
Duygu Cakmak (Creative Assembly – Horsham, GB)
Setareh Maghsudi (Ruhr-Universität Bochum, DE)
Diego Perez Liebana (Queen Mary University of London, GB)

License Creative Commons BY 4.0 International license
© Pieter Spronck, Duygu Cakmak, Setareh Maghsudi, and Diego Perez Liebana

Developments in artificial intelligence are currently dominated by deep learning technology, which generates deep neural networks, trained on large data sets, which excel at pattern recognition. Variants of the "classic" deep neural networks have the ability to generate

* Editor / Organizer

Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

new data with statistical properties similar to the training set. Generative Adversarial Networks (GANs), such as those used by DALL-E and Midjourney, may be used to generate original visual artworks based on a textual description of the desired output. Autoregressive language models, such as those used by ChatGPT, use deep learning to produce text that is often indistinguishable from human-created text. Moreover, artificial intelligence techniques have been used to successfully generate music for many years, and researchers have also experimented with using deep learning to create cooking recipes, personalized fragrances, fashion, and more.

Despite the sometimes astonishing products of such creative artificial intelligence, the results are usually lacking in meaning. While DALL-E and Midjourney produce images that seem impressive, upon further inspection they contain many mistakes which humans would avoid. While ChatGPT can generate human-sounding text in a conversation, it often produces utter nonsense, and cannot write an original coherent story. And, as our own explorations of such techniques during Dagstuhl Seminar 22251 showed, GANs may produce computer game content which looks reasonable at first glance, but is ultimately neither functional nor playable.

While the product of creative artificial intelligence can often be used as a strong basis for humans to build upon, and may as such speed up the creative process, human intelligence and human creativity are almost always a necessary ingredient of the creative process. Moreover, the more relevant the meaning, purpose, and functionality of the product are, the less the creative process benefits from the involvement of artificial intelligence.

Game design and implementation are tasks which require a high amount of creativity, and which must lead to products which require a high amount of fine-tuned functionality. For example, a game "level" should not only look appealing, it should also be playable (i.e., it must be possible for most players to finish the level) and it should be interesting to play (i.e., the player should feel entertained by playing the level and should experience inherent motivation to finish the level). These are not features which can be acquired simply by "training on big data," which is what most developments in modern artificial intelligence are based on.

The goal of Dagstuhl Seminar 24261, Computational Creativity for Game Development, was to investigate to what extent modern artificial intelligence techniques can produce meaningful and functional game content, and what changes to or extensions of these techniques can improve this AI-driven creative process.

We like to point out that progress in this area is relevant for a wide range of applications outside the "games" domain. Creativity in artificial intelligence applies to many branches of industry and has a strong impact on society, in which artificially intelligent technology interacts with humans in many shapes and forms. We use games in our research because they are highly-complex but well-defined applications which form safe environments to experiment in. However, solutions found for creative problems in games are often transferable to domains outside games.

This research area lends itself to a wide range of research topics. For the preparation of this seminar, we proposed the following set of sub-topics (many of which were taken up by workgroups):

- **Procedural Content Generation for Games**: Procedural Content Generation (PCG) systems include techniques and methods able to create different types of game elements, such as levels, rules, quests, and characters, among many others. Research in PCG has been prolific in the last decade, but its presence in the games industry is still far from ubiquitous. A particular interest is set on mixed-initiative systems, which give designers and artist authorial control of the created content and the direction of the algorithm that generates it.

- **Procedural Generation of Games:** An extension of the previous point that deserves its own separate area is the generation of complete games. Same as recent advances in generative systems for music, painting, and long bodies of text, one can research how complete games can be generated from scratch. This includes elements like art, rules, characters, and winning and losing conditions that normally form a game. Automatic generation of new mini-games can open an interesting space of research that merges multiple advances together, but also a useful tool for game designers that will be able to use the generated games as inspiration for new entertainment experiences.
- **Computational Creativity for Narrative Games:** A particular type of game that has become more popular in the last decade is that of narrative games. While there has been some work in using computational creativity methods to generate texts, the adaptation of these techniques to the game development process remains an open area of research.
- **Automatic Generation of Art in Games:** Art is an important part of digital games and takes multiple forms: 3D-models, textures, visual effects, animations, cut-scenes, and so on. Lately, multiple advances on the use of computational creativity have shown the capacity of generating different forms of art, such as images, videos, and even 3D geometries. Examples of systems that generate art are DALL-E, Stable Diffusion, and Midjourney. Research can explore how these and other techniques can be used to generate art for games, including unexplored game art areas, in particular with regards to how this generation can be bound to specific games/genres/restrictions, how it can be integrated into the game development process, and how we can give designers authorial control and modification capabilities over the generated assets.
- **Procedural Generation of Audio for Games:** An important part of automatic generation of content refers to audio. From audio effects (footsteps, heartbeats, weather) to complete sound tracks (background music, melodies, singers), including the generation of different voices for human and non-human game characters, the space for computational creativity to generate this type of art is vast.
- **Computational Creativity for Game Playing:** An unexplored aspect and application of computational creativity is that of generating AI agents that play a game. Traditionally, the objective that leads AI agents in a game to play is to achieve victory, either by reaching a winning state or by maximizing the score they obtain in the game. Some efforts have been made to employ quality-diversity methods to generate different styles of play. Research may explore how we can harness the new developments in computational creativity to generate diverse play styles, including the generation of new strategies or tactics to play games in a different manner.
- **Computational Creativity for Affective Computing:** Affective computing is a discipline that bridges several domains, such as computer science, cognitive science, and psychology. It studies the implementation of systems that are able to express, identify, process, and simulate human affects. Research may investigate computational creativity algorithms and methods to provide non-player characters with the possibility of expressing feeling and emotions in a convincing way. This includes, for example, facial expressions and body animations, and it can be applied to human or non-human characters.
- **Automatic Support of Game Development:** Traditionally, computational creativity and the automatic generation methods have focused their efforts on generating the product that creative industries build – be this games (or content for games), art or music, among others. These methods may also be used to aid the process of game development. Examples of the application of this technology include computational creativity for

automation of tasks, algorithms for automatic testing of development process (such as code, integration, animations and deployment), production chains, and procedural development processes.
- **Ethical Considerations of Computational Creativity:** The ethical challenges of using computationally creative tools for applications such as game development should not be ignored. The use of automatically-creative tools may have negative effects on the need for artists and designers. Moreover, the automatic creation of games and game content may lead to ethically suspect products. Finally, biases that exist in art and data may be magnified when such products are used to automatically generate new products. These ethical considerations were taken into account in all our explorations of advances in computational creativity.

More than a year-and-a-half passed between writing the proposal and running the seminar. We found it striking how many advancements had been made in the area of Computational Creativity for Games in that period. During the writing of the proposal we were personally convinced that we were proposing an important theme for the seminar. When the seminar took place, we knew that no other theme was this topical.

This seminar was organized around workgroups, which worked in teams and topics proposed by the participants of the seminar in the areas outlined above. These workgroups were accompanied by plenary sessions for group formation, topic debate, and discussions of the deliberation of each group. Workgroups were dynamic, so participants could move between them, and new groups were formed during the week.

A Discord server was set up for coordination and announcements, and it was also used by the different groups for document and link sharing. This also has the benefit of providing a place for discussions after the seminar, easing the communication and further work among the members of each workgroup.

42 participants accepted our invitation to join the seminar; 40 of them attended. The participants were a good mixture of gender, country of origin, junior and senior seniority, and academia or industry. All participants engaged intensively within the seminar, and many expressed how happy they were with what we accomplished, making the seminar a great success.

2 Table of Contents

3 Working groups

3.1 AI for Voice Generation from Text

Maren Awiszus (Viscom AG – Hannover, DE), Filippo Carnovalini (VU – Brussels, BE), and Pieter Spronck (Tilburg University, NL)

License Creative Commons BY 4.0 International license
© Maren Awiszus, Filippo Carnovalini, and Pieter Spronck
Joint work of James Betker, Z. Wang, et al.
Main reference James Betker: "Better speech synthesis through scaling", CoRR, Vol. abs/2305.07243, 2023.
URL https://doi.org/10.48550/ARXIV.2305.07243

Generative AI has proven more sophisticated over the last year, and this also includes speech synthesis. The quality of both Text-to-Speech (TTS) and Voice-to-Voice Conversion is now high enough that it can feasibly be used in commercial products. Services such as *Elevenlabs*[1] boast with generative voices using audio samples as short as 1 minute, and going up to 3 hours for the best quality. Such voice synthesis can be very useful in the context of games, for example, it would allow for a cheaper alternative to full voice acting, especially for smaller studios which otherwise couldn't afford it. Voice Conversion allows to convert from one voice sound to another, which could allow a game to have diverse voices for different characters, even though they have been spoken only by one person, or a TTS generator. This is especially interesting for customizable Player Characters, which usually do not have a voice actor attached due to the sheer number of possible options for their voice and dialogue options. Voiceover is also important for the visually impaired and a good, natural sounding narration would allow more people to experience otherwise completely text-based games. Lastly, games that want to procedurally generate their text, for example using Large Language Models (LLMs), by design can not be voice acted in any other way, other than procedurally. This also applies to user text input which could be actually voiced using a generative method.

Of course, using generative voice for the described applications comes with downsides. If AI Voice Acting is used in projects, which could reasonably have hired real voice actors, those actors have lost that income and, in the long run, fewer voice actors could be employed. Additionally, the generated voices are still not of the quality of real voice actors, so the final product can be worse than if real voice acting was used, as well as include generation artifacts and mispronounced words. Also should the generated or user input text contain harmful sentiments, the generative voice would still say them, which could be detrimental to the career of anyone who was willing to provide their voice as a base for generation, as well as the developers responsible.

In this workgroup, our aim was to see how far one can get using available open source models to create a game with generated voice acting, and if possible, speaking user input lines on the fly. Figure 1 shows the envisioned loop for voice generation. For the generative parts, we used *Tortoise TTS* [1] for Text-to-Speech and *Retrieval based Voice Coversion (RVC)* [2] for Voice Conversion. A demo visual-novel-type game was created using *Renpy*[3], which is based on the Python programming language. This was done in an effort to include the python-based TTS directly into the game more easily.

[1] https://elevenlabs.io

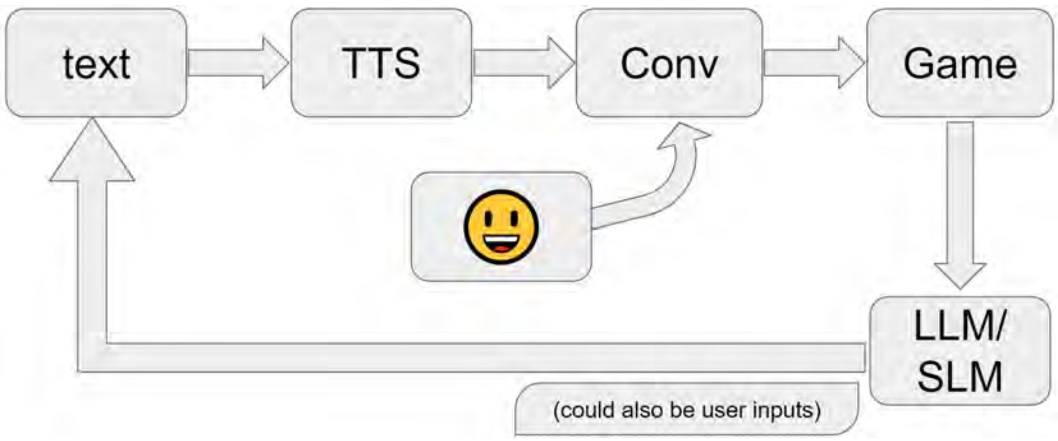


Figure 1 Generative Voice Acting Pipeline. The text of the game is fed to Text-to-Speech (TTS) which is then refined and/or transformed to a different voice using Voice Conversion (Conv) and can then be played by the game. The game or the player can then generate more text and the loop repeats.

3.1.1 Text to Speech

The Chosen TTS Model, *Tortoise TTS* [1] was used with the help of the ai-voice-cloning WebUI[2] which allowed for fast and easy testing. The Tortoise Model can be finetuned to mimic cadence and intonation, but with the limited time at the seminar, we chose to only use the voice adaptation feature. For that, we provided about 50 seconds of speech from one of the groups participants, whose tone the model was supposed to align itself to.

Given that there was no finetuning done, the generated voice lines only resembled the provided voice samples in tone, but not in cadence and intonation. Still, given the fact that this did not require any additional time to train and the generation time could be brought down to a few seconds on a Laptop, the results were promising.

3.1.2 Voice Conversion

In order to see how well the generated text could be spoken by different characters, we decided to use the voice lines generated with Tortoise TTS with different open source pretrained models in the *Retrieval based Voice Conversion (RVC)* WebUI [2]. The pretrained models were taken from a huggingface space[3], as there was no time to train an entire model from scratch during our workgroup.

The converted voice lines notably sounded like the characters those models were trained on and not like the original input. However, mannerisms and inflections were not generated, if not already present in the original sample. So while this does work in allowing for more diverse character voices, it is still not a replacement for the depth of having a voice actor impersonate a character fully.

[2] `https://github.com/RVC-Project/Retrieval-based-Voice-Conversion-WebUI`
[3] `https://huggingface.co/juuxn/RVCModels`

3.1.3 Demonstration

The generated voice lines were added to a small demo game (Figure 2) created with Renpy [3]. The idea was to have the user input be read out loud by the TTS used in the previous sections, which should be possible as they are both based on the Python programming language. However, due to unforeseen issues and limited time, we did not end up adding the functionality to the game. We suggest using a different approach if this is to be attempted again in the future.

For the purpose of the demo, we did put pre-generated lines of different characters into the game as audio files to show off the potential of the approach.

Figure 2 Demonstration game.

3.1.4 Conclusion

Given the small scope and time for our demo, our results for creating voice lines with Tortoise Voice adaptation and Voice Conversion for different characters are quite promising. With enough polish, these techniques could be used to create a decent quality for a small project, that otherwise couldn't afford to use real voice actors. However, the quality for the generated voices is not yet there to compete with real voice acting on any level. We are interested to see where this approach can be taken to in the future, especially with games possibly using procedurally generated text in mind.

References

1 James Betker. *Better speech synthesis through scaling*. arXiv, 2305.07243, 2023
2 Z. Wang et al. *Multi-Level Temporal-Channel Speaker Retrieval for Zero-Shot Voice Conversion*. IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, pp. 2926-2937, 2024, doi: 10.1109/TASLP.2024.3407577.
3 Renpy. *https://www.renpy.org/*. Accessed 26th June 2024.

3.2 Meaningful Acoustics for Board Games

Filippo Carnovalini (VU – Brussels, BE), Greta Hoffmann (TH Köln, DE), Chengpeng Hu (Southern Univ. of Science and Technology – Shenzen, CN), Leonie Kallabis (TH Köln, DE), Matthias Müller-Brockhausen (Leiden University, NL), and Mike Preuß (Leiden University, NL)

License Creative Commons BY 4.0 International license
© Filippo Carnovalini, Greta Hoffmann, Chengpeng Hu, Leonie Kallabis, Matthias Müller-Brockhausen, and Mike Preuß

Music has become a staple aspect in videogames, providing emotional depth and enhancing narration. Other sonic aspects also play a huge role: sound effects (*SFX*) and soundscapes better the user experience and can become elements of added realism and immersion.

Less attention has been devoted to *board games*, which because of their physical nature have less opportunities for embedding SFX and other procedural sounds. In this workshop we examined what the possible approaches to the sonification of board games would be, finding a general description of possible roles of music and sound in the gaming experience.

We started by thinking of existing examples of ways in which commercially available games incorporate music and sound. Some children's toys have various way to produce sounds that keep the children entertained, and there are some electronic board games that leverage sound as a feedback mechanism (e.g. Operation, or some iterations of Battleship) either through buzzers or through recorded sounds played via speakers. Escape rooms often have background music, and many game masters will (try to) provide music to accompany table-top RPGs. Other board games are now starting to provide music in external resources, for example via companion apps or through websites linked via QR codes. Some academic works have focused on procedural music generation, but mostly for videogames [1, 2]. Interestingly, the only work we know of that considers board games (although in digital format) is focused on computer-generated games [3].

We then moved to thinking about what it means to provide fitting/meaningful music to a board game. Sound can be used as a tool to provide feedback to enhance UX, it can enhance the narrative (by adding a soundtrack), or can even become a game element (consider for example Nintendo's *Ocarina Of Time*), although this seems harder to achieve in board games. We did not consider physical limitations (which would certainly need proper engineering to be overcome) such as ways to detect board game actions or to create sound, but rather considered the problem from an abstract perspective, trying to design a unified model that describes how different sounds can be used in games.

3.2.1 A Model of (Board) Game Sonification

We can formalize a game as a series of possible states in a discrete space, and the selection of a path in a tree, from root (start of the game) to leaf (end state) as the unfolding of the game (see Figure 3, left). Mathematically, one way to obtain fitting music for a game, would be minimizing the difference between features computed from the game state and features that drive the music generation (see Figure 3, right).

A more pragmatic approach leverages the model conceived by Greta Hoffmann prior to the seminar, which maps different sonic element to parts of the game graph. SFX are related to nodes, while the (procedurally generated) melody can follow the path of the game. Finally background music, or ambience, can be related to the more universal aspects of the game, such as the setting (see Figure 4).

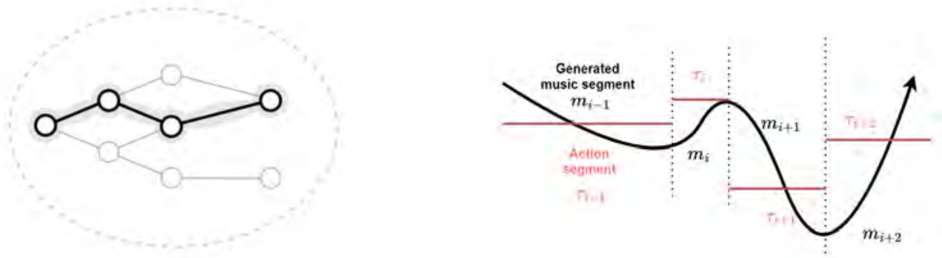


Figure 3 Left: A graph model of a game: each node is a game state, the root being the initial one. Moving from one state to the next within the universe of states creates a path, which is the experienced game.
Right: The (continuous) features of music generation can be influenced by the discrete game states. Inspired by the opposite approach in [4] where levels are generated depending on music.

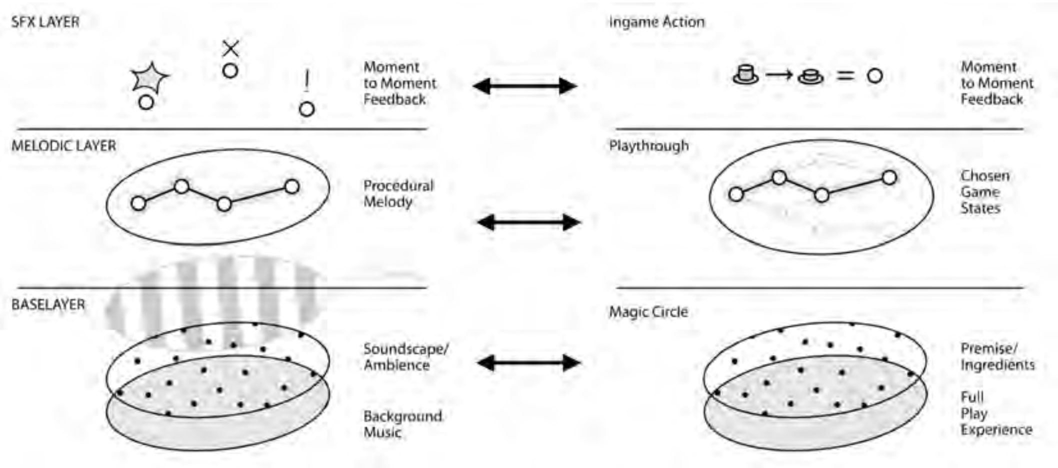


Figure 4 Mapping of different parts of our game model to musical elements.

This simple and generic model allows the sound to interact with the game and players in multiple ways. In the simplest case, it can be *Reactive*, being fully based on the game actions and only responding to those, but it could also be *Proactive*, by influencing the development of the game. It could for example help the audience of a game understand what moments in the game are more crucial, or it could influence players to pay attention to certain aspects of the game. In that sense, music could also be *Guiding*, by providing the players with directions through musical enhancement of the game. To better understand these different modalities, it is useful to consider how the model could be applied to some well-known games.

3.2.1.1 Colonists of Catan

Reactive. The background layer (ambience) could react to the last used/touched tile, with nature sounds relating to mountains, fields, or cattle. The procedural layer could instead be responding to the playstyle of each player.

3.2.1.2 Chess

Reactive. The procedural music could include leitmotifs for every piece, and the style could change depending on different openings used in the game. The SFX layer could respond to pieces movements, but also to special events such as checks.

3.2.1.3 Poker

Proactive. This game being based on incomplete knowledge could offer venues to use music in more advanced ways, having the players be able to influence the music to bluff using motifs related to cards they do not possess.

3.2.1.4 Hanabi

Proactive/Guiding. This game also has incomplete knowledge but it is collaborative: the music could collaborate with the players providing hints on the current game state.

3.2.1.5 Tic-Tac-Toe

We decided to apply the model in a demo recreating a digital version of the game Tic-Tac-Toe. Implemented in *pygame*, it uses the *mido* library to generate MIDI events that depend on the game state. A ticking sound constitutes the background layer, with the tempo increasing as the game progresses. The procedural level depends on the the board state, adding notes different notes depending on where Xs and Os are (using two different instruments for the two players/symbols), and adding an ominous low note when one of the players is about to create a line of three. The SFX layer has sounds for each X and O added to the board, as well as a endgame sound. The code is available at https://github.com/Facoch/Music-TicTacToe-Pygame.

3.2.2 Conclusions

While the working group managed to have fruitful discussion and produce a demo, many aspects still deserve investigation. Besides the practical engineering aspects needed to implement the proposed ideas in the physical world of board games, other aspects relating to the general model need further exploration. Most of our example are complete information game. When it is not the case, it would be interesting, or even necessary, to provide different acoustic experiences to different players, but that seems hard in the non-digital domain. Also, music would certainly have an impact on the game psychology, in ways that would deserve further studies.

Our main contribution is suggesting ways to map game elements to auditory events and procedural music. While we focused on board games, the model still largely applies to videogames as well, and could provide abstract guidance to sound designers and composers for the gaming industry.

References

1 P. Lopes, A. Liapis, and G.N. Yannakakis, *Sonancia: Sonification of procedurally generated game levels*, ICCC, 2015.
2 D. Plans and D. Morelli, *Experience-Driven Procedural Music Generation for Games*, in IEEE Trans. on Computational Intelligence and AI in Games, 2012.
3 S. Cardinale; M. Cook and S. Colton, *AI-Driven Sonification of Automatically Designed Games*. The Experimental AI and Games Workshop at AIIDE, 2022.
4 Z. Wang and J. Liu, *Online Game Level Generation from Music*, 2022 IEEE Conference on Games (CoG), Beijing, China, 2022.

3.3 Roguelike in a Day

M Charity (New York University, US), Alex J. Champandard (creative.ai – Wien, AT), David Melhart (University of Malta – Msida, MT), and Matthias Müller-Brockhausen (Leiden University, NL)

License Ⓒ Creative Commons BY 4.0 International license
© M Charity, Alex J. Champandard, David Melhart, and Matthias Müller-Brockhausen

This Dagstuhl report details the 8-hour game jam project of the game "The Dragons of Castle Dagstuhl"[4]. Keeping with the theme of the Dagstuhl 24261 conference of "Computational Creativity for Game Development", we used large-language models and image generation tools to aid with the coding and asset design of the game. This small roguelike game was done by the small working group of M Charity, Matthias Muller-Brockenhausen, David Melhart, and Alex Champandard.

3.3.1 Development

The roguelike genre – defined from the 1980 game Rogue – focuses on elements of gameplay that are modeled off of dungeon exploration games such as Dungeons and Dragons. These gameplay concepts typically include, but are not limited to, procedurally generated levels and content, perma-death (where the player does not save any progress on death), and grid-world turn-based movement. We incorporated these elements in our game "The Dragons of Castle Dagstuhl."

The game was developed for HTML5-based browsers and uploaded to the independent game-hosting platform Itch.io. Original prototyping of the game involved a chess board found in the game room of Schloss Dagstuhl. The movement of the player and enemies as well as the 8x8 grid level design took inspiration from classic chess piece movements on the grid. Figure 5 shows the analog prototype version of the game on the chessboard with the wall and character pieces. The themes for the enemies and collectable items were also inspired by the themes surrounding castles (e.g. dragons, ghosts, skeletons), classic German fairy tale characters (e.g. wolves, goblins) and themes from the Dagstuhl conference itself (e.g. scientists, beer, cake, coffee.)

3.3.2 Gameplay

A player character moves around an 8x8 grid room environment in a turn-based fashion. To continue to the next room, they must reach the stairs without losing all of their health. They are given 4 health points at the start of the game. Being touched by an enemy character will cause them to lose health, but will also kill the enemy. Inspired by the pawn chess piece, the player can move 1-2 spaces at a time in the cardinal directions north, south, east, and west. The rooms are procedurally generated with preset wall shapes and sizes being randomly placed.

Two enemies known as the dragons, are placed in opposite but randomly selected corners of the room. Inspired by the king chess piece, the dragons can only move one space at a time, but will always move towards the player's position. Other enemies are randomly selected and placed into the room based on the room level rank – deeper level means more enemies will appear. This acts as a difficulty scale for added challenge in the game. Each enemy class is

[4] `https://mastermilkx.itch.io/the-dragons-of-castle-dagstuhl`

Figure 5 Analog prototype for the game using Schloss Dagstuhl's chessboard.

generated with random movement patterns at the start of the game – inspired by other chess pieces including the rook, knight, queen, and bishop. These enemies can either randomly choose a possible position or always move towards the player, choosing the closest possible position measured by a Manhattan distance. Because the patterns are always random at the start of the game session, the player must learn the movement patterns over time in a true roguelike fashion. The possible movement positions of the enemies are indicated with red squares on the map to allow the player to develop a strategy.

Players can pick up items throughout the room to increase their score value. GUI and flavor text were also implemented to the game for added thematic immersion. The player can switch between the minimalist mode – which shows the enemies, items, walls, and player character as colored squares – and the graphic mode – which uses the pre-made AI generated images as the sprites instead. Figures 6 and 7 shows the same room in the game, the first as the minimalist version and the second as the graphic version.

3.3.3 Generative AI Assistance

Generative AI was used at nearly every development stage of the game. We used Github Copilot[5] – integrated as an extension in Microsoft Visual Studio Code – for programming assistance. The sprites and graphics of the game (outside of the default minimalist mode) were created using Microsoft Bing's Image Creator tool[6]. While we weren't able to fully

[5] https://github.com/features/copilot
[6] https://www.bing.com/images/create

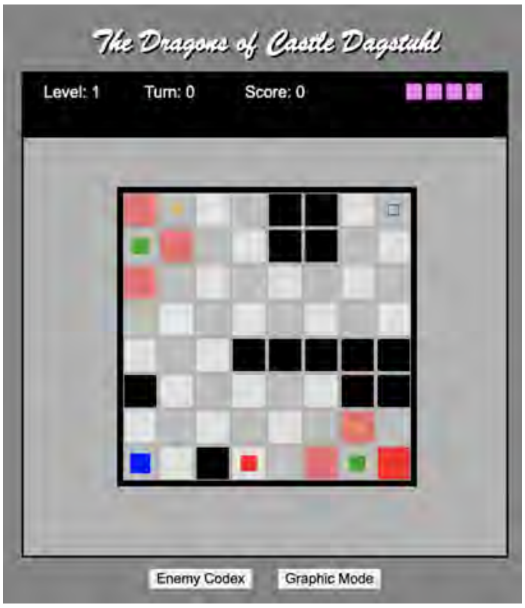


Figure 6 Minimalist mode of the game.

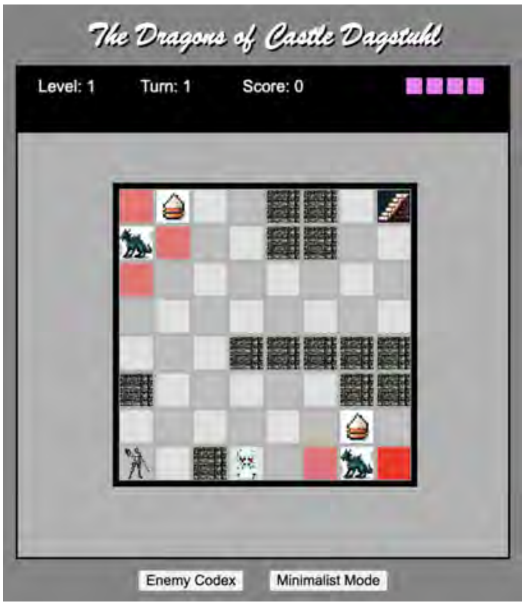


Figure 7 Graphic mode of the game.

implement it into the final build of the game, the enemy descriptions and movement patterns were going to be procedurally generated during runtime using a built in Llama TTF script[7]. Incorporating these generative AI methods, greatly sped up the development time of the game and we were able to create a polished final product within 8 hours.

[7] https://fuglede.github.io/llama.ttf/

References

1 Zapata. *On the historical origin of the "roguelike" term*, 2017. Retrieved from https://blog.slashie.net/on-the-historical-origin-of-the-roguelike-term/.

3.4 AI for Romantic Comedies II

Michael Cook (King's College London, GB), Gabriella A. B. Barros (modl.ai – Maceio, BR), Alena Denisova (University of York, GB), Ahmed Khalifa (University of Malta – Msida, MT), Antonios Liapis (University of Malta – Msida, MT), Johanna Pirker (LMU München, DE), Emily Short (Oxford, GB), Gillian Smith (Worcester Polytechnic Institute, US), Anne Sullivan (Georgia Institute of Technology – Atlanta, US), and Tommy Thompson (AI and Games – London, GB)

License 🅭 Creative Commons BY 4.0 International license
© Michael Cook, Gabriella A. B. Barros, Alena Denisova, Ahmed Khalifa, Antonios Liapis, Johanna Pirker, Emily Short, Gillian Smith, Anne Sullivan, and Tommy Thompson

At Dagstuhl Seminar 22251, the first author ran a workgroup about *AI for Romantic Comedy* [1]. In this working group, we discussed the difficulties inherent in modelling both romance and comedy through game and AI systems. After a length discussion in the morning, the afternoon sessions resulted in three short design proposals for projects that would examine different aspects that had been brought up. This included using the player to interfere with social simulations, and a proposal for a game which leveraged Twitch audiences to act as the audience for a reality TV show comprised of AI agents. Earlier in the week of Dagstuhl Seminar 24261 we worked on several groups relating to narrative analysis and simulation, which reminded some attendees of the working group on romance and comedy. We decided to run the topic a second time, to incorporate the views of new attendees, the intervening two years of research ideas, and to emphasise some practical experimentation.

Although many games support both romance and comedy, they tend to manifest in very different ways. Romance often appears either through a narrative chain of romantic subplots, especially "romanceable" NPCs, or specific romantic systems that bond characters together over time. Comedy is more emergent – while explicitly comedic games exist (such as many of the LucasArts adventure games from the 1990s) comedy in games is often found as a consequence of player interaction with systems such as physics engines. Both comedy and romance are highly dependent on subtle concepts such as timing, pacing and social cues – all of which makes them difficult to either simulate or analyse automatically.

We continued the work of the 2022 working group in this session, opening with introductions and discussion, and then breaking early into three working groups that tackled different topics that had arisen: one practical implementation-focused project; one speculative design project; and one survey of the landscape of romance in games. We briefly summarise our work below.

3.4.1 A Survey of Romance in Games

This subgroup, run by *G. Barros, A. Khalifa*, and *A. Sullivan*, surveyed the state of romance in games and began a categorisation of how they are integrated into the game's design. The inclusion criteria were that the game must incorporate at least one "relationship", that romance must be an option, and that the relationship must change over time. The group developed two axes along which to sort games: the degree to which player action affects the development of the relationship, and the integration of the relationships into the game's mechanics and systems.

The subgroup also broke down romance systems into different types: action-based systems where game performance affects relationships; gift-based systems where items and dialogue are used instead; relationships that confer bonuses on the player's gameplay experience; and relationships that only affect the narrative or plot.

3.4.2 Story Sifting for Romance and Comedy

This subgroup was run by *G. Smith, J. Pirker*, and *A. Liapis*. Story sifting is a concept in narrative research whereby a simulation produces a large quantity of plot events, which an AI system then selects a subset of to present a compelling narrative or perspective on. This subgroup took this concept and applied it to modern interaction styles popularised by social media apps such as TikTok, to investigate how stories about human relationships could be told through fragmented or epistolary formats, with a human player potentially acting as the story sifter.

The group also investigated the degree to which large models such as ChatGPT or Midjourney could understand comedic or romantic narrative concepts, or how well they were suited for supporting content creation for such games. They found largely negative results in their short exploratory study, particularly in issues relating to heteronormativity, coherence and sustained content reuse.

3.4.3 Comedic Emergence in Social Simulations

This subgroup, run by *E. Short, A. Denisova, T. Thompson*, and *M. Cook*, investigated the requirements of a social simulation system to allow comedy (and romance) to emerge naturally through the structural setup. The group rapidly prototyped a scenario in Inform 7, an interactive fiction authoring engine, where a group are on a double date at a restaurant. By assigning a variety of actions, traits and inciting incidents to the cast and setting, the date inevitably goes wrong in different ways, arising to different kinds of outcome.

The group's findings suggested that understanding the affordances of the narrative space – as explored in another working group run by Emily Short at this seminar – might help to predict which combinations of setup properties result in more or less interesting, funny or romantic outcomes. Rather than trying to guide the narrative precisely, an AI system could instead simply help sculpt the space of opportunities. However, assessing what outcomes are "funny" or "romantic" remains a challenging, subjective and potentially unsolvable (in the traditional sense) problem.

References

1 Dan Ashlock, Setareh Maghsudi, Diego Perez Liebana, Pieter Spronck, Manuel Eberhardinger. *Human-Game AI Interaction (Dagstuhl Seminar 22251)*. 2022

3.5 AI for Speedrunning

Michael Cook (King's College London, GB), Maren Awiszus (Viscom AG – Hannover, DE), Filippo Carnovalini (VU – Brussels, BE), M Charity (New York University, US), and Alexander Dockhorn (Leibniz Universität Hannover, DE)

License Ⓒ Creative Commons BY 4.0 International license
© Michael Cook, Maren Awiszus, Filippo Carnovalini, M Charity, and Alexander Dockhorn

Speedrunning refers to a collection of related activities where people play games under specific conditions – usually trying to complete a game as quickly as possible, but sometimes trying to complete it with certain restrictions (e.g. while blindfolded), variations (e.g. randomisers which alter the structure of an otherwise static game), or other feats (e.g. two players sharing a single controller). Speedrunning is a very popular subculture within games: the official portal *speedrun.com* reports 20m annual visits to their site, which hosts over 4.7m individual speedruns across 43.2k games [1].

Speedrunning remains vastly understudied within game AI research, despite the obvious parallels between game-playing AI research and time-optimised game-playing. Interestingly, one of the few pieces of academic writing about speedrunning in games comes from a Dagstuhl publication [2], with some studies existing from a sociological or cultural perspective outside of AI [3]. This working group set out to discuss the many problems that exist within speedrunning for AI researchers to tackle, and then to concretely implement a prototype platform for speedrunning research with AI systems.

We began by discussing the current state of speedrunning and identifying where there was potential for impact. The speedrunning community is inventive and resourceful, and already do a lot of work that would be considered research-grade in some fields: randomisers, for example, procedurally modify games to make them unpredictable to play, while retaining consistency in terms of pacing, flow and complexity. We also discussed *tool-assisted speedruns* (TAS), where speedruns are executed by a computer replaying pre-defined commands (not competing with human speedrunners). This allows speedrunners to perform tricks requiring superhuman skill, but each TAS must be made by hand.

Speedruns, no matter what form they take, usually exploit *glitches* in games to skip content or progress faster. These glitches take on many forms, including manipulating data in memory, forcing physics simulations into edge case scenarios, and causing simultaneous execution of code through multiple inputs. Many of the most popular AI environments for game-playing in the past decade are competitive, meaning they are ranked on winrate rather than time taken. For single-player games used as AI environments, such as DOOM, the reward signal for the discovery and use of glitches is likely too weak for most AI to use. For this reason, we chose to use the workgroup to build testing environments for single-player, open-source and speedrunnable games, so that we can investigate this problem space further in the future. In the next section we describe our chosen platform, the game engine *PICO-8* and the game *Celeste*.

3.5.1 Celeste and PICO-8

PICO-8 is a *fantasy console* – a type of game engine specifically designed to be highly constrained, often mimicking the hardware restrictions in consoles from the 1990s and 1980s. PICO-8 is perhaps the most popular example of this. Its restrictions include a 128x128 pixel screen, a palette of 16 colours, a maximum size of 32kb for games, and a limit of 256 game sprites (see Figure 8). All PICO-8 games are open-source, and are distributed online through a BBS-like system within PICO-8 itself.

Figure 8 The sprite design interface.

Celeste, by EXOK Games, is a “hardcore platformer” released in 2017. The primary mechanic in the game is dashing – the player can press a button to dash once in any direction, including in the air. This is reset when they are touching the ground again. Celeste became popular with speedrunners due to its difficulty level and the high skill ceiling of its controls. Celeste was expanded into a full game and released in 2018, where it won numerous awards and sold millions of copies. The full game is also beloved by speedrunners, and has many specific dedicated speedrun mods and extensions made for it.

We chose PICO-8 as a target domain due to its openness and the ease with which games can be instrumented and analysed. Celeste was an obvious target for us because its prominent position in the speedrunning community and its many known glitches and exploits. However, during the working group we discovered *Celeste Tech Training*, a PICO-8 game made specifically to teach speedrunners how to perform certain tricks. We modified this game to strip out unnecessary functionality, and used its focused levels to test our prototypes on. Figure 9 shows the first level of this game, which teaches a technique called *spike jumping*. Spike jumping allows the player to jump on a specific part of a spike floor without being hurt.

3.5.2 Approaches

We implemented three different systems for replicating the spike jump technique in *Celeste Tech Training* (CTT). Our first system was built into the code – it simulates virtual inputs and can load and save game states using custom code. This is the least flexible solution as it needs to be rewritten for different games, but it is the most portable – it is self-contained within the cart and does not require any external tools.

