Report from Dagstuhl Seminar 25432

Deep Continual Learning in the Foundation Model Era

Christopher Kanan[*1], Martin Mundt[*2], Tinne Tuytelaars[*3], Joost van de Weijer[*4], and Timm Felix Hess[†5]

1 **University of Rochester, US.** chriskanan@gmail.com
2 **Universität Bremen, DE.** mundtm@uni-bremen.de
3 **KU Leuven, BE.** tinne.tuytelaars@esat.kuleuven.be
4 **Computer Vision Center – Barcelona, ES.** joost@cvc.uab.es
5 **KU Leuven, BE.** TimmFelix.Hess@esat.kuleuven.be

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 25432 "Deep Continual Learning in the Foundation Model Era". This seminar brought together 23 researchers to discuss research at the intersection of continual learning and foundation models. The discussion centered on the major challenges arising from continual training of foundation models, including the need for new benchmarks, new opportunities for memory-based continual learning, emerging application domains, and the development of efficient metrics to quantify forgetting of foundation model knowledge. In addition, the report contains a summery of the talks of the partipants.

Seminar October 19–24, 2025 – https://www.dagstuhl.de/25432
2012 ACM Subject Classification Computing methodologies → Artificial intelligence; Computing methodologies → Neural networks; Computing methodologies → Learning settings; Computing methodologies → Learning paradigms; Computing methodologies → Bio-inspired approaches; Computing methodologies → Learning latent representations; Computing methodologies → Computer vision; Computing methodologies → Natural language processing; Computing methodologies → Distributed artificial intelligence
Keywords and phrases continual learning, deep learning, foundation models
Digital Object Identifier 10.4230/DagRep.15.10.119

1 Executive Summary

Christopher Kanan (University of Rochester, US, chriskanan@gmail.com)
Martin Mundt (Universität Bremen, DE, mundtm@uni-bremen.de)
Tinne Tuytelaars (KU Leuven, BE, tinne.tuytelaars@esat.kuleuven.be)
Joost van de Weijer (Computer Vision Center – Barcelona, ES, joost@cvc.uab.es)

This summary presents the main outcomes of the Dagstuhl Seminar "Deep Continual Learning in the Foundation Model Era" (25432). Foundation models have transformed artificial intelligence and generated major socio-economic impact. However, they face significant challenges: the prevailing training paradigm, which relies on retraining models from scratch as new data arrives, is unsustainable, models must also be adapted to better align with human values and mitigate biases, and adaptation often leads to catastrophic forgetting. The field of deep continual learning focuses on enabling systems to acquire knowledge over time

[*] Editor / Organizer
[†] Editorial Assistant / Collector

Deep Continual Learning in the Foundation Model Era, *Dagstuhl Reports*, Vol. 15, Issue 10, pp. 119–134
Editors: Christopher Kanan, Martin Mundt, Tinne Tuytelaars, Joost van de Weijer, and Timm Felix Hess
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

from non-stationary data streams. It provides theoretical and methodological tools that can help address key limitations of foundation models, including efficient updating, adaptation at scale, alignment with human values, and bias mitigation.

The objective of the seminar was to bring together world-class researchers in the fields of deep continual learning and foundation model training to examine their interplay, identify key challenges, and outline promising directions for future research.

As a major outcome of the seminar, the following promising research directions were identified:

- New memory usage paradigms are emerging for continual learning in foundation models. Approaches that combine parametric learning with in-context learning are a promising path toward rapid adaptation and experience accumulation. In-context learning reduces the need for costly and forgetting-prone parameter updates, while periodic consolidation of in-context knowledge through distillation into model parameters may limit the computational overhead associated with growing context.
- New methods are needed for the efficient evaluation of catastrophic forgetting in foundation models during adaptation, as full-model evaluation is computationally prohibitive in iterative adaptation settings.
- Model scale, pre-training, and architectural choices fundamentally alter the stability–plasticity trade-off. Systematic studies across model scales and objectives are needed to understand when scale inherently mitigates forgetting and when explicit continual learning mechanisms remain necessary.
- Continual learning for foundation models requires new benchmarks that go beyond synthetic task sequences and better reflect real-world usage. Promising benchmark domains include personalization over time, multi-agent interaction, model self-improvement loops, and unlearning for privacy or safety. These benchmarks should impose realistic compute constraints.