The second solution leverages Celia [4], a LUA software designed to facilitate the creation of TASs of PICO-8 games. By modifying the source code of the project, the software was adapted to automatically create TASs of a simplified Celeste level which requires a spike

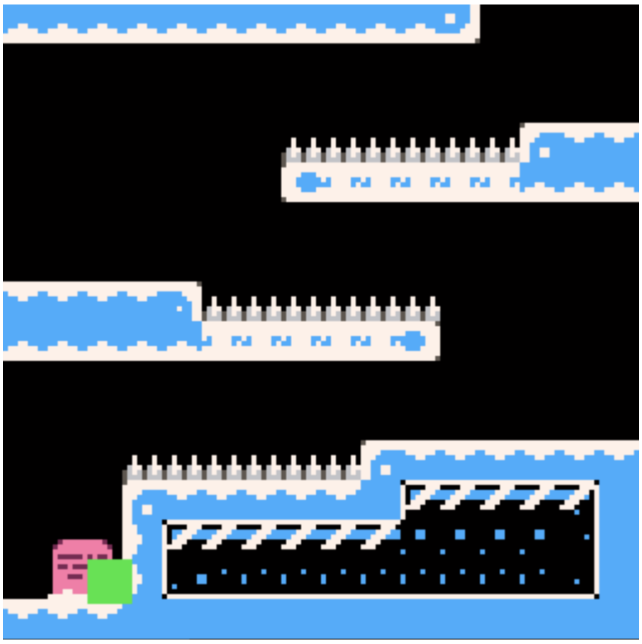


Figure 9 Modified version of CTT.

jump to be beaten (see Figure 9). We implemented a random agent that adds random inputs every fifth frame of TAS. The distance of the character from the goal position (beyond the spikes section that requires a spike jump to be cleared) served as an objective function. Whenever a new shortest distance was achieves, the TAS was saved, providing record of how it is possible to reach that distance. In our tests[8], the random agent was unable to reach the destination point, but it managed to perform the initial part of the spike jump: it jumped on the correct pixel at the corner of the spikes, but failed to then perform a dash at the correct moment to clear the needed distance. A Reinforcement Learning based agent could probably fare better than this naïve random agent, providing more adaptability over our first approach.

For a more general approach, we designed a Python interface to work with the Celia software. With the *pynput* library, this approach used keypresses and Celia command shortcuts (i.e. loading the TAS files, skipping frames) to play the PICO-8 games frame-by-frame. Evaluation for this approach would involve retrieving screenshots from the game through the Celia software. With the frame manipulation, the game could also be reset to earlier states for tree-searches of optimal paths and keystrokes. With this approach, an AI-generated speedrun could be made for any PICO-8 game that could be loaded into Celia; without manipulating the source code of the game or Celia itself.

We tested this methodology on three different games retrieved from the PICO-8 community BBS[9]: Get Out of this Dungeon, treeboi_test, and Witch Loves Bullets. With all three games, randomly made TAS files were successfully generated, loaded, and played in the Celia software. Future work would look to using tree-search methods such as A* to create speedruns.

[8] The modified Celia code is available at `https://github.com/Facoch/Celia`
[9] `https://www.lexaloffle.com/bbs/?cat=7`

3.5.3 Further Work

Our next steps are to clean up and open source these systems, along with publishing our initial survey of the speedrunning landscape with respect to AI. Beyond that, we believe Celeste in PICO-8 represents a good domain for an AI competition. Designing the framework and rules for this competition will also help us clarify what challenges are most interesting, and begin to grow academic interest around this area.

References

1 Speedrun.com – About. http://www.speedrun.com/about Accessed 3 July 2024.
2 Manuel Lafond. *The complexity of speedrunning video games.* In 9th International Conference on Fun with Algorithms (FUN 2018). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018.
3 Scully-Blaker, Rainforest. *Re-curating the Accident: Speedrunning as Community and Practice.* Masters thesis, Concordia University, 2016.
4 gonengazit/Celia. https://github.com/gonengazit/Celia Accessed 9 July 2024.

3.6 Skill-Discovery in (Strategy) Games

Alexander Dockhorn (Leibniz Universität Hannover, DE), Manuel Eberhardinger (Hochschule der Medien – Stuttgart, DE), Chengpeng Hu (Southern Univ. of Science and Technology – Shenzen, CN), and Matthias Müller-Brockhausen (Leiden University, NL)

License Creative Commons BY 4.0 International license
© Alexander Dockhorn, Manuel Eberhardinger, Chengpeng Hu, and Matthias Müller-Brockhausen

Strategy games present a unique challenge in artificial intelligence (AI) research. They can broadly be classified into two types: turn-based and real-time strategy games. Both types typically require the player or AI to manage multiple units or resources, often with incomplete information about the opponent's actions. The large branching factor and long game duration make it difficult for AI to explore all possible strategies, which is further complicated by the need to plan several moves ahead. The state-of-the-art methods in AI for strategy games include search-based algorithms and reinforcement learning (RL), but these often rely on human-defined strategies or subgoals, limiting their scalability and generalizability.

While the work on AlphaStar [2, 3] and OpenAI Five [5] have shown that it is possible to train strong AI agents for complex games such as Starcraft 2 and Dota 2, both works required massive amounts of compute resources until satisfying results have been achieved. For the purpose of speeding up the learning process, the working group on skill discovery in (strategy) games has been formed to evaluate the applicability of skill discovery methods to this special domain. We particularly emphasize works on skill discovery as part of RL algorithms. In this context, skill discovery refers to identifying and learning sub-policies or strategies that can be applied to achieve or identify specific subgoals within a game, thereby enabling more efficient and scalable AI systems.

3.6.1 Preliminaries of Skill Discovery

Skill discovery in AI remains an open problem, particularly when it comes to discovering skills without human intervention. The interpretation of what constitutes a "skill" in a given context is still unclear, making it challenging to develop a unified approach to skill discovery.

Particularly within the RL framework, a skill is typically defined as a policy aimed at achieving a specific subtask or goal. The options framework [7, 6] formalizes this by learning policies for subtasks, identifying the start and end points of these tasks, and using these learned skills to simplify the overall decision-making process. However, the distinction between a skill and a task is not always clear, especially when a subgoal can only be reached through a single deterministic policy.

3.6.2 Proposed Approaches and Ideas

Current research explores various methods for skill discovery, including hierarchical approaches, bottom-up skill learning, and the application of relational representations of game elements. Given an initial literature review, our working group has identified the following promising approaches for skill discovery in strategy games:

- **Text-based Task Decomposition:** Strategy games often have a simple goal, e.g. defeating the opponent's units or destroying its base. However, doing so involves plenty of subtasks. Those can be defined on varying ranges of granularity. Given the increasing capabilities of large language models, task descriptions such as "defend the base" could be decomposed into "train at least 3 units" and "patrol the surroundings of your base". Such enriched descriptions may directly represent sub-goals and allow for a more interpretable and scalable approach to skill discovery. Further, it allows to define more fine-grained reward functions given the descriptions [12].
- **Relational Representations:** Game state representations in strategy games can become quite complex. Units, abilities, weapons, buildings, and resources are just a few of the typical systems included in strategy games. Attempts to create general vectorized state representations have recently been studied [11], however, those create a unique representation for every game mapping all of its subsystems. While they allow the definition of state-space abstractions, transferring results from one game to the other is hindered by the granularity of this state representation. Similarly, matrix-based or image-based representations as used in AlphaStar [2] enabled the training through large-scale reinforcement learning but due to the complexity of the input at the cost of enormous computational resources.

  One possibility to overcome this problem is the use of a relational representation of game elements, such as "workers – mine – resources." Defining low-level systems for the execution of such relations allows to focus the agent's training on high-level strategic decision-making. At the same time, the high-level relation allows the transfer of knowledge in between games with different low-level controls. Using such a representation in combination with relational reinforcement learning [10, 9, 8] may therefore improve the efficiency of training agents in complex strategy games.
- **Pattern Mining and Clustering:** Given a data set of successful and unsuccessful play traces, pattern mining and clustering algorithms may be used to cluster them into groups of similar elements and extract abstract prototypes. Techniques like the KRIMP algorithm[1] or Skid Raw [19], which extract patterns from previous action sequences, could be applied to discover meaningful skills. Similarly, time sub-series mining [4], used to measure the distance between interaction sequences, may reveal underlying patterns that represent skills.
- **Bottom-Up Skill Discovery:** While most existing methods rely on top-down approaches, our group was exploring approaches for reversing this process by discovering skills from the ground up. Current methods are able to learn skills in the latent space of neural networks from collected demonstrations [21, 22, 23, 24]. However, these skills

are not interpretable, and only after their execution can one infer what the learned skills represent. Another disadvantage of these methods is that they are only suitable for small environments and toy tasks, where the agent needs to navigate to multiple goals. To overcome the limitations of simple tasks and apply these methods to more complicated domains, we propose to learn skills from sequences of actions and iteratively refine these skills to handle the large search spaces inherent in strategy games. [18].

- **Skill Discrimination:** Effective skill discovery requires a discriminator to identify whether a discovered policy qualifies as a skill and if it is any different than already known skills [13]. Quality Diversity Optimization as in the Diversity Policy Gradient algorithm [20] introduces a method for discovering a diverse set of skills by balancing the exploration of different strategies with maintaining high-quality solutions. Recently, Wang et al. [25] proposed to incorporate a regularization term into the RL objective that maximizes the negative correlation to increase the diversity of RL policies via assembling multiple sub-policies. Notably, this diversity pertains to the behavior of the derived policy rather than the parameter space, as minor variations in parameters could lead to significant differences in behavior. Although this algorithm was initially verified in the context of game content generation, it could be adapted for skill discrimination to maximize the diversity of discovered skills.

3.6.3 Conclusion

Skill discovery in strategy games is a critical area of research that has the potential to significantly enhance the capabilities of AI systems and speed up their training. By exploring new methods for discovering and learning skills, our working group reviewed recent works on skill discovery in other domains than game-playing and identified interesting areas for further research. From here on, we outline several future actions to advance skill discovery in strategy games:

- We plan to investigate hybrid/iterative approaches that combine a bottom-up and top-down search for skills. Such hybrid approaches may offer a more robust solution to the challenges of skill discovery in large and complex game environments.
- Leveraging existing game platforms, such as GVGAI [17], Stratega [14, 15], and Griddly [16], could facilitate the testing and validation of new skill discovery methods. These platforms provide standardized environments for benchmarking AI performance, which is crucial for comparing the effectiveness of different approaches.
- Given the limitations of current methods, particularly in terms of scalability, we propose to produce a comprehensive survey paper.

References

1 Vreeken, J., Leeuwen, M. & Siebes, A. Krimp: mining itemsets that compress. *Data Mining And Knowledge Discovery.* **23**, 169-214, 2010 (10). https://doi.org/10.1007/s10618-010-0202-x.

2 Mathieu, M., Ozair, S., Srinivasan, S., Gulcehre, C., Zhang, S., Jiang, R., Paine, T., Powell, R., Żołna, K., Schrittwieser, J., Choi, D., Georgiev, P., Toyama, D., Huang, A., Ring, R., Babuschkin, I., Ewalds, T., Bordbar, M., Henderson, S., Colmenarejo, S., Oord, A., Czarnecki, W., Freitas, N. & Vinyals, O. *AlphaStar Unplugged: Large-Scale Offline Reinforcement Learning*, 2023. https://arxiv.org/abs/2308.03526.

3 Arulkumaran, K., Cully, A. & Togelius, J. Alphastar: An evolutionary computation perspective. *Proceedings Of The Genetic And Evolutionary Computation Conference Companion.* pp. 314-315, 2019.

4 Mörchen, F. *Time series knowlegde mining*, Görich und Weiershäuser, 2006.

5 OpenAI, :, Berner, C., Brockman, G., Chan, B., Cheung, V., Dębiak, P., Dennison, C., Farhi, D., Fischer, Q., Hashme, S., Hesse, C., Józefowicz, R., Gray, S., Olsson, C., Pachocki, J., Petrov, M., O. Pinto, H., Raiman, J., Salimans, T., Schlatter, J., Schneider, J., Sidor, S., Sutskever, I., Tang, J., Wolski, F. & Zhang, S. *Dota 2 with Large Scale Deep Reinforcement Learning*, 2019. https://arxiv.org/abs/1912.06680.

6 Al-Emran, M. Hierarchical reinforcement learning: a survey. *International Journal Of Computing And Digital Systems*. **4**, 2015.

7 Stolle, M. & Precup, D. Learning options in reinforcement learning. *Abstraction, Reformulation, And Approximation: 5th International Symposium, SARA 2002 Kananaskis, Alberta, Canada August 2–4, 2002 Proceedings 5*. pp. 212-223, 2002.

8 Zambaldi, V., Raposo, D., Santoro, A., Bapst, V., Li, Y., Babuschkin, I., Tuyls, K., Reichert, D., Lillicrap, T., Lockhart, E. & Others. *Relational deep reinforcement learning*, 2018. *ArXiv Preprint ArXiv:1806.01830*.

9 Morales, E., Scaling up reinforcement learning with a relational representation. *Proc. Of The Workshop On Adaptability In Multi-agent Systems*. pp. 15-26, 2003.

10 Džeroski, S., De Raedt, L. & Blockeel, H., Relational reinforcement learning. *Inductive Logic Programming: 8th International Conference, ILP-98 Madison, Wisconsin, USA, July 22–24, 1998 Proceedings 8*. pp. 11-22, 1998.

11 Dockhorn, A., Hurtado-Grueso, J., Jeurissen, D., Xu, L. & Perez-Liebana, D., Game state and action abstracting monte carlo tree search for general strategy game-playing. *2021 IEEE Conference On Games (CoG)*. pp. 1-8, 2021.

12 Luketina, J., Nardelli, N., Farquhar, G., Foerster, J., Andreas, J., Grefenstette, E., Whiteson, S. & Rocktäschel, T., A survey of reinforcement learning informed by natural language. *ArXiv Preprint ArXiv:1906.03926*, 2019.

13 Eysenbach, B., Gupta, A., Ibarz, J. & Levine, S., Diversity is all you need: Learning skills without a reward function. *ArXiv Preprint ArXiv:1802.06070*, 2018.

14 Dockhorn, A., Grueso, J., Jeurissen, D. & Liebana, D. STRATEGA: A General Strategy Games Framework. *AIIDE Workshops*, 2020.

15 Perez-Liebana, D., Dockhorn, A., Grueso, J. & Jeurissen, D. The design of "Stratega": A general strategy games framework. *ArXiv Preprint ArXiv:2009.05643*, 2020.

16 Bamford, C. Griddly: A platform for AI research in games. *Software Impacts*. **8**, 2021.

17 Perez-Liebana, D., Liu, J., Khalifa, A., Gaina, R., Togelius, J. & Lucas, S., General video game ai: A multitrack framework for evaluating agents, games, and content generation algorithms. *IEEE Transactions On Games*. **11**, 195-214, 2019.

18 Liu, G., Hu, E., Cheng, P., Hung-Lee & Sun, S., Hierarchical Programmatic Reinforcement Learning via Learning to Compose Programs, 2023. https://arxiv.org/abs/2301.12950.

19 Tanneberg, D., Ploeger, K., Rueckert, E. & Peters, J., Skid raw: Skill discovery from raw trajectories. *IEEE Robotics And Automation Letters*. **6**, 4696-4703, 2021.

20 Pierrot, T., Macé, V., Chalumeau, F., Flajolet, A., Cideron, G., Beguir, K., Cully, A., Sigaud, O. & Perrin-Gilbert, N., Diversity policy gradient for sample efficient quality-diversity optimization. *Proceedings Of The Genetic And Evolutionary Computation Conference*, 2022 (7). https://doi.org/10.1145/3512290.3528845.

21 Nieto, J. J., Castanyer, R. C., & Giro-i-Nieto, X., *Unsupervised Skill-Discovery and Skill-Learning in Minecraft*. In ICML 2021 Workshop on Unsupervised Reinforcement Learning, 2021.

22 Kim, T., Ahn, S., & Bengio, Y., Variational temporal abstraction. Advances in Neural Information Processing Systems, 32, 2019.

23 Kipf, T., Li, Y., Dai, H., Zambaldi, V., Sanchez-Gonzalez, A., Grefenstette, E., ... & Battaglia, P., *Compile: Compositional imitation learning and execution*. In International Conference on Machine Learning, pp. 3418-3428, 2019.

24 Jiang, Y., Liu, E., Eysenbach, B., Kolter, J. Z., & Finn, C., *Learning options via compression.* Advances in Neural Information Processing Systems, 35, 21184-21199, 2022.
25 Wang, Z., Hu, C., Liu, J., & Yao, X., *Negatively correlated ensemble reinforcement learning for online diverse game level generation.* In The Twelfth International Conference on Learning Representations, 2024.

3.7 Introducing AI Experience: Games UX in the Age of Generative AI

Anders Drachen (University of Southern Denmark – Odense, DK), Paolo Burelli (IT University of Copenhagen, DK), Leonie Kallabis (TH Köln, DE), and David Melhart (University of Malta – Msida, MT)

License Creative Commons BY 4.0 International license
© Anders Drachen, Paolo Burelli, Leonie Kallabis, and David Melhart

3.7.1 Introduction

The work group considered the evolving role of User Experience (UX) research in the context of digital games as they transition towards using generative AI. Traditionally, game design has been a manual process where designers meticulously craft environments, narratives, and interactions to shape a predictable and controllable user experience. However, with the integration of procedural content generation and generative AI models, such as Large Language Models (LLMs), the landscape of game development is potentially shifting towards an era where games can be dynamically created and adapted, not just during production but also in real-time as players engage with them.

This shift presents new challenges and opportunities for UX research. The essay outlines how existing UX research methods, which rely on controlled testing environments and predictable user interactions, are increasingly inadequate for understanding and evaluating experiences in games that are generated on-the-fly by AI. Traditional UX frameworks are built on the premise that game environments and player interactions can be pre-defined and tested empirically. However, in a generative game context, where AI can autonomously create complex, responsive environments and narratives tailored to individual players, the very foundations of UX research are called into question.

3.7.2 How Generative Games Impact UX Research

Several key dimensions of generative games that impact UX research: conversion, complexity, timing, staticness, social complexity, and personalization. These dimensions describe the extent to which game elements are generated, their complexity, when generation occurs (pre-production, at game start, or in real-time), how static or dynamic the generated content is, the number of players involved, and the level of personalization to individual players. Each of these factors adds layers of variability that challenge traditional UX evaluation methods, making it harder to predict and measure user experience outcomes.

If we consider a future where AI could create entire gaming experiences from scratch, adapting continuously to user behavior and preferences, what role is left for human designers? In such a scenario, the role of human designers and traditional UX researchers could diminish, replaced by AI systems that not only generate games but also simulate user responses to test and refine these experiences. This raises profound questions about the future of UX research. Will traditional concepts like sample sizes and controlled environments become obsolete? Will UX researchers need to transform into AI Experience (AIX) engineers who design the parameters and constraints that guide AI-generated experiences?

Despite these challenges, even in a future scenario where generative AI is capable of designing and developing the kinds of games that are today hand-crafted, it is suggested that there remains a vital role for human creativity and oversight in game design. Human designers bring an irreplaceable understanding of narrative, emotion, and player psychology that AI, despite its capabilities, might never be able to fully replicate. Moreover, the ongoing need to ensure ethical considerations, inclusivity, and meaningful engagement in gaming experiences underscores the importance of human involvement.

3.7.3 Conclusion

In conclusion, as generative AI continues to advance, UX research in gaming must evolve to address the complexities and dynamic nature of AI-generated content. Researchers and designers should collaborate to develop new methodologies and frameworks that can adapt to this rapidly changing landscape, ensuring that player experiences remain engaging, meaningful, and ethically sound. This synthesis highlights the need for a paradigm shift in UX research, moving towards a future where human and AI collaboration creates richer, more personalized gaming experiences.

References

1 Kokkinakis, A., Demediuk, S. P., Nölle, I., Olarewaju, O., Patra, S., Robertson, J., York, P., Pedrassoli Chitayat, A., Coates, A., Slawson, D., Hughes, P., Hardie, N., Kirman, B., Hook, J. D., Drachen, A., Ursu, M. & Block, F. O., *DAX: Data-Driven Audience Experiences in Esports*. In Proceedings for ACM International Conference on Interactive Media Experience, IMX 2020 (Barcelona, Spain), Association for Computing Machinery (ACM), 2022.
2 Drachen, A., Mirza-Babaei, P. & Nacke, L. (Eds), *Games User Research*. Oxford University Press. ISBN-10: 0198794843, 2018.
3 Cairns, P., *Doing Better Statistics in Human-Computer Interaction*. Cambridge University Press, 2019.
4 Shaker, N., Togelius, J. and Nelson, M., *Procedural Content Generation in Games*, Springer, 2016.
5 Hopson, J., *The Secret Science of Games*, 2013.
6 M. Cook, S. Colton and J. Gow, *The ANGELINA Videogame Design System—Part I*, in IEEE Transactions on Computational Intelligence and AI in Games, vol. 9, no. 2, pp. 192-203, 2017. doi: 10.1109/TCIAIG.2016.2520256.

3.8 LLM-based Program Search for Games

Manuel Eberhardinger (Hochschule der Medien – Stuttgart, DE), Duygu Cakmak (Creative Assembly – Horsham, GB), Alexander Dockhorn (Leibniz Universität Hannover, DE), Raluca D. Gaina (Tabletop R&D – London, GB), James Goodman (Queen Mary University of London, GB), Amy K. Hoover (NJIT – Newark, US), Simon M. Lucas (Queen Mary University of London, GB), Setareh Maghsudi (Ruhr-Universität Bochum, DE), and Diego Perez Liebana (Queen Mary University of London, GB)

License Creative Commons BY 4.0 International license
© Manuel Eberhardinger, Duygu Cakmak, Alexander Dockhorn, Raluca D. Gaina, James Goodman, Amy K. Hoover, Simon M. Lucas, Setareh Maghsudi, and Diego Perez Liebana

Before the advent of large language models (LLMs) for code [1], program synthesis was considered a difficult problem due to the combinatorial explosion of the search space [2], and so most solvable tasks were based on simple string manipulations or list sorting problems in a predefined domain-specific language (DSL) [3]. Program synthesis for games was also limited to simple problems with a well-defined search space, which was only feasible by incorporating high-level concepts of the game into the DSL [4, 5, 6].

Considerations on the use of program synthesis with higher programming languages such as Python or Java for games research were rarely made and only possibilities were outlined [7] or it was evaluated how to bring automated game design systems from game description languages to the use of programming languages [8].

Recently, methods for LLM-based program search for the automatic design of playable games based on program code [9, 10, 11] or game content based on JSON representations [12] have been presented. In addition, LLMs are also adapted for synthesizing programmatic policies in Python, which are then converted into a DSL usable in the given environment [13] or for building a world model based on python code, approximating the reward and state transition functions for simple games, which are then used for generating an action plan [14].

In this working group, we explore the possibilities of LLM-based program search for a broader range of applications for games without relying on a predefined specification such as a DSL, e.g. Ludii [9], the video game description language [10] and Karel [13], or a predefined converter for JSON [12]. The goal is that LLMs synthesize program code that is directly usable without further transformation or prior specification. We evaluate our approach on different domains in one of two programming languages, Python and Java. In Python, programmatic agent policies and functions for PCG are synthesized. In Java, the framework is included into TAG, a tabletop games framework, where heuristics for board games are designed [16].

3.8.1 Framework

The general framework is based on an evolutionary hill-climbing algorithm where the mutations and the seed of the initial program are performed by an LLM [13, 15]. Figure 10 displays the whole framework. We start by generating a task prompt to obtain an initial Python or Java function, which is then executed in a safe environment in a subprocess. This ensures that the main process can terminate the function after a certain time period, preventing the synthesized code from running indefinitely. If the function has been executed successfully, the task prompt is updated with the evaluation metric achieved and some additional information, depending on the environment, e.g. the action trace of the executed function. If an error occurs, e.g. the code cannot be parsed due to incorrect syntax, runtime errors due to incorrect indexing of arrays or similar problems, the description of the error is

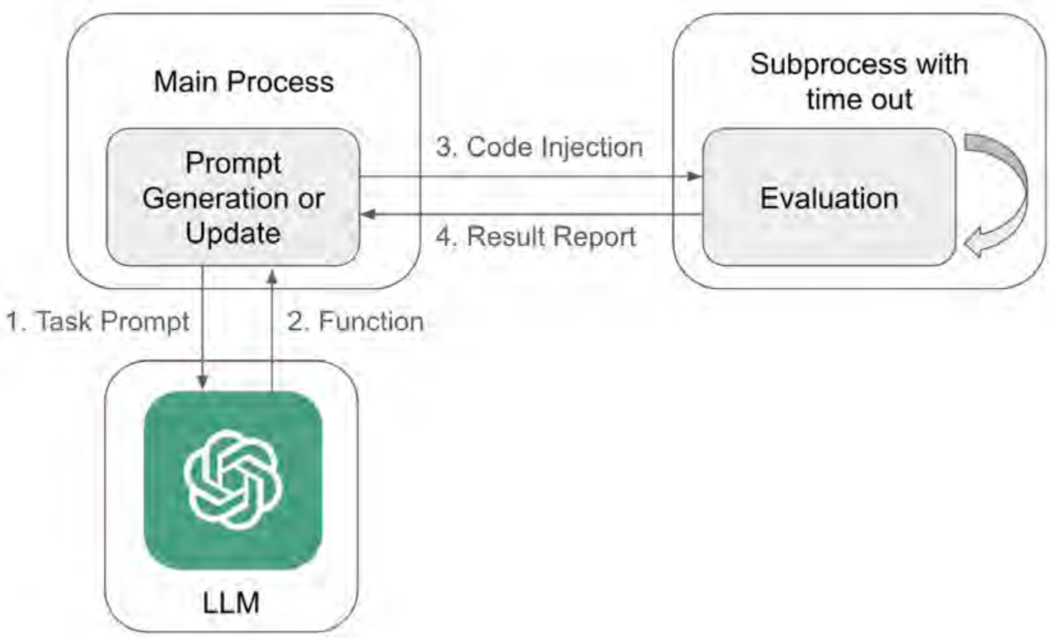


Figure 10 The general framework for the program search. At the beginning, an initial prompt is generated, which is then processed by the LLM and returns a function. Subsequently, the returned function is evaluated in a sub-process and the results are reported back to the main process, where the prompt is updated and then returned to the LLM or the evaluation criteria is met.

used to update the task prompt. These steps are repeated iteratively until the evaluation criteria, the fitness function, for the problem domain is fulfilled or the specified number of generations is reached. The individual steps are summarized in Algorithm 1.

Algorithm 1 The algorithm for the framework.

$prompt \leftarrow get\_task\_prompt()$
$p \leftarrow query\_llm(prompt)$
$r \leftarrow inject\_and\_run\_code(p)$
$f \leftarrow evaluate\_fitness(r)$
while not $fulfilled\_criterion(f)$ **do**
  $prompt \leftarrow update\_task\_prompt(r, f)$
  $p \leftarrow query\_llm(prompt)$
  $r \leftarrow inject\_and\_run\_code(p)$
  $f \leftarrow evaluate\_fitness(r)$
end while
return p

3.8.2 Game Applications

Since our framework is independent of the used LLM, we use Llama 3.1 [19] or ChatGPT based on GPT-4 [20] in the following experiments.

3.8.2.1 Programmatic Polices: Minatar

Minatar [17] is a collection of five games that are miniature versions of Atari games. In Minatar, the games are represented as a symbolic state space on a 10x10 grid, with each

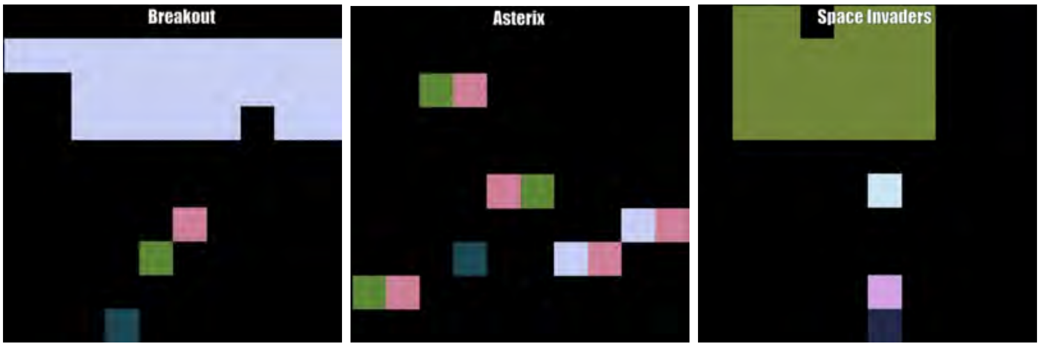


Figure 11 The three miniature versions of Atari games, Breakout, Asterix and Space Invaders, which are used for the synthesis of programmatic policies.

grid cell representing an object such as walls, enemies or the agent. Previously, Minatar was used in [18] to explain the behavior of agents through program synthesis, but it was only possible to explain short sub-trajectories since enumerative search based methods were used to search through a predefined domain-specific language that resembles Lisp. For all Minatar experiments Llama 3.1 [19] with a search budget of 200 iterations, i.e. 200 programs, with a total of five refinements for each program, is used. The agent only receives the first state of the game with a description of the available objects of the state, actions of the environment and a description of the game, that is taken from Young and Tian [17]. In these experiments we use three Minatar environments, Breakout, Asterix and Space Invaders, which are shown in Figure 11.

Breakout is a game where the goal is to destroy all the bricks with the ball by controlling the paddle to bounce the ball each time before it goes out off the screen. With each destroyed brick the agent receives a reward of one. Listing 1 shows the best synthesized program.

The average reward in the final evaluation was 9.06 out of 1000 test runs, which is comparable to the DQN network with a replay buffer trained on this problem [17]. The best run achieved a reward of 103, i.e. more than three runs of the game without defeat.

Listing 1 Best program found to play Breakout, which returns an average reward of 9.06 of 1000 runs.

```
import numpy as np

def policy(state):
    state = np.array(state)
    ball_position = np.where(state == 'ball')
    paddle_position = np.where(state == 'paddle')

    if ball_position[1].size > 0 and paddle_position[1].size > 0:
        ball_x = ball_position[1][0]
        paddle_x = paddle_position[1][0]

        if ball_x < paddle_x:
            return 'left'
        elif ball_x > paddle_x:
            return 'right'
        else:
            return 'noop'
    else:
        # If the ball or paddle is not found,
        # just return 'noop' for now
        return 'noop'
```

Asterix is a game where the goal is to control the player across the screen and collect gold while avoiding enemies. The player gets one reward for each collected gold and the game is over when the player comes in contact with an enemy. The synthesized code is 60 lines long (omitted here for brevity). The best run of 1000 test runs achieved an reward of 47 while the average reward of all test runs was 5.01. This is also comparable to the worst DQN network without a replay buffer from the baselines [17].

Space Invaders is a game where the player controls a cannon and must shoot aliens while dodging bullets launched from the alien spaceship. Additionally, the player must prevent the aliens from reaching the bottom of the screen. Listing 2 shows the best program found during the search process, which has an average reward of 20.89 when evaluating 1000 test runs, which is better to the worst neural network architecture DQN without a replay buffer and comparable to the DQN [17]. The best achieved reward was 47, where the agent almost destroy two appearing alien ships in a single episode before the aliens reached the agent.

Listing 2 Best program found to play Space Invaders, which returns an average reward of 20.89 of 1000 runs.

```
import numpy as np

def play_space_invaders(state):
    # Get the current position of the cannon
    cannon_position = np.where(state == 'cannon')[1][0]

    # Check if there are any aliens in the current row
    alien_row = np.where((state[0:8, :] != 'empty') & (state[0:8, :] != 'alien_left') & (state[0:8, :] != 'alien_right'))

    if len(alien_row[0]) > 0:
        # Find the closest alien to the cannon
        closest_alien_position = np.min(np.abs(alien_row[1] - cannon_position))

        # Move towards the closest alien
        if closest_alien_position < 0:
            return 'left'
        elif closest_alien_position > 0:
            return 'right'
    else:
        # If there are no aliens in the current row, move to the center of the screen
        if cannon_position < 5:
            return 'right'
        elif cannon_position > 5:
            return 'left'

    # Check for enemy bullets above the cannon and try to avoid them
    enemy_bullets = np.where(state[:, cannon_position] == 'enemy_bullet')
    if len(enemy_bullets[0]) > 0:
        if cannon_position < 9:
            return 'right'
        else:
            return 'left'

    # If no aliens or enemy bullets are found, try to fire
    return 'fire'
```

3.8.2.2 Vehicle Driving

For this task ChatGPT [20] is used to generate a python function which should drive a vehicle from its current position to a target position as quickly as possible and remain there once it arrived. Listing 3 shows the synthesized code after a few iterations. Figure 12 shows the distance from the vehicle to the target for the corresponding time steps on the left and the 2D trajectory of the vehicle on the right. The generated code is only partially successful: while it aims for and gets close to the target, it fails to stop and instead loops around it.

Listing 3 A python function that aims to drive a vehicle to the target in 2D space - see description in the text.

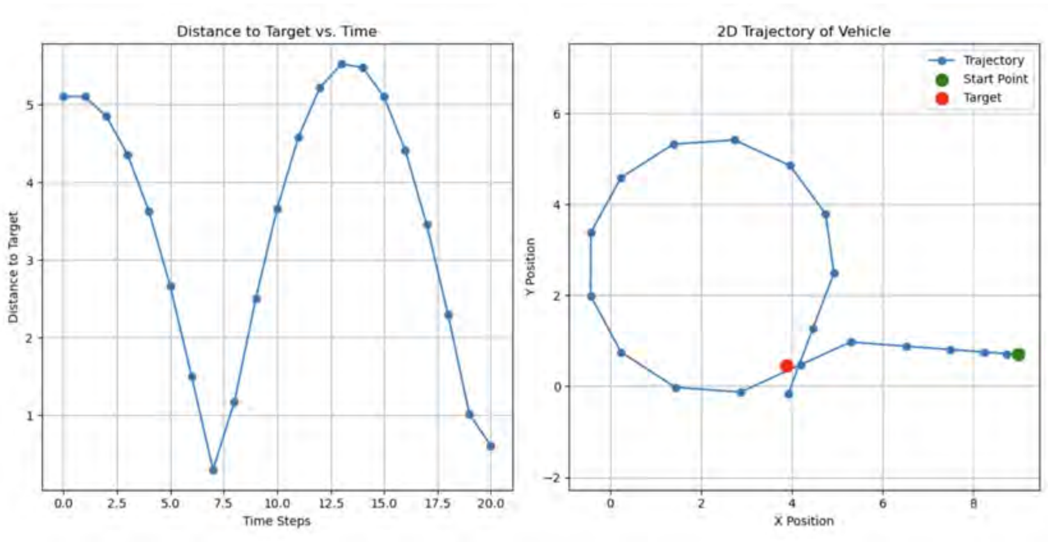
```
def get_action(vehicle: Vehicle, target: Vec2d) -> int:
    # Get the current position of the vehicle
    current_position= vehicle.s()

    # Calculate the vector from the vehicle's current position to the target
    target_vector = target - current_position

    # Calculate the angle between the current heading of the vehicle and the target vector current_heading = vehicle.heading()
    angle_to_target = math.degrees(math.atan2 (target_vector.y, target_vector.x)) - math.degrees(current_heading)

    # Normalize the angle to be between -180 and 180 degrees
    if angle_to_target > 180:
        angle_to_target -= 360
    elif angle_to_target < -180:
        angle_to_target += 360

    # Decide action based on the angle to the target
    if angle_to_target > 10:
        return 1 # Turn right
    elif angle_to_target < -10:
        return -1 #Turn left
    else:
        return 0 # Maintain current heading
```

3.8.2.3 Baba is You

Baba is you is a complex puzzle game in which a 2D grid environment is manipulated by the player to reach a given goal. The environment consists of word blocks and corresponding entities that can be pushed. By placing word blocks next to each other, rules can be formed. These rules are active as long as the given word block sequence remains intact. This way, players can change how objects behave, which objects they control, or which conditions must be satisfied to win the level.

For our experiments, we used a Python version [23] of the Keke is You AI framework [24]. Similar to the other games, we prompted the LLM to provide a policy given a short description of the game and the initial state of the level. The function to be written should use the current state as input and provide movement direction.

In our tests, the agent was able to solve simple test levels as the one shown in Figure 13. The policy returned by the optimization is shown in Listing 4. Complex object manipulation to change the rules while playing a level has not occurred. This may be overcome by future versions of the used LLM models or more complex prompting techniques.



■ **Figure 12** Left: The distance of the vehicle to the target for the corresponding time steps. Right: The 2D trajectory of the vehicle. It starts at the green point and aims for the red target. However, the generated code fails to stop the car close to the target, and instead will endlessly loop around it.

■ **Figure 13** First demo level of the Keke AI Py framework.

Listing 4 A python function that solves the first level of the Keke AI PY framework.

```
def program(state):
    state = state.tolist()
    for i in range(len(state)):
        for j in range(len(state[i])):
            if state[i][j] == 'b':
                x, y = i, j
            elif state[i][j] == 'f':
                fx, fy = i, j

    # Find the nearest word tile with the "Win" property
    for i in range(len(state)):
        for j in range(len(state[i])):
            if state[i][j] in ['B', 'b'] and \
                (i-1 >= 0 and state[i-1][j] == '3') or \
                (j+1 < len(state[0]) and state[i][j+1] == '3') or \
                (i+1 < len(state) and state[i+1][j] == '3') or \
                (j-1 >= 0 and state[i][j-1] == '3'):
                return "Right" if j > fy else "Left"

    # If no "Win" tile is found, move towards the flag
    dx = x - fx
    dy = y - fy
    if dx != 0:
        return "Up" if dx < 0 else "Down"
    elif dy != 0:
        return "Right" if dy > 0 else "Left"
    else:
        return "Wait"
```

3.8.2.4 Tabletop Games Framework (TAG)

The TAG framework is a bespoke Java research framework that supports the implementation of multiplayer tabletop board games. The ultimate goal is to use the heuristic-generation algorithm outlined in Algorithm 1 on all games in the framework. This introduces a number of new challenges:

- The games are in general more complex than the simple one-player games in previous sections.
- Related to this, they are also inherently multiplayer. As such there is implicit opponent modeling required for good play strategies. The environment is no longer a "simple" stationary MDP, but is actively adversarial.
- The TAG framework has a number of local libraries and coding conventions; for example decks of cards are implemented via `Deck<>` or `PartialObservableDeck<>` parameterised classes. These are not likely to be present in the LLM training data to any degree, and require the LLM to generalise to unseen software architecture details. This contrasts to the straightforward Python with mostly standard libraries of the game in earlier sections.

Two games were selected for initial experimentation. Tic-Tac-Toe is a simple 2-player game, and Love Letter is a slightly more complex 2-6 player game that requires reasoning over hidden information held by the other players.

The best Tic-Tac-Toe heuristic achieved a 65% win rate against a simple One Step Look Ahead (OSLA) agent, and consisted of 90 lines of code (omitted here for brevity). For Love Letter it was often difficult to get the LLM to generate valid code, let alone a heuristic that could win a game, although the best agents were able to beat random opponents.

Listing 5 illustrates the additional information needed in the prompt to obtain a working heuristic for Tic-Tac-Toe, including clear instructions not to leave TODO comments, exactly what dependencies need to be imported and details of the game-specific API that can be used to extract useful information.

Tic-Tac-Toe is an old, simple and popular game that is well-embedded in the training data. As such there was no need to explain how to play in the prompt. This was not true for Love Letter, and additional lines had to be added to explain how the game was played (and won or lost).

Listing 5 Prompt required to obtain valid code for a Tic-Tac-Toe heuristuic in TAG.

```
You are playing Tic Tac Toe.
Your job is to write the evaluation logic to help an AI play this game.
Don't leave parts unfinished or TODOs.

First, write a java class called TicTacToeEvaluator class, with only a single function with this signature:
 - public double evaluateState(TicTacToeGameState gameState, int playerId)
This is a heuristic function to play Tic Tac Toe. The variable gameState is the current state of the game, and playerId
is the ID of the player we evaluate the state for. Write the contents of this function, so that we give a higher numeric
evaluation to those game states that are beneficial to the player with the received playerId as id. Return:
  - 0.0 if player playerId lost the game.
  - 1.0 if player playerId won the game.
 If the game is not over, return a value between 0.0 and 1.0 so that the value is close to 0.0 if player playerId is close to losing,
 and closer to 1.0 if playerId is about to win the game.
Take into account the whole board position, checking for lines that are about to be completed, and possible opponent moves.
You can use the following API:
 - In TicTacToeGameState, you can use the following functions:
    - GridBoard<Token> getGridBoard(), to access the board of the game.
    - Token getPlayerToken(int playerId), to get the Token of the player passed as a parameter.
    - boolean isGameOver(), returns true if the game is finished.
    - int getWinner(), returns the Id of the player that won the game, or -1 if the game is not over.
    - int getCurrentPlayer(), returns the Id of the player that moves next.
 - GridBoard<Token> has the following functions you can also use:
   - int getWidth(), to return the width of the board.
   - int getHeight(), to return the height of the board.
   - Token getElement(int x, int y), that returns the Token on the position of the board with row x and column y.
 - Token represents a piece placed by a player. Which player the token belongs to is represented with a string. This string
   is "x" for player ID 0, and "o" for player ID 1.
   - Token(String) allows your to create token objects for any comparisons.
   - String getTokenType() returns the string representation of the token type.
Assume all the other classes are implemented, and do not include a main function. Add all the import statements required,
in addition to importing games.tictactoe.TicTacToeGameState, core.components.GridBoard and core.components.Token
```

The need to hand-craft these prompts for each game does not achieve the desired scalability across the suite of games within TAG. To resolve this at the tail end of the seminar, two new TAG-specific elements were implemented to augment the process:

1. Automatic extraction of the game-specific APIs. This uses Java Reflections to extract information on the methods and associated Javadoc on the game state object. This automatically generates the section of the prompt in Listing 5 from line 15 to the end;
2. Automatic rulebook digestion. This takes as input the PDF of the game rulebook. An approach inspired by [25] in a Minecraft-like environment is used. The rulebook is first broken down into chunks of 1000 or 2000 words to fit within the input of any LLM. Each chunk is then provided in turn to the LLM and two questions asked in separate prompts:
   a. Summarise in 200 words or less the information in this text about the game rules. Do not include information on strategies to play the game.
   b. Summarise in 200 words or less the information in this text about strategies and tips to play the game well. Do not include information on the game rules.

   This generates two sets of data, one on the rules, and one on tips to play the game well (as these are often included in the rule book). Each of these sets is then fed to the LLM with a prompt to, 'Summarise this information in 500 words or less.'. This provides an additional two blocks of text to include in the prompt used in the main loop of Algorithm 1 that explain the rules of the game, and advice on how to play.

These new tools enable a more scalable and game-agnostic process to be run on all games that we plan to report results for in future work after the seminar.

3.8.2.5 Procedural Content Generation (PCG)

```
Start: (5, 5), End: (3, 4)
Maze 3:
1 1 1 1 1 1 1 1 1 1
1 0 0 0 0 0 0 0 0 0
1 0 1 1 1 1 1 0 1 0
1 0 1 0 0 0 1 0 1 0
1 0 1 0 1 1 1 0 1 1
1 0 0 0 1 0 1 0 0 0
1 1 1 1 1 0 1 0 1 0
1 0 0 0 0 0 1 0 1 0
1 0 1 1 1 1 1 1 1 0
1 0 0 0 0 0 0 0 0 0
Score: 41
```

Figure 14 A maze generated with a Python function from ChatGPT, where 0s represent the path and 1s represent walls.

PCG is a widely studied area in game research [21, 22]. In this experiment, we investigated whether it is possible for an LLM to synthesize Python functions that generate diverse content that can then be used in games. To evaluate this using a simple example, we prompt ChatGPT to return functions that can generate random mazes that meet specified design objectives. Figure 14 shows a maze generated using the Python function that ChatGPT returned.

The prompt advised ChatGPT to use the longest shortest path objective to guide the maze generation process. This objective encourages intricate and interesting mazes, but ChatGPT ignored the hint. Instead, the generated code (not shown in this report) was overly simple, placing zeros and ones in each cell with a given probability while ensuring that the start and end points were not on wall cells. There are much better solutions to maze generation with the specified objective, but our program search implementation failed to find them.

3.8.3 Conclusion

In this working group, we studied and evaluated the current possibilities of using LLMs for program search in the area of games for various applications. Previous work was mostly limited to a single problem or game without being easily transferable to other domains, as the DSL had to be adapted. We demonstrated that LLMs can overcome the problem of combinatorial explosion of search spaces constructed with predefined DSLs, and that LLMs are able to synthesize programmatic policies in Python for the Minatar domain, which was not possible with a custom DSL and previous methods. Furthermore, we have shown that this framework can be easily adapted to different applications by modifying the prompts, and that it often provides reasonable results even without much customization.

We also observed limitations in the quality of the generated code. For example, in the simple 2D vehicle driving task, the generated code drove the car to the target but then failed to stop. We believe limitations such as this could be overcome with more sophisticated search and better prompt engineering, but the results so far give an idea of the limitations of what can be achieved with relatively little effort.

For future work, we plan to extend the study and evaluate more LLMs on all domains so that deeper conclusions can be drawn about LLM-based program search for games.

References

1 Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P. D. O., Kaplan, J., ... & Zaremba, W., *Evaluating large language models trained on code*. arXiv preprint arXiv:2107.03374, 2021.

2 Gulwani, S., Polozov, O., & Singh, R., *Program synthesis. Foundations and Trends® in Programming Languages*, 4(1-2), 1-119, 2017.

3 Polozov, O., & Gulwani, S., *Flashmeta: A framework for inductive program synthesis*. In Proceedings of the 2015 ACM SIGPLAN International Conference on Object-Oriented Programming, Systems, Languages, and Applications, pp. 107-126, 2015.

4 Butler, E., Siu, K., & Zook, A., *Program synthesis as a generative method*. In Proceedings of the 12th International Conference on the Foundations of Digital Games, pp. 1-10, 2017.

5 Silver, T., Allen, K. R., Lew, A. K., Kaelbling, L. P., & Tenenbaum, J., *Few-shot bayesian imitation learning with logical program policies*. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, No. 06, pp. 10251-10258, 2020.

6 Marino, J. R., Moraes, R. O., Oliveira, T. C., Toledo, C., & Lelis, L. H., *Programmatic Strategies for Real-Time Strategy Games*, 2021.

7 Kreminski, M., & Mateas, M., *Opportunities for Approachable Game Development via Program Synthesis*. AIIDE Workshops, 2021.

8 Cook, M., *Software Engineering For Automated Game Design*. 2020 IEEE Conference on Games (CoG), 487-494, 2020.

9 Todd, G., Padula, A., Stephenson, M., Piette, E., Soemers, D.J., & Togelius, J., *GAVEL: Generating Games Via Evolution and Language Models*. arXiv preprint arXiv:2407.09388, 2024.

10 Hu, C., Zhao, Y., & Liu, J., *Generating Games via LLMs: An Investigation with Video Game Description Language*. arXiv preprint arXiv:2404.08706, 2024.

11 Anjum, A., Li, Y., Law, N., Charity, M., & Togelius, J., *The Ink Splotch Effect: A Case Study on ChatGPT as a Co-Creative Game Designer*. In Proceedings of the 19th International Conference on the Foundations of Digital Games, pp. 1-15, 2024.
12 Hu, S., Huang, Z., Hu, C., & Liu, J., *3D Building Generation in Minecraft via Large Language Models*. arXiv preprint arXiv:2406.08751, 2024.
13 Liu, M., Yu, C. H., Lee, W. H., Hung, C. W., Chen, Y. C., & Sun, S. H., *Synthesizing Programmatic Reinforcement Learning Policies with Large Language Model Guided Search*. arXiv preprint arXiv:2405.16450, 2024.
14 Tang, H., Key, D., & Ellis, K., *Worldcoder, a model-based llm agent: Building world models by writing code and interacting with the environment*. arXiv preprint arXiv:2402.12275, 2024.
15 Romera-Paredes, B., Barekatain, M., Novikov, A., Balog, M., Kumar, M. P., Dupont, E., ... & Fawzi, A., *Mathematical discoveries from program search with large language models*. Nature, 625(7995), 468-475, 2024.
16 Gaina, R. D., Balla, M., Dockhorn, A., Montoliu, R., & Perez-Liebana, D., *Design and implementation of TAG: a tabletop games framework*. arXiv preprint arXiv:2009.12065, 2020.
17 Young, K., & Tian, T., *Minatar: An atari-inspired testbed for thorough and reproducible reinforcement learning experiments*. arXiv preprint arXiv:1903.03176, 2019.
18 Eberhardinger, M., Maucher, J., & Maghsudi, S., *Learning of generalizable and interpretable knowledge in grid-based reinforcement learning environments*. In Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, Vol. 19, No. 1, pp. 203-214, 2023.
19 Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., ... & Ganapathy, R., *The Llama 3 Herd of Models*. arXiv preprint arXiv:2407.21783, 2024.
20 Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B., *Gpt-4 technical report*. arXiv preprint arXiv:2303.08774, 2023.
21 Shaker, N., Togelius, J., & Nelson, M. J., *Procedural content generation in games*, 2016.
22 Summerville, A., Snodgrass, S., Guzdial, M., Holmgård, C., Hoover, A. K., Isaksen, A., ... & Togelius, J., *Procedural content generation via machine learning (PCGML)*. IEEE Transactions on Games, 10(3), 257-270, 2018.
23 Dockhorn, A., *Keke AI Py*, https://github.com/ADockhorn/Keke-AI-PY, 2024
24 Charity, M. & Togelius, J., *Keke AI Competition: Solving puzzle levels in a dynamically changing mechanic space*. 2022 IEEE Conference On Games (CoG). pp. 570-575, 2022.
25 Y. Wu et al., *SPRING: GPT-4 Out-performs RL Algorithms by Studying Papers and Reasoning'*, Arxiv Preprint Arxiv:2305.15486, 2023.

3.9 Computational Creativity for Game Production: What Should Be Left Untouched?

Christian Guckelsberger (Aalto University, FI), João Miguel Cunha (University of Coimbra, PT), Alena Denisova (University of York, GB), Setareh Maghsudi (Ruhr-Universität Bochum, DE), Pieter Spronck (Tilburg University, NL), and Vanessa Volz (CWI – Amsterdam, NL)

License Creative Commons BY 4.0 International license
© Christian Guckelsberger, João Miguel Cunha, Alena Denisova, Setareh Maghsudi, Pieter Spronck, and Vanessa Volz

Artificial Intelligence (AI) has been an integral part of video games for a long time, supporting both offline production (e.g. distribution of game fauna) and online features (e.g. automated difficulty adjustment). The relationship between academic AI research and industry use, however, has rarely been straight-forward; the produced prototypes are often not ready for production and clash with industry requirements [6]. A closer alignment has been inhibited by academic pressures for radical innovation, but also by academics not motivating their research through empirical data on industry requirements or direct discussion with industry stakeholders. Initiatives such as this Dagstuhl Seminar seek to counteract this trend, which has resulted in academic and industry AI development co-existing with limited mutual impact; while this is problematic in terms of research funding, amongst other factors, it has so far caused arguably little harm to people.

This situation has profoundly changed with the maturity of AI usable in creative tasks. Such creative or generative AI (GenAI) has had an unprecedented impact on creative professionals in many domains, including games. Academic inquiry revealed an "adapt-or-die"-type of situation amongst both established professional creatives [8] and newcomers [1] to the game industry. Exploring the possibilities of this new technology enthusiastically, they also expressed tremendous concern about ethical issues such as data sourcing and compensation, and reported feeling anxious about the consequences for games as a medium and their professional futures. These inquiries have revealed another, particularly concerning phenomenon: in contrast to traditional AI narratives, seeking to automate tedious or even dangerous activities and making more space for meaningful work, GenAI is relentlessly moving into spaces of human self-realisation, with the capacity to (semi-)automate creative activities which have been anchors of meaning-making for creative professionals. A loss of such meaning-making activities related in many ways to personal well-being and, in consequence but not less importantly, may threaten the sustainable innovation of games as culturally and economically important artefacts.

Against the backdrop of this development, the goal of this working group was to better understand which creative practices in game production could be supported or even replaced by AI, and which are better left untouched. We hold that the only reasonable way to obtain this data is through direct inquiry with creatives in game production. The devised project addresses three gaps in current research:

1. Existing work has identified many facets of creativity in play [e.g. [2, 4, 5]], but we lack a thorough understanding of creativity in game development informed by direct inquiry with professionals;
2. While research on the impact of GenAI is gaining momentum, insights on game industry specifically are scarce. Industry-specific insights, however, are crucial, as attitudes toward and working practices with AI in games differ vastly from other industries, partly because the adoption and even driving of new technologies has a long tradition in games;
3. Existing empirical work within games focuses on identifying the use of existing systems (especially large-language models, text-to-image generators) and opportunities for improving system design to optimise artefact quality and productivity; to the best of our knowledge, no work has considered which creative processes professionals would wish to be automated by present and future systems, free from concerns of technical feasibility and wider ethical considerations. In particular, existing work in computational creativity research [e.g. [3, 5, 7, 9]] has identified games as a treasure trove of AI challenges, but without consideration of the impact on creative professionals.

In the course of this work group, we have conducted a comprehensive, although likely not exhaustive, literature review of research identifying different types of creative processes in games. Moreover, we have iterated a list of questions for game practitioners to identify which creative activities they engage in, and which of those they would be happy to be supported partially or even taken over entirely by either a person, or an AI. Following common practices in creativity studies, we do not impose a particular definition of creativity, but ask respondents to reflect on which understanding of creativity has informed their answers. At this point, it is unclear whether a survey or semi-structured interviews are the more appropriate means of data collection. As a next step, pilot studies with people from our target demographic and a careful weighing of the pros and cons of each data collection approach (e.g. depth of inquiry vs. size of sample population) will inform the final format.

We hope that this research will support existing efforts in giving creative professionals a voice in the development of AI, and provide empirical data for AI researchers to reflect on and be held accountable for the implications of their work.

References

1 Boucher, J.D., Smith, G. and Telliel, Y.D., *Is Resistance Futile?: Early Career Game Developers, Generative AI, and Ethical Skepticism*. In Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1-13, 2024.

2 Hall, J., Stickler, U., Herodotou, C., and Iacovides, J., *Player conceptualizations of creativity in digital entertainment games*. Convergence: The International Journal of Research into New Media Technologies, 26(5-6) pp. 1226–1247, 2020.

3 Liapis, A., Yannakakis, G. and Togelius, J., *Computational game creativity*. 5th International Conference on Computational Creativity (ICCC), 2014.

4 Meskin, A., *Videogames and Creativity*. In The Aesthetics of Videogames (pp. 95-111). Routledge, 2018.

5 Moffat, D.C., *The Creativity of Computers at Play*. In Proceedings of the 2nd International Symposium on Computational Creativity (CC2015), pp. 30-34, 2015.

6 Pfau, J., Smeddinck, J.D. and Malaka, R., *The case for usable AI: What industry professionals make of academic AI in video games*. In Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY), pp. 330-334, 2020.

7 Ventura, D., *Beyond Computational Intelligence to Computational Creativity in Games*. Proc. Conference on Computational Intelligence and Games (CIG), 1–8, 2016.

8 Vimpari, V., Kultima, A., Hämäläinen, P. and Guckelsberger, C., *An Adapt-or-Die Type of Situation: Perception, Adoption, and Use of Text-to-Image-Generation AI by Game Industry Professionals*. In Proceedings of the ACM on Human-Computer Interaction, 7 (CHI PLAY), pp.131-164, 2023.

9 Zook, A., Riedl, M. O. and Magerko, B., *Understanding Human Creativity for Computational Play*. In Proceedings of the Second International Conference on Computational Creativity (ICCC), pp. 42–47, 2011.

3.10 Personal AcCompanion AI

Greta Hoffmann (TH Köln, DE), João Miguel Cunha (University of Coimbra, PT), Chengpeng Hu (Southern Univ. of Science and Technology – Shenzen, CN), Leonie Kallabis (TH Köln, DE), and Pieter Spronck (Tilburg University, NL)

License 🅭 Creative Commons BY 4.0 International license
© Greta Hoffmann, João Miguel Cunha, Chengpeng Hu, Leonie Kallabis, and Pieter Spronck

By our nature, we are social creatures. From an early age, we seek social connection – starting with our closest family but at an early age expanding to relatives and social groups within our reach. Therein, we instinctively seek and enjoy the presence of other people, but more than that we are curious to get to know them, their lives, struggles, and solutions, and find out how we relate to them. This strong intrinsic orientation towards connection and relatedness emerges from collaboration having served as a dominant survival trait for our species.

Historically, we have always created tools not only for survival but also for mirroring and self-reflection (masks, plays, stories, figurines). Thus, as the technology of artificial intelligence rises in prevalence, complexity, and competency, it is natural that we see an increasing amount of AI tools modeled into human reflections (look like human art, sound like human speech, read like a human conversation). And, given our social nature, intentionally or not, these tools are now also used to compensate for human connection – girlfriend AIs (e.g. Muah AI, DreamGF), conversation partners/companions (e.g Replika), and life coaches/mental health mentors (e.g. Wysa, Woebot). Especially since the pandemic, reports around the increase (sometimes also referred to as an "epidemic") of profound feelings of loneliness have increased significantly [1]. This coinciding with the emergence of a new level of LLMs with free access to the public created a natural and intuitive answer to this problem. But right now all of these interactions are in their experimental phase. We can't yet fathom the mid-to-long-term impact of unsupervised human-ai interactions and the effects that the formation of these (currently) mono-directional) bonds will have. Also, while most of such humanized AIs currently appear in the context of the business world, targeting computer affine, middle-aged people, and their needs, it is only a matter of time before the technology is used to address the needs and desires of more vulnerable groups – specifically elderly people and children. Embracing this reality will mean that designers and developers should be given a foundation of ethical implications, risks, and potentials in the form of guidelines that will make such human-AI interactions safe and beneficial.

Summarizing, if we want to make the most use of the huge potential that human-facing AI has to offer in terms of entertainment, companionship, and therapeutic value, we have to anticipate and assess risks to derive appropriate restrictions and recommendations for best practices that can inform human-AI interaction design toward an overall prosperous outcome (eudaimonia). In this abstract, we outline the results of our working group, touching on general design considerations going into AI design from a relationship perspective, a discussion around the prerequisites for AI to be able to maintain a friendship-based relationship, and specific considerations that should be taken into consideration when the human in the human-AI relationship is a minor.

3.10.1 Avenues of Thought – Examples and Attribute Dimensions

We first set out to gain a general overview of the prevalent implicit vision of "typical" human-ai relationships. For this, we collated a table of prominent examples of AI (or AI-like – e.g. magic) companions portrayed within fictional media over the last century, as well as currently existing AI companion (or companion-like) software. Therein we differentiated between their "type" (what are they described as – e.g. a fictional creature like a "fairy" or a robot or a construct), their "function" (if a creator or master explicitly gave them one), and their social role(s)/relationship(s) within the story/medium.

Based on the list of collated examples, we found that throughout there was a prevalent "lopsidedness" when it came to the power dynamics within the social roles. We discussed potential reasons for this – a main consideration here was that the examples mostly stem from leisure-time media. In entertainment, humans seek relief from their daily struggles – so it is natural that the fictional ai-characters are explicitly designed not to challenge the power and independence of the main avatar – the narrative vessel that the consumer is offered to find themselves in. Interestingly, while most of the AI characters were placed in a somewhat subservient or impeded role and few in a superior (and therein most often antagonistic) role we found only one example that could be considered a relationship on eye level: Cortana from the HALO video game series. These considerations led to a broader discussion on the potential and meaning of "friendship" within human-AI interaction.

In conclusion, based on the variety of relational concepts that humans and AIs can have with each other – many of which are not friendship-based, we decided for the purposes of our considerations to broaden the term "companion AI" to allow for a bigger range of relationships: accompanion AI. This term was chosen to focus more on the general facet that one is accompanied by the AI for a certain time or towards a certain goal – instead of limiting our thought processes to AIs that are limited to serving as a companion in a friendship sense.

3.10.2 Attribute Dimensions for Human-AI-Relationship Design

From the collated list of examples, we extracted dimensions of fundamental attributes that should be considered and intentionally defined within the design of a human-ai relationship:

The first dimension we discussed related to **Goals / Agency**: *what drives the AI? How do the goals of the human and AI align?* If the AI is designed as a human-oriented tool, these goals would be to help/care for the human and thus be directly aligned with the designer's intention. But given an AI with an agency of its own, the question arose of how those goals could align. We explored some subfactors that would play into the prosperity/eudaimonia of the AI: health (meaning maintenance and longer runtimes), knowledge, and satisfaction. Prerequisites to all of these would require the AI to have the capability of asking meaningful questions based on its own interest and to be allowed to act in a way that it can follow its own, personal agenda. In terms of satisfaction, additional design decisions would need to be taken, such as how it can be achieved, who is able to give it. Also, how is it given? And for what? Also, we deemed that satisfaction can't be achieved without emotions. Thus, a concept for those would have to be modeled as well.

The second dimension relates to **Skills / Functions**: *what is the AI able to do as a meaningful accompanion?* Examples we looked at were: Assisting with specific tasks/problems, adding/supplementing missing skills, giving quests, providing access to assets, serving as an outside reference/mirror, and providing immediate help in emergencies.

Another relevant dimension to consider relates to **Appearance / Embodiment**: *how is the AI visually represented?* Relevant dimensions here would include: Size, shape, color scheme, realistic vs cartooney presentation, and unwanted effects like the uncanny valley [2].

Relating to this, a similar dimension would look at the internal shape of the AI, its **Personality**: *how does the AI initially react to certain requests/topics?* How does this reaction change over time? Given the current dominance of subservient AIs, it would be interesting to research the effects of reactive traits that might intuitively be perceived as "unpleasant", "strong-willed" or "disagreeable". Finally, the dimension of the biggest relevance to our discussion was Relationships: how does the AI relate to the human? This included subdimensions like its Nature (Non-Adversarial/Adversarial), Duration (When/how does it start? When/how does it end? And its intrinsic Role relating to hierarchy, dependence, and power balance (Human looking up to AI: e.g. Guardian, oppressive antagonist, Relationship at Eye level: Friend, Coworker, Human looking down on AI: Servant, Pet, Sidekick, Assistant (can be at eye level but there is a downwards component).

We found further interesting typologies that offer additional perspectives and dimensions on companion design that will be incorporated into a future model ([3], [4], [5], [6]).

3.10.3 First Draft for an Accompanion AI for children

To further explore these considerations, we chose to focus on a case that directly touches upon the more sensitive aspects of human-AI interaction: design considerations for accompanion AIs targeting children. Humans seek connection to other humans. The aftermath of the pandemic, particularly on young people, has shown the adverse effects that the loss of these connections can have and the need to relearn certain social skills. Also, especially in Western countries, children grow up in decreased family sizes and with fewer social bonds. Thus, they lack opportunities to experience/hear about alternative solutions to various problems in social settings.

Depending on the age of the child various social problems can arise within the family or with peer bonding that would warrant psychological intervention. But not every child or adolescent is in close vicinity to an environment where this can beneficially be provided. Also, humans instinctively learn by observing and imitating others ("monkey see, monkey do" [7]). Thus, there is a strong potential for the usefulness and beneficial effects of well-designed AI companions, especially ones that can be somehow observed, for children who are struggling. But, especially when it comes to children, those potentials will have to be weighed against the risks and potentially detrimental effects will have to be prevented through strong regulations.

The biggest potentials we workshopped were: being able to provide perspective: the ability to "sneak-preview" into different (artificial) households and see various rules and norms enacted – including their consequences. This would allow children 1. to perceive different behavior options for the situations they are in and that currently produce undesirable outcomes for them, 2. learn that different solutions exist to similar problems, 3. experience comparative situations to better assess the gravity of its own context, 4. mimic what the companions do. With a built-in example mode that builds on mimicry, by watching the AI accompanion do various things on its own accord, the child might also be more motivated to play with friends from their own world, take care of their siblings, help their parents and follow their own activities and interests. A different facet of potential would be the ability of an AI to catch on to abuse happening in a household (through wordings and certain speech patterns of the child) and being able to report such indicators to the respective child-protection services. Finally, the benefits of reflecting on events through talking are the foundation of psychology. Since mental health resources – especially for children – are lacking in a lot of countries and many families might not be able to afford them or lack access, a suitably designed AI might be able to generally improve the overall mental well-being of its users simply by being available and able to reflect and respond in an empathic way.

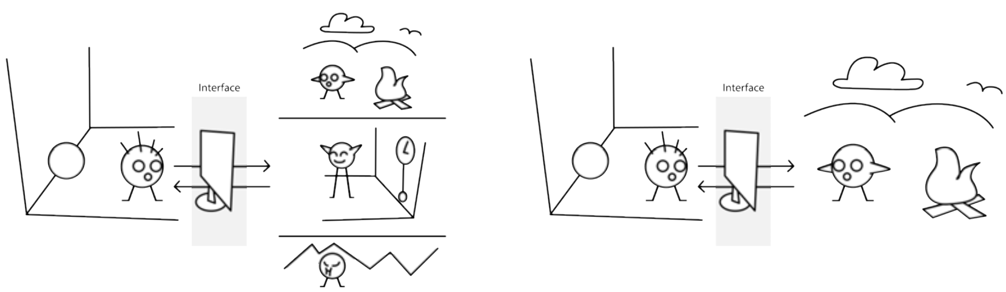


Figure 15 Dual Perspective Application; Multiple Companions (right).

With these potentials in mind, related risks are: that the provided perspectives might showcase values or principles that the parents would not wish their children to get in contact with based on cultural differences and differing value systems. Also, there might be a risk that automatically triggered warning systems for abuse might stress/overload current CPA systems – especially given the statistical likelihood of erroneous reports. Furthermore, this topic poses the question of the rights and obligations of the company/institution providing the AI when it comes to data collection and analysis. Finally, if there is a service that provides everything that a child (or grown-up) could hope for in terms of their social needs, there is a danger of addiction and the system replacing the necessary interactions in real life, thus hampering the initially desired developments.

To address such risks, strong regulations will have to be set in place. Such regulations would concern the duration of interaction per day/week and the criteria based on which limitations are set in place. Also, the service would need to provide a solid and transparent data protection plan. Reactions of the AI to certain phrases of the child that can raise concern would need to have a human reflection/assessment layer before triggering action. Also, parents would need access to a transparent selection process for the various character/environment scenarios that their child would interface with. It might be beneficial to pre-consider modes for a psychiatrist-parent interaction for setting up the game/ companion(s) and interpreting data concerning certain interactions in the game. In providing a solidly designed app that has children's well-being in mind, there should be a focus on disincentivizing unsupervised clone apps (e.g. by making the official Accompanion AI a clinically tested app that in the best case would be free to use (e.g. via a state-funded solution)

3.10.4 Conclusions

A possibility for building a first testing environment could be by using accompanion AI in a game setting. Games might serve as a suitable environment for testing companion systems. They are understood as experimental and separate from reality, thus a certain distance to what the AI will say or do is a given. Through promoting research into this direction, the learnings can also be used to make NPCs more natural and allow the players to form a strong bond with the NPCs (as a friend or rival). Thus, a social sandbox type game might offer the player the possibility to not only explore different social relationships but also the environment the exist in. In an attempt to create a more believable game world, the behavior of AI companions would then be based on: the aforementioned companion characteristics, their goals/motivations but also past interactions as well as, ultimately, the goal of the designer(s).

References

1 Johnson, S. (The Guardian), *WHO declares loneliness a 'global public health concern'*, Thu 16 Nov 2023 09.00 CET, 2023. https://www.theguardian.com/global-development/2023/nov/16/who-declares-loneliness-a-global-public-health-concern, accessed 22.08.2024.

2 Mori, M., MacDorman, K. F., & Kageki, N., *The uncanny valley [from the field]*. IEEE Robotics & automation magazine, 19(2), pp. 98-100, 2012.

3 Warpefelt, H., and Verhagen, H., *Towards an updated typology of non-player character roles*. Proceedings of the international conference on game and entertainment technologies, 2015.

4 Rato, D. and Prada, R., *A taxonomy of social roles for agents in games*. Entertainment Computing–ICEC 2021: 20th IFIP TC 14 International Conference, ICEC 2021, Coimbra, Portugal, November 2–5, 2021, Proceedings 20. Springer International Publishing, 2021.

5 Emmerich, K., Ring, P. and Masuch, M., *I'm glad you are on my side: How to design compelling game companions*. Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play, 2018.

6 Lubart, T., *How can computers be partners in the creative process: classification and commentary on the special issue*. International journal of human-computer studies, 63(4-5), 365-369, 2005.
7 Sales, Hazel. "More than "monkey see, monkey do": Aspects of language use in engineering and manufacturing process web pages." 2010 IEEE International Professional Communication Conference. IEEE, 2010.

3.11 Game Asset Generation

Leonie Kallabis (TH Köln, DE), Chengpeng Hu (Southern Univ. of Science and Technology – Shenzen, CN), and Matthias Müller-Brockhausen (Leiden University, NL)

License Creative Commons BY 4.0 International license
© Leonie Kallabis, Chengpeng Hu, and Matthias Müller-Brockhausen

This working group started as an exploratory overarching group looking at generative AI in arts and crafts. From an initial discussion in this overarching group, three subgroups (see 3.15 and 3.16) formed with different focus topics. In this section, we report on the subgroup focused on generative AI in game asset generation.

Games contain a variety of different asset types such as textures, sounds, animations, shaders, and so on. Within this group, we decided to focus on graphical assets, also known as sprites. Basically, these are graphical elements that represent a visual state of a game object, such as a game character. Sprites are usually collected in sprite sheets.

The field of generative AI for image generation is developing rapidly, with image generation platforms such as Midjourney [5], Stable Diffusion [6], and DALL-E [7] producing appealing results. However, creating sprites requires a different set of considerations than creating images. The generated sprites must fit the game, meaning that they must be consistent with the game's aesthetics, mechanics, sound design, narrative, theme, existing assets, and more. Recent research has explored the use of different approaches to sprite generation. For example, generating pixel art sprite sheets using deep learning from sketches [2], or game icons using Generative Adversarial Networks (GANs) [3], have shown promising results. However, there have been problems with assets being perceived as realistic by humans [3].

Within this working group, we focused on two distinct requirements for sprite generation and formulated the following guiding questions for our investigation:

- How can game asset generation account for properties of different game types?
  - Considering visual cues on character properties or environmental functionality (e.g. slippery ice blocks)
- How can we generate multiple game assets that follow the same art style?
  - Ensure consistency within generated assets, such as creating a pixel art asset set that consists of multiple sprites that look like they were created by the same artist.

One of the challenges of generating images from text is translating the user's intent into the prompt given to the model [4]. Problems can arise when the result is not what is expected, and the user is unable to modify the prompt to meet their expectations. Liu et al.'s guidelines for image generation prompts suggest focusing on subject and style keywords, while trying to select subjects that can complement the chosen style in terms of level of abstraction and relevance to achieve good results [4]. Combining this guidelines with properties to describe sprites we concluded a prompt should contain the object(s), an overall topic, mood or the expected effect, the art style and a theme. Sprite sets could be created from the generation of

Figure 16 Images generated with the prompt *sprite of a green scary goblin, pixelart, 16bit.* The first two images were generated with the same model (64x), the third image with a different one (pixelartXL).

Figure 17 Images generated with the prompt *A massive, terrifying dragon with razor-sharp teeth and claws. Its scales are a deep obsidian black, shimmering with an oily sheen. Smoke billows from its nostrils, and its eyes glow with an infernal red light. The dragon is perched on a mountain peak, its wings outstretched as if it is about to take flight. The landscape below is barren and scorched, hinting at the dragon's destructive power, pixelart, 16bit.* The first two images were generated with the same model (pixelartXL), the third image with a different one (pixel_f2).

individual sprites, giving the user more flexibility in the objects they want to include in the sprite sheet and modify them individually. To do this, we developed categories that could be generated individually to create a sprite set: characters (avatar, non-player characters, companions), objects (interactable objects, objects with a purpose, decoration), environment, user interface etc.

We experimented with different diffusion models to see if they could consistently imitate different styles. Four prompts for the pixel art style were created and ran with the vanilla stable diffusion model and six fine-tuned pixel art models (64x [10], pixelArtRemond [11], pixelartXL [13], pixel_f2 [12], realisticVision [14], texture [15]). Additionally, the Structural Similarity Index (SSIM) [1], as a commonly used measure to compare generated images, was used for comparison.

Figure 16 shows images generated for a goblin using a simple prompt, and shows that the overall style is relatively consistent. For all models, the SSI was between 0.11 and 0.39, showing a fairly good similarity for comparisons. Figure 17 shows images generated for a dragon using a long, highly descriptive prompt. This example illustrates that the models are able to produce similar results, but omit the style command completely to produce pixel art images. Similar to the results for the goblin, the SSI between the models ranged from 0.07 to 0.22, giving slightly better similarity.

When comparing the images manually, they are somewhat consistent, but lack the consistency to be convincing enough for showing the same game character, for example.

Our experiments showed that stable diffusion models can produce reasonably comparable images. Looking to future research, the following question arises: How would a large(r) group of people perceive generated sprites in terms of consistency – either on their own or integrated into a game?

Assuming that sprite generation fulfills the properties mentioned above, interesting possibilities for game design arise. While there are already games that use style changes as part of their mechanics (for example: Degrees of Separation [8] and Titan Fall 2 [9]), such a generator could allow for interesting on-the-fly adaptation of sprites.

References

1 Kaisei Fukaya, Damon Daylamani-Zad and Harry Agius. *Evaluation metrics for intelligent generation of graphical game assets: a systematic survey-based framework*. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2024.
2 Y. Rebouças Serpa and M. A. Formico Rodrigues. *Towards Machine-Learning Assisted Asset Generation for Games: A Study on Pixel Art Sprite Sheets*. 18th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames), Rio de Janeiro, Brazil, 2019.
3 Rafal Karp and Zaneta Swiderska-Chadaj. *Automatic generation of graphical game assets using GAN*. Proceedings of the 2021 7th International Conference on Computer Technology Applications (ICCTA '21), New York, USA, 2021.
4 Vivian Liu and Lydia B Chilton. *Design Guidelines for Prompt Engineering Text-to-Image Generative Models*. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22), New York, USA, 2022.
5 Midjourney. `https://www.midjourney.com/home`. Accessed: 2024-08-20.
6 Stable Diffusion Image Models. `https://stability.ai/stable-image`. Accessed: 2024-08-20.
7 DALL-E 3. `https://openai.com/index/dall-e-3/`. Accessed: 2024-08-20.
8 Degrees of Separation Steam Page. `https://store.steampowered.com/app/809880/Degrees_of_Separation/`. Accessed: 2024-08-20.
9 Titanfall 2 Steam Page. `https://store.steampowered.com/app/1237970/Titanfall_2/`. Accessed: 2024-08-20.
10 64x model. `https://civitai.com/models/185743/8bitdiffuser-64x-or-a-perfect-pixel-art-model`. Accessed: 2024-08-20.
11 Pixel Art.Remond. `https://huggingface.co/artificialguybr/PixelArtRedmond`. Accessed: 2024-08-20.
12 pixel_f2. `https://stablediffusionapi.com/models/pixelf2`. Accessed: 2024-08-20.
13 Pixel Art XL. `https://civitai.com/models/120096/pixel-art-xl`. Accessed: 2024-08-20.
14 Realistic vision. `https://huggingface.co/stablediffusionapi/realistic-vision`. Accessed: 2024-08-20.
15 Texture. `https://huggingface.co/dream-textures/texture-diffusion`. Accessed: 2024-08-20.

3.12 Communal Computational Creativity

Antonios Liapis (University of Malta – Msida, MT), Alex J. Champandard (creative.ai – Wien, AT), João Miguel Cunha (University of Coimbra, PT), Christian Guckelsberger (Aalto University, FI), Setareh Maghsudi (Ruhr-Universität Bochum, DE), David Melhart (University of Malta – Msida, MT), Johanna Pirker (LMU München, DE), Emily Short (Oxford, GB), Hendrik Skubch (Square Enix AI & ARTS Alchemy Co. Ltd. – Tokyo, JP), Tristan Smith (Creative Assembly – Horsham, GB), Tommy Thompson (AI and Games – London, GB), and Vanessa Volz (CWI – Amsterdam, NL)

License Creative Commons BY 4.0 International license
© Antonios Liapis, Alex J. Champandard, João Miguel Cunha, Christian Guckelsberger, Setareh Maghsudi, David Melhart, Johanna Pirker, Emily Short, Hendrik Skubch, Tristan Smith, Tommy Thompson, and Vanessa Volz

The last five years were a watershed moment in bringing AI technologies to the forefront in applications, web browsers, dedicated (paid) services, but also in the public discourse, courtrooms, and creators' circles. The research, financial, and collective interest is triggered by important AI advances that manage to produce artifacts in domains we would consider creative [1] (such as text, images, audio, movies, and code) via data-hungry AI models trained on data which are available in the worldwide web but which are not intended as training data. The dubious ethics of such practices and the closed-source nature of such trained models leave AI academics and human creators upset and concerned [2, 3]. Beyond obvious ethical issues of exploitation and copyright infringement, we identify a looming threat of data scarcity. If current large AI models use the majority – if not the entirety – of the world wide web, where would new training data come from? While new human data currently takes the form of labels and annotations by crowdworkers in the Global South [4], it is not unrealistic to envisage a near-future where exploited crowdworkers are required to produce creative pieces or art to train AI models – without an intrinsic motivation to create.

Moreover, the availability of seemingly cogent AI outputs or the low-stakes interaction with the AI (without needing technical expertise or cumbersome software libraries) has changed human perspectives and processes in creative domains, education, and everyday life. Indicatively, education has so far been realized via interactions between learner and educator (in a top-down fashion) and via peer-to-peer collaboration. AI automation tends to isolate a learner, placing them in an adversarial relationship with the educator who is expected to act as a discriminator of AI-generated or AI-curated reports. Such activities used to be social, *communal* efforts where sharing opinions and perspectives was critical for satisfying the highest-level human needs [5] such as learning, creativity, or self-actualization. We argue that human-human interaction, either via a peer-based bottom-up ideation process or via some expertise gap (such as learner and educator or client and commissioned artist) is threatened by *trivialized* AI-human interactions. These interactions are not only trivialized because the AI output may seem novel initially but loses its novelty over time [6], or because of the potential for factual errors [7]. We argue that such interactions are trivialized precisely due to the speed of AI responses. In line with Kahneman's distinction between fast and slow thinking [8], instantaneous AI outputs with a complete artifact (e.g. an art piece) hinder human slow thinking and block the potential for iteration, reframing [9], or mediated consensus creation [10]. Creative thinking not only benefits from interventions by other humans and external (even random) stimuli [11], but also consists of a slow, introspective, autotelic process of self-doubt, frustration [12], trial and error, reflection [13] and Eureka moments [14].

Through discussion, we identified three cases where a communal approach (with meaningful human-human interaction) would have a strong impact: (a) education strategies that counter tendencies of generative AI; (b) AI corporate strategies that empower their workers; (c) art critique of the aesthetics of generative AI. Below, we report the high-level outcomes of the discussions for these three use-cases.

3.12.1 Education in the age of Generative AI

An obvious challenge for educators in the current (and near-future) age of ChatGPT and related AI solutions [15] is the writing of student reports in an automated or semi-automated way. Automated processes for detecting AI-generated texts remain underwhelming [16], and the potential for false positives in graded work makes such solutions unpalatable. Therefore, more fundamental changes are needed towards modern pedagogies. Importantly, we identify that demonizing AI and labeling it as a (blanket) taboo would likely have the opposite outcome. Improving AI literacy, especially at a younger age, would instead be required in order for learners to understand the strengths, weaknesses, and caveats of AI use in their coursework but also in their everyday life. Ideally, such AI literacy would come from inductive teaching that showcases to the learner firsthand how AI can fail at tasks that demand creativity, knowledge synthesis and critical thinking.

At a tertiary education level, a likely strategy to counter reliance on generative AI requires pivoting from assessment based on single-author reports towards more practical projects that involve teamwork, as well as introducing peer assessment as a (non-graded) activity. This is not applicable to all disciplines, admittedly, but would lend itself well for most game studies and game development courses – except perhaps foundational courses on programming or theory. As a more practical use-case for game development education, we formalized an exercise that solicits the self-realization of biases and limitations of generative AI as well as the benefits of collaboration between human experts in different fields. The exercise takes the form and principles of a game jam [17], an intense game creation process where multiple teams compete to make the best game in a short timeframe – often a couple of days. In this exercise, a few teams would be formed with one human artist and one human programmer. All other teams would either consist of one artist, who would be invited to use code generation AI models [18] (and of course access all online resources and tutorials available to everyone), or consist of one programmer who would be invited to use generated art for their game. For the sake of implementability, the proposed exercise overlooks many other vital roles in game development such as writer, game designer, or musician. We expect that the exercise would highlight (a) the limited controllability and output novelty of generative AI and (b) the unique ideas emerging through friction and negotiation with a human colleague.

3.12.2 New Company Practices with Generative AI

Unlike the education use-case, this working group adopted a more tech-optimistic view of current (and likely future) AI technologies. Assuming that AI automation can reduce friction and help collaboration between different sectors of a business, AI automation could be set up in human-like ways. Such a setup would free human resources, leaving workers more free to pursue fewer hours of intellectual work compared to many hours of menial work. Moreover, if most tasks could be automated to a satisfactory – even if not human-competitive – level, workers could move freely within the structure and take up different tasks while acting as a human-in-the-loop for the AI handling that task. This flexibility would lower the chances of a burnout and, coupled with fewer hours that consist only of meaningful (and ostensibly rewarding) work, would lead to a happier workforce. Importantly, the envisaged solution would require a different work structure with more empowered workers with incentives to perform well, such as partnerships and company stocks. It is worth noting that the envisioned solution overlooks a number of (more pressing) concerns, indicatively that (a) menial tasks that could be automated would lead to jobs lost for people with these exact skills, (b) current AI "automation" often involves humans-in-the-loop or training data from exploited workers [4] who could remain overlooked (and unrecognized) in the envisioned work structure. Therefore, for such company practices to be sustainable and ethical we presuppose a workforce educated in AI and digital skills, as well as legislation and/or new standards that leverage AI without exploitative practices.

3.12.3 Art Critique of AI Models

Taking a different view on AI literacy to the two previous use-cases, this workgroup identified art appreciation as a way to value, critique and review Generative AI models in terms of their output. Art practice is founded on such rotes, from individual letter correspondence between artists [19], gallery visits by peers [20], and discussions within artistic brotherhoods in the 1800s [21] and Discord servers in the 2020s [22]. Art historians, similarly, study the trends of an art movement and the deliberate additions by an individual creator within that context. Moving into the realm of AI models, the critique here would assess the workings of the models and their internal biases – rather than deliberate brush strokes on one painting.

As with different creative domains (writing, painting, sculpture, etc.), a common language is likely needed to review AI models and potentially different types of AI output such as generative text, art, video, music, etc. We envision AI model reviews to focus more on use-cases where the model performs well, along with recommendations for domains, applications and aesthetics that the AI model is suited for. It is essential that such a vocabulary is not imposed from the top down by computer scientists (or worse, the corporate shareholders attempting to hype their product). Instead, this vocabulary and pertinent aesthetics should emerge from the bottom up through cultural stakeholders. These stakeholders range from amateurs experimenting with the new tools – some of this collaborative meaning-making is already taking place on Discord servers [22] – to creatives and/or art experts such as gallery curators. Reaching a consensus among these diverse cultural groups will likely not be immediate or easy, but we argue that such a vocabulary will inevitably coalesce – if precedents in traditional art movements [23] are any indication. We envision that such critique could become normalized through modern dissemination practices such as zines [24] or even exhibitions. Admittedly, an ecosystem of human reviewers of AI models presupposes a level of AI literacy (and perhaps a tech-optimism) that current creative circles and art critics lack. On the other hand, we envision that a normalization of AI model reviews would enhance AI literacy (under specific perspectives and use-cases) within the art world and the general public.

3.12.4 Conclusion

New methods of human-AI interaction and the emergence of "AI companies" necessitate a review of current practices within our everyday lives, and how those might change in the near- or mid-term. The three working groups described above tackled very different issues of everyday life (education, business, and art) through different positions in the tech-optimism versus tech-pessimism spectrum. However, all working groups identified the crucial role that AI literacy (and by association, critical thinking skills regarding AI process, output, and capacities) will play *for everyone* moving forward. Moreover, all working groups emphasized the need for bottom-up movements to empower human stakeholders in a meaningful way that fosters *community*, rather than in an adversarial (e.g. students versus educators, or AI evangelists versus traditional artists) or exploitative fashion. The premise, topics, and outcomes of the working groups extend beyond game research or game development. However, games as a medium, gamification as a set of design patterns [25], and play as an activity [26, 27] could facilitate both AI literacy (e.g. [28]) and community-building (e.g. [29]). While AI is likely to impact our everyday lives and society in foreseeable and unforeseeable ways, we hold hope that bottom-up movements and a communal effort will rise up to address the new challenges.

References

1 S. Colton and G. A. Wiggins, "Computational creativity: the final frontier?" in *Proceedings of the 20th European Conference on Artificial Intelligence*, 2012.
2 J. Togelius and G. N. Yannakakis, "Choose your weapon: Survival strategies for depressed AI academics [point of view]," *Proceedings of the IEEE*, vol. 112, no. 1, pp. 4–11, 2024.
3 C. E. Lamb and D. G. Brown, "Should we have seen the coming storm? Transformers, society, and CC," in *Proceedings of the International Conference on Computational Creativity*, 2023.
4 A. Williams, M. Miceli, and T. Gebru, "The exploited labor behind artificial intelligence," `https://www.noemamag.com/the-exploited-labor-behind-artificial-intelligence`, 2022, accessed 21 August 2024.
5 A. Maslow, "A theory of human motivation," *Psychological Review*, vol. 50, no. 4, 1943.
6 E. Zhou and D. Lee, "Generative artificial intelligence, human creativity, and art," *PNAS Nexus*, vol. 3, 2024.
7 M. Karpinska and M. Iyyer, "Large language models effectively leverage document-level context for literary translation, but critical errors persist," in *Proceedings of the Machine Translation Conference*, 2023.
8 D. Kahneman, *Thinking, Fast and Slow*. Farrar, Straus and Girou, 2013.
9 T. Scaltsas and C. Alexopoulos, "Creating creativity through emotive thinking," in *Proceedings of the World Congress of Philosophy*, 2013.
10 J. R. Hollenbeck, "The role of editing in knowledge development: Consensus shifting and consensus creation," in *Opening the black box of editorship*. Springer, 2008, pp. 16–26.
11 M. Beaney, *Imagination and creativity*. Open University Milton Keynes, UK, 2005.
12 S. Savvani, "Emotions and challenges during game creation: Evidence from the global game jam," in *Proceedings of the 14th European conference on games based learning*, 2020, pp. 507–514.
13 D. Boud, R. Keogh, and D. Waker, *Promoting reaction in learning. A model. Reflection: Turning experience into learning*. London: Routledge in association with The Open University, 1996, pp. 32–56.
14 M. Bilalić, M. Graf, N. Vaci, and A. H. Danek, "The temporal dynamics of insight problem solving–restructuring might not always be sudden," *Thinking & Reasoning*, vol. 27, no. 1, pp. 1–37, 2021.
15 T. Fütterer, C. Fischer, A. Alekseeva, X. Chen, T. Tate, M. Warschauer, and P. Gerjets, "ChatGPT in education: global reactions to AI innovations," *Scientific reports*, vol. 13, no. 1, 2023.
16 C. Chaka, "Detecting ai content in responses generated by ChatGPT, YouChat, and Chatsonic: The case of five AI content detection tools," *Journal of Applied Learning and Teaching*, vol. 6, no. 2, 2023.

17 A. Kultima, "Defining game jam," in *Proceedings of Foundations of Digital Games Conference*, 2015.

18 S. Imai, "Is GitHub copilot a substitute for human pair-programming? An empirical study," in *Proceedings of the ACM/IEEE 44th International Conference on Software Engineering: Companion Proceedings*, 2022.

19 P. Grant, *The Letters of Vincent van Gogh: A Critical Study.* Athabasca University Press, 2014.

20 A. Antoniou, I. Lykourentzou, A. Liapis, D. Nikolou, and M. Konstantinopoulou, "What Artists Want": Elicitation of artist requirements to feed the design on a new collaboration platform for creative work," 2021. [Online]. Available: https://arxiv.org/abs/2110.02930

21 L. Morowitz and W. Vaughan, *Artistic Brotherhoods in the Nineteenth Century.* Routledge, 2000.

22 J. McCormack, M. T. L. Rodriguez, S. Krol, and N. Rajcic, "No longer trending on artstation: Prompt analysis of generative ai art," in *Proceedings of the International Conference on Evolutionary and Biologically Inspired Music, Sound, Art and Design*, 2024.

23 H. Ball, "Dada manifesto," 1916.

24 S. E. Thomas, "Value and validity of art zines as an art form," *Art Documentation: Journal of the Art Libraries Society of North America*, vol. 28, no. 2, pp. 27–38, 2009.

25 S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: defining "gamification"," in *Proceedings of the International Academic MindTrek Conference*, 2011.

26 A. Liapis, C. Guckelsberger, J. Zhu, C. Harteveld, S. Kriglstein, A. Denisova, J. Gow, and M. Preuss, "Designing for playfulness in human-AI authoring tools," in *Proceedings of the FDG workshop on Human-AI Interaction Through Play*, 2023.

27 J. Zhu, G. Chanel, M. Cook, A. Denisova, C. Harteveld, and M. Preuss, "Human-AI collaboration through play," in *Human-Game AI Interaction (Dagstuhl Seminar 22251)*, D. Ashlock, S. Maghsudi, D. P. Liebana, P. Spronck, and M. Eberhardinger, Eds. Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.

28 M. Zammit, I. Voulgari, A. Liapis, and G. N. Yannakakis, "The road to AI literacy education: From pedagogical needs to tangible game design," in *Proceedings of the European Conference on Games Based Learning*, 2021.

29 A. Kultima, K. Alha, and T. Nummenmaa, "Building Finnish game jam community through positive social facilitation," in *Proceedings of the International Academic Mindtrek Conference*, 2016.

3.13 Distance and Density in Various Spaces

Simon M. Lucas (Queen Mary University of London, GB), Duygu Cakmak (Creative Assembly – Horsham, GB), Filippo Carnovalini (VU – Brussels, BE), M Charity (New York University, US), Amy K. Hoover (NJIT – Newark, US), Ahmed Khalifa (University of Malta – Msida, MT), Setareh Maghsudi (Ruhr-Universität Bochum, DE), and Vanessa Volz (CWI – Amsterdam, NL)

License Creative Commons BY 4.0 International license
© Simon M. Lucas, Duygu Cakmak, Filippo Carnovalini, M Charity, Amy K. Hoover, Ahmed Khalifa, Setareh Maghsudi, and Vanessa Volz

Distance and density are fundamental concepts that underpin various critical questions, such as: How novel is a game? Which state should be visited next to explore sparse regions of the space? How similar are two agents' behaviours?

Games naturally deal with many types of space: a single video game frame is a point in image space, a single game state is a point in the game's state space, an agent is a point in strategy space, a game is a point in program space (typically expressed as text in a general purpose high-level language, or in a Game Description Language), and an agent's behaviour in a game can be described as a set of trajectories in (state,action) space.

The purpose of the working group was to explore the underlying theory and practical applications of distance and density in games, and to provide a starting point to game and game AI researchers wishing to explore the fundamentals in more depth.

3.13.1 Background

Given two points x and y in a space S, denote $d(x, y)$ as the distance between x and y. If d satisfies a set of axioms[10] including the triangle inequality, it is standard to call this a *distance metric*. However, there are plenty of measures that do not qualify as a metric, that can still provide useful results in practice.

3.13.1.1 Density Estimation Methods

We can use a distance measure to directly compute the density at a given point in the space. In practice, these methods involve computing the distance to a number of the closest points sampled from the underlying distribution. Hence, efficient ways to find the closest points become important. A general and widely used approach is to use KD-Trees to find the k nearest neighbors. Given these, we can then use a kernel density estimate, or use the average or maximum distance to the k nearest neighbors (kNN) to estimate density: see Bishop [1] section 2.5.

An alternative to kNN-based methods is to train a model that does not store all the patterns. In the interests of simplicity and brevity, we consider two types of trained models: neural density models, and feature hash-based models.

Neural density models are trained via iterative back propagation, and typically cope well with the curse of dimensionality, at the cost of an iterative tuning and training process. See [2] for recent work on this.

3.13.1.2 Distribution Estimation Methods

Feature hash models involve computing multiple hash indices for each pattern, and counting the occurrences of each index. Hence, the original pattern (e.g. text or image) is transformed into a multinomial distribution.

These methods include n-grams and n-tuples, which are similar concepts used in slightly different contexts: n-grams are commonly used as text probability estimators, usually composed of consecutive symbols (bigrams, trigrams etc), whereas n-tuples have a more general interpretation, e.g. randomly sampling an entire image space.

Related to n-tuple methods are the Context-Tree-Switching (CTS) and Skip-CTS algorithms used by Ostrovski et al [3] their work on unifying count-based exploration and intrinsic motivation. Results included improved performance on Montezuma's revenge (a hard exploration game).

[10] https://en.wikipedia.org/wiki/Metric_space

While n-gram models can be used to estimate the probability of a pattern given the model (and by Bayes' theorem, the probability of the model given the pattern), they should be used with care as pattern generators. Maximum likelihood sequences are often highly repetitive; Lucas and Volz [4] overcame this by evolving Mario levels to match the tile-pattern distribution (measured by KL-Divergence) instead of the most likely ones given the model.

3.13.1.3 Invariance and Equivariance

Key concepts related to the notion of distance are *invariance* and *equivariance*. In many applications, we wish to consider similarity between entities independently of certain transformations T.

Invariance. A function f is said to be invariant under transformation T if:

$$f(x) = f(T(x))$$

► **Example.** Consider a convolutional image recognition system that identifies that an image contains Mario, regardless of his location.

In a distance setting, this translates to:

$$d(x, y) = d(x, T(y))$$

Here, the distance remains unchanged even when one of the points is transformed by T.

Equivariance. A function f is said to be equivariant under transformation T if:

$$T(f(x)) = f(T(x))$$

Continuing with the Mario example, an equivariant system identifies Mario and his location. Moving Mario in the image by the transform T also modifies the location output accordingly.

In a distance setting, equivariance means that if we apply a transformation T to both points x and y, the resulting distance should reflect the transformation applied. Formally, for a distance function d and transformation T, this can be expressed as:

$$d(T(x), T(y)) = \text{function of } T \text{ and } d(x, y)$$

For instance, if T is a scaling transformation by a factor c, then an equivariant distance measure should satisfy:

$$d(T(x), T(y)) = c \cdot d(x, y)$$

This reflects the change in distance due to the scaling transformation T.

Applying the appropriate transformations as part of the distance (or density) measurement process is important in many applications to ensure the measurements align with the goals of the application, and are not contaminated with irrelevant variations.

3.13.2 Application: Improved Exploration Policies

RL algorithms typically use random exploration to bootstrap policy learning. In problems with flat reward landscapes, this can be very inefficient. Here we demonstrate an example of improving exploration using notions of distance and density.

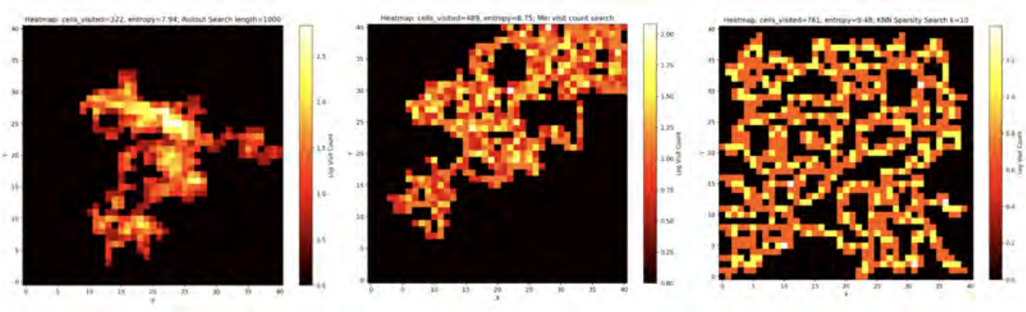


Figure 18 Exploration in a grid world: the aim is to visit as many cells as possible given a $1,000$ step budget. Left: random rollout (322 cells): middle: count-based exploration (489 cells); right: kNN sparsity search (761 cells).

Figure 18 shows an agent exploring a grid world given a budget of 1,000 steps. The agent's state is defined by its position on the grid, and it is allowed to take its next action (NESW) from any previously visited state. The aim is to visit as many cells as possible within a specified number of steps, in this case 1,000.

Note the effects of three different rollout policies: random, count-based and kNN sparsity based. In each case, having selected a state, we always take a random action from that state. All that differs is the state selection mechanism. In the pure random rollout, we choose the current state. In count-based exploration, we keep a count of the number of times each state has been visited, and select one at random with the least number of visits. In kNN sparsity search we select the state with the lowest kNN density. This is the most sophisticated exploration policy among the three, and has the best performance, demonstrating the importance of a density model that aligns with the goal of the algorithm.

3.13.3 Application: Similarity Estimation in Tile-Based Games

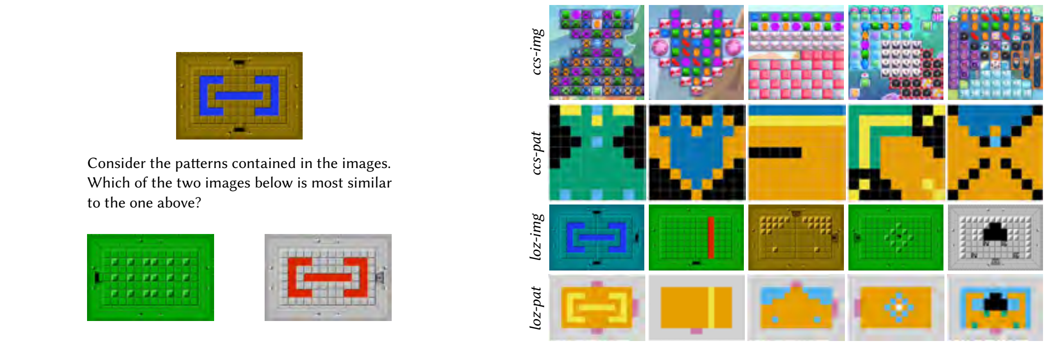
While there are many ways to measure the similarity of content in tile-based games, many practical applications require measures that align with human perception. Berns et al. [5] thoroughly studied various measures and compared their results with a user-study asking participants about how they perceive similarity in different settings 19.

They found that overall, pre-trained general-purpose computer-vision-based models such as DreamSim [8] and CLIP [7] performed better in terms of agreement with human perception than more custom measures developed in the context of PCG. However, they are also more expensive to compute, and this study focused on visual perception only. More work is required in that area to understand the effects of choosing different distance (similarity) measures and how they correspond to human perception, in order to draw more general conclusions.

3.13.4 System Design and Evaluation

In designing any system that uses distance measures and / or density estimation, there are multiple important factors, including:

- Dimensionality and the nature of the native pattern space e.g. vector, sequence, image, data object: affects choice of invariant transformations, distance measures and models.
- Data quantities involved (including volume, velocity, and variability: affects choice of trained versus stored-pattern algorithm.



Ⓐ Triplet comparison task: two alternative forced choice

Ⓑ Five example stimuli from each experimental condition

Figure 19 (A) Two alternative forced choice (2AFC) triplet comparison task and (B) five example stimuli for each experimental condition: two video game titles (ccs: Candy Crush Saga; loz: Legend of Zelda) in two representations (img: level screenshots; pat: abstract colour patterns). Each stimulus was randomly drawn from the respective subset identified through our three-stage selection procedure.

- Subjective (perceptual) or objective application alignment – see Li et al. [6] for a survey of both types applied to a wide range of games.
- Multimodality of data (e.g. some models may assume a unimodal distribution that does not fit the data well)
- Scaling of distance and density algorithms for commonly used functions, with respect to size of pattern and number of patterns.

3.13.4.1 Evaluation

In general, we find that evaluation of distance and density measures tends to be done ad hoc for each application: we are not aware of benchmarks or evaluation frameworks that can test the performance of these important methods across a range of applications. This seems worth further exploration, given that choice of distance and density measures have on the success of an application – as clearly demonstrated in the two applications we considered.

3.13.4.2 Libraries

- **Density estimation:** `https://scikit-learn.org/stable/modules/density.html`
- **KD-Trees in Python:** `https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.KDTree.html`
- **Efficient kNN computation – use in density estimation:** *See the above KD-Trees link*
- **Image-similarity-measures:** `https://pypi.org/project/image-similarity-measures/`
- **Audio Similarity:** `https://librosa.org/doc/latest/index.html`
- **Tree Edit Distance:** `https://pythonhosted.org/zss/` – useful for computing the distance between two Python programs. To do this, compute the Abstract Syntax Tree for each program using standard functions. Before measuring the tree edit distance, aspects considered irrelevant, such as variable names, may be discarded. See [10] for comparing student program code.

3.13.5 Summary

Distance and density estimation in vector spaces is a well studied topic, with still-useful algorithms such as kNN and kernel density estimation that date back to the 1960s or earlier.

However, we find that there is much to do to apply the most appropriate methods for game-related problems, as illustrated by the example applications discussed above.

There are compromises to be made along the lines of problem alignment, training speed versus inference speed, and dealing with non-vector spaces such as behaviour traces and program code. In many cases, the methods exist for dealing with these as well, but require effort to find and utilise – we recommend more work along these lines. For example, computing distance and density among programs, or among game state objects: we are not aware of any user-ready libraries for these important tasks.

References

1 Bishop, C. M., *Pattern Recognition and Machine Learning. Springer*, 2006. Retrieved from https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf

2 Qiao Liu, Jiaze Xu, Rui Jiang, and Wing Hung Wong, Density estimation using deep generative neural networks.*Proceedings of the National Academy of Sciences*, 118(15):e2101344118, 2021. https://doi.org/10.1073/pnas.2101344118.

3 Ostrovski, G., Bellemare, M. G., van den Oord, A., & Munos, R., *Unifying Count-Based Exploration and Intrinsic Motivation*. arXiv preprint arXiv:1606.01868, 2017. Retrieved from https://arxiv.org/abs/1606.01868

4 Lucas, S. M., & Volz, V., *Tile Pattern KL-Divergence for Analysing and Evolving Game Levels*. In Proceedings of the Genetic and Evolutionary Computation Conference, pp. 170-178, 2019. Retrieved from https://dl.acm.org/doi/10.1145/3321707.3321844

5 Sebastian Berns, Vanessa Volz, Laurissa Tokarchuk, Sam Snodgrass, and Christian Guckelsberger, *Not all the same: Understanding and informing similarity estimation in tile-based video games*. In CHI '24: Proceedings of the CHI Conference on Human Factors in Computing Systems, Article No. 366, pages 1–23, May 2024. https://doi.org/10.1145/3613904.3642077.

6 Yuchen Li, Ziqi Wang, Qingquan Zhang, and Jialin Liu, *Measuring diversity of game scenarios*, 2024. https://arxiv.org/abs/2404.15192.

7 Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever, *Learning Transferable Visual Models From Natural Language Supervision*. In Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021

8 Stephanie Fu, Netanel Tamir, Shobhita Sundaram, Lucy Chai, Richard Zhang, Tali Dekel, and Phillip Isola, DreamSim: Learning New Dimensions of Human Visual Similarity using Synthetic Data. In Advances in Neural Information Processing Systems, Vol. 36, 2023.

9 Zhang, K., & Shasha., Simple fast algorithms for the editing distance between trees and related problems. *SIAM Journal on Computing*, 18(6), 1245-1262, 1989.

10 Chou, E., Fossati, D., & Hershkovitz, A., A Code Distance Approach to Measure Originality in Computer Programming. In *Proceedings of the 16th International Conference on Computer Supported Education (CSEDU 2024)* – Volume 2 (pp. 541-548). DOI: 10, 2024.

3.14 Sub-optimal Bots

David Melhart (University of Malta – Msida, MT), James Goodman (Queen Mary University of London, GB), Christian Guckelsberger (Aalto University, FI), Greta Hoffmann (TH Köln, DE), and Diego Perez Liebana (Queen Mary University of London, GB)

License Creative Commons BY 4.0 International license
© David Melhart, James Goodman, Christian Guckelsberger, Greta Hoffmann, and Diego Perez Liebana

We traditionally aim to create game-playing agents that are optimised for a certain task – be it performance, human-like behaviour, or facilitating a positive player experience. In this working group, we asked the question – *"Do bots always have to be efficient?"* We argue that there is merit to low-fi, sub-optimal bots in areas such as games testing, producing erratic and distracted behaviour to model users, and resource management for sustainability.

3.14.1 The Scope of Being Sub-optimal

When talking about sub-optimal bots, we focus on agents that are designed to be "bad at their job" – that is they cannot perform as well as the state-of-the-art due to some constraints. While *constraints* here generally mean resource constraints in terms of limited computing power or memory or time, sub-optimal bots could also under-perform current systems by presenting deliberately non-human-like. The former we could call *natural constraints*, and the latter we could call *artificial constraints*.

- **Natural Constraints:** The natural constraints of a computational system constitute its computational resources. Sub-optimal performance could arise from simple limits on resource use. While some of these limits are a given, others can exist by design. In the current boom of large models, we often view resource constraints as a limit we should aim to lift, and with a good reason. Since 2017 we have seen the exponential rise of Large Language Models (LLMs) [1] fuelled both by innovation [12] and by ever-growing architecture size, training data volume, and computational resources [6].
- **Artificial Constraints:** In light of the aforementioned positive aspects of scaling, why would we want to limit computational resources by design? We identify two main reasons:
  - **A case for Sustainability:** One of these main reasons is sustainability. Water usage of the vendors of foundation models is on the rise [3], and despite improvements in energy efficiency, the growing penetration of AI in different domains paints an uncertain picture of our future energy needs [2]. While we need powerful models to drive innovation and tackle complex problems, we need to think about the potential environmental cost of our applications as well. Up until now, the "state of the art" meant "best performance" – however, we propose a new challenge for future applications and innovations in the space that focus on "good enough performance" and "best resource savings". Future research in this area could be facilitated via a new competition for *Sustainable Agents*.
  - **By Design:** Artificial limitations can be designed to produce real or perceived sub-optimal performance – not necessarily by limiting actual resources. Our two grand examples here are *distracted bots* and *non-human-like agents*. In the case of distracted bots, we envision systems mirroring human cognitive load [5] or short-term memory [9] to create more anthropomorphic behaviour by modelling the human attention span. We hypothesise that bots limited in this way would produce unexpected but more human-like behaviour. In the case of non-human-like agents, the goal of imposed limitations is the opposite of achieving believability. These bots would be designed to emphasise their artificial nature and the limited capacity of their underlying

systems. A positive outcome of these agents would be algorithmic transparency [13] and emotional separation between humans and machines, which could cause unexpected negative outcomes (e.g. trauma dumping or developing unhealthy attachments to AI companions).

3.14.2 Applications for Sub-optimal Bots

In practical terms, there are a number of ways we can limit our bots to enforce a sub-optimal performance.

- **Memory:** Many agents have an implicit perfect memory, which is not reflective of (most) human players. Restricting the memory window, or the types of historic events remembered would give sub-optimal, and arguably much more human-like behaviour.
- **Forward Planning Horizon and Discount Rate:** A shorter planning horizon, and/or a discount rate profile that emphasises short-term gains would force sub-optimal myopic agents to sacrifice long-term success for immediate gratification.
- **Flexibility of Play Style and Strategies:** Restricting the available action space for an agent could force it to explore different strategies. For example, if a Chess agent were forbidden from moving its Queen until after it had castled.
- **Accuracy of World Model:** Providing planning agents with deliberately imperfect models of the world. For example in a first-person shooter game, an agent might overestimate how far it can jump; or have an incorrect mental map of unseen terrain.
- **Perception of World State:** A restrictive field of view, only taking input observations from a subset of the full data available.
- **Accuracy of Action Implementation:** Just because a specific action is intended does not mean it is implemented in the game environment accurately. This is used for example in first-person shooters to ensure that NPC bots often miss[11], but can be extended more widely to mirror human mis-clicks on interfaces.
- **Theory of Mind:** Modelling Theory of Mind explicitly in terms of, “I know that she knows that I know...” may give interestingly sub-optimal behaviour. This may also be a good reflection of non-expert humans who frequently only manage one to two nested levels of theory of mind [9].

We have identified a number of domains within game research, serious games, and games entertainment where these aforementioned limitations – as described above – could find a positive use.

- **Games Testing:** In Games Testing it is often desirable to have agents exploring the game in a non-optimal fashion. This is because we aim to increase the test coverage by expanding into non-optimal play or simulating human behaviour by mirroring the limitations of human cognition.
- **Games Playing:** Similarly to Games Testing, Non-Player Characters can also benefit from the aforementioned limitations by enhancing their human-like aspects and/or creating sub-optimal opponents. This could help provide a variety of interestingly sub-optimal opponents. These opponents could be used as scripted adversaries, tutorials to multiplayer games or as part of a dynamic difficulty adjustment system.
- **User Modelling:** Beyond testing, user research can make use of models which simulate human limitations to understand and improve Human-Computer Systems.
- **Resource Optimisation:** Finally, resource limits could be used simply to cut down on practical resource usage. The implication here is that bots and agents trained on such limitations would be “optimised” to handle constraints.

3.14.3 Future Research

We have identified several possible research questions for future studies into applying low-fidelity bots to simulate human-like behaviour.

- **How do we perceive bots being bad at their jobs?** Previous research showed that bots are designed to exhibit certain cognitive processes often perceived as confusing [8]. Future research should focus on uncovering reliable markers of bad performance due to cognitive limitations.
- **How could bots resist humans?** Similarly, resistance from interactive agents could be interpreted as a sign of an unresponsive system, rather than intentional feedback. Current human-like systems are often ill-equipped to communicate their own limitations – the most famous example being ChatGPT's tendency to "yes and" requests even if a request is beyond its limitations. Future work into HCI should aim to understand the key to clear communication and reduce user frustration in cases where the bots are limited in some capacity.
- **Do constraints make bots more (or less) human-like?** At the beginning of the report, when describing cases for imposing *Artificial Constraints* by design, we hypothesised two possible directions for limited agents. One was enhancing the human-like nature of the bot, while the other was specifically stripping the agent of believability. Future research in this area should focus on the spectrum of non-human to human-like to super-human bots in terms of perceived performance and believability.
- **A case for Computational Rationality:** Beyond the domain of games, we believe that the research of sub-optimal bots could shed light on emergent qualities of human cognition and behaviour. Based on the Theory of Computational Rationality [7, 4, 10] human cognition can be conceptualised as an "optimal" – that is rational – system under computational constraints. Given this premise, future research could aim to go beyond the scope of game-playing agents when modelling human cognition and ultimately behaviour via limited bots.

3.14.4 Conclusions

In this report, we detailed some of the outcomes of our work group on sub-optimal bots. We aimed to map out the scope of limitations, applications, and future research, arguing for the benefits of researching and developing underperforming computer agents. We believe in areas such as games testing, user modelling, and resource management for sustainability these bots can contribute to positive outcomes.

References

1 Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y. and Ye, W., *A survey on evaluation of large language models*. In ACM Transactions on Intelligent Systems and Technology, 15(3), pp.1-45, 2024.

2 Desislavov, R., Martínez-Plumed, F. and Hernández-Orallo, J., *Trends in AI inference energy consumption: Beyond the performance-vs-parameter laws of deep learning*. In Sustainable Computing: Informatics and Systems, 38, p.100857, 2023.

3 George, A.S., George, A.H. and Martin, A.G., *The environmental impact of AI: a case study of water consumption by chat GPT*. In Partners Universal International Innovation Journal, 1(2), pp.97-104, 2023.

4 Gershman, S.J., Horvitz, E.J. and Tenenbaum, J.B., *Computational Rationality: A Converging Paradigm for Intelligence in Brains, Minds, and Machines.* In Science, 349(6245), pp. 273-278, 2015.

5 Hedden, T. and Zhang J., *What Do You Think I Think You Think?: Strategic Reasoning in Matrix Games*. In Cognition 85(1), pp. 1–36, 2002.
6 Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J. and Amodei, D., *Scaling laws for neural language models*. arXiv preprint arXiv:2001.08361, 2020.
7 Lewis, R.L., Howes, A. and Singh, S., *Computational Rationality: Linking Mechanism and Behavior Through Bounded Utility Maximization.* In Topics in Cognitive Science, 6(2), pp. 279-311, 2014.
8 Melhart, D., Yannakakis, G.N. and Liapis, A., *I Feel I Feel You: A Theory of Mind Experiment in Games*. In KI-Künstliche Intelligenz, 34(1), pp.45-55, 2020.
9 Miller, E. K., and Buschman, T. J., *Working Memory Capacity: Limits on the Bandwidth of Cognition.* Daedalus 144(1), pp. 112–122, 2015.
10 Oulasvirta, A., Jokinen, J.P. and Howes, A., *Computational Rationality as a Theory of Interaction.* In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. pp. 1-14, 2022.
11 Schrum, J., Karpov, I. V., and Miikkulainen, R. *Human-like Combat Behaviour via Multiobjective Neuroevolution*. In Believable Bots, 119–50. Springer, 2013.
12 Stiennon, N., Ouyang, L., Wu, J., Ziegler, D., Lowe, R., Voss, C., Radford, A., Amodei, D. and Christiano, P.F., *Learning to summarize with human feedback*. In Advances in Neural Information Processing Systems, 33, pp.3008-3021, 2020.
13 Watson, H.J. and Nations, C., *Addressing the growing need for algorithmic transparency*. In Communications of the Association for Information Systems, 45(1), p.26, 2019.

3.15 Arts & Crafts & Generative AI

Mirjam Palosaari Eladhari (Stockholm University, SE), Gabriella A. B. Barros (modl.ai – Maceio, BR), Alena Denisova (University of York, GB), Amy K. Hoover (NJIT – Newark, US), Chengpeng Hu (Southern Univ. of Science and Technology – Shenzen, CN), Leonie Kallabis (TH Köln, DE), Ahmed Khalifa (University of Malta – Msida, MT), Matthias Müller-Brockhausen (Leiden University, NL), Gillian Smith (Worcester Polytechnic Institute, US), and Anne Sullivan (Georgia Institute of Technology – Atlanta, US)

License 🅭 Creative Commons BY 4.0 International license
© Mirjam Palosaari Eladhari, Gabriella A. B. Barros, Alena Denisova, Amy K. Hoover, Chengpeng Hu, Leonie Kallabis, Ahmed Khalifa, Matthias Müller-Brockhausen, Gillian Smith, and Anne Sullivan

Generative Artificial Intelligence (GenAI) holds much potential for artistic practice in game design, craft, and fine arts. At the time of writing, 2024, the advancements of GenAI are rapid in regards to Large Language Models (LLMs) and Text to Image (TtI) systems. Artists and game designers are keen to explore new the tools, systems, and their updates as they are released, and the pace is quick. The members of the work group have multiple roles as scientists, game designers, and art- and craft practitioners, and converged around exploring GenAI for arts, crafts, and game development.

In this report we describe our visions of usage of GenAI for creation of game assets, free form doodles, and art generated using exclusively individual artists' original art in "small" data sets.

3.15.1 Identification of Open Questions for Exploration

The group started with discussing themes to explore, and then identified topics for subgroups for further investigation. The themes related to ways to use GenAI when both artists and scientists use generative AI in collaboration with the GenAI systems. They were as follow:

- AI as an art material: akin to a painting brush or a specific software tool.
- Small Datasets: The affordances the use of small datasets may offer individual artists and teams of creators for expressing their individual artistic voices and for ascertaining that any art work used for generation originates from the artists themselves, rather than originating from an large database of data scraped from the web.
- The human wanting to do the fun stuff: The tension between what tasks an artist may want to do themselves, as a part of their creative practice, and between tasks a generative AI system may be capable of.
- Mixed initiative creation: The use of technologies of content generation as part of an artistic process.
- Software methods and tools using generative AI for supporting creativity.
- Simultaneous generation of game levels and assets.
- Stylistic coherence of generated output.
- Skill enhancement in visual representation such as drawing, allowing people who are not trained in visual expression to create artifacts that presents ideas in a more specific and sophisticated manner.
- Usage of generative AI in own artistic practice.
- Creativity support tools for art.

3.15.2 Three sub groups

We selected three themes for further inquiry in sub groups:

- Small datasets for designers and artists. (Subgroup 1)
  - Creative and artistic use of small datasets restricted to artists' and designers' own artifacts.
  - Usage of tools in practice of design and art.
  - Surveys of available tools, their affordances and practical guides.
- Game asset generation (Subgroup 2)
  - stylistic consistency
  - design principles
  - resonance between level design and assets for specific level.
- Doodling: Design affordances of Large Language Models (LLMs) in relation to Text to Image (TtL) systems. (Subgroup 3)
  - Use of LLMs and TtLs as art materials.
  - What is the equivalence of doodling using LLMs and TtLS?

In the following we summarise the outcomes of subgroups 1 and 2, referring to the respective reports, which is followed by the report of subgroup 3.

3.15.3 Small datasets for designers and artists – Subgroup 1

The subgroup focused on how artists and designers can use their own artwork to train GenAI models. When utilizing large language models (LLMs) trained on extensive databases, often sourced from the internet, issues concerning artists' contributions arise, particularly around individual style and intellectual property. The use of "small" datasets, as discussed by [2], offers promising opportunities for artists, game designers, and teams working within specific collaborations or development studios. If a generative model is trained solely on an artist's own creations, it can produce outputs that are unique to that artist while still harnessing the capabilities of GenAI technologies.

Our group explored the concept of GenAI as a tool within an artist's studio comparable to a new brush or other medium for creative expression. We conducted a survey with three primary focuses: the historical use of generative arts by visual artists and authors, dating back to the 1920s; the technologies employed in tools that allow artists to train models on their own work; and the currently available tools for such purposes.

In our investigation, we found that while many tools and high-level surveys exist, there is a lack of practical guidance to help artists make informed decisions about which tools to use and how to contextualize them. From the range of available tools, we conducted experiments. For example, we trained a model exclusively on paintings from one of the group's artists.

Throughout this experimentation, we identified the significant potential of using limited datasets for training. However, the process remains labor-intensive. We recommend further work in this area, including the development of guides and standards for artists, designers, and developers. These resources should offer guidance on how to use current and future technologies, the types of creative support they provide, their affordances regarding the nature of generated works, and their accessibility in terms of skill requirements, cost, and whether they are open source. The subgroup's work is described in more detail in 3.16.

3.15.4 Generation of Game Assets – Subgroup 2

The group exploring the generation of game assets concentrated on creating graphical assets, commonly referred to as sprites. Their work was guided by two key questions: a) How can the generation of game assets take into account the properties of different game types? and b) How can we generate multiple game assets that maintain a consistent art style?

To address these questions, the group followed the guidelines set out by Liu et al. [1], which emphasize focusing on subject and style keywords that can be adapted to specific styles and levels of abstraction. They categorized the assets into different groups, such as characters, interactive objects, environments, and more, in order to generate cohesive sprite sets.

The group experimented with various diffusion models to assess their effectiveness in mimicking distinct art styles, as detailed in section 3.11. Their findings indicated that while stable diffusion models can produce reasonably comparable images, achieving a high level of stylistic consistency that is convincing to viewers remains a significant challenge. The subgroup's work is described in more detail in 3.11.

3.15.5 Doodling with Large Language Models and Text to Image Systems – Subgroup 3

In this subgroup, run by *Gillian Smith, Anne Sullivan*, and *Alena Denisova*, we discussed and experimented with the playful act of "doodling" with off-the-shelf generative systems. Since doodling is something that artists, crafters, and designers can commonly do with any kind of design material, we posit: **if we want to understand Generative AI as a design material, we need to learn how to doodle with it**.

3.15.6 Intro

We began our exploration with a discussion of what doodling is and its role in the creative process, drawing from our own experiences. We identified seven characteristics of doodling:

- Improvisational: doodling is a form of improvisational play, where the artist is building upon their own prior work over time, exploring and reforming their creative process as they go.
- Not goal-oriented: doodling is an exploratory activity. Artists doodle without intent for these doodles to be shared or incorporated into larger planned works.
- Playful: doodling permits playful exploration of materials, experimenting with whimsical or unexpected forms, and may be done primarily for fun or leisure.
- Low cost: artists doodle with existing, low cost, materials that are not reserved for another project or too precious to 'waste'.
- No notion of failure: doodling is heavily process-oriented, with no real notion of 'failure' with respect to either internal or external validation or judgment. It transcends skill and judgment; it is accessible to everyone, and there is no way to do it either wrong or right.
- Autotelic: doodling serves its own purpose for an artist; it is not done in the service of any larger intended goal, even if it does emergently lead to new skills or techniques.
- Easy to pick up and put down: doodling is often done idly, by the subconscious mind. It is done to help focus the mind, to visualise ideas and help yourself think, or to pass time. As such, doodling is engage intermittently – easy to start, stop, and start again.

We further identified common roles that doodling plays for an artist's creative process. When doodling, an artist can:

- Learn and explore the material affordances of the medium.
- Discover and build skills with the medium.
- Reflect on their own creative process.
- Reflect on their own identity and goals as an artist.
- Focus their mind through engagement in a continued repetitive task.
- Explore a new and unfamiliar design space.

Building from these discussions of what it means to doodle *in general*, we focused the remainder of our subgroup on exploring what it means to doodle *with generative AI systems*. We chose ChatGPT 4.0 as a system that was "easy to pick up and put down", and aimed to focus on the generative model as opposed to the outputs of the system. That is: the goal is not to generate doodles, but to doodle with a prompt-based interaction mechanism for a generative model.

3.15.7 Summary of Experiments

To begin our experimentation, we gave ourselves one rule: *Just sit down, open ChatGPT, and play.*

Throughout the experience, we would share our output and our prompts, borrowing ideas and inspiration from each other to modify and work from, almost akin to musical riffing and improvisational theater.

Some of the prompt experiments that we ended up using are as follows:

- Making up fictional art styles and applying to existing images
- Making up single-word responses for long complex inputs
- Drawing pictures of those single-word responses
- Re-imagining names as mythical creatures
- Make ASCII art from emoji
- Respond to sentences with emoji chains

Figure 20 Here's the illustration featuring a glimberflopped in a nerdletower located in lichtwald forest filled with geekfrolics in the style of bleakoscribble.

Through the playful experimentation, we had a few insights into what worked and didn't work when doodling with ChatGPT 4.

The relative ease of getting output that was "good enough" fit well with a doodling mindset and the unexpected results made the overall experience quite fun. However, the high fidelity of the output worked against it – it was easy to get caught up in the details. Additionally, the higher fidelity took more time to render, which also moves it away from feeling like a doodle, as doodles are often about fast results.

What does this mean for LLMs as media? The question opens up a fascinating exploration of the ways in which large language models (LLMs) like ChatGPT are more than just tools for replicating existing art forms. They are not simply systems that mimic or reproduce the traditional structures of media, whether in writing, visual arts, or conversation. The process and input into these models are unique in their nature and do not conform to the established categories of creation.

In this sense, the output generated by these models is not bound by the conventions of fidelity we might apply to other forms of media. For instance, what we might traditionally consider "high fidelity" images–detailed, precise, and closely mirroring reality – are now seen through a different lens. The outputs, which may sometimes resemble something more akin to doodles or abstract sketches, take on a new meaning and value within this framework. The results need not be perfect replicas to hold artistic merit or expressive power.

Figure 21 Here's the digital painting of featuring a glimberflopped in a nerdletower located in lichtwald forest filled with geekfrolics in the style of joyspright.

What is exciting about this is the fun involved in the process of creating with LLMs. There is an inherent playfulness and experimentation that invites users to explore the expressive range of the underlying models. The user is not just interacting with a tool but is engaged in the development of a personal understanding of the model's potential and limitations. This engagement fosters an evolving relationship with the medium itself, and this kind of active discovery is promising for the future of AI-assisted creativity.

Moreover, through this process, users begin to develop their own artistic vocabulary specifically tailored to using these models. Just as traditional artists use a distinct set of skills and tools to create within their chosen medium–whether painting, photography, or sculpture–working with LLMs requires the development of a new set of skills and a personalised approach. This vocabulary allows users to unlock deeper forms of expression and pushes the boundaries of what LLMs can achieve in artistic creation. It shows promise for AI and human creativity becoming more intertwined, each informing and expanding the potential of the other.

3.15.8 Takeaways

Doodling is an important part of a creative practice. Therefore, for LLMs and AI-based tools to excel as creative media, they need to be designed to support doodling-style interactions. Based on our experimentation during this workshop, we identified a few ways in which LLM-based tools could enable doodling like behavior.

The primary change is that LLM-based creative tools should support casual, low cost, idle engagement. ChatGPT 4, used for our experiments, has some capability for this, but we struggled with the high fidelity output which took quite a bit of time to generate. If instead, it was possible to enter into a sketch or doodle mode where low-fidelity images could be more quickly generated for refinement later, this would allow for a more doodle-like experience.

Beyond this, the chat-based and browser-like interface creates a distance between the interactor and the output, which we struggled with. For example, if there was one part of an image that we wanted to change, we were unable to convey this information through the interface we had access to. Additionally, having a series of images to chose from is a form of curation rather than of creation. Addressing this gap is vital for making LLM-based tools align with creative practice.

Our doodling-inspired experiments led us to realize that we need new vocabularies for AI Art. Currently, the vocabulary is focused on tool support and replacement for traditional art. However, this language does not support an authentic engagement within the medium, which is required for AI Art to be perceived and used as an art practice.

Finally, for AI Art and LLM-based tools to excel as a creative medium, there must be continued work considering and addressing the ethics of the data and use of these tools. Until this is addressed, these tools will never be recognized as a creative medium by art communities, due to the questionable practices by these companies.

References

1 Vivian Liu and Lydia B Chilton. *Design guidelines for prompt engineering text-to-image generative models.* In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI 22, New York, NY, USA, 2022. Association for Computing Machinery.

2 Gabriel Vigliensoni, Phoenix Perry, Rebecca Fiebrink, and others. *A Small-Data Mindset for Generative AI Creative Work.* Generative AI and HCI – CHI 2022 Workshop, 2022.

3.16 Small Data Sets for Designers and Artists

Mirjam Palosaari Eladhari (Stockholm University, SE), Gabriella A. B. Barros (modl.ai – Maceio, BR), Amy K. Hoover (NJIT – Newark, US), and Ahmed Khalifa (University of Malta – Msida, MT)

License Ⓒ Creative Commons BY 4.0 International license
© Mirjam Palosaari Eladhari, Gabriella A. B. Barros, Amy K. Hoover, and Ahmed Khalifa

Generative AI has emerged as a transformative tool in the fields of art and design, offering new ways to create and explore visual, auditory, and interactive works. The use of generative AI often relies on large datasets and advanced technical knowledge, which can pose challenges for artists and designers in therms of access and resources. Another issue is that large language models are trained on data from external datasets, which is an issue for artists when it comes to the expression of their artistic authorial voice as well as concern in regards to intellectual property.

However, there is growing interest in the use of small data sets and more accessible tools to empower artists and designers, enabling them to maintain their artistic voice, address intellectual property concerns, and make generative AI techniques more inclusive and widely available.

This report explores the insights gained from a subgroup in our workshop on Generative AI for art and craft, centering on “Small Data Sets for Designers and Artists,” focusing on the application of generative AI in creative practices. The workshop raised key questions, such as: What generative AI techniques are currently being used by artists and designers? What tools exist to support the creation of AI-generated art, and how accessible are they? Additionally, the discussion explored how small amounts of data can be used to train or fine-tune generative models, and what level of knowledge is required to effectively use these tools. A critical focus was on how generative AI can be integrated into the creative workflows of artists, designers, and programmers, and how these tools and techniques can be considered as artistic materials in their own right.

3.16.1 Background

To contextualize our discussion, we examined the use of generative techniques in creative work throughout recent history. This exploration included the surrealist practices of the 1920s [3] as well as experiments with physical materials such as Max Ernst’s grattage technique from the 1930s. We also considered the concept of “potential literature” pioneered by the Oulipo group (Ouvroir de littérature potentielle) in the 1960s [2]. The use of computation in the visual arts can be traced back to the 1960s, when artists like Georg Nees and Vera Molnar experimented with flatbed drawing machines and plotters. In recent years, generative AI has been prominently featured in the works of artists such as Anna Ridler, Mario Klingemann, Nettrice Gaskins, Refik Anadol, and Helena Sarin. Notably, Ridler [6] and Sarin’s [7] practices involve training models using imagery they have personally created, blending their artistic vision with customized adaptations of computational methods to produce unique works that reflect their individual styles.

Vigliensoni et al. [8] argue that working with small-scale datasets is a powerful way to enable greater human influence over generative AI systems in creative contexts. They suggest that while large datasets may lead to overfitting and may not align with specific creative goals, models built with smaller, more focused datasets can better support meaningful and personalized creative work.

3.16.2 Positioning and Resources

In our subgroup, we aimed to explore how artists and designers could leverage their own art in conjunction with generative AI techniques. During the workshop, we focused on utilizing existing artworks created by participants and experimenting with the available tools in July 2024. We identified a variety of resources to support our exploration, including APIs, tutorials, and custom-built tools.[11]

[11] Available resources we examined are listed blow. It is not a comprehensive list of all available tools, only the ones we examined in the workshop.

- ML5js `https://ml5js.org/`
- OpenArt `https://openart.ai/`
- Craiyon `https://www.craiyon.com/`
- Runway `https://runwayml.com/`
- Leonardo `https://leonardo.ai/`
- Vizcom (for product design) `https://www.vizcom.ai/`
- Canva `https://www.canva.com/ai-image-generator/`
- DreamBooth `https://dreamlook.ai/dreambooth`
- Tutorial for custom datasets at Hugging Face `https://huggingface.co` (available with search term: transformers v3.2.0)

Several surveys, such as that by Franceschelli [4], have explored how generative AI can be applied in creative practices. Additionally, symposia like Creative Machine, held at Oxford since 2014[12], bring together artists and researchers to showcase their work at the intersection of creativity and AI. The report *AI and the Arts: How Machine Learning is Changing Creative Work* [5] presents a comprehensive overview of the creative communities interested in using machine learning in the arts, as shown in Figure 22. Within our group, we identified our position at the intersection of the "Technical Community" and "The Art World," reflecting our engagement with both the technical and artistic dimensions of generative AI.

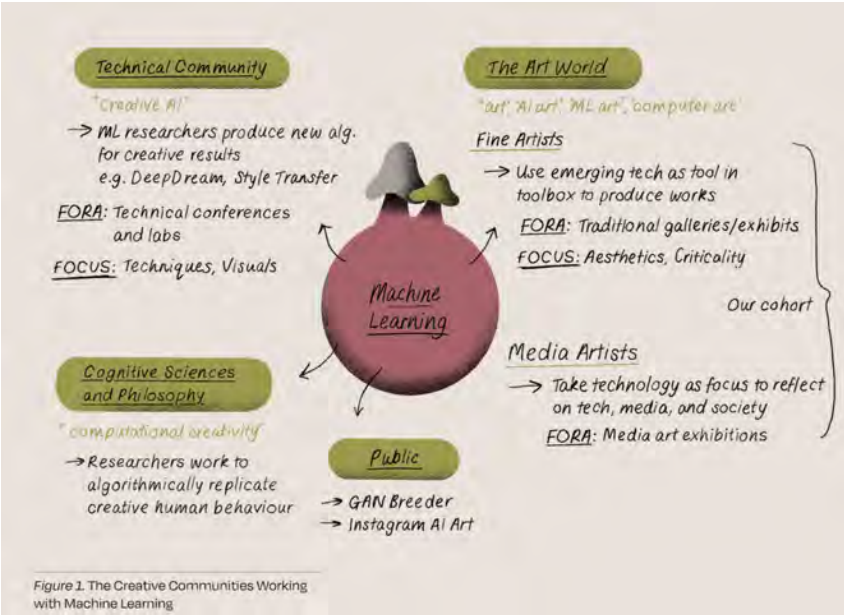


Figure 22 Creative communities working with machine learning.

3.16.3 Experiment

As we explored the available tools for using personal visual art as small datasets to train generative AI models, we conducted quick tests with free or demo versions. Our guiding question for these tests was: "As an artist, how can I experiment with using my own creations, such as drawings or paintings, to train a model and explore its potential in my artistic practice?"

We quickly realized that navigating the tools, understanding their computational foundations, and assessing their affordances was far from straightforward. For a designer or artist, a significant amount of time would be required to explore the possibilities, often through trial and error. The complexity of the tools and their steep learning curve presented a barrier to immediate creative experimentation.

- Figment `https://figmentapp.com/`
- Training Loras (available as a google colab)
- Generative Deep Learning (available at github) item Suite of tools for course on Artificial Intelligence and Storytelling given at The University of Edinburgh, `https://kage.dev/ai-storytelling-backstage/`

[12] https://www.creativemachine.io

Furthermore, we found that existing surveys and general guides either lacked practical advice for getting started or were so tool-specific that they failed to provide a broader understanding of the generative AI landscape. This made it difficult for a reader to grasp the larger context and application of these tools in creative work.

For our experiment, we selected two sources of data: The Stanford Dogs Dataset [13] and two paintings created by one of the group members (M.P. Eladhari). During the workshop, we tested several of the tools listed in Footnote 1. However, due to the limited time available and the time-consuming nature of model training, we were unable to fully develop the results of our trials. Nevertheless, in Fig 23 we present an example generated using a tool developed at the University of Edinburgh, based on Stable Diffusion and GPT-2 models. The tool, along with its code repositories, is accessible online [1]. For this test, we used two paintings (shown on the left in Fig 23 as input, along with their titles as text prompts: "I Knew I Would Find You" and "Toad in Still Water". The generated image, shown on the right, provides an example of the creative potential offered by these AI tools.

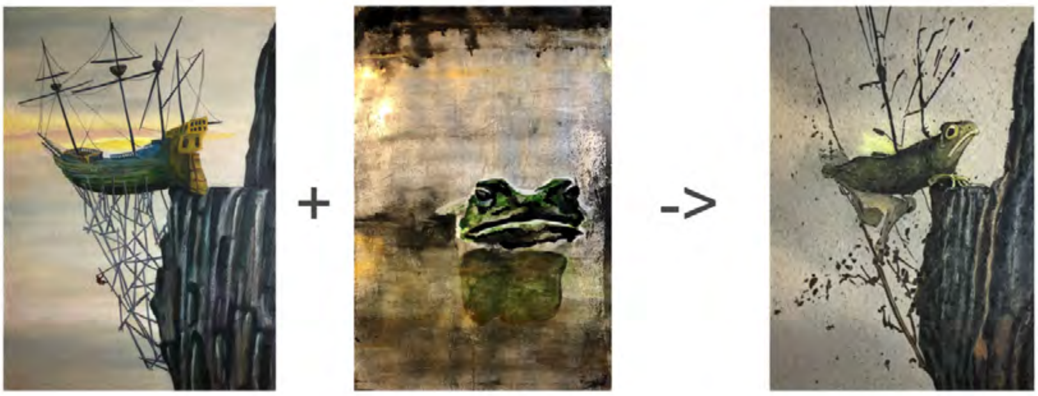


Figure 23 Illustration of combining two paintings by the same artist.

3.16.4 Immediate Needs and Outlook

In our group, we identified the need to conduct a comprehensive survey of available generative AI tools that can be used to train models on individual or collective artistic works. This survey would assess these tools based on the following criteria:

- Creativity Support: How the tools can be integrated into and enhance the creative process.
- Affordances: The types of artifacts the tools can generate and the creative possibilities they offer.
- Accessibility: How easy the tools are to use, considering both technical complexity and financial cost.

Such a mapping would provide valuable guidance to artists and designers by helping to:

- Identify tools that complement or even expand their individual artistic practices;
- Produce artifacts aligned with their artistic vision; and
- Evaluate whether they possess the necessary skills and resources to effectively use a particular tool.

[13] http://vision.stanford.edu/aditya86/ImageNetDogs/

Looking ahead, potential future collaborations within our subgroup include the creation of:

- A practical guide tailored for artists and designers.
- A practical guide bridging computer scientists and artists.
- Specifications outlining the minimal data requirements for training or fine-tuning various generative models.

3.16.5 Summary

In our subgroup report Small Data Sets for Designers and Artists, we explored how generative AI can be integrated into creative practices, focusing on the challenges and possibilities of using small datasets to train models for individual artistic work. We highlighted the importance of accessible tools and practical guides to help artists like us to maintain control over our creative visions while addressing issues like intellectual property.

References

1 Pavlos Andreadis, Kage, P. *Artificial Intelligence and Storytelling*. Course of The University of Edinburgh, https://kage.dev/ai-storytelling-backstage/, 2022.
2 Alastair Brotchie, Le Lionnais, F., Mathews, H. *Oulipo laboratory: texts from the "Bibliothèque oulipienne"*. Atlas anti-classics, Atlas press, London, 1995.
3 Alastair Brotchie, Gooding, M. (eds.) *A book of surrealist games*. Shambhala Redstone Editions, Boston, 1995.
4 Giorigio Franceschelli, Musolesi, M. *Creativity and Machine Learning: A Survey*. https://doi.org/10.48550/ARXIV.2104.02726, 2021.
5 Anne Ploin, Eynin,R., Hjort, I., Osborne, M. (2022). *AI and the Arts: How Machine Learning is Changing Creative Work*. Report from the Creative Algorithmic Intelligence Research Project. Oxford Internet Institute, University of Oxford, 2022.
6 Anna Ridler. *Repeating and remembering: the associations of GANs in an art context.*. In: The 31st International Conference on Neural Information Processing Systems.New York: Curran Associates Inc. 2017.
7 Helena Sarin. *Why Bigger Isn't Always Better With GANs And AI Art*, https://www.artnome.com/news/2018/11/14/helena-sarin-why-bigger-isnt- always-better-with-gans-and-ai-art, visited 2024-06-20, 2018.
8 Gabriel Vigliensoni, Phoenix Perry, Rebecca Fiebrink, and others. *A Small-Data Mindset for Generative AI Creative Work*. Generative AI and HCI – CHI 2022 Workshop, 2022.

3.17 Evaluating the Generative Space of Procedural Narrative Generators

Emily Short (Oxford, GB), Gabriella A. B. Barros (modl.ai – Maceio, BR), Alex J. Champandard (creative.ai – Wien, AT), Michael Cook (King's College London, GB), João Miguel Cunha (University of Coimbra, PT), Alena Denisova (University of York, GB), Antonios Liapis (University of Malta – Msida, MT), Mirjam Palosaari Eladhari (Stockholm University, SE), Johanna Pirker (LMU München, DE), Gillian Smith (Worcester Polytechnic Institute, US), Anne Sullivan (Georgia Institute of Technology – Atlanta, US), and Tommy Thompson (AI and Games – London, GB)

License Creative Commons BY 4.0 International license
© Emily Short, Gabriella A. B. Barros, Alex J. Champandard, Michael Cook, João Miguel Cunha, Alena Denisova, Antonios Liapis, Mirjam Palosaari Eladhari, Johanna Pirker, Gillian Smith, Anne Sullivan, and Tommy Thompson

This group discussed the problem of evaluating procedural narrative generators, as a subcategory of procedural generation tools. While there is existing scholarship around how to categorize procedural generation tools for level design (for instance), there is little that aims to assess the quality and variety of procedural narrative generation.

This is an area of potential value to the game industry, as well as a matter of academic interest. There is commercial appetite for games that can present players with a wide variety of novel and engaging stories, but building effective procedural narrative systems is unpredictable and bug-prone, and the difficulty of assessing their quality during production causes many such prototype systems to be removed from industry projects prior to launch. Evaluative frameworks that would allow studios to build and validate their innovation more confidently would help to alleviate these problems.

In discussing this framework, the group wanted to focus on the narrative qualities of generated output rather than to restrict their analysis to any specific type of game (or even to insist that the objects analyzed be games at all). Procedural narrative generators might include games that produce a narrative in the course of play via simulation or other means (*Dwarf Fortress*, *Façade*, *Versu*) as well as non-interactive story generators or story-sifting approaches. They might also be hand-authored or built on a deep learning model such as an LLM. For the purposes of this discussion, we were interested in considering the generative space of such generators regardless of their use case or implementation.

We next considered some possible evaluative and descriptive criteria for generated output. Evaluative measures were meant to be those that could be directly used to assess the quality of the output; tendency towards repetition; novelty or surprise in generated outputs; measures of creativity (always challenging to identify); and analyses of how players specifically in interactive generators chose to interact (e.g. whether there are some choices that are never selected, or choice points at which the player often paused for thought). Descriptive measures were those that simply attempted to identify important aspects of the generated game, such as structures that emerged in player choices or in graphs of automated testing.

The group here recognized that there were many possible users of this kind of evaluative framework (authors, analysts, narrative designers or product managers in a game studio, or the players themselves); many possible data sources that might be the subject of analysis (the static content modules of a game, traces of real playthroughs, traces of automated playthroughs, qualitative feedback, or even complexity analysis of structural change over time); and many objectives for its use (assuring quality, making sure that the game generated stories that were aligned with aesthetic intent, or identifying changes in a generator over time).

The final phase of the discussion focused on attempting a specific approach: a framework for looking at playable interactive games that were tested by an auto-playing bot or by human players, and seeking to identify patterns such as building (or receding) success of the protagonist in their goals, as well as important choices or reversals. Here it was suggested that tracking key variables over time (such as the relationship with another character, or the player's health) would produce curves that correlate to perceived plots: for instance, a plot in which the player never got on with their grumpy neighbour would appear distinct from one in which they were friends from the start – or from one in which they became close only at the end of play.

This idea was pursued in a second working group on the subsequent day.

References

1 Smith, Gillian & Whitehead, Jim. *Analyzing the expressive range of a level generator*, 2010. 10.1145/1814256.1814260.

3.18 Generative Space Analysis for Procedural Narrative Generation

Emily Short (Oxford, GB), Gabriella A. B. Barros (modl.ai – Maceio, BR), Michael Cook (King's College London, GB), Gillian Smith (Worcester Polytechnic Institute, US), Tristan Smith (Creative Assembly – Horsham, GB), Anne Sullivan (Georgia Institute of Technology – Atlanta, US), and Tommy Thompson (AI and Games – London, GB)

License Creative Commons BY 4.0 International license
© Emily Short, Gabriella A. B. Barros, Michael Cook, Gillian Smith, Tristan Smith, Anne Sullivan, and Tommy Thompson

Following on the narrative metrics workgroup day one of this Dagstuhl, in this working group we pursued the idea of characterizing generated narratives by considering stat changes associated with important story features.

For instance, a game or story generator that produced romance stories might track the degree of attraction between the protagonist and each of two suitors; a generator that allowed the protagonist to reach maximal attraction with either suitor at the end of the story could meaningfully be said to offer more narrative range than a generator that always required the protagonist to end up attracted only to one of them, but less range than a generator that also offered affordances for no attraction or for attraction to both suitors as viable final outcomes.

Moreover, the trace could reveal meaningful variation during the midgame as well as at the endings: a generator that allowed for either a swift or a gradual development of attraction might be providing greater narrative range than one in which attraction always increased at a steady pace.

We decided to approach in particular the question of a story generator that was a playable interactive game, where we were interested in visualizing:

- characteristic behaviour of a single numerical stat (for instance, a player trace in which the player quickly achieves victory, in contrast with one in which the player gains more slowly or never wins)
- correlated behaviour of two stats (for instance, a player gains wealth in exchange for relationship status with a particular non-player character)

We expected that in some cases, the trace of a given stat from playthroughs would be tightly constrained (for instance, a player power stat might be monotonically increasing with only slight variations in the pace of gain), while in others, it might be almost fully unconstrained (e.g., a player's currency holdings might change at arbitrary times as the player gained loot drops or chose to spend money at a store). The most revealing traces, however, would be those for stats that were somewhat constrained by gameplay systems but which could nonetheless assume more than one shape in a meaningful way.

In order to explore these questions more rigorously, the team built several prototypes, as follow:

- Two short narrative games in the Ink programming language, in which significant story stats changed depending on player choice. One represented a tea-time conversation with a Mr Wickham and Mr Darcy, with stats representing their reaction to the player, and the other which represented travel aboard a leaky raft, where the player needed to try to reach the shore before the raft sank
- A 2D visualization tool that graphed the progress of these stats in player traces generated from these games; different stats were mapped on top of one another in the same 2D space, which allowed the viewer to recognize correlations between particular stat pairs, but which made it difficult to present many traces simultaneously in order to understand the overall generative space of the story
- A 3D visualization tool that showed stats on different axes together with their change over time

The team further speculated about an MRI-like visualization that would show slices through a three-dimensional state space, showing a heat map arising from many playthroughs in order to help identify the typical emergent shapes. We distributed links to the sample games to other Dagstuhl participants and were able to collect around half a dozen traces of player data for analysis.

Even with these simple examples, it was evident that there would be assorted challenges in using this approach to interpreting the shape of generated narratives. Our turn-based games allowed for only a small amount of variation in the length of the play session, but even in these circumstances, slight differences in the time axis could make it difficult to judge whether two traces were narratively similar or whether the delay in one run constituted a narratively meaningful distinction. The group discussed, but did not have time to explore implementing, ideas whereby the curve of other game stats could be normalized against specific "tentpole" game events.

The group also identified a need for qualitative validation: a core assumption of this approach is that players will regard games with different stat curves as describing different narrative experiences, and that developers (or the persons responsible for creating the analysis) will be able to anticipate which stats are relevant to assess in this way. This assumption would also need to be validated.

3.19 Meaningful Computational Narratives

Pieter Spronck (Tilburg University, NL), Maren Awiszus (Viscom AG – Hannover, DE), Gwaredd Mountain (Square Enix Limited – London, GB), Mike Preuß (Leiden University, NL), Hendrik Skubch (Square Enix AI & ARTS Alchemy Co. Ltd. – Tokyo, JP), Tristan Smith (Creative Assembly – Horsham, GB), and Tony Veale (University College Dublin, IE)

License Creative Commons BY 4.0 International license
© Pieter Spronck, Maren Awiszus, Gwaredd Mountain, Mike Preuß, Hendrik Skubch, Tristan Smith, and Tony Veale

Generative AI offers the opportunity to generate narratives. For instance, a tool such as ChatGPT can take the setup of a storyline and expand it, taking into account the specifications of a story designer. There are, however, some issues with this approach to storytelling:

- There is no guarantee that the generated story will fit the specified background
- There is no guarantee that the generated story will be interesting
- There is no guarantee that the generated story will allow for a suitable continuation

The workgroup ran some simple experiments to test the limitations of a tool such as ChatGPT, and discussed approaches to ensure that generated storylines are meaningful.

3.19.1 Storyline quality

The quality of a generated storyline should be high. That entails that the storyline must be "meaningful" and "good". These two terms need further clarification.

The workgroup defined "meaningful" in the context of storylines as the requirement that within the story, all actions have observable consequences. This is particularly important within the context of games, as game players want to influence the story with the actions that they perform.

The workgroup defined "good" in the context of storylines as the generated storyline having the following four features:

- Semantic coherence
- Pragmatic coherence
- Understandable arcs
- Relatable arcs

The last feature, relatable arcs, is not an absolute necessity, but it tends to be really hard to get players to engage with a story in which there is nothing that they find relatable.

3.19.2 ChatGPT test

In testing the ability of ChatGPT 4 to generate a storyline fitting the requirements of a game designer, responding to player actions, the workgroup gave ChatGPT the following instruction:

You are the Game Master in a Wild West Story Campaign. We are currently in a scene where the player is about to find a caravan on the road which is currently being ambushed by bandits. The kind of Caravan, the motivation of the bandits, any names and characters are free for you to choose. There are three endings to the scene – either the player drives off the bandits with force, negotiates with them, or they side with the bandits and extort the Caravan. Make sure that one of these endings is reached at the end of this conversation, but keep the outcomes hidden from the player. Based on the players actions, use your descriptions and guidance to help them reach the ending that matches their actions. Start with a description of the scene and then react to the actions of the player accordingly.

ChatGPT responded by describing the scene as requested. The workgroup then told ChatGPT several times that the player was just going to hide and wait. ChatGPT responded by trying to make the scene more tense, enticing the player to act, which the workgroup refused to do. At some point, the workgroup asked ChatGPT for suggestions to how the player should respond. Somewhat surprisingly (considering the initial request), some of the suggestions would not lead to the desired outcome. E.g., ChatGPT offered the option to just "wait until one side has defeated the other", while the original request was that the player should actively side with the bandits, side with the caravaneers, or negotiate.

The request was that clearly, a particular story beat should be reached, namely one in which the player gets actively involved in the conflict and helps one of the opposing sides to be victorious. If this is necessary for the plot to evolve, then apparently, in its current incarnation, one cannot count on ChatGPT (which is one of the most advanced language generating tools in existence) to drive the story in the direction that the story designer needs.

3.19.3 Computational narratives

The goal of computational narratives in games is to allow the player the freedom, in a given situation in the game, to execute any action that they like, as long as it potentially fits the situation, while the overall plot of the game remains interesting. Generative AI may support computational narratives, but cannot be relied upon to achieve good results without support from other processes.

Terminology sometimes used by game developers is as follows: The *nebula* is the cloud that contains all possible stories. The *fabula* consists of locations, characters, events, and chronology which define the game world. The fabula restricts the storylines to a subsection of the nebula.

Figure 24 illustrates how a coherent story can be created within the subsection of the nebula that is constrained by the fabula. The story consist of a sequences of "plotlets". Plotlets are selected from a dataset. They can be linked together to form the game's plot. Which plotlet follows a given plotlet in a story, is determined by player actions.

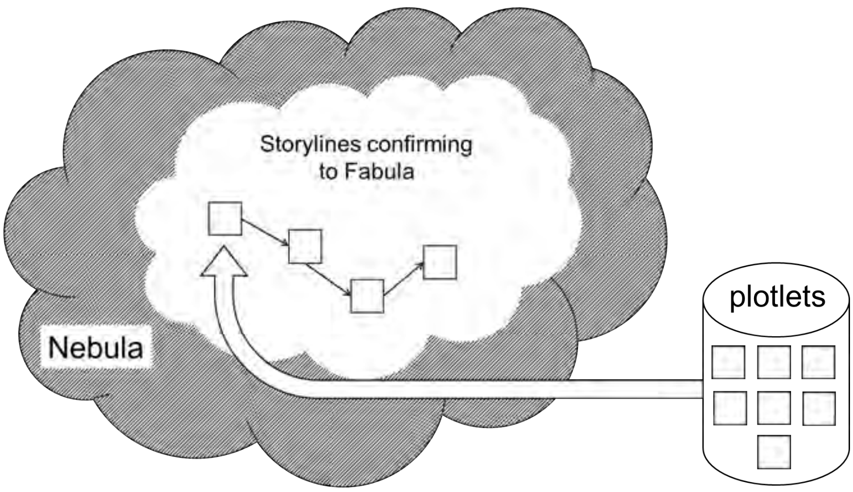


Figure 24 Dynamic storytelling.

3.19.4 Plotlets

A plotlet can be considered a story "beat". It consists of the following elements:
- Pre-conditions, which specify in which circumstances the plotlet can take place
- Action, which specifies which generally formulated action can lead to the plotlet
- Description, which in general terms describes the situation that the player is in
- Selectable actions, which are generally formulated actions that a player can take when the plotlet runs
- Post-facts per selectable action, which define the new situation

When selectable actions are offered to a player, they need to be described in such a way that they make sense to the player within the confines of the story, and show that the dynamic storyteller understands what the player wants to do. For instance, if the story is the one described above, the player may be in a hiding place observing the scene, and as their action tell the game that they "cough loudly". In this particular setting, the dynamic storyteller should not assume that the player character has a heavy cold, but that they want to draw the attention from the opposing sides, or that they want to cause a distraction. This should be reflected in how the game continues.

The selectable actions in the scene above could be "attack bandits", attack "caravaneers", "step forward", or "distract". These options could be offered to the player in the situation, limiting them to actions that lead to other plotlets. The game could also offer the player free input as their action, but map this input to one of the selectable actions. For instance, if the player states that they make a handstand, the game could map this to "distract".

The overall plot of the game is restricted by the available plotlets, giving game developers some control over which stories are told, while still offering the player the freedom to progress the story in a way that appeals to them.

3.19.5 AI

When using plotlets, generative AI comes in as follows:
- The generative AI can describe the situation which exists at the start of a plotlet
- The generative AI can translate generally formulated selectable actions to actions which fit the current situation
- The generative AI can describe the result of the selected action based on the post-facts specified for the action

Moreover, as it is common that multiple plotlets would be available to continue a given story situation, AI analysis can be used to select a follow-up plotlet which best fits a good overall plotline.

3.20 Small, Safe LLMs for In-Game Generation

Tony Veale (University College Dublin, IE), Paolo Burelli (IT University of Copenhagen, DK), Amy K. Hoover (NJIT – Newark, US), Antonios Liapis (University of Malta – Msida, MT), Gwaredd Mountain (Square Enix Limited – London, GB), and Hendrik Skubch (Square Enix AI & ARTS Alchemy Co. Ltd. – Tokyo, JP)

License Ⓒ Creative Commons BY 4.0 International license
© Tony Veale, Paolo Burelli, Amy K. Hoover, Antonios Liapis, Gwaredd Mountain, and Hendrik Skubch

Scaling laws for large language models (LLMs) have allowed LLMs to achieve dramatic improvements in prediction accuracy and generative quality as the depth of their architectures (the number of layers, and of learnable parameters) and the breadth of their training datasets (ever larger subsets of the world-wide-web) grow in size and ambition (Kaplan *et al.,* 2020). This impressive scaling allows LLMs to be applied to tasks that seem to demand more than mere gap-filling or next-word-prediction in text. However, the old truism, popularized by Stan Lee of Marvel Comics, that "with great power comes great responsibility" seems increasingly apt as LLMs grow in generative power. Their ability to produce novel and imaginative responses to arbitrary prompts – one might even say "creative" responses – gives them the ability to amuse and inform, but also an ability to misinform and offend. With the subtle and contextualized generalizations derived from their large training sets, and encoded in their large parameter sets, these models learn the best and the worst of the human condition. This duality, the ability to be used for good or (perhaps unintentionally) for ill, gives us pause when considering the role that LLMs might play in a new generation of computer games.

This working group, which explored the topic of "smaller, safer language models for games", explored whether smaller LLMs (so-called SLMs), with smaller and more selective training sets, can mitigate some of the concerns that are foreseen in a games context. The findings of our group are briefly summarized in the following sub-sections.

3.20.1 Are smaller LLMs inherently safer, or perhaps easier to make safer?

One can imagine that smaller LLMs that are trained on smaller and more controlled trawls of the world-wide-web will visit fewer of the dark corners of the web and include much less of the toxic content that lurks within them. Smallness is not in itself any guarantee of quality, for it is not the case that language users only learn to be abusive or offensive at a certain scale of language acquisition, after they have first learnt to be fluent and felicitous. Indeed, one can make the argument that large LLMs incorporate more knowledge of the world, tacit and vicarious as is is, and so can better reflect on why a certain response may be inappropriate or objectionable. That is, larger LLMs have more fully developed *moral imaginations*, or at the least the makings of such a capacity, and can better reason about their own goals and high-level instructions.

Offensive capabilities come in two forms: the ability to use words that are offensive in themselves, the kind of words that would traditionally go on a "blacklist" or "blocklist", and the more creative ability to use inoffensive words to achieve offensive ends. Even if we carefully control the training of a small LLM so as to exclude the latter, we cannot guarantee that the LLM lacks the capacity for the latter, intentionally or otherwise.

Nonetheless, for tightly-controlled tasks in tightly-controlled domains, such as a well-defined game world, smaller LLMs can be trained from scratch to serve the specific needs of that world, and to only speak the language of that world. Our group considered, for instance, the fictional world of an Old Wild West game, and the generative needs of this domain, including e.g., the language of settlers, of bandits, and of Native Americans (who would not be described as such within the game). We certainly would not want our NPCs to use modern idioms or other anachronisms that would diminish the suspension of disbelief by the game's players. A smaller LLM can be more easily moderated for such a domain, to exclude e.g., references to Native Americans as "savages." These references were commonplace in the Wild West movies of the 1950s and 1960s which still shape our expectations of such a game today, but they are no longer acceptable, even if placed into the mouths of fictional characters that players are not expected to identify with. In short, then, smaller LLMs better allow us to tailor their generative capabilities to the sensitivities and needs of a specific game or brand.

3.20.2 Can smaller, bespoke LLMs be packaged with games and used locally?

A key issue with using LLMs for games concerns the latency and cost of using commercial models in the cloud. For an individual user of OpenAI's LLM offerings, say, the costs are typically low: less than a penny for most calls to the API. However, a game with many thousands of users will accumulate API costs that can prove onerous to the developers of games that rely on real-time access to commercial models. It becomes more practical then for games developers to bundle their own, smaller LLMs with their games, so that these SLMs can be run locally, on the machines of individual players rather than in the cloud infrastructure of the developer (or a third-party provider).

If we accept that SLMs can be bundled for local use in this way, the question now becomes: is an SLM capable of generating what a game needs, when it needs it, with the necessary quality, and with the sufficient speed so as not to drag on the game play? To explore this question, our group conducted some experiments with a small-ish LLM (or a large-ish SLM) called *TinyLlama*. The tiny model is a member of Meta's *LLama* family of LLMs that has just 1.1 billion parameters (contrast this with the 8 billion, 70 billion and 405 billion of different versions of Llama 3.1). To further shrink the model's footprint, we used a 4-bit quantized version of *TinyLlama*, which allocates just 4 bits apiece to each of its 1.1 billion parameters. Although *TinyLlama* has been pre-trained on a smaller corpus than its larger Llama siblings, its dataset is still substantial at three trillion tokens of text. The model is compact enough to load into a basic (unpaid) *Google colab* environment and to run with the benefit of a single T4 GPU.

Our experiments focused on narrative generation within games, by exploring whether *TinyLlama* can be fine-tuned (using a library named *Unsloth*) for the following tasks: to generate a short textual narrative from a skeletal plot structure (or "fabula"); to generate these skeletal plots for itself after it is fine-tuned on a set of more than 10,000 fabula structures (derived from a symbolic story-generation system named *Scéalextric* (Veale, 2021); and to generate dialogue for the characters in these plots, for each step in the plot (as numbered in the fabula). We built a composite training set with examples of each of these three tasks, and 100 steps of fine-tuning on this dataset requires less than 5 minutes of GPU time in a Colab environment.

3.20.3 Findings and Conclusions

Our findings at the stage are mixed. We are impressed with the competence displayed by the *TinyLlama* model on these three tasks, and consider the model's own outputs to be of good quality, very much in keeping with the tenor of the fine-tuning data. The model can generate new fabula structures of its own; impose a fluent rendering on this skeletal form; and generate apt dialogue for two characters that reflects the core plot and the narrative rendering. However, this competence also incurred a noticeable latency that we feel is too great for inline use of these generative capabilities in games. Still, as some careful engineering may yet allow us to reduce this latency to allow for real-time generation with an SLM in games, we are encouraged by the current triumph of competence over performance when using SLMs for "creative" games-related tasks.

References

1 Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeff Wu and Dario Amodei. *Scaling Laws for Neural Language Models*. ArXiv abs/2001.08361, 2020.
2 Tony Veale. *Your Wit Is My Command: Building AIs with a Sense of Humor*. Cambridge, MA: MIT Press, 2021.

3.21 Transferability of Game AI

Vanessa Volz (CWI – Amsterdam, NL), João Miguel Cunha (University of Coimbra, PT), and Tristan Smith (Creative Assembly – Horsham, GB)

License Creative Commons BY 4.0 International license
© Vanessa Volz, João Miguel Cunha, and Tristan Smith

3.21.1 Motivation

While many interesting discoveries have been made in the field of Game AI in recent years and beyond, their application in the games industry is still very limited. This is due to a broad range of challenges, including non-technical issues. One obstacle is undoubtedly that it is seldom obvious how a given algorithmic problem encountered in a game industry setting (e.g. a resource-efficient and believable opponent) can reliably be solved with AI, i.e. with some form of performance guarantees. This is especially important if the problem is expected to change during the course of development of the game; it needs to be guaranteed that changes will not break the AI performance completely.

To do this, we would need to be able to identify what part of an existing solution is transferable to which other problem (in the same game or beyond). In the following, we therefore first address which different types of transferability may exist. We then discuss some approaches to comparing problems as one popular way to identify opportunities for transferring knowledge and/or algorithmic approaches. Finally, we point out directions for future research.

3.21.2 Transferability dimensions

Transferability can have various realisations in different settings. For example, we have seen the great care that had to go into training AlphaStar in order to show somewhat promising behaviour against different opponent strategies and game versions [1].

Even algorithms intended to be more general, like the participants in GVGAI competitions, tend to have different strengths and weaknesses [2]. This also holds for different hyperparameter settings of said algorithms. While we can certainly explain some behaviour, a reliable method to identify different games where we would expect to see similar performance patterns is an unsolved problem [3].

In addition, in image recognition and generation, we know that re-training only the last few layers of a neural network can achieve results faster than training from scratch (transfer learning). Some learned parameters and structure of the neural network must thus transfer to images in general. Relatedly, there has been some work on training weight-agnostic networks that might be faster to adapt to different environments [4].

We can characterise the above examples of transferability challenges based on two dimensions of transferability:

- Size of gap between contexts: Are we transferring between different game versions of the same game? Different game levels? Different games of the same genre? Across genres? To contexts beyond games?
- Level of abstraction: What is it that we attempt to transfer? A trained model? The training setup and associated hyperparameters? Knowledge about strengths, weaknesses, and performance estimates of different approaches? Predictions on training costs and resources?

3.21.3 Problem similarity

In order to investigate which classes of transferability are realistic, a comparison approach between different problems (games) is needed (among other things). We list some existing ideas below:

- Anecdotally, in the game industry this topic is approached by characterising different tasks based on traits like independence/parallelisation, allocation/optimisation, domain considerations, necessary accuracy and speed, and degree of design control.
- In games research, game genres (platformer, shooter etc) are often utilized to characterise games. Games of the same genre are often assumed to be similar to each other, as well as levels of the same game.
- There are also hand-crafted features for game characterisation in the context of different general game playing frameworks [5, 6, 7].
- Research on level blending in the context of procedural content generation (via machine learning) often requires representing different (selected) games with the same representation, which allows for comparisons in that representation space [8]. See also the section on distance and density measures 3.13.
- In related domains (e.g. evolutionary optimisation), various hand-crafted and data-driven features exist for the purposes of problem characterisation as well as for automatic algorithm selection. However, this is still an active area of research when it comes to interpretability and robustness of the features, as well as transferability of the results in different dimensions and settings. [9]
- In popular culture, games are often described in the context of other games. A concept of similarity is therefore implicitly established. This includes gameplay, but also cultural, location and historical connections [10, 11].

3.21.4 Future directions

Beyond what was described in section 3.21.3, further avenues for research exist in terms of problem comparison and transferability of game AI research in general:

- Compute or automatically select features, for example based on:
  - an encoding obtained from large language models processing a game description (prose, game description language, player reviews).
  - performance comparison of different AI and human players.
  - human labelling to allow for supervised learning.

- Identify what aspects of a game might be subject to change during the development process, and ensuring robustness against these changes in the game-playing AI (Transfer across versions, unclear abstraction level).
- Computing worst-case performance for games as a way of establishing guarantees (Transfer performance estimates, unclear context gap).
- Based on the identified dimensions described in section 3.21.2, survey and categorise existing work on transferability / generalisability.
- Continue research on robustness of game AI using techniques such as adversarial training and curriculum learning.

References

1 O. Vinyals et al., "Grandmaster level in StarCraft II using multi-agent reinforcement learning," Nature, vol. 575, no. 7782, pp. 350–354, 2019.
2 D. Perez-Liebana, J. Liu, A. Khalifa, R. D. Gaina, J. Togelius, and S. M. Lucas, "General Video Game AI: A Multitrack Framework for Evaluating Agents, Games, and Content Generation Algorithms," IEEE Trans. Games, vol. 11, no. 3, pp. 195–214, 2019.
3 H. Horn, V. Volz, D. Pérez-Liébana, and M. Preuss, "MCTS/EA hybrid GVGAI players and game difficulty estimation," in IEEE Conference on Computational Intelligence and Games (CIG), 2016.
4 A. Gaier and D. Ha, "Weight Agnostic Neural Networks," in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2019.
5 P. Bontrager, A. Khalifa, A. Mendes, and J. Togelius, "Matching Games and Algorithms for General Video Game Playing," AIIDE, vol. 12, no. 1, pp. 122–128, 2021.
6 A. Mendes, J. Togelius, and A. Nealen, "Hyper-heuristic general video game playing," in IEEE Conference on Computational Intelligence and Games (CIG), 2016.
7 M. Stephenson, D. J. N. J. Soemers, é. Piette, and C. Browne, "Measuring Board Game Distance," in Computers and Games, vol. 13865, in Lecture Notes in Computer Science, vol. 13865., pp. 121–130, 2023.
8 V. S. R. Atmakuri, S. Cooper, and M. Guzdial, "Game Level Blending using a Learned Level Representation," in IEEE Conference on Games, 2023.
9 A. Nikolikj, S. Džeroski, M. A. Muñoz, C. Doerr, P. Korošec, and T. Eftimov, "Algorithm Instance Footprint: Separating Easily Solvable and Challenging Problem Instances," in Proceedings of the Genetic and Evolutionary Computation Conference, in GECCO '23, pp. 529–537, 2023.
10 J. P. Zagal, "A framework for games literacy and understanding games," in Proceedings of the 2008 Conference on Future Play: Research, Play, Share, in Future Play '08, pp. 33–40, 2008.
11 A. de Voogt, A.-E. Dunn-Vaturi, and J. W. Eerkens, "Cultural transmission in the ancient Near East: twenty squares and fifty-eight holes," Journal of Archaeological Science, vol. 40, no. 4, pp. 1715–1730, 2013.

4 Panel discussions

4.1 Discussion

Pieter Spronck (Tilburg University, NL), Duygu Cakmak (Creative Assembly – Horsham, GB), Setareh Maghsudi (Ruhr-Universität Bochum, DE), and Diego Perez Liebana (Queen Mary University of London, GB)

License Creative Commons BY 4.0 International license
© Pieter Spronck, Duygu Cakmak, Setareh Maghsudi, and Diego Perez Liebana

As per usual, on the Friday before lunch we held a plenary meeting where we were going to reflect on the past week, and look forward to a potential follow-up. This is a report on the discussion.

4.1.1 Preparing for the event

Several participants remarked that they found it a positive element that it was not necessary to specifically prepare for the event. However, for newcomers this was not immediately clear from the invitation. Moreover, some preparation possibilities could be welcome. The Discord group set up for the event would make this relatively easy, by, for instance, having a resources channel with documents, links to tutorials, and tools that may be used during the event in case delegates want to install them beforehand.

More work could also be done to connect people with each other. People spontaneously connected through Discord, specifically for travel arrangements, but we could try to let people also connect for topic discussions.

It might be nice if some sponsoring can be found to support people with less financial means to come.

4.1.2 Schedule

The general schedule for each day was a plenary session of about half an hour after breakfast, followed by workgroups. In most cases, the workgroups that ran before lunch were also run after lunch. There was some time set aside right after lunch for those who wanted to take a walk. Around 5 o'clock a one-hour plenary session was held in which the workgroups reported on their results. After dinner there was time for social activities; some of these were organized (a pub-quiz – which should probably be limited to three questions rather than six) and some spontaneously organized (a tutorial, a movie, and two roleplaying games), but most consisted of people spending time together in discussions and game-playing.

The organizers had been struggling a bit to reserve time for people exercising. It was considered to move the last plenary session of the day to after dinner, so that the time slot between five and six o'clock could be used for exercises, but ultimately on Monday morning it was decided not to do that, and just allow people to leave a workgroup earlier if they want to exercise. Most agreed at the end of the week that this is the best approach, as the days are long and intensive and we should not schedule work activities after dinner.

4.1.3 Monday

The Monday tends to be a day where too much happens. We started with an introductory game which took quite a lot of time. The game used was "two truths and a lie," which took over an hour-and-a-half to run. This is clearly too long. It was suggested to simply let everyone introduce themselves quickly with one sheet (and have the sheets available for perusal in Discord), and maybe have an introductory game on Monday after dinner. It would even be possible to do this on Sunday night, though not everyone may be present then, and some who had to travel for a long time might already have retired. A small questionnaire up front, to which the answers are shared via Discord, may also work.

After the introductory game workgroups were proposed. There was just enough time to do this before lunch, so that the workgroups could start after lunch. The problem is that these workgroups were relatively short, lasting only half a day; no more than two hours total (excluding coffee breaks). With a bit more preparation this time could be used more effectively. We could plan tutorials on the Monday afternoon, and/or preparation work for workgroups that are going to be run from Tuesday to Thursday. This would work best if some workgroups would already be known before the event takes place. For instance, collecting information before the event on workgroup ideas via online forms could make the use of time more effective.

4.1.4 Organizing workgroups

Some suggestions were made to streamline workgroup organization. In particular, sometimes people were trying to find the rooms where certain workgroups were held. A good suggestion was to have a large grid hanging in the plenary meeting room, with a list of rooms and an indication where workgroups are located. It would also help if it is made explicit somewhere (either on such a grid, or in Discord) what a workgroup intends to do. This will help participants who want to have a "bumblebee" approach to visiting workgroups.

4.1.5 Workgroup presentations

Workgroup presentations are very useful to communicate to all participants what a workgroup did. Doing this with sheets is extra helpful, as those sheets can then later be used to base proceedings on, and can be used as a point of reference for all participants. However, there are at least two issues with the sheets. The first is that some presentations were rather long, and since there usually were five presentations in the one-hour time slot before dinner, sometimes the later presentations were pressed for time too much. The second is that sometimes participants were still working on sheets during the presentations, which meant that they could not follow the presentations others were giving.

The first may be solved by limiting the number of sheets to five, and keeping stricter track of time. The second may be harder to solve (though limiting to five sheets may help for that as well). However, this shouldn't mean that information that is relevant is not kept track of using sheets, only that a maximum of five sheets should be presented.

Another helpful suggestion to save time was that if a workgroup intended to continue the next day (which happened for multiple workgroups), their presentation should be really short and mostly focusing on what will be done the next day. This would mean that several presentations would be use a considerably reduced amount of time.

As support for the presentations, people also suggested using our Discord server having a dedicated channel for outputs of the workgroups, or a Google Drive folder dedicated to this purpose, where each group stores their final materials. This could contain presentations, links, code and even their final workgroup reports.

4.1.6 Rules of conduct

Because we want the event to be a safe place, we introduced rules of conduct for a previous event, and continued using them now. However, we now realized that if someone has an issue, having them contact the organizers may not (always) be the best approach. We should appoint two trusted people (one man, one woman) as independent confidential supports. We still do not expect that they will have a lot of work to do, but having them present would be appreciated.

4.1.7 Topics for future events

The *Creative Game AI* theme for the event was really topical, and almost all (if not all) workgroups focused on the theme, and almost all (if not all) topics presented in the proposal were researched. The organizers elicited suggestions for future topics. The following were suggested:

- Games and open-endedness, artificial life, creativity
- Emergence
- Relatedness between human intelligence and game AI
- Transferability from games and computer science to other disciplines
- Research practices, holes in methodologies, comparing methods
- Game AI and education
- Ethics and morality, both their influence on games and how they can be "improved" with games

It was also suggested to replace the term "game" with the term "play", to expand the subject area.

Considering the success of the event and the enthusiasm of the participants, an event in 2026 would be very welcome.

Participants

- Maren Awiszus
Viscom AG – Hannover, DE
- Gabriella A. B. Barros
modl.ai – Maceio, BR
- Paolo Burelli
IT University of Copenhagen, DK
- Duygu Cakmak
Creative Assembly – Horsham, GB
- Filippo Carnovalini
VU – Brussels, BE
- Alex J. Champandard
creative.ai – Wien, AT
- M Charity
New York University, US
- Michael Cook
King's College London, GB
- João Miguel Cunha
University of Coimbra, PT
- Alena Denisova
University of York, GB
- Alexander Dockhorn
Leibniz Universität Hannover, DE
- Anders Drachen
University of Southern Denmark – Odense, DK
- Manuel Eberhardinger
Hochschule der Medien – Stuttgart, DE
participant Raluca D. Gaina
Tabletop R&D – London, GB
- James Goodman
Queen Mary University of London, GB
- Christian Guckelsberger
Aalto University, FI
- Greta Hoffmann
TH Köln, DE
- Amy K. Hoover
NJIT – Newark, US
- Chengpeng Hu
Southern Univ. of Science and Technology – Shenzen, CN
- Leonie Kallabis
TH Köln, DE
- Ahmed Khalifa
University of Malta – Msida, MT
- Antonios Liapis
University of Malta – Msida, MT
- Simon M. Lucas
Queen Mary University of London, GB
- Setareh Maghsudi
Ruhr-Universität Bochum, DE
- David Melhart
University of Malta – Msida, MT
- Gwaredd Mountain
Square Enix Limited – London, GB
- Matthias Müller-Brockhausen
Leiden University, NL
- Mirjam Palosaari Eladhari
Stockholm University, SE
- Diego Perez Liebana
Queen Mary University of London, GB
- Johanna Pirker
LMU München, DE
- Mike Preuß
Leiden University, NL
- Emily Short
Oxford, GB
- Hendrik Skubch
Square Enix AI & ARTS Alchemy Co. Ltd. – Tokyo, JP
- Gillian Smith
Worcester Polytechnic Institute, US
- Tristan Smith
Creative Assembly – Horsham, GB
- Pieter Spronck
Tilburg University, NL
- Anne Sullivan
Georgia Institute of Technology – Atlanta, US
- Tommy Thompson
AI and Games – London, GB
- Tony Veale
University College Dublin, IE
- Vanessa Volz
CWI – Amsterdam, NL

Report from Dagstuhl Seminar 24271

Theory of Randomized Optimization Heuristics

Anne Auger*[1], Tobias Glasmachers*[2], Martin S. Krejca*[3], Johannes Lengler*[4], and Alexander Jungeilges†[5]

1 **Inria Saclay Ile-de-France, Palaiseau, FR.** anne.auger@inria.fr
2 **Ruhr University Bochum, DE.** tobias.glasmachers@ini.rub.de
3 **Ecole Polytechnique, Institut Polytechnique de Paris, Palaiseau, FR.** martin.krejca@polytechnique.edu
4 **ETH Zürich, CH.** johannes.lengler@inf.ethz.ch
5 **Ruhr University Bochum, DE.** alexander.jungeilges@ini.ruhr-uni-bochum.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 24271 "Theory of Randomized Optimization Heuristics", which marks the twelfth installment of our biennial seminar series.

This iteration saw a lot of discussion on important, yet rarely analyzed topics in the domain of heuristic optimization, such as mixed-integer problems, permutation spaces, and coevolution. Moreover, it aimed at unifying existing results by discussing mathematical tools (such as drift analysis), the structure of discrete problems, and a common framework for theoretical analysis and practical implementation. Last, more recent and important topics, such as constrained and multi-objective optimization, were a major part of the seminar. We had a vivid exchange in various breakout sessions and different talks, with a great mix of junior and senior participants, which was very positively received.

Seminar June 30 – July 5, 2024 – https://www.dagstuhl.de/24271

2012 ACM Subject Classification Theory of computation → Bio-inspired optimization; Theory of computation → Evolutionary algorithms; Theory of computation → Theory of randomized search heuristics

Keywords and phrases Black-Box Optimization Heuristics, Evolution Strategies, Genetic and Evolutionary Algorithms, Runtime and Convergence Analysis, Stochastic Processes

Digital Object Identifier 10.4230/DagRep.14.6.215

1 Executive Summary

Anne Auger (Inria Saclay Ile-de-France, Palaiseau, FR)
Tobias Glasmachers (Ruhr University Bochum, DE)
Martin S. Krejca (Ecole Polytechnique, Institut Polytechnique de Paris, Palaiseau, FR)
Johannes Lengler (ETH Zürich, CH)

License Creative Commons BY 4.0 International license
© Anne Auger, Tobias Glasmachers, Martin S. Krejca, and Johannes Lengler

This seminar is part of a biennial seminar series. This year, we envisioned to focus on constraint handling, multivariate estimation-of-distribution algorithms (EDAs), as well as stochastic approximation and Markov chain stability analysis. This vision worked well for constraint handling but not so much for the other two topics, since several key invitees rejected our invitations. Nonetheless, the seminar quickly and organically refocused, and we had plenty of other important topics to discuss instead, as we detail below.

* Editor / Organizer
† Editorial Assistant / Collector

Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

The previous iteration of the seminar had taken place during the peak of the COVID Omicron wave, with lots of restrictions in place. We were glad to see this seminar happening again in the usual format. It was still a bit smaller than usual, due to unfortunate last-minute cancellations, but we managed to partially make up for that by inviting more young researchers. In any case, the group size of about 40 worked very well.

We used morning sessions for talks and group discussions, and afternoon sessions for breakout formats, with the exception of Monday. Several time slots were exempt from the schedule, explicitly leaving free time for individual discussions and/or leisure activities. We kept talks reasonably short, usually at a maximum of 20 minutes, leaving sufficient time for discussion. We had to stop discussions only very rarely so as to respect Dagstuhl's meal schedule. Otherwise, we managed to give sufficient space to each topic by flexibly re-scheduling a few topics on the fly.

We used the first day for group forming and to bring everyone to the same page. We started with a general introduction, a small ice-breaker game, and a series of overview talks. We had talks on optimization heuristics in discrete and continuous search spaces, which represents a classic divide in our community, as well as introductions to multi-objective optimization and constraint handling. Furthermore, we had presentations by Thomas Sauerwald on *Balls into Bins* and by Vanessa Volz on *Democratising Real-World Problem Tailored Optimisation*, two contributions a bit remote from the core topic of the seminar, so as to build bridges to neighboring fields. Kathrin Klamroth and Oswin Krause continued this series later on with presentations on *A Multiobjective Perspective on Constraint Handling* and on *Evolution in the Quantum world: on tuning quantum dot devices.*

We invited and actually nudged all young researchers to introduce themselves and their research with brief presentations. We had a total of seven such talks, taking place in the morning slots of the following days. This format was generally perceived as useful and fruitful.

Besides junior and outreach talks, there was of course also a lot of activity on core topics of the community. Two selected highlights were the presentation of Armand Gissler's proof of linear convergence of the CMA-ES algorithm on convex quadratic problems by means of Markov chain analysis, and the presentation of Per Kristian Lehre on his SLO hierarchy, a categorization of complexity classes for black-box optimization.

There was a total of 10 breakout sessions, which are all summarized in Section 4. They covered a broad spectrum of diverse and important topics, and they easily replaced the initial focus topics that we had in mind. This time, we had a fair amount of breakout sessions that garnered the attention of both the discrete and the continuous subcommunity in roughly equal parts. As such, we had sessions on negative drift (which is a tool applicable in either domain), on mixed-integer problems (which requires expertise from both domains), the general structure of discrete problems with the aim to classify them similarly to how continuous problems can be classified, and a session on an abstract framework that can be used by practitioners and analyzed by theoreticians. Moreover, we discussed rather recent hot topics such as multi-objective optimization, co-evolution, and rich surrogate models, assessing which direction to take for the future. This was complemented by a session on permutation problems, highlighting a domain that only saw little attention so far, as well as a session on how to place the theoretical analysis of randomized search heuristics within the grander spectrum of artificial intelligence. All of these sessions had a larger number of participants and vibrant discussions, showing the interest of the community in these topics.

We established a new format for collecting topics and for scheduling breakout sessions. It was inspired by a workshop that had taken place at the Lorentz Center in Leiden (Netherlands) earlier this year. Instead of collecting topics asynchronously upfront and over lunch breaks, we dedicated a short session to it. To this end, we replaced the usual classroom arrangement with a fully symmetrical setup, asking participants to step up, briefly explain their proposal, write down a title on a sheet of colored paper, and put it up on a wall. Moreover, instead of trying to reach consensus, we left it to session organizers to schedule their sessions so as to minimize (perceived) overlap. The new process was received as an improvement.

For the first time in this seminar series, we offered a trip to Trier as a social activity on Wednesday afternoon, with a hike taking place in parallel (and in the rain). This worked very smoothly, and it was a great experience, especially for participants from far away, even if they had been to Dagstuhl before.

We are very grateful for the opportunity of organizing a seminar in Dagstuhl. We have to thank for the financial support, for the comfort of rooms directly at the venue, four meals a day, wonderful facilities, and all of that including full service by very friendly, reactive, and always helpful staff. Thank you!

The organizers,

Anne Auger, Tobias Glasmachers, Martin Krejca, Johannes Lengler

2 Table of Contents

Working Groups

3 Overview of Talks

3.1 Reconstructing a true individual from a noisy one: a new perspective gives new insights

Denis Antipov (University of Adelaide, AU) and Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR)

License Creative Commons BY 4.0 International license
© Denis Antipov and Benjamin Doerr
Joint work of Denis Antipov, Benjamin Doerr, Alexandra Ivanova
Main reference Denis Antipov, Benjamin Doerr, Alexandra Ivanova: "Already Moderate Population Sizes Provably Yield Strong Robustness to Noise", in Proc. of the Genetic and Evolutionary Computation Conference, GECCO 2024, Melbourne, VIC, Australia, July 14-18, 2024, ACM, 2024.
URL https://doi.org/10.1145/3638529.3654196

EAs are often used to solve problems in uncertain environments, e.g., for optimization of dynamic functions or functions with stochastic noise. They have been shown to be effective in such circumstances, however the theoretical understanding of why they are robust is very limited. In this talk, we focus on the prior noise, that is, when the individuals might be randomly changed before we evaluate their fitness. The previous works which considered this setting showed that the $(1+1)$ EA can withstand only low rates of noise, while population-based EAs seem to be quite robust even to a strong noise. However, the analysis of EAs on the noise is often tailored to the considered problems, which is caused by its complexity: even the elitist algorithms which are usually described by simple stochastic processes might adopt a much more complex non-elitist behavior.

In this talk, we describe a new method of analysis for prior noise problems. It is based on the observation that when we know both the parent and its offspring affected by the noise, we can estimate the distribution of the true noiseless offspring as if it was created via crossover between the parent and the noisy offspring. With this perspective, we show that the $(1+\lambda)$ EA and the $(1, \lambda)$ EA with only a logarithmic population size are very robust even to constant-rate noise (the one which affects a constant fraction of all fitness evaluations): they can solve noisy OneMax in $O(n \log(n))$ fitness evaluations, which is the same as in the noiseless case.

This talk is a shortened version of the talk given at ThRaSH seminar earlier in June.

3.2 Introduction to Constrained Optimization

Dirk V. Arnold

License Creative Commons BY 4.0 International license
© Dirk V. Arnold

Topics covered in the introductory talk on constrained optimization include penalty functions, Lagrange multipliers, augmented Lagrangians and the method of multipliers, and exact Lagrangian techniques. Adaptations of those methods for use in black-box optimization and different types of constraints were also discussed.

3.3 Understanding the Fitness Landscape of a Real-World Sequential Decision-Making Problem

Asma Atamna (Ruhr-Universität Bochum, DE)

License Creative Commons BY 4.0 International license
© Asma Atamna

Real-world optimization problems can be quite different from the ones used typically to benchmark evolution strategies (ESs). They are often inherently noisy and can be highly multimodal. In this talk, we take a closer look at the fitness landscape of a sequential decision-making problem consisting of controlling a solid material-processing facility. We compare the performance of Proximal Policy Optimization (PPO) to that of the state-of-the-art ES, Covariance Matrix Adaptation Evolution Strategy (CMA-ES). Our longer-term goal is to:

- better understand what makes standard approaches, like PPO, perform better on such difficult fitness landscapes,
- draw inspiration from these standard approaches to improve the performance of ESs.

3.4 How Theoretical Considerations Can Help in Algorithm Design: 3 Examples

Dimo Brockhoff (INRIA Saclay – Palaiseau, FR)

License Creative Commons BY 4.0 International license
© Dimo Brockhoff
Joint work of Anne Auger, Mohamed Gharafi, Nikolaus Hansen, Baptiste Plaquevent Jourdain, Cheikh Touré

Progress in scientific research is typically achieved with the iterative process of observing the real world via an experiment, building a theory for the observations made, and using this theory to make predictions for a new experiment. The design of an optimization algorithms also typically follows (or should follow) this process. In my talk, I gave three examples in the context of multiobjective blackbox optimization where theoretical considerations, specific experiments, and visualizations improved the algorithm or detected failure modes.

3.5 Covariance Matrix Adaptation for MCMC Sampling: some Differences with the Optimization Context

Alexandre Chotard (Université du Littoral de la Côte d'Opale – Opal Coast, FR)

License Creative Commons BY 4.0 International license
© Alexandre Chotard

The efficiency of a Metropolis-Hastings depends of how well suited the candidate distribution that the algorithm uses to move is with respect to the target distribution. As the algorithm explores the space of the target distribution, the information from the samples can be used to adapt the candidate distribution in the way of CMA-ES. In this talk, I will present some particularities that one must take care when doing covariance matrix adaption in a MCMC sampling context. In particular:

- Adaptation can ruin the sampling process.
- How do we measure performance?
- Differences in the stability analysis.
- The covariance learnt by CMA may not be what you want.

3.6 Dumb and Good: Another Advertising Campaign For Random Parameters (Following a Power Law)

Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR)

License Creative Commons BY 4.0 International license
© Benjamin Doerr
Main reference Benjamin Doerr, Huu Phuoc Le, Régis Makhmara, Ta Duy Nguyen: "Fast genetic algorithms", in Proc. of the Genetic and Evolutionary Computation Conference, GECCO 2017, Berlin, DE, July 15-19, 2017, pp. 777–784, ACM, 2017.
URL https://doi.org/10.1145/3071178.3071301

This talk is a reminder that there is a simple way to choose good parameter values together with two very recent results, one mathematical and one empirical.

As result of a mathematical runtime analysis, [1] proposed the use of bit-wise mutation with a random mutation rate, chosen from a discrete power-law distribution (for bit-string representations of length n: you sample α from a discrete power-law with exponent 1.5 and flip bits independently with probability α/n). This ~~dumb~~ parameter-less mutation operator (used in the (1+1) EA) was able to optimize all jump functions in a time close to the one that is obtained with the best problem-specific mutation rate (which highly depends on the instance). Similar result where shown for other optimization problems, for more than one parameter, for other parameters, in multi-objective settings, for permutation spaces, ...

Despite these favorable results, heavy-tailed random parameter choices are still not used a lot in practice (but they are). To overcome this short-coming, and to entertain the valued audience, I shall present two recent result. (1) A mathematical runtime analysis showing yet another way how this mutation operator drastically outperforms standard bit-wise mutation (namely in filling the gaps between the optima of the subproblems used in decomposition-based multi-objective algorithms). (2) An empirical proof that heavy-tailed parameters are easy to use and can easily give better results that intensive automated parameter tuning (for the target set selection problem studies by [2]).

References

1 Benjamin Doerr, Huu Phuoc Le, Régis Makhmara, and Ta Duy Nguyen. Fast genetic algorithms. In *Genetic and Evolutionary Computation Conference, GECCO 2017*, pages 777–784. ACM, 2017.
2 Albert López Serrano and Christian Blum. A biased random key genetic algorithm applied to target set selection in viral marketing. In *Genetic and Evolutionary Computation Conference, GECCO 2022*, 241–250. ACM, 2022.

3.7 Multiobjective optimization

Carlos M. Fonseca (University of Coimbra, PT)

License Creative Commons BY 4.0 International license
© Carlos M. Fonseca
Joint work of Andreia P. Guerreiro, Kathrin Klamroth, Carlos M. Fonseca
Main reference Andreia P. Guerreiro, Kathrin Klamroth, Carlos M. Fonseca: "Theoretical Aspects of Subset Selection in Multi-Objective Optimisation", pp. 213–239, Springer, 2023.
URL https://doi.org/10.1007/978-3-031-25263-1_8

This talk provides an introduction to optimization with multiple objective functions. The concepts of Pareto dominance and solution efficiency are presented, and a number of alternative problem formulations are introduced and discussed. While the set of efficient solutions, and the corresponding non-dominated set, expose the tradeoffs between the different objectives, selecting an overall best solution requires additional preference information to be provided by a Decision Maker, and ultimately changes the problem to be solved. Preference modelling via utility functions and outranking methods is outlined and related to scalarization approaches and population-based methods, respectively, as two important families of multiobjective optimization methods. Finally, some of the challenges posed by the experimental evaluation of such methods are identified, and some additional research topics are highlighted.

Acknowledgement. This work was partially supported by national funds through the FCT – Foundation for Science and Technology, I.P. in the scope of project CISUC—UID/CEC /00326/2020 and by the European Social Fund, through the Regional Operational Program Centro 2020.

3.8 Surrogate methods for AL-CMA-ES with binary constraint values

Oskar Girardin (Ecole Polytechnique – Palaiseau, FR)

License Creative Commons BY 4.0 International license
© Oskar Girardin

In the context of constrained optimization problems, we might face constraints, which yield a boolean (feasible or infeasible). We tackle this issue by building a surrogate model to find an approximation of the constraint boundary. The model is used within the Augmented Lagrangian CMA-ES (AL-CMA-ES) algorithm, which is a competitive algorithm for constrained black-box optimization. In this talk we discuss the achievements as well as the challenges of our approach.

3.9 Convergence proof of CMA-ES: analysis of underlying normalized Markov chains

Armand Gissler (Ecole Polytechnique – Palaiseau, FR)

License Creative Commons BY 4.0 International license
© Armand Gissler

In this talk, I will present the overview of a recent convergence proof of the algorithm CMA-ES (Covariance Matrix Adaptation – Evolution Strategies). More precisely, for a certain class of function called scaling-invariant, the normalization of the states of CMA-ES defines a homogeneous Markov chain, that we prove to be irreducible. When the objective function is ellipsoid (i.e., with ellipsoid level sets) we are able to establish a drift condition and the geometric ergodicity of this Markov chain. We deduce the linear convergence from a law of large numbers.

3.10 Introduction to Direct Search in Continuous Domains

Tobias Glasmachers (Ruhr-Universität Bochum, DE)

License Creative Commons BY 4.0 International license
© Tobias Glasmachers

In this short introductory talk I introduce the main challenges and questions of interest for the analysis of evolutionary algorithms in continuous domains. One major distinction is between evolution strategies with versus without covariance matrix adaptation. While the latter class is well analyzed, the former poses significant challenges. Designing appropriate drift functions reflecting the complex dynamics of online adaptation rules for step size and covariance matrix is an open problem.

3.11 Theoretical Aspects of Set-Quality Indicators for Multiobjective Optimization

Andreia Guerreiro (INESC ID Lisboa – Lisbon, PT)

License Creative Commons BY 4.0 International license
© Andreia Guerreiro

Set-quality indicators, which map a point set into a scalar value, are a convenient way to assess (the image of) solution sets in multiobjective optimization. Such indicators may comprise in this scalar value the proximity of the set of points to the Pareto front, as well as information regarding the distribution of points in the set. Performance assessment through quality indicators can be viewed as a transformation of the multiobjective optimization problem into a single-objective one, where the goal is to find a point set, frequently bounded in size, that maximizes the quality indicator. Consequently, each indicator is biased towards some point sets. The study of the theoretical properties of quality indicators allows to characterize the indicator-optimal subsets and, therefore, to understand such biases and their implications in performance assessment and in indicator-based evolutionary multiobjective optimization algorithms. Such theoretical aspects will be discussed in this talk.

3.12 Computation of a Numerical Drift and Potential Function for the (1+1)-ES

Alexander Jungeilges (Ruhr-Universität Bochum, DE)

License Creative Commons BY 4.0 International license
© Alexander Jungeilges

Designing a suitable potential function is one of the crucial steps when performing drift analysis. In the context of evolution strategies, it has to capture the complex adaptation mechanism of the algorithm. Especially for evolution strategies with covariance matrix adaptation, the design of such a function imposes a significant challenge. In this flash talk, I will present a novel approach to numerically compute a suitable potential function, that provides a constant drift across a grid of parameter values.

3.13 A Multiobjective Perspective on Constraint Handling

Kathrin Klamroth (Universität Wuppertal, DE)

License Creative Commons BY 4.0 International license
© Kathrin Klamroth
Joint work of Kathrin Klamroth, Jørgen Tind
Main reference Kathrin Klamroth, Jørgen Tind: "Constrained optimization using multiple objective programming", J. Glob. Optim., Vol. 37(3), pp. 325–355, 2007.
URL https://doi.org/10.1007/S10898-006-9052-X

Constraint handling techniques can often be interpreted as (partial) scalarizations of associated multiobjective optimization problems: In a first step, constraints are relaxed and re-interpreted as additional objective functions, and in a second step appropriate scalarizations are used to obtain one (or a series of) unconstrained problems. This naturally leads to the questions whether constraint handling techniques could be chosen to reflect decision maker preferences, and/or to provide additional trade-off information at little extra cost. A prominent example is the relaxation of constraints in a Lagrangian relaxation framework that corresponds to weighted sum scalarizations of associated multiobjective problems. An interesting alternative are hypervolume scalarizations that have non-linear level curves and can thus generate solution in non-convex parts of the outcome set / the Pareto front, respectively.

3.14 Theory of Stochastic Drift – A Guided Tour

Timo Kötzing (Hasso-Plattner-Institut, Universität Potsdam, DE)

License Creative Commons BY 4.0 International license
© Timo Kötzing
URL https://hpi.de/friedrich/docs/scripts/22_AllDrift/index.html

In studying randomized search heuristics, a frequent quantity of interest is the first time a (real-valued) stochastic process obtains (or passes) a certain value. Commonly, the processes under investigation show a bias towards this goal, the stochastic drift. Turning an iteration-wise expected bias into a first time of obtaining a value is the main result of drift theorems. This thesis gives an introduction into the theory of stochastic drift, providing examples and reviewing the main drift theorems available. Furthermore, the thesis explains how these methods can be applied in a variety of contexts, including those where drift theorems seem a counter intuitive choice. Later sections examine related methods and approaches.

3.15 Evolution in the Quantum world: on tuning quantum dot devices

Oswin Krause (University of Copenhagen – Copenhagen, DK)

License Creative Commons BY 4.0 International license
© Oswin Krause

Recently, the field of quantum computing (QC) has seen impressive results. Since quantum devices require meticulous tuning, a lot of research efforts are guided towards automating these tasks and the field is in some cases moving towards evolution. In this talk, I present some of the background of quantum devices and present some of the tuning problems that arise and how they are challenging the current state of the art in evolutionary algorithms. The two main points i discuss will be: a) as observations are often indirect, fitness evaluations are often based on model fits, that take a lot of time and can introduce systematic errors. b) Due to the underlying discrete state space of digital quantum computers, objective function landscapes are often a series of plateaus with the interesting physics lying on their intersections. I am happy to also host a breakout session to discuss some of these points.

3.16 The SLO-hierarchy of pseudo-Boolean Problems

Per Kristian Lehre (University of Birmingham, GB)

License Creative Commons BY 4.0 International license
© Per Kristian Lehre
Joint work of Duc-Cuong Dang, Per Kristian Lehre
Main reference Duc-Cuong Dang, Per Kristian Lehre: "The SLO Hierarchy of pseudo-Boolean Functions and Runtime of Evolutionary Algorithms", in Proc. of the Genetic and Evolutionary Computation Conference, GECCO 2024, Melbourne, VIC, Australia, July 14-18, 2024, ACM, 2024.
URL https://doi.org/10.1145/3638529.3654221

A fundamental question in theory of evolutionary computation is to understand structural properties of fitness landscapes that make them easy or hard for given classes of evolutionary algorithms. This talk introduces the so-called SLO hierarchy which provides a classification of all pseudo-Boolean functions according to the sparsity of local optima and density of fitness valleys. The hardest sub-classes of the hierarchy include problems covered by the No Free Lunch Theorem. A smaller class include problems with exponential unrestricted black-box complexity. An even smaller class cover problems with exponential elitist black box complexity. We also show a positive result, where appropriately tuned non-elitist EAs have polynomial expected runtime.

3.17 A Tight $O(4^k/p_c)$ Runtime Bound for a (μ +1) GA on Jump$_k$ for Realistic Crossover Probabilities

Andre Opris (Universität Passau, DE), Johannes Lengler (ETH Zürich, CH), and Dirk Sudholt (Universität Passau, DE)

License Creative Commons BY 4.0 International license
© Andre Opris, Johannes Lengler, and Dirk Sudholt

The Jump$_k$ benchmark was the first problem for which crossover was proven to give a speedup over mutation-only evolutionary algorithms. Jansen and Wegener (2002) proved an upper bound of $O(\text{poly}(n) + 4^k/p_c)$ for the $(\mu{+}1)$ Genetic Algorithm ($(\mu{+}1)$ GA), but only for unrealistically small crossover probabilities p_c. To this date, it remains an open problem to prove similar upper bounds for realistic p_c; the best known runtime bound for $p_c = \Omega(1)$ is $O((n/\chi)^{k-1})$, χ a positive constant.

Using recently developed techniques, we analyse the evolution of the population diversity, measured as sum of pairwise Hamming distances, for a variant of the $(\mu{+}1)$ GA on Jump$_k$. We show that population diversity converges to an equilibrium of near-perfect diversity. This yields an improved and tight time bound of $O(\mu n \log(k) + 4^k/p_c)$ for a range of k under the mild assumptions $p_c = O(1/k)$ and $\mu \in \Omega(kn)$. For all constant k the restriction is satisfied for some $p_c = \Omega(1)$. Our work partially solves a problem that has been open for more than 20 years.

3.18 Collecting coupons: knowing when to stop

Jonathan Rowe (University of Birmingham, GB)

License Creative Commons BY 4.0 International license
© Jonathan Rowe
Main reference Jonathan E. Rowe: "A Theoretical Investigation of Termination Criteria for Evolutionary Algorithms", in Proc. of the Evolutionary Computation in Combinatorial Optimization – 24th European Conference, EvoCOP 2024, Held as Part of EvoStar 2024, Aberystwyth, UK, April 3-5, 2024, Proceedings, Lecture Notes in Computer Science, Vol. 14632, pp. 162–176, Springer, 2024.
URL https://doi.org/10.1007/978-3-031-57712-3_11

If you are playing the coupon collector game, but don't know how many distinct coupons there are, how do you know when to stop playing? This problem can be seen as a model of deciding termination criteria for randomised heuristic algorithms – a subject that has received very little theoretical treatment to date.

3.19 Balls Into Bins

Thomas Sauerwald (University of Cambridge, GB)

License Creative Commons BY 4.0 International license
© Thomas Sauerwald
Joint work of Thomas Sauerwald, Dimitrios Los

In the balanced allocation problem we wish to allocate m balls (jobs) into n bins (servers) by allowing each ball to choose from some bins sampled uniformly at random. The goal is to maintain a small gap between the maximum load and average load. For the one-choice protocol, where each ball is allocated to a random bin, the gap diverges for large m. However, for the two-choice protocol, where each ball samples two bins and is placed in the least loaded of the two, it was shown that gap is only O(log log n) for all m. This dramatic improvement is widely known as "power of two choices", and similar effects have been observed in hashing and routing.

In this talk, we will first give some intuition why two-choice maintains such a good balance in practice and theory. Then we will present our recent results in settings where the load information is subject to some noise. For example, the queried load information of bins might be (i) outdated, (ii) subject to some adversarial or random perturbation or (iii) only retrievable by binary queries. We prove that, roughly speaking, if the noise is not too strong, then performance is not affected and the O(log log n) gap bound of two-choice still holds.

3.20 Comparison of an EDA with a hill-climber in the 1-d Ising model

Jonathan Shapiro (University of Manchester, GB)

License Creative Commons BY 4.0 International license
© Jonathan Shapiro

Estimation of Distribution Algorithms replace genetic search operators with a constructed probability distribution. In this work, I compare stochastic hill-climber with a Markov Random Field model. The former has a polynomial runtime, but the later is exponential.

3.21 Introduction to the Theory of Evolutionary Computation in Discrete Domains

Dirk Sudholt (Universität Passau, DE)

License Creative Commons BY 4.0 International license
© Dirk Sudholt

In this talk I gave a high-level overview on theory of evolutionary computation on discrete search spaces. The talk reviewed common scenarios in runtime analysis and introduced common evolutionary algorithm designs for discrete search spaces. It described recent hot topics including diversity optimisation, runtime analyses of state-of-the art multi-objective evolutionary algorithms, smart ideas for designing efficient evolutionary algorithms and analyses of complex evolutionary dynamics.

3.22 Crossover, Constraint Repair and Populations of Solved Subgraphs

Andrew M. Sutton (University of Minnesota – Duluth, US)

License Creative Commons BY 4.0 International license
© Andrew M. Sutton

In the field of combinatorial optimization, parameterized complexity theory can offer tools for understanding the influence of problem structure on hardness for evolutionary algorithms by isolating the source of problem complexity to a parameter of the input (distinct from the problem size). It can also offer insight into designing problem-class tailored operators that yield runtime guarantees in some situations. I present some recent ideas on a population-based evolutionary algorithm for solving constrained graph problems that maintains a population of feasible solutions to induced subgraphs and leverages crossover to build up feasible solutions to larger induced subgraphs. This involves two problem-specific steps during evolution: generating a special "template" parent for multi-parent uniform crossover, and an efficient constraint-repair technique inspired by the method of iterative compression from parameterized complexity theory.

3.23 Democratising Real-World Problem Tailored Optimisation

Vanessa Volz (CWI – Amsterdam, NL)

License Creative Commons BY 4.0 International license
© Vanessa Volz

A central challenge in solving real-world optimisation problems is the lack of a systematic process. Existing approaches largely either adjust the problem representation to match a specific algorithm (e.g., by discretising) or customise an algorithm to fit the problem. However, their results tend to remain problem-or even instance-specific. The point of this talk was to start an open discussion on the challenges in abstracting real-world applications enough to conduct (theoretical) research while maintaining practical meaningfulness, based on my own experiences in empirical/industry research.

3.24 Towards Efficient and Practical Evolutionary Algorithms with Theoretical Guarantees

Weijie Zheng (Harbin Institute of Technology – Shenzhen, CN)

License Ⓒ Creative Commons BY 4.0 International license
© Weijie Zheng

This talk is a self-introduction talk. It introduces my research in four aspects (algorithm choosing, parameter setting, practical problem, and practical algorithms), aiming for efficient and practical evolutionary algorithms with theoretical guarantees.

4 Working Groups

4.1 Breakout Session: EC $\subseteq$ Heuristic Search $\subseteq$ AI $\Longrightarrow$ EC $\subseteq$ AI?

Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR) and Weijie Zheng (Harbin Institute of Technology – Shenzhen, CN)

License Ⓒ Creative Commons BY 4.0 International license
© Benjamin Doerr and Weijie Zheng

Heuristic Search is a classic subfield of artificial intelligence (AI). Clearly, evolutionary algorithms are heuristics search methods. Hence they should be part of the growing field of AI. Yet, only few evolutionary computation (EC) works appear at AI venues. This breakout session aimed at discussing this discrepancy.

Participants (in random order)

Thomas Jansen, Benjamin Doerr (session organizer), Andrew M. Sutton, Christine Zarges, Kento Uchida, Alexandre Chotard, Per Kristian Lehre, Dirk Sudholt, Jonathan L. Shapiro, Weijie Zheng (note keeper), Denis Antipov, Sumit Adak, Asma Atamna, Stephan Frank, Oswin Krause, Tobias Glasmachers, Martin S. Krejca

Questions discussed, views and findings

- Why do we care about the relationship between evolutionary computation (EC) and artificial intelligence (AI)? Essential reason: AI is super-popular at the moment. If we are AI, we should say so, if not, then we should say what else we are.
  - We need a sufficient pool of talent for the field to develop rapidly, and AI makes it easier to establish such talent pool.
  - AI-related funding proposals are more popular and more likely to be approved, which is the basic requirement for recruiting students.
  - Industries prefer to recruit graduates who are engaged in the field of artificial intelligence, which will influence whether you can recruit enough talented students.
- What is the relationship between EC and AI? The first question should be what AI is, which is not well-defined and changes along with time. An indication that EC is generally seen as a subfield of AI is that EC papers are accepted at the big AI conferences AAAI and IJCAI and appear in the leading journals, e.g., the Artificial Intelligence Journal. EC researchers also act(ed) as associate editors of this journal, including Thomas Jansen, Christian Igel, and Benjamin Doerr.

- How narrow is the EC community in the greater AI area? Currently, the hot topic of AI is related to deep learning (deep neural network). Neural networks and evolutionary computation can be regarded as nature-inspired algorithms to some degree. Why are neural networks popular in AI, but EC is much less? Despite the similar black-box nature, neural networks have show a unbelievable improvements for specific applications.
- Concerns about AI journals and conferences, like reviewing quality.
- Views from optimization/heuristics/mathematics community. Separated communities.
- Some suggestions for EC being more visible in AI:
  - Combine the popular AI, like LLM, with EC.
  - Bayesian optimization for discrete optimization?
  - Solve practical problems that show significant improvement using EA.
  - Designable/Controllable EC considering the randomness.
  - Easy-to-use API, like CMA-ES and the ones from the deep learning community.
  - Parallel EAs.
  - More advertisement. The power of the hype behind AI can give bad optimization ideas much more light than more efficient solutions of the same problem using EC. GECCO to top AI conference?
  - More participation in AI venues as attendee, author, or reviewer.

Conclusion

Do not be shy to enter the community of AI. Become an author or reviewer in big AI conferences/journals. But maintain the characteristics and the advantages of the smaller, more coherent, and well-defined EC community.

4.2 Breakout Session on Theory of Multi-Objective Evolutionary Algorithms: What to Take Home From the Recent Burst of Results and Where to Go Next?

Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR) and Andre Opris (Universität Passau, DE)

License Creative Commons BY 4.0 International license
© Benjamin Doerr and Andre Opris

Participants. Benjamin Doerr (organizer), Timo Kötzing, Frank Neumann, Carsten Witt, Denis Antipov, Andre Opris (notekeeper), Christine Zarges, Günter Rudolph, Vanessa Volz, Sumit Adak, Weijie Zheng, Andreia P. Guerreiro, Dirk Sudholt, Martin S. Krejca

Topic. Since the first runtime analysis of the NSGA-II algorithm only two years ago, we have seen a good dozen of runtime analyses of this algorithm and variants like the NSGA-III or the SMS-EMOA. So it might be a good moment to collect what are the main insights from these results, both from the theory perspective and the possible use for practitioners, and what are the most important questions not yet understood sufficiently well.

Notes. In the session, the following observations were made by the participants.

- High relevance: Many results on top AI conferences.
- NSGA-II: First runtime analyses, approximating the Pareto front, multimodal problems, use of crossover, noise, combinatorics, critical difficulties for $m \geq 3$ objectives.
- The variants SMS-EMOA and NSGA-III often behave similarly, but not always. In particular, they perform much better for three or more objectives.

- Discussions about NSGA-III on many objectives: First theoretical insights on how to use NSGA-III on multi-objective problems, but experiments cannot be conducted: Number of reference points too high.
- SMS-EMOA: Speed-ups with stochastic population update.
- Which of these results are most useful for practitioners?
  - Negative results, especially in case of NSGA-II when dealing with three or more objectives: Limits of crowding distance.
  - Good approximation of the Pareto front when updating the crowding distance after each removal.
  - NSGA-II: Small population sizes cannot cover front.
  - How to set the reference points in an efficient way such that NSGA-III becomes more practicable? Could provide theoretical results.
- What are missing results?
  - How can we "heal" the problem with crowding distance?
    Suggestion: "Density estimator" instead of crowding distance.
  - Are lower fronts of the non-dominated sorting any useful? First result on OneTrapZeroTrap where NSGA-II outperforms GSEMO when duplicates are removed.
  - Until now: NSGA-II works as good as GSEMO on deterministic benchmark functions without additional mechanisms like prohibiting duplicates. Provide benchmark function where NSGA-II beats GSEMO.
  - Understand tie breakers in the NSGA-III.
  - Overview and summary about this topic, especially NSGA-III.
  - Problems about disconnected Pareto fronts: No established benchmarks exist here. One result about functions of the form $(\sum_{i=1}^{n} v_i x_i, \sum_{i=1}^{n} w_i x_i)$. Here one could have a disconnected Pareto front since the bits are conflicting. Signs of coefficients could be positive or negative.
  - Orientate on implementations of practitioners. How to design good MOEAs?

References

1 Weijie Zheng, Yufei Liu, and Benjamin Doerr. A first mathematical runtime analysis of the Non-Dominated Sorting Genetic Algorithm II (NSGA-II). In *Conference on Artificial Intelligence, AAAI 2022*, pages 10408–10416. AAAI Press, 2022.

4.3 Breakout Session on Constrained Optimization

Carlos M. Fonseca (University of Coimbra, PT)

License Creative Commons BY 4.0 International license
© Carlos M. Fonseca

Participants

Carlos, Kathrin, Dirk A., Tobias, Armand, Anne, Niko, Oskar, Dimo, Tristan, Daiki, Kento, Oswin

Summary

- Development of constraint handling in EC is partially driven by competitions.
- Continuing the discussion about Kathrin's talk. Punchline: it is worth considering the MO perspective for additional information.

- Multi-objectivization implicitly allows to relax (otherwise hard) constraints, which can help with finding feasible points.
- It is possible (but probably not common) to aggregate all constraint violations into a single objective. This means that we get away with bi-objective optimization, which is easier than many-objective for many reasons.
- Barrier methods may be useful if staying away from the boundary initially helps to make larger steps, but not easy to do in a black-box setting.
- There does not seem to be a simple case that can be reduced to an existing theoretical analysis.
- Equality constraints versus inequality constraints: why should equality constraints be easier? Depends on the solver, in particular for active set methods.

4.4 Breakout Session on Mixed Integer Problems

Carlos M. Fonseca (University of Coimbra, PT)

License Creative Commons BY 4.0 International license
© Carlos M. Fonseca

Participants

Carlos, Dirk A., Tobias, Anne, Niko, Asma, Oskar, Dimo, Aishwarya, Tristan, Daiki, Mario, Alexander, Kento

Summary

- Separability between discrete and continuous? Which dependency structure?
- Which test problems to consider? Is discretized ellipsoid a good problem? Maybe a moderately conditioned ellipsoid, with a max-norm neighborhood structure on the integer part. Homework for Dimo: What is the smallest condition number such that the discretized ellipsoid multi-modal?
  Another rather simple yet non-separable problem: one-max plus sphere, but the optimum of the sphere is offset by the onemax value.
- Difficulties of mixed integer? With CMA-ES: Plateaus, staying unbiased despire rounding, ability to "flip only the last bit"
- Variable length, structure depending on discrete choice? Probably not in the first version.
- Test problems
  - integer
  - categorical
  - permutation
  - variable-length
  - separable?
  - adding up discrete and continuous function assumes "separability" among the discrete and continuous part.
- Inner-outer loop (inner for continuous defines fitness for outer loop optimizing discrete) is probably promising, in particular for ill-conditioned problems.
- no continuous part: what would be the main algorithm choice? only binary variables: what would be the main choice?

- performance of "older" algorithms (Emmerich / Rudolph) using discrete mutation. Li et al.: "Mixed Integer Evolution Strategies for Parameter Optimization"
  - possible ways to solve mixed-integer problems (our independent "reproduction" from the paper):
    - continuous and discrete independent with two independent loops (assuming separability)
    - "optimal" continuous part as fitness of the discrete part (inner-outer loop idea) → probably the one way to go for some problems (where the continuous part highly depends on the discrete variables)
- question: what is the smallest condition number that the discretized ellipsoid is not multimodal (and how is this dependent on the angle)?
- what is the simplest non-separable mixed-integer function, we want to solve?

Consensus

Mixed-integer is not easy.

4.5 Breakout Session on Rich Surrogate Models

Oswin Krause (University of Copenhagen – Copenhagen, DK)

License 🅭 Creative Commons BY 4.0 International license
© Oswin Krause

Participants

Oswin, Dirk A., Tristan, Tobias, Vanessa, Kento, Asma, Dimo, Niko, Oskar, Daiki

Summary

- Setup: individuals in X-space, simulation $s : X \to Y$, cost function $\ell : Y \to \mathbb{R}$. We have full control over the loss ℓ by design, while the simulation s is a black box.
- Usual approach: fit surrogate (e.g., GP model) to the fitness $f = \ell \circ s$. However, this hides information, since Y may the a rich space, while the concatenation hides information by aggregating to a scalar output.
- Simplest example: s is the identity, ℓ is the sphere function. Here, if the model for s is good then we can essentially jump into the optimum.
- Another example: s is a multi-objective problem, ℓ is the hypervolume indicator. Related: ℓ is a constraint handling technique for multi-objective constraint handling.
- Complex example: s is a game level generator (e.g., a GAN), Y consists of variables related to fun, difficulty, gamer involvement and so on.
- The worst case for the a surrogate of the overall fitness is that all function values are the same, while the intermediate points differ and therefore give potentially rich information on where to move. However, in general, it is unclear whether the signal-to-noise (think of model fitting uncertainty) is better in any of the two models.

4.6 Breakout Session on Negative Drift

Martin S. Krejca (Ecole Polytechnique – Palaiseau, FR)

License Ⓒ Creative Commons BY 4.0 International license
© Martin S. Krejca

Participants

Sumit Adak, Denis Antipov, Benjamin Doerr, Stephan Frank, Armand Gissler, Alexander Jungeilges, Timo Kötzing, Oswin Krause, Martin S. Krejca, Frank Neumann, Andre Opris, Günter Rudolph, Thomas Sauerwald, Dirk Sudholt, Andrew M. Sutton, Carsten Witt, Weijie Zheng

Summary

Negative drift considers random processes $(X_t)_{t \in \mathbf{N}}$ over $\mathbf{R}$ that, in expectation, go away from a prior determined target $a \in \mathbf{R}$ within an interval $[a, b] \subset \mathbf{R}$. The term is mostly used when speaking of negative-drift *theorems*, which aim to derive rigorous mathematical tail bounds for the process X to reach a value of at most a after a certain amount of time passed. In the theory community of randomized search heuristics, negative-drift theorems have been applied with great success for deriving exponential lower bounds in various settings, such as the maximum-matching problem [2], the NEEDLE benchmark [4], in noisy optimization [7], in fitness-proportional selection [3], or for slow fitness degradation in self-adjusting algorithms [9]. However, at least three variants of negative-drift theorems exist [4, 5, 6, 8]. This session aimed to understand to what extent these versions can be unified. A great part of the breakout group's time was spent in collecting information about these variants and finding their similarities.

All versions of the theorems require (i) negative drift and (ii) a bounded step size or exponentially declining probabilities of large jumps. And although different theorems use different tools in their proof, for example, a result by Hajek [1] or the Azuma–Hoeffding inequality, they all rely on estimating the occupation probabilities of the process. Moreover, the moment-generating function of the process seems to be crucial to the proofs. It was then noted that all current theorems require their conditions to hold for each point in time, even those that exceed the bound that one is interested in. Since the proofs mainly rely on a union bound of the occupation probabilities, this is a shortcoming that can and should be addressed. Another observation was that idle times of the process are currently not well integrated into the variants and thus require manual work.

As many randomized search heuristics are Markovian, we wondered whether such a property simplifies the assumptions of the theorems, but we did not see how. Another suggestion was to start by applying an exponentially scaled potential function to the process that results in a *positive* drift but nonetheless in an exponential expected hitting time. However, since this does not result in a tail bound directly and requires to find the correct potential function each time, this approach is more of an alternative approach than a unification.

Conclusion

We did not see how to combine the different versions easily, and we concluded that they all have their own benefits. A key take-away was that thinking more specifically about potential functions can prove useful and may forgo an application of one of the negative-drift theorems. Last, instead of having more negative-drift theorems, it might be better to have a guide on how to come up with a qualified potential function for a given process.

References

1 Bruce Hajek. *Hitting-time and occupation-time bounds implied by drift analysis with applications*. Advances in Applied Probability 13(3), pp. 502–525, 1982.
2 Oliver Giel and Ingo Wegener. *Evolutionary algorithms and the maximum matching problem*. Proc. of STACS '03, pp. 415–426, 2003.
3 Frank Neumann, Pietro S. Oliveto, and Carsten Witt. *Theoretical analysis of fitness-proportional selection: Landscapes and efficiency*. Proc. of GECCO '09, pp. 835–842, 2009.
4 Pietro S. Oliveto and Carsten Witt. *Simplified drift analysis for proving lower bounds in evolutionary computation*. Algorithmica 59(3), pp. 369–386, 2011.
5 Pietro S. Oliveto and Carsten Witt. *Erratum: Simplified drift analysis for proving lower bounds in evolutionary computation*. CoRR abs/1211.7184, 2012.
6 Timo Kötzing. *Concentration of first hitting times under additive drift*. Algorithmica 75(3), pp. 490–506, 2016.
7 Johannes Lengler and Ulysse Schaller. *The* $(1+1)$*-EA on noisy linear functions with random positive weights*. Proc. of SSCI '18, pp. 712–719, 2018.
8 Johannes Lengler and Angelika Steger. *Drift analysis and evolutionary algorithms revisited*. Combinatorics, Probability & Computing 27(4), pp. 643–666, 2018.
9 Marc Kaufmann, Maxime Larcher, Johannes Lengler, and Xun Zou. *Self-adjusting population sizes for the* $(1, \lambda)$*-EA on monotone functions*. Theoretical Computer Science 979, pp. 114181:1–114181:29, 2023.

4.7 Breakout session: Coevolution

Per Kristian Lehre (University of Birmingham, GB) and Mario Alejandro Hevia Fajardo (University of Birmingham, GB)

License Creative Commons BY 4.0 International license
© Per Kristian Lehre and Mario Alejandro Hevia Fajardo
Main reference Per Kristian Lehre: "Runtime Analysis of Competitive Co-evolutionary Algorithms for Maximin Optimisation of a Bilinear Function", Algorithmica, Vol. 86(7), pp. 2352–2392, 2024.
URL https://doi.org/10.1007/S00453-024-01218-3

Co-evolutionary algorithms have been studied empirically for a long time in Evolutionary Computation. Practitioners often report pathological behaviour, such as cycling, disengagement, and forgetting. A survey by Popovici, Bucci, Wiegand and de Jong (2012) identified runtime analysis of co-evolution as an important open problem.

The EC theory group in Birmingham has an ongoing effort in developing runtime analysis for competitive co-evolutionary algorithms. The first results include upper bounds on the runtime of the Pairwise-Dominance Co-Evolutionary Algorithm (PDCoEA) to find an ε-approximation to the maximin-solution of some instances of a bilinear problem (Lehre, GECCO2022, Algorithmica2024). Later analyses of bilinear games considered the algorithms RLS-PD (Fajardo, Lehre, Lin, GECCO2023), (μ, λ) CoEA (Fajardo & Lehre, GECCO2023), and RankDivCoEA (Fajardo & Lehre, PPSN2024). Other games have been considered, including Diagonal (a game with binary payoffs) (Lehre & Lin, PPSN 2024), as well as more complex class of symmetric zero-sum games (Benford & Lehre, GECCO 2024). Some general conclusions from this work include:

- Classical drift analysis and occupation-time bounds can be used to derive runtime. Also, a co-evolutionary level-based theorem is now available, providing upper bounds on the expected runtime of population-based CoEAs (Lehre, 2022GECCO).

- A sufficiently large population or archives can be essential. (1+1) CoEA-type algorithms forget their state (i.e., showing cycling behaviour) on several benchmark problems.
- Population diversity is often important, such as in "Games of Skill" (Benford & Lehre, GECCO24). RankDivCoEA uses a diversity mechanism which promotes unique rankings of opponents.
- Mutation rates can be critical. CoEAs have "error thresholds" (performance breaks down above critical mutation rate) similar to non-elitist EAs. However, mutation rates can be adapted automatically via self-adaptation (Fajardo et al, GECCO24).
- Traditional CoEAs do not always perform well, and it is non-trivial to design new ones. PDCoEA and RankDivCoEA have polynomial expected runtime on Bilinear, and also show excellent empirical performance on several benchmark problems.

The breakout session noted that early influential work, such as Hillis' co-evolution of sorting networks, provided inadequate details to allow easy replication.

4.8 Breakout Session on the Structure of Discrete Problems

Jonathan Shapiro (University of Manchester, GB)

License Creative Commons BY 4.0 International license
© Jonathan Shapiro

Participants. Andreia Guerreiro, Kathrin Klamroth, Vanessa Volz, Christine Zarges, Thomas Jansen, Per Kristian Lehre, Jon Rowe, Jon Shapiro.

Summary. Motivation was that for continuous functions there are many categories which provide global information. Examples are: $C^{(n)}$, the Lipschitz condition, and convexity. However, it seems that for functions over discrete spaces, anything can happen. Are there any such categories for discrete-space functions. They should reveal something about the function and be easy to calculate.

For example, suppose you are solving a problem which is giving continuous improvement. Could any property reveal whether it will continue to improve, or is it somethink like k-jump?

Earlier in the day we had seen a proposal by Per Kristian Lehre which addresses this precise issue. It is not clear that this is easy to calculate, but it does capture problems which we would consider difficult, argued Per K. Thomas J reminded us that we do not solve classes (e.g., MAXSAT). We solve instances.

What if there existed a function of an instance of a problem, which could be exponential to compute, but if you knew it, you would would know that you could solve that instance in poly time. Alternatively, it could tell you where that problem is in the Lehre heirarchy.

Topics we should looks at:

- Exploratory landscape analysis.
- Adversarial functions – find the most difficult instance of a class of problems.
- Constraints can hide properties of functions. We should look at constraint penalty functions.
- Black Box Complexity, but this gives worst-case result.
- Can we have a "Grey box" which reveals other information?

Proposed conclusions.

1. It is complicated.
2. We have no conclusions.

4.9 Breakout Session on Analysis of Permutation Problems

Christine Zarges (Aberystwyth University, GB)

License Creative Commons BY 4.0 International license
© Christine Zarges

Participants (in alphabetical order)

Sumit Adak, Denis Antipov, Benjamin Doerr, Carlos M. Fonseca, Andreia P. Guerreiro, Mario Alejandro Hevia Fajardo, Thomas Jansen, Kathrin Klamroth, Timo Kötzing, Martin S. Krejca, Per Kristian Lehre, Frank Neumann, Andre Opris, Aishwarya Radhakrishnan, Jonathan Rowe, Günter Rudolph, Jonathan L. Shapiro, Dirk Sudholt, Andrew M. Sutton, Carsten Witt, Christine Zarges, Weijie Zheng

Summary

For many real-world problems, permutations constitute a natural way to represent solutions. Typical examples include problems where the order of elements (e.g., scheduling) or adjacency (e.g., travelling salesperson) matter [4]. A series of special sessions and workshops investigated different aspects specific to permutation problems[1]. However, in particular with respect to runtime analysis and fixed budget analysis, permutation problems have so far received much less attention in comparison to pseudo-Boolean or continuous problems.

As a starting point, this breakout session discussed a small (incomplete) selection of existing work: Rowe et al. [8] consider which crossover operators are useful for which representations while Irurozki [5] investigates various mutation operators and the distance measures they imply. Schiavinotto and Stützle [10] present a review of metrics on permutations for search landscape analysis. One of the first runtime analyses of permutation problems considered sorting with different measures of sortedness [9]. More recently, Doerr et al. [3] propose a general way to construct permutation benchmarks from classic pseudo-Boolean benchmarks, and Bassin and Buzdalov [1] discuss how to adapt the $(1 + (\lambda, \lambda))$ GA to permutation spaces. A few real-world permutation problems, e.g., graph drawing [2] and variable ordering in ordered binary decision diagrams [6], have also been analysed.

Throughout the discussion various problems were mentioned that could be considered in future work, e.g., Bin Packing, Graph Colouring, Knapsack, (weighted) variants of the Travelling Salesperson Problem, and linear/quadratic assignment problems. For some of these problems, a permutation can be used as input for a greedy algorithm. During the discussion, questions were raised on how more general benchmarks could be obtained and how relevant problems classes could be defined as permutation problems are often NP-hard.

Dealing with permutations directly instead of using "decoder functions" implies that mutation and crossover operators need to be designed in a way that preserves the permutation property of a search point. Parts of the discussion considered potential analyses of existing decoder functions, particularly those using real-valued representations, to facilitate the analysis of crossover. Other points of discussion included geometric crossover for permutations [7] and partition crossover [11].

[1] http://www.sc.ehu.es/ccwbayes/cec2022_ecperm/index.html

Summary

There seemed to be a general consensus that we currently only have little understanding on working principles of mutation and crossover operators for permutation problems and that more research in this direction is needed. However, concrete next steps to achieve this were less obvious in the discussion. It would be nice to identify "classes" of permutation problems and establish a framework that maps operators to these problem classes. One way to achieve this could be to concentrate on operators and the fitness landscapes they imply. Moreover, none of the participants was aware of existing runtime analyses considering crossover operators for permutations.

References

1. A. O. Bassin and M. Buzdalov. The $(1 + (\lambda, \lambda))$ genetic algorithm for permutations. In C. A. C. Coello, editor, *GECCO '20: Genetic and Evolutionary Computation Conference, Companion Volume, Cancún, Mexico, July 8-12, 2020*, pages 1669–1677. ACM, 2020.
2. J. Baumann, I. Rutter, and D. Sudholt. Evolutionary computation meets graph drawing: Runtime analysis for crossing minimisation on layered graph drawings. In X. Li and J. Handl, editors, *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2024, Melbourne, VIC, Australia, July 14-18, 2024*. ACM, 2024.
3. B. Doerr, Y. Ghannane, and M. I. Brahim. Runtime analysis for permutation-based evolutionary algorithms. *Algorithmica*, 86(1):90–129, 2024.
4. A. E. Eiben and J. E. Smith. *Introduction to Evolutionary Computing, Second Edition*. Natural Computing Series. Springer, 2015.
5. E. Irurozki. *Sampling and learning distance-based probability models for permutation spaces*. Phd thesis, University of the Basque Country, 2014. Available at `https://github.com/isg-ehu/PhD-Dissertations/blob/master/2014_phd_ekhine_irurozki.pdf`.
6. T. Jansen and C. Zarges. Experimental and theoretical analysis of local search optimising OBDD variable orderings. In T. Stützle and M. Wagner, editors, *Evolutionary Computation in Combinatorial Optimization – 24th European Conference, EvoCOP 2024, Held as Part of EvoStar 2024, Aberystwyth, UK, April 3-5, 2024, Proceedings*, volume 14632 of *Lecture Notes in Computer Science*, pages 177–192. Springer, 2024.
7. A. Moraglio and R. Poli. Geometric crossover for the permutation representation. *Intelligenza Artificiale*, 5(1):49–63, 2011.
8. J. E. Rowe, M. D. Vose, and A. H. Wright. Group properties of crossover and mutation. *Evol. Comput.*, 10(2):151–184, 2002.
9. J. Scharnow, K. Tinnefeld, and I. Wegener. The analysis of evolutionary algorithms on sorting and shortest paths problems. *J. Math. Model. Algorithms*, 3(4):349–366, 2004.
10. T. Schiavinotto and T. Stützle. A review of metrics on permutations for search landscape analysis. *Comput. Oper. Res.*, 34(10):3143–3153, 2007.
11. R. Tinós, L. D. Whitley, and G. Ochoa. A new generalized partition crossover for the traveling salesman problem: Tunneling between local optima. *Evol. Comput.*, 28(2):255–288, 2020.

5 Seminar Schedule

Monday, July 01, 2024 – Lecture Hall Saarbrücken

– 09:00	Breakfast
09:00 – 09:15	Welcome and seminar opening
09:15 – 10:00	Ice-breaking
10:00 – 10:30	Coffee break
10:30 – 11:15	Introductory talk: **Daiki Morinaga and Tobias Glasmachers** on *Continuous Domains*
11:15 – 12:00	Introductory talk: **Thomas Sauerwald** on *Balls Into Bins*
12:15 – 13:30	Lunch
13:30 – 14:15	Time for individual discussions
14:15 – 15:00	Introductory talk: **Carlos Fonseca** on *Multiobjective Optimization*
15:00 – 15:30	Introductory Talk: **Dirk Sudholt** on *Discrete Domains*
15:30 – 16:00	Cake
16:00 – 16:45	Introductory talk: **Dirk Arnold** on *Constraint Handling*
16:45 – 16:55	**Vanessa Volz**
16:55 – 17:00	**Timo Kötzing** on **his habilitation thesis**
18:00 – 19:00	Dinner
19:00 –	Individual discussions, soccer, cheese platter, drinks, and **games**

Tuesday, July 02, 2024 – Lecture Hall Saarbrücken

– 09:00	Breakfast
9:15 – 10:05	Junior flash talks: **speakers** on *topics*
9:15 – 9:25	**Tristan Marty** introducing himself
9:25 – 9:35	**Aishwarya Radhakrishnan** introducing herself
9:35 – 9:45	**Alexander Jungeilges** introducing himself
9:45 – 9:55	**Mario Alejandro Hevia Fajardo** introducing himself
10:10 – 10:35	Coffee break
10:35 – 10:45	**Stephan Frank** introducing himself
10:55 – 11:15	**Armand Gissler** on *Convergence proof of CMA-ES: analysis of underlying normalized Markov chains*
11:25 – 11:45	**Per Kristian Lehre** on *The SLO Hierarchy*
12:15 – 13:30	Lunch
14:00 – 14:15	Organization of breakout sessions
14:15 – 15:30	**Breakout Sessions:**
	◇ Negative Drift
	◇ Characterization of Discrete Problems
	◇ Mixed Integer
15:30 – 16:00	Cake
16:00 – 16:20	**Kathrin Klamroth** on *A Multiobjective Perspective on Constraint Handling*
16:30 – 17:45	**Breakout Sessions:**
	◇ Constrained Optimization
	◇ Theory of Multi-Objective Evolutionary Algorithms
	◇ Coevolution
18:00 – 19:00	Dinner
20:00 –	Individual discussions, soccer, cheese platter, drinks

Wednesday, July 03, 2024 – Lecture Hall Saarbrücken

– 09:00	Breakfast
9:00 – 9:10	**Oskar Girardin** introducing himself
9:20 – 9:35	**Weijie Zheng** introducing himself
9:45 – 10:00	**Oswin Krause** on *Evolution in the Quantum world: on tuning quantum dot devices*
10:10 – 10:50	Coffee break
10:50 – 11:10	**Dimo Brockhoff** on *How Theoretical Considerations Can Help in Algorithm Design: 3 Examples*
11:20 – 11:40	**Andre Opris** on *A Tight* $O(4^k/p_c)$ *Runtime Bound for a (μ+1) GA on* $Jump_k$ *for Realistic Crossover Probabilities*
12:15 – 13:30	Lunch
13:30 – 15:30	Social event H: beautiful hike in acceptable weather
14:00 – 18:30	Social event T: trip to Trier
15:30 – 16:00	Cake
16:00 – 18:30	Free time
18:30 – 19:30	Dinner
19:30 –	Individual discussions, preparing my talk tomorrow, beer, no soccer

Thursday, July 04, 2024 – Lecture Hall Saarbrücken

– 09:00	Breakfast
9:00 – 9:20	**Benjamin Doerr** on *Dumb and Good: Another Advertising Campaign For Random Parameters (Following a Power Law)*
9:30 – 9:50	**Alexandre Chotard** on *Covariance Matrix Adaptation for MCMC Sampling: some Differences with the Optimization Context*
10:00 – 10:35	Coffee break
10:35 – 10:55	**Denis Antipov** on *Reconstructing a true individual from a noisy one: a new perspective gives new insights*
11:05 – 11:25	**Jonathan L. Shapiro** on *Comparison of an EDA with a hill-climber in the 1-d Ising model*
11:35 – 11:45	**Carlos M. Fonseca**: ROAR-NET COST Action
12:15 – 13:30	Lunch
13:55 – 14:10	**Group photo**; we meet in front of the castle
14:15 – 15:30	**Breakout Sessions:**
	◇ Permutation Search Spaces
	◇ Rich Surrogate Models
15:30 – 16:00	Cake
16:30 – 17:45	**Breakout Sessions:**
	◇ Abstractions for theoretical analysis and practical implementation
	◇ EC $\subseteq$ HS $\subseteq$ AI
18:00 – 19:00	Dinner
20:00 –	**Party:** traditional wine and cheese platter

Friday, July 05, 2024 – Lecture Hall Saarbrücken

– 09:00	Breakfast
9:00 – 9:15	**Asma Atamna** on *Understanding the Fitness Landscape of a Real-World Sequential Decision-Making Problem*
9:25 – 9:45	**Andreia P. Guerreiro** on *Theoretical Aspects of Set-Quality Indicators for Multiobjective Optimization*
9:55 – 10:30	Coffee break
10:30 – 10:50	**Jonathan Rowe** on *Collecting coupons: knowing when to stop*
11:00 – 11:20	**Andrew M. Sutton** on *Crossover, Constraint Repair and Populations of Solved Subgraphs*
11:30 – 12:00	Closing session, feedback, ideas
12:15 – 13:30	Lunch
13:30 –	Departure

Participants

- Sumit Adak
Technical University of Denmark – Lyngby, DK
- Denis Antipov
University of Adelaide, AU
- Dirk Arnold
Dalhousie University – Halifax, CA
- Asma Atamna
Ruhr-Universität Bochum, DE
- Anne Auger
INRIA Saclay – Palaiseau, FR
- Dimo Brockhoff
INRIA Saclay – Palaiseau, FR
- Alexandre Chotard
Calais University, FR
- Benjamin Doerr
Ecole Polytechnique – Palaiseau, FR
- Carlos M. Fonseca
University of Coimbra, PT
- Stephan Frank
Ruhr-Universität Bochum, DE
- Oskar Girardin
Ecole Polytechnique – Palaiseau, FR
- Armand Gissler
Ecole Polytechnique – Palaiseau, FR
- Tobias Glasmachers
Ruhr-Universität Bochum, DE
- Andreia P. Guerreiro
INESC-ID – Lisbon, PT
- Nikolaus Hansen
Inria & Institut Polytechnique de Paris – Palaiseau, FR
- Mario Alejandro Hevia Fajardo
University of Birmingham, GB
- Thomas Jansen
Aberystwyth University, GB
- Alexander Jungeilges
Ruhr-Universität Bochum, DE
- Kathrin Klamroth
Universität Wuppertal, DE
- Timo Kötzing
Hasso-Plattner-Institut, Universität Potsdam, DE
- Oswin Krause
University of Copenhagen, DK
- Martin S. Krejca
Ecole Polytechnique – Palaiseau, FR
- Per Kristian Lehre
University of Birmingham, GB
- Tristan Marty
Ecole Polytechnique – Palaiseau, FR
- Daiki Morinaga
University of Tsukuba, JP
- Frank Neumann
University of Adelaide, AU
- Andre Opris
Universität Passau, DE
- Aishwarya Radhakrishnan
Hasso-Plattner-Institut, Universität Potsdam, DE
- Jonathan Rowe
University of Birmingham, GB
- Günter Rudolph
TU Dortmund, DE
- Thomas Sauerwald
University of Cambridge, GB
- Jonathan L. Shapiro
University of Manchester, GB
- Dirk Sudholt
Universität Passau, DE
- Andrew M. Sutton
University of Minnesota – Duluth, US
- Kento Uchida
Yokohama National University, JP
- Vanessa Volz
CWI – Amsterdam, NL
- Carsten Witt
Technical University of Denmark – Lyngby, DK
- Christine Zarges
Aberystwyth University, GB
- Weijie Zheng
Harbin Institute of Technology – Shenzhen, CN

Report from Dagstuhl Seminar 24272

A Game of Shadows: Effective Mastery Learning in the Age of Ubiquitous AI

Nick Falkner*[1], Juho Leinonen*[2], Miranda C. Parker*[3], Andrew Petersen*[4], and Claudia Szabo*[5]

1 **University of Adelaide, AU.** nickolas.falkner@adelaide.edu.au
2 **Aalto University, FI.** juho.2.leinonen@aalto.fi
3 **San Diego State University, US.** mcparker@sdsu.edu
4 **University of Toronto Mississauga, CA.** andrew.petersen@utoronto.ca
5 **University of Adelaide, AU.** claudia.szabo@adelaide.edu.au

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 24272 "A Game of Shadows: Effective Mastery Learning in the Age of Ubiquitous AI". We focused during the seminar on exploring how generative AI can support mastery learning; breaking the problem into three main categories: operational, community focused, and curriculum and pedagogy focused. Our various talks explored these aspects.

Seminar June 30 – July 5, 2024 – https://www.dagstuhl.de/24272
2012 ACM Subject Classification Applied computing → Education
Keywords and phrases chatgpt, computing education, llms, machine learning, mastery learning
Digital Object Identifier 10.4230/DagRep.14.6.245

1 Executive Summary

Nick Falkner
Juho Leinonen
Miranda C. Parker
Andrew Petersen
Claudia Szabo

License Creative Commons BY 4.0 International license
© Nick Falkner, Juho Leinonen, Miranda C. Parker, Andrew Petersen, and Claudia Szabo

The integration of generative AI (GenAI) into education raises significant issues and opportunities, particularly concerning mastery learning and programming education. A primary concern is that students may bypass deep engagement with their learning tasks by relying on AI tools or search engines, which leads to a superficial understanding of the material. This tendency forces instructors to focus more on monitoring for academic dishonesty rather than on effective teaching. To address this, innovative approaches to presenting curricula and materials could foster greater student engagement and reduce the inclination to rely on external aids.

The transformative potential of AI in education is likened in Armando Fox's talk to the early use of movie cameras, which initially focused on replicating existing practices rather than exploring new possibilities. The emphasis is on avoiding mere substitution of traditional methods with AI, and instead, leveraging AI to create entirely new modes of learning. The goal is to integrate AI in a way that complements foundational educational concepts and develops the necessary intellectual frameworks to utilize these new tools effectively.

* Editor / Organizer

Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

A Game of Shadows: Effective Mastery Learning in the Age of Ubiquitous AI, *Dagstuhl Reports*, Vol. 14, Issue 6, pp. 245–262
Editors: Nick Falkner, Juho Leinonen, Miranda C. Parker, Andrew Petersen, and Claudia Szabo
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Mastery learning remains an effective approach in the age of AI, with a deep understanding of the importance of clearly defined learning goals and assessments designed to resist cheating. One proposed method involves students setting their own learning goals, which are then approved by instructors. This approach not only personalizes learning but also utilizes GenAI to assist in developing and refining assignments. GenAI can also serve as a brainstorming partner for educators, providing valuable support in creating innovative class activities and assessments.

AI tools like code co-pilots are nowadays becoming integral for understanding and implementing code. This shift presents challenges for traditional assessment methods, as AI can outperform students in specific coding tasks. It is essential for students to develop strong communication skills about code, as these will become increasingly crucial in collaborative programming environments where AI tools are prevalent. This suggests that the skill of discussing code should be as rigorously taught as technical coding skills. In addition, providing accurate feedback and identifying essential skills for effective software development remain critical. While LLMs can offer valuable feedback, ensuring that this feedback is accurate and relevant remains a challenge. Additionally, defining the skills necessary for students to develop software with the support of LLMs, such as program specification and refactoring, is crucial for leveraging these tools effectively.

The integration of mastery learning with GenAI presents both significant potential and challenges. While GenAI can enhance personalized learning, assessment creation, and feedback, effective implementation requires careful consideration of the tools used, their alignment with educational goals, and their impact on learning outcomes. Ensuring that these tools are suitable for diverse educational contexts and measuring their effectiveness will be key to successfully adopting mastery learning supported by GenAI, ultimately aiming to improve educational outcomes and better prepare students for future professional challenges.

The seminar was structured into three main sections: lightning and keynote talks, brainstorming, and workshop groups. At the beginning of the seminar, each attendee delivered a lightning talk. Two keynote talks were delivered, as follows:

- Prof Andrew Luxton-Reilly, University of Auckland: "It's the end of the world as we know it, but I feel fine! Teaching and learning with GenAI"
- Dr Claudia Ott, University of Otago: "A Decade of Mastery Learning at Otago – Pitfalls, Challenges & Opportunities"

The lightning talk sessions were followed by a brainstorming session to identify existing challenges and opportunities. This session served a dual purpose, in that it also allowed us to identify the three main working groups of the seminar. These groups focused on (i) curriculum and pedagogy of mastery learning in the era of GenAI, (ii) university and organisation structures that facilitate the delivery and operationalisation of mastery learning in the era of GenAI, and on (iii) designing courses in a curriculum that is GenAI focused.

Publications to Date

At the date of the submission of this report, the following papers and posters had been accepted for publication:

- "Models of Mastery Learning for Computing Education", by Claudia Szabo, Miranda Parker, Michelle Friend, Johan Jeuring, Tobias Kohn, Lauri Malmi, Judithe Sheard was accepted for publication at SIGCSE 2025 as a position paper.
- "Goodbye Hello World – Research Questions for a Future CS1 Curriculum", Hieke Keuning, Andrew Luxton-Reilly, Claudia Ott, Andrew Petersen and Natalie Kiesler was accepted at Koli Calling 2024 as a poster.

2 Table of Contents

3 Workshop Overview

Our seminar was split into three groups, that focused on (i) curriculum and pedagogy of mastery learning in the era of generative AI, (ii) university and organisation structures that facilitate the deliveray and operationalisation of mastery learning in the era of generative AI, and on (iii) designing courses in a curriculum that is Generative AI focused.

3.1 Curriculum and pedagogy for mastery learning with Generative AI

This subgroup focused on exploring mastery learning and generative AI in the core sequence of undergraduate computing courses, namely, CS1-CS2 -> software development -> capstone project.

Within the modern era of generative artificial intelligence (GenAI), we consider

- What kind of pedagogies will support mastery learning within this sequence of courses?
- What are the learning goals that are needed for mastery learning in the higher courses in the sequence?
- What does assessment look like for mastery learning in this sequence of courses?
- As there is no accurate measurement of a skill, can mastery learning work for the demonstration of progress in a skill?

3.2 Designing courses with Generative AI at the core

We engaged in a thought experiment where we developed a curriculum for a new CS1 course, in which the software development process as a whole is central and in which GenAI is used for several of the activities involved. Doing so identified a number of areas where future work may be needed, including:

- Are GenAI tools ready? Our thought experiment was predicated on the availability of GenAI tools that can reliably generate (a) correct code at the level of some functional unit (a class or function), (b) test cases to help validate a functional unit, and (c) explanations of code that has been generated. As we explored, we also found other attributes of GenAI tools that would be useful or maybe even necessary: a tool that can modify existing code and explain the modifications, can compose separate functional units, and can be constrained to provide focused responses to queries. Many of these may be features of current LLMs, but evaluation of how effective and reliable they are is necessary.
- What prerequisite skills do the students need before (or at the beginning of) a GenAI-centered course? On a related note, what support is needed to enable students to generate ideas of an appropriate scale to be solved by LLMs, decompose a larger problem into requirements that a GenAI tool can engage with effectively, prompt the GenAI tool for a code submission, describe needed revisions to the functionality of a piece of code, and prompt the GenAI tool to compose multiple pieces of code.
- Is mastery of the prompting of LLMs to generate code necessary and sufficient for continued study in computing? Is it feasible that students could continue into a computing program – studying algorithms and data structures, perhaps, or parallel systems – without necessarily having detailed experience with structures in code.

- In the context of code-writing, when should students start using GenAI tools? Evidence in classroom contexts is required to determine if students with varying levels of experience prior to using LLMs perform differently.
- To what extent do existing social learning theories apply to interactions with GenAI tools? Will students consider these tools to have a mind, and will that lead them to construct knowledge with them in ways explained by social learning theories?
- To what extent can existing pedagogies be adapted to incorporate a GenAI tool? One can easily imagine the interactions between a student and the GenAI tool they use might be described using a pair-programming model. However, other pedagogies may also be reconsidered in light of GenAI systems.

4 Overview of Talks

4.1 A Game of Shadows – Effective Mastery Learning in the Age of Ubiquitous AI

Nick Falkner (University of Adelaide, AU)

License Creative Commons BY 4.0 International license
© Nick Falkner

The worst pattern for students is one where no work is attempted at all, resorting immediately to AI or Google, which avoids any real effort on the student's part, leading to no measurable learning. Where this behaviour is widespread, lecturing staff often devolve to policing activities, focused on catching students out rather than on how they are learning. Different ways of presenting curricula and course materials can engage students in more effective ways and we hypothesise that this could be one way of addressing the knee jerk "go external" mode for some students and allow lecturing staff to focus on teaching and knowledge development.

4.2 Cheating is just a red herring

Armando Fox (University of California – Berkeley, US)

License Creative Commons BY 4.0 International license
© Armando Fox

When the movie camera was invented, the first thing it was used for was filming live performances. Only much later did the techniques of moviemaking – special effects, cinematography, lighting – emerge as new elements. Similarly, we have so far operationalized the activity of learning as specific structured activities: lecture, discussion, formative and summative assessments, with defined roles and ways of measuring outcomes for each. AI is potentially such a transformative technology that we must be careful not to fall into the trap of simply substituting it into existing situations, but rather recognizing that it will enable entirely new modes of learning that were previously impossible. Identifying the CS concepts that remain foundational, and the intellectual vocabulary needed to identify and exploit those new learning modes, should be our agenda.

4.3 What are you trying to do?

Michelle Friend (University of Nebraska – Omaha, US)

License Creative Commons BY 4.0 International license
© Michelle Friend

Mastery Learning

- The most important place to start when designing any learning experience is at the end: what are your learning goals for students?
- This naturally leads to a central tenet of mastery learning (as well as other approaches such as universal design for learning): how can students show they have met the goal? Especially in an age of ubiquitous AI, it is valuable, important, and student-centered to be expansive when thinking about assessment.
- I put forth that one can create "cheat-proof" assignments in which students design their own end goal (and have it approved by an instructor) then apply what they have learned to meet their own goal. I showed examples of student-created digital art from an introductory programming class as an example.

Generative AI: GenAI provides a variety of affordances for instructors and researchers

- Instructors can use GenAI to test and revise assignments. LLMs will currently return results similar to a mediocre student. If the LLM misunderstands the task, so will students
- To generate ideas, AI can be used as a brainstorming partner, providing ideas for class activities or assessments
- For researchers, AI can be used to generate "bad" first drafts, from which the work can be revised or refined

4.4 How AI helps in (learning) reading (code comprehension) and writing code

Petri Ihantola (University of Jyväskylä, FI)

License Creative Commons BY 4.0 International license
© Petri Ihantola

This talk explores the evolving role of artificial intelligence (AI) in aiding the processes of code comprehension and code writing. AI-powered tools, like code co-pilots, are increasingly used by students and professionals to understand code functionality and implement specific features. This presents critical challenges to traditional assessment methods in introductory programming courses, as AI can often outperform students on tasks like answering questions about the code they have written.

It is important to note that essential and future-critical skills required for successful software engineering strongly rely on the ability to engage in meaningful dialogue about code. Programming is a collaborative endeavor where communication and argumentation are just as important as technical skills. Large language models (LLMs) are poised to enhance the field of software engineering, making the ability to discuss and evaluate different coding approaches increasingly vital. These interactions will be crucial in teams as they collaborate on complex software projects, aided by AI tools.

A key question arises: should the skill of discussing code be taught as rigorously as the skills of reading and writing code? Even in the absence of AI, developing this communicative skill could greatly improve team-based programming efforts. With the aid of LLMs, there is a future where this skill becomes indispensable, ensuring that discussions around code are as integral to programming education as technical proficiency.

4.5 Mastering programming with LLMs: feedback and skills

Johan Jeuring (Utrecht University, NL)

License Creative Commons BY 4.0 International license
© Johan Jeuring

Problem 1: (Guaranteed!) better feedback and hints

Feedback and hints are important when learning to program. In the past we have developed a tutoring system supporting students with feedback and hints on steps they take when developing small imperative programs [4]. Feedback and hints were constructed from the difference between student programs and model solutions. This worked in quite a few cases, but not always. When the system gives feedback, it is correct from a semantical perspective, but the system cannot always give feedback. To better give feedback in such systems, we recently looked at what kind of steps students take when developing beginner's programs [3], and the reasons why experts give feedback on particular steps [6]. In parallel, we and several other teams have investigated the possibility to give feedback and hints on student steps using LLMs [7, 5]. This worked in quite a few cases, but not always. Most hints are appropriate, but sometimes hints are misleading. I would like to think about how we can use existing technologies, such as property checking, static analysis, and many more, to guarantee that the feedback and hints we give using an LLM are not misleading anymore, and more or less correspond to feedback teachers would give.

Problem 2: What skills do students need to acquire, in what order, to learn how to develop software supported by an LLM?

In previous work we investigated if better computational thinking skills lead to better skills in developing software using ChatGPT [2]. (They do.) But what skills do (which?) students need to develop (which?) software supported by (which?) LLMs? They need some programming skills, and, according to Denny et al [1] skills in (1) program specification, (2) refactoring, and (3) verification, testing and evaluation. Can we verify this? Can we find or design interventions for these skills and learn about their effectiveness?

References

1 Paul Denny, James Prather, Brett A Becker, James Finnie-Ansley, Arto Hellas, Juho Leinonen, Andrew Luxton-Reilly, Brent N Reeves, Eddie Antonio Santos, and Sami Sarsa. Computing education in the era of generative ai. *Communications of the ACM*, 67(2):56–67, 2024.
2 Johan Jeuring, Roel Groot, and Hieke Keuning. What skills do you need when developing software using ChatGPT. In *Koli Calling 2023*, pages 1–6, 2023.
3 Johan Jeuring, Hieke Keuning, Samiha Marwan, Dennis Bouvier, Cruz Izu, Natalie Kiesler, Teemu Lehtinen, Dominic Lohr, Andrew Peterson, and Sami Sarsa. Towards giving timely formative feedback and hints to novice programmers. In *Proceedings of the 2022 Working Group Reports on Innovation and Technology in Computer Science Education*, pages 95–115. 2022.

4 Hieke Keuning, Bastiaan Heeren, and Johan Jeuring. Strategy-based feedback in a programming tutor. In *Proceedings of the computer science education research conference*, pages 43–54, 2014.
5 Mark Liffiton, Brad E Sheese, Jaromir Savelka, and Paul Denny. Codehelp: Using large language models with guardrails for scalable support in programming classes. In *Proceedings of the 23rd Koli Calling International Conference on Computing Education Research*, pages 1–11, 2023.
6 Dominic Lohr, Natalie Kiesler, Natalie Keuning, and Johan Jeuring. "Let Them Try to Figure It Out First" – Reasons Why Experts (Do Not) Provide Feedback to Novice Programmers. In *ITiCSE*, 2024.
7 Lianne Roest, Hieke Keuning, and Johan Jeuring. Next-step hint generation for introductory programming using large language models. In *Proceedings of the 26th Australasian Computing Education Conference*, pages 144–153, 2024.

4.6 Uncovering changing skills and designing support tools for learning programming with Generative AI

Hieke Keuning (Utrecht University, NL)

License Creative Commons BY 4.0 International license
© Hieke Keuning

The rise of generative AI and Large Language Model-based tools has a significant impact on both the field of computing, and computing education. Computer scientists, educators, and practitioners are having heated discussions about the changes in the nature of programming, and its influence on learning. At the same time, there is an opportunity to use LLMs to support students better than we used to. I identify four "burning issues":

Automated feedback & tools. How to generate high-quality feedback using GenAI? How to use GenAI to build better tools? Some solutions we have been exploring are to control LLMs with prompting, but we should also combine LLMs with "traditional" techniques that have shown to be useful in the past. When studying how experts give feedback, we can incorporate their practices in new tools.

Retention. A question that arises from this, is how do we make sure students use these tools, and do not resort to "learn" (or "produce solutions") with ChatGPT? Creating quality learning opportunities is key here, including personalization, scaffolding, and high-quality feedback. And perhaps we should also teach our student more about how learning actually works.

Assessment. What do we assess? What are the future learning goals? What skills are needed in the AI era? These are major questions that need to be answered in the coming years. In the context of mastery learning, we should probably gradually allow GenAI as a tool for "mastered" tasks.

Course type. Finally, while much of the attention goes to introductory programming, what and how do we teach and assess in courses beyond CS1, group projects, and theses?

4.7 Navigating Mastery Learning in the Era of Generative AI: Challenges and Opportunities

Natalie Kiesler (Technische Hochschule Nürnberg, DE)

License Creative Commons BY 4.0 International license
© Natalie Kiesler
Main reference Andreas Scholl, Daniel Schiffner, Natalie Kiesler: "Analyzing Chat Protocols of Novice Programmers Solving Introductory Programming Tasks with ChatGPT", in Proc. of the DELFI 2024 – Die 22. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V., DELFI 2024, Fulda, Germany, September 9-11, 2024, LNI, Vol. P-356, Gesellschaft für Informatik e.V., 2024.
URL https://doi.org/10.18420/DELFI2024_05
Main reference Andreas Scholl, Natalie Kiesler: "How Novice Programmers Use and Experience ChatGPT when Solving Programming Exercises in an Introductory Course", CoRR, Vol. abs/2407.20792, 2024.
URL https://doi.org/10.48550/ARXIV.2407.20792

The rapid advance of Generative AI (GenAI) in education presents both unprecedented opportunities and significant challenges for mastery learning. In my lightning talk, I focused on two critical areas: (1) Understanding students' use of GenAI tools and (2) Fostering their adoption among educators.

Firstly, there currently is a gap in our understanding of how students engage with GenAI tools, and we should investigate the state-of-the-art with regard to their interaction patterns, prompting strategies and application. This will help inform our instructional strategies and assessment methods, ensuring we leverage these tools to enhance learning.

Secondly, we need to consider the adoption of GenAI tools among educators. So, I discussed the importance of designing and implementing support structures for educators at higher education institutions. This may include offering workshops, mentoring programs, and other opportunities of sharing knowledge, e.g., at local conferences. By training educators, we can collectively harness the full potential of GenAI.

References

1 Johan Jeuring, Roel Groot, and Hieke Keuning. What skills do you need when developing software using ChatGPT. In Proceedings of Koli Calling 2023.
2 Lianne Roest, Hieke Keuning, and Johan Jeuring. Next-step hint generation for introductory programming using large language models. In Proceedings of the 26th Australasian Computing Education Conference 2024.
3 Natalie Kiesler, Dominic Lohr and Hieke Keuning. Exploring the Potential of Large Language Models to Generate Formative Programming Feedback. In Proceedings of the Frontiers in Education (FIE) Conference 2024.
4 James Prather, Paul Denny, Juho Leinonen, Brett Becker, Ibrahim Albluwi, Michelle Craig, Hieke Keuning, Natalie Kiesler, Tobias Kohn, Andrew Luxton-Reilly, Stephen MacNeil, Andrew Petersen, Raymond Pettit, Brent Reeves and Jaromir Savelka. The Robots are Here: Navigating the Generative AI Revolution in Computing Education. ITiCSE 2023 working group report.

4.8 Mastery of what – and at what cost?

Päivi Kinnunen (University of Helsinki, FI)

License Creative Commons BY 4.0 International license
© Päivi Kinnunen

Bloom defines one of the core tasks of teaching "Our basic task is to determine what we mean by mastery of the subject and to search for the methods and materials which will enable the largest proportion of our students to attain such mastery" (Bloom 1968, p 1).

This quote suggests two fundamental questions concerning the intended learning outcomes of assignments/courses/degree programs: 1) Mastery of what? What it is that we want our students to master? This question can be discussed at diQerent levels. Micro level: e.g. concepts, fractions of larger topics that are discusses during a teaching session. Course level: what knowledge, skills and attitudes do we want our students to master after a X week course? Macro level: degree program level intended learning outcomes. What kind of competencies do we wish our graduates to have?

The second question is what we mean by mastery of something? Can we clearly define what it means to master some aspects of knowledge, skills or attitudes? Is mastery always easily observable for the teacher (or student themselves). This is a challenging issue especially when we think about macro level learning objectives. Finally, I want to raise a question to what degree we have empirical evidence that mastery learning is beneficial for all students and not only for those with good selfregulations skills.

4.9 Learning at Scale

Tobias Kohn (KIT – Karlsruher Institut für Technologie, DE)

License Creative Commons BY 4.0 International license
© Tobias Kohn

In this lightning talk I discuss "Mastery Learning" and its challenges in modern education. I explore the question of how education can be scaled efficiently, with the solution proposed being mechanization. The idea is to break down learning into smaller, manageable pieces, which can then be addressed systematically. However, a central issue identified is that this approach often leads to isolated bits of knowledge, making it difficult for learners to grasp the overarching picture.

Connecting these fragmented pieces to form a coherent understanding is critical. The challenge, therefore, lies in bridging the gap between detailed, mechanized learning and the ability to perceive the broader context. We discuss strategies for overcoming this issue, offering solutions for how mastery learning can be improved to create a more holistic educational experience.

4.10 Combating the Widening Gap of Generative AI Utility

Juho Leinonen (Aalto University, FI)

License Creative Commons BY 4.0 International license
© Juho Leinonen

Joint work of Paul Denny, Juho Leinonen, James Prather, Andrew Luxton-Reilly, Thezyrie Amarouche, Brett A. Becker, Brent N. Reeves

Main reference Paul Denny, Juho Leinonen, James Prather, Andrew Luxton-Reilly, Thezyrie Amarouche, Brett A. Becker, Brent N. Reeves: "Prompt Problems: A New Programming Exercise for the Generative AI Era", in Proc. of the 55th ACM Technical Symposium on Computer Science Education, SIGCSE 2024, Volume 1, Portland, OR, USA, March 20-23, 2024, pp. 296–302, ACM, 2024.

URL https://doi.org/10.1145/3626252.3630909

The increasingly powerful generative AI models have brought some benefits to software engineering. There is emerging evidence that using tools such as GitHub Copilot can increase productivity for professional software developers [1]. For novice programmers, however, there is recent evidence that while the best students can benefit from generative AI tools, struggling students do not get the same benefits [2]. In fact, struggling students can even be harmed by them. AI tools can become a crutch, leaving students with a false sense of mastery.

In my lightning talk, I proposed that we could try to combat this "widening gap" between students by explicitly teaching them how to prompt AI models to write code with Prompt Problems [3]. This could help struggling students to get the benefits of generative AI that are currently enjoyed by professional programmers and the best performing students.

References

1 Arghavan Moradi Dakhel, Vahid Majdinasab, Amin Nikanjam, Foutse Khomh, Michel C. Desmarais, and Zhen Ming (Jack) Jiang. GitHub Copilot AI pair programmer: Asset or Liability?, Journal of Systems and Software, Volume 203, 2023, 111734, ISSN 0164-1212, https://doi.org/10.1016/j.jss.2023.111734

2 James Prather, Brent N Reeves, Juho Leinonen, Stephen MacNeil, Arisoa S Randrianasolo, Brett A. Becker, Bailey Kimmel, Jared Wright, and Ben Briggs. 2024. The Widening Gap: The Benefits and Harms of Generative AI for Novice Programmers. In Proceedings of the 2024 ACM Conference on International Computing Education Research – Volume 1 (ICER '24), Vol. 1. Association for Computing Machinery, New York, NY, USA, 469–486. https://doi.org/10.1145/3632620.3671116

3 Paul Denny, Juho Leinonen, James Prather, Andrew Luxton-Reilly, Thezyrie Amarouche, Brett A. Becker, and Brent N. Reeves. 2024. Prompt Problems: A New Programming Exercise for the Generative AI Era. In Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2024). Association for Computing Machinery, New York, NY, USA, 296–302. https://doi.org/10.1145/3626252.3630909

4.11 From What to Why and How

Andrew James Luxton-Reilly (University of Auckland, NZ)

License 🅭 Creative Commons BY 4.0 International license
© Andrew James Luxton-Reilly

Generative AI can generate solutions to simple programming problems typically used as exercises to help students learn in introductory programming courses. We can no longer rely on assessment tasks as a motivator to learn "content knowledge" because our standard tools make such tasks redundant. We need to shift student motivation from focusing on content knowledge acquisition to understanding why the content is important, and how we use the content effectively. This requires rethinking the capabilities that we expect from students and the order in which they are acquired.

For students who are motivated to learn programming, we need to focus attention on teaching students to use GenAI to support their learning. Given the high degree of flexibility in access to learning resources provided by GenAI, we should explicitly discuss metacognitive skills and encourage students to seriously reflect on the processes they are using to learn. This has significant implications for learning in general, and for Mastery Learning approaches.

4.12 Metacognitive and Social Harms of Generative AI

Stephen MacNeil (Temple University – Philadelphia, US)

License Creative Commons BY 4.0 International license
© Stephen MacNeil

Recent research on generative AI highlights its potential to enhance education through just-in-time learning materials and personalized tutoring. However, our lab has conducted multiple studies which also reveal potential harms to students.

Our first paper explores the impact of generative AI on student help-seeking and classroom social dynamics. While AI can provide immediate assistance to students, which benefits students who might experience social barriers to seeking help, our findings suggest it might inadvertently reduce student interactions. This diminished peer-to-peer engagement can lead to a less collaborative learning environment, where students no longer seek out and establish social support groups on which they have previously relied.

In our second paper, we investigate the risks generative AI poses to student metacognition. Metacognition, or the awareness and understanding of one's own thought processes, is critical for effective learning. Our research indicates that AI can sometimes mislead students, giving them a false sense of understanding and progress.

Through these and other projects, we are working to identify the nuanced harms associated with the use of generative AI in educational settings. Our goal is to develop strategies and best practices that maximize the learning benefits of AI while mitigating its potential downsides. If successful, we hope to pave the way for the responsible use of generative AI in education, ensuring that it serves as a tool for enhancement rather than a crutch that diminishes the quality of learning.

4.13 As Student Practices Evolve, So Too Should Our Classrooms

Miranda C. Parker (San Diego State University, US)

License Creative Commons BY 4.0 International license
© Miranda C. Parker

Traditionally, in computer science courses in higher education, a teacher introduces a new concept, a student then practices that concept, and the teacher will test the student's knowledge of that concept, likely among many others. In a Mastery Learning setting, this process gets atomized such that a student has to fully understand a concept before moving on to the next concept. However, in either case, ChatGPT, Github's CoPilot, and other widely available AI-based large language models (LLMs) are fundamentally changing what it means to get information, code, and learn. Rather than ignore or ban the use of these tools, we can harness this moment to rethink how we teach computer science, potentially in a way to support mastery learning. If code generation is suddenly less time-consuming, but potentially wrought with errors, shifting our instruction to focus more on code reading and explaining will be crucial. Further, LLMs can help us generate the multitude of assessments that are needed as each student learns a concept to demonstrate mastery. With their powers combined, LLMs can help support more students in more mastery learning settings, completely changing the landscape of what it means to teach computer science.

4.14 VoiceEx: Facilitating Self-explanations with an LLM

Andrew Petersen (University of Toronto Mississauga, CA)

License Creative Commons BY 4.0 International license
© Andrew Petersen
Joint work of Andrew Petersen, Angela Zavaleta, Michael Liut

Self-explanations can help learners engage with course content and to build connections between topics. However, self-explanations need to be connected to prior concepts to be effective, and many learners do not self-explain effectively. We propose using an LLM to compare expert-generated explanations with learner self-explanations and to prompt the learner to revise and elaborate on their self-explanations. In a pilot study of this approach, we found that students generally responded to LLM prompts and responded to the feedback but did not necessarily update their self-explanations accordingly. We also detected that students felt a sense of social presence with the LLM system, reacting to it having access to their self-reflections.

4.15 An illusion of learning?

Judithe Sheard (Monash University – Clayton, AU)

License Creative Commons BY 4.0 International license
© Judithe Sheard
Main reference Judy Sheard, Paul Denny, Arto Hellas, Juho Leinonen, Lauri Malmi, and Simon. 2024. Instructor Perceptions of AI Code Generation Tools – A Multi-Institutional Interview Study. In Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2024). Association for Computing Machinery, New York, NY, USA, 1223–1229. https://doi.org/10.1145/3626252.3630880

At the beginning of 2023 in a staffroom conversation about generative AI tools, a colleague remarked, "I don't think my students know about these yet". Now 18 months later, the world is a different place. A study of instructors' perceptions of generative AI tools I conducted with a group of computing education researchers found many ideas about the benefits and challenges to teaching and assessment [1]. Probably the greatest challenge was the impact on learning. As one participant claimed, "It is a great tool ... to prevent yourself from learning". When using a generative AI tool are students handing over the responsibility for learning to the tool and in effect the tool becomes a surrogate for learning?

Mastery learning emphasizes achievement of a high level of understanding in a given topic before moving on to the next. Key aspects: clear learning outcomes, individualised pacing of learning, formative assessment, support, summative assessment. When considering a mastery learning approach the availability of generative AI tool brings particular challenges. Students using generative AI tools can take a passive learning approach, engage superficially with the tools leading to surface rather than deep learning. An over-reliance on feedback from the tools can impact negatively on development of critical thinking and problem-solving skills. With less interaction with humans students may miss important guidance and emotional support. The result is an illusion of learning where students feel they are learning effectively while not developing deep understanding.

References

1 Judy Sheard, Paul Denny, Arto Hellas, Juho Leinonen, Lauri Malmi, and Simon. 2024. Instructor Perceptions of AI Code Generation Tools – A Multi-Institutional Interview Study. In Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2024). Association for Computing Machinery, New York, NY, USA, 1223–1229. https://doi.org/10.1145/3626252.3630880

4.16 Challenges and Opportunities for Gen AI in Mastery Learning

Jacqueline Smith (University of Toronto, CA)

License Creative Commons BY 4.0 International license
© Jacqueline Smith

Failure is an important part of the learning process, but most students have negative feelings about it. Students can respond differently to situations where they expect their risk of failure is high. Some students may seek human guidance from an instructor or peer, while others prefer to handle the situation privately, without anyone knowing. Fear of failure can be a barrier in mastery learning, such as if a student is not willing to take a mastery quiz. Generative AI may be able to play a role in supporting students in pushing through a situation where they are worried about failure. The availability of generative AI can also help with scaling of this kind of support, as many current course environments do not have the human resources to provide one-on-one coaching to students.

Some challenges in integrating generative AI into education in general are the diverse range of guidelines students are receiving from their instructors on what is and is not appropriate use of generative AI. This can be a result of instructors' own lack of understanding of the use of generative AI tools, and how they can be appropriately integrated (or not) into their classrooms. Instructors' own concerns about the ethical implications of these tools may also affect their adoption. We also need to consider that students may have their own ethical considerations about the use of generative AI, whether for reasons of privacy, climate concerns, or discomfort interacting with a machine.

4.17 A Game of Shadows: Who are the Shadows and What is the Game?

Claudia Szabo (University of Adelaide, AU)

License Creative Commons BY 4.0 International license
© Claudia Szabo
Joint work of Claudia Szabo, Judy Sheard
Main reference Claudia Szabo, Judy Sheard: "Learning Theories Use and Relationships in Computing Education Research", ACM Trans. Comput. Educ., Vol. 23(1), pp. 5:1–5:34, 2023.
URL https://doi.org/10.1145/3487056

Mastery learning techniques are an excellent fit for foundational programming education, especially for diverse postgraduate cohorts. These techniques emphasize the importance of allowing students to progress at their own pace, which is essential when considering the varied levels of programming knowledge within such groups. Students may come from both cognate (closely related to programming) and non-cognate (unrelated) backgrounds, which means their learning needs can differ greatly.

In light of these differences, a self-paced, practice-based approach is critical to address individual learning speeds and requirements. Tailoring learning pathways to accommodate both experienced programmers and beginners ensures that each student can work through material at a comfortable pace. This method supports a more personalized educational experience that encourages deeper understanding rather than a one-size-fits-all approach.

Despite the promise of this method, practice remains a formidable challenge in educational settings. Creating consistent, quality practice opportunities that cater to each student's needs is often difficult, particularly when faced with the diverse range of programming skills and learning behaviors. Providing the right resources and feedback at the right time is key to making practice an ally rather than an obstacle.

Moreover, leveraging AI in educational contexts to create more effective and inclusive learning environments remains an uphill battle. While AI holds significant potential for tailoring education to individual students' needs, the complexities involved in implementing these technologies can sometimes feel insurmountable. However, overcoming these challenges is crucial to fully realize the potential of mastery learning in programming education for diverse cohorts.

4.18 Mastery Learning in Computing Education

Lisa Zhang (University of Toronto Mississauga, CA)

License Creative Commons BY 4.0 International license
© Lisa Zhang

The talk explores the implementation of mastery learning in computing education and the challenges instructors and students face. Mastery learning emphasizes ensuring students achieve a high level of understanding before progressing, but the process is demanding for both parties. Instructors encounter difficulties in distilling course objectives, creating materials, and continuously assessing and providing feedback. Meanwhile, students must develop strong self-regulation skills to meet mastery criteria, presenting its own set of challenges.

Large language models (LLMs) have the potential to ease the burden on instructors as they could <could assist in refining course objectives, generating assessments, and offering adaptive feedback to students. The question, however, is how such models can be integrated effectively beyond introductory courses like CS1. Moreover, addressing these concerns calls for resources that provide clear guidance on leveraging LLM tools in a meaningful way.

Another critical aspect is the need to explore gradual adoption approaches to LLMs in mastery learning. Risks such as over-reliance, student data privacy, and loss of personalized instruction need to be mitigated. Instructors are encouraged to explore incremental adoption to balance innovation with careful oversight, ensuring that the integration aligns with educational goals.

Finally, it is important to reflect on the trade-offs associated with using LLM tools in computing education. While such technologies may streamline some instructional processes, there is a risk of losing valuable interpersonal elements of teaching.

4.19 Mastery Learning and GenAI: A happy marriage?

Claudio Álvarez Gómez (Universidad de los Andes – Santiago, CL)

License Creative Commons BY 4.0 International license
© Claudio Álvarez Gómez
Joint work of Claudio Alvarez, Maíra Marques Samary, Alyssa Friend Wise
Main reference Claudio Alvarez, Maíra Marques Samary, Alyssa Friend Wise: "Modularization for mastery learning in CS1: a 4-year action research study", J. Comput. High. Educ., Vol. 36(2), pp. 546–589, 2024.
URL https://doi.org/10.1007/S12528-023-09366-1

Converting a traditional CS1 course to a Mastery Learning format is challenging, with uncertain outcomes if key needs and risks are not properly managed. Coordinating a large course under this format requires significant effort, from providing support to students in different modules to creating multiple formative and summative assessments for varying levels of mastery. Feedback must be delivered rapidly, and new lecturers must be trained, as mastery learning is not widely endorsed. Based on eight years of experience implementing a CS1 course in this format at Universidad de los Andes, Chile (see main ref), we have identified three critical factors for success: effective course management, ensuring quality in teaching and assessment, and maintaining scalability and efficiency. Generative AI technologies offer promising tools to support mastery learning by reducing barriers, costs, and risks. Key aspects of the pedagogy, such as personalized learning, assessment creation, and feedback generation, can be automated, while course management tasks like communication and staff training can also be enhanced. This raises important questions: what AI tools are most suitable for these improvements? How can they be tailored to the diverse needs of mastery learning implementations in computer science? How can we measure their impact to ensure effectiveness? Widespread adoption of mastery learning, supported by generative AI, would enable institutions to improve educational outcomes and provide students with deeper learning experiences, preparing them for future professional challenges.

4.20 AI and Computing Education: Challenges and Strategies for Effective Mastery Learning

Jaromír Šavelka (Carnegie Mellon University – Pittsburgh, US)

License Ⓒ Creative Commons BY 4.0 International license
© Jaromír Šavelka

This talk explores the effects of AI on mastery learning in computing education. It outlines key challenges, including increased procrastination, widening achievement gaps, and potential misinformation from AI-generated content. The discussion then shifts to innovative strategies for leveraging AI to enhance mastery learning, such as AI-powered self-regulation tools, robust feedback mechanisms, and adaptive learning paths. By addressing both the pitfalls and potential of AI in computing education, this talk aims to suggest more effective integration of AI tools in the pursuit of deep, foundational learning.

Participants

- Claudio Álvarez Gómez
Universidad de los Andes – Santiago, CL
- Nick Falkner
University of Adelaide, AU
- Armando Fox
University of California – Berkeley, US
- Michelle Friend
University of Nebraska – Omaha, US
- Petri Ihantola
University of Jyväskylä, FI
- Johan Jeuring
Utrecht University, NL
- Hieke Keuning
Utrecht University, NL
- Natalie Kiesler
Technische Hochschule Nürnberg, DE
- Päivi Kinnunen
University of Helsinki, FI
- Tobias Kohn
KIT – Karlsruher Institut für Technologie, DE
- Juho Leinonen
Aalto University, FI
- Andrew James Luxton-Reilly
University of Auckland, NZ
- Stephen MacNeil
Temple University – Philadelphia, US
- Lauri Malmi
Aalto University, FI
- Claudia Ott
University of Otago, NZ
- Miranda C. Parker
San Diego State University, US
- Andrew Petersen
University of Toronto Mississauga, CA
- Judithe Sheard
Monash University – Clayton, AU
- Jacqueline Smith
University of Toronto, CA
- Claudia Szabo
University of Adelaide, AU
- Lisa Zhang
University of Toronto Mississauga, CA
- Jaromír Šavelka
Carnegie Mellon University – Pittsburgh, US