2 Table of Contents

3 Overview of Talks

3.1 Three forms of continual learning from foundation models

Rahaf Aljundi (Toyota Motor Europe – Zaventem, BE)

Continual learning enables models to adapt to streaming data while retaining knowledge, but traditional parameter updates risk catastrophic forgetting. Foundation models exhibit in-context learning capability, offering a complementary approach to adaptation. I propose a system combining a core reasoning model with short- and long-term memory. Short-term memory captures immediate context and reasoning, while long-term memory aggregates these for slow, informed updates. This architecture supports rapid adaptation and sustained improvement without full retraining. Empirical insights from presented work show in-context learning outperforming textual memory and fine-tuning, suggesting memory-centric strategies could further enhance performance, including in vision-language tasks.

3.2 How green is continual adaptation of Foundation Models, really?

Andrew D. Bagdanov (University of Florence, IT)

As research in Continual Learning continues gaining momentum, it is essential to understand its implications in the Foundation Model Era. The typically massive scale of foundation models demands efficient methods for continual adaptation, which has prompted a shift from monolithic approaches to more modular techniques such as prompt learning and adapter-based methods. While energy efficiency is often touted as an advantage of Continual Learning, a systematic comparative analysis of these methods regarding their efficiency-accuracy trade-offs is still missing in the literature. In this talk, I will explore the architectural differences between representative monolithic and modular approaches, and present an empirical evaluation of their energy consumption during training and inference. By uncovering the true energy costs of these methods, we hope to better assess their sustainability and impact on future AI research and practice.

3.3 Modular Continual Learning

Lucas Caccia (Microsoft Research – Montréal, CA)

In this talk I present different approaches to modular continual learning. Modularity in parameter, model and data space offer different tradeoffs, each with key properties that are required to enable long term, efficient, continual learning systems.

3.4 The Myth of Catastrophic Forgetting

Antonio Carta (University of Pisa, IT)

In this talk I discuss some limitations of continual learning setups, and how they impacted the design space of continual learning methods. Then, I highlight some open challenges such as robustness to noise, better measures of representation forgetting, and continual optimization.

3.5 Toward "Model-based" Continual Learning

Laurent Charlin (HEC Montréal, CA)

Joint work of Laurent Charlin, Yipeng Zhang, Hafez Ghaemi, Jungyoon Lee, Shahab Bakhtiari, Eilif B. Mulle
Main reference Anonymous: "Self-Supervised Learning from Structural Invariance", in Proc. of the Submitted to The Fourteenth International Conference on Learning Representations, 2025.
URL https://openreview.net/forum?id=r3JUDAYjIH

We suggest using "natural" processes as a source of continuous data. Such data has interesting properties, including temporal consistency, multi-level periodicity, and the appearance of novel elements. We also observe that "close by" data (for example, two photos of the same landscape taken at slightly different times) might be modelled as being generated from two representations that change sparsely. We instantiate this idea for learning image representation (using SSL) from natural pairs, demonstrating that this type of continuous data is beneficial for representation learning. We also propose a model inspired by the generative data process.

3.6 Three questions/issues on continual learning

Jonghyun Choi (Seoul National University, KR)

Joint work of Jonghyun Choi, Minhyuk Seo, Tinne Tuytelaars
Main reference Minhyuk Seo, Hyunseo Koh, Jonghyun Choi: "Budgeted Online Continual Learning by Adaptive Layer Freezing and Frequency-based Sampling", in Proc. of the The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025, OpenReview.net, 2025.
URL https://openreview.net/forum?id=dOAkHmsjRX

I start with three questions: What is continual learning (CL) beyond class-incremental learning (CIL)? Are online CIL methods being compared fairly? And why does continual-like training improve multimodal foundation models? I introduce CIL setups we proposed so far. I then argue that considering computational cost among CIL methods is crucial for fair ground, introducing our computationally efficient CIL method. For the third question, I made a connection between federated learning (spatially distributed multitask learning) and continual learning (temporally distributed multitask learning). Finally, I introduce our recent work for federated learning with heterogeneous models for personalization.

3.7 XAI in incremental learning

Barbara Hammer (Universität Bielefeld, DE)

How can persons deal with continuously learning models? I will argue that explainability is a crucial issue to enable persons to understand how and why a model has changed. I will have a glimpse at incremental explanations and explanation of drift, and I will argue that we need to move towards XAI technologies which also address interactions when dealing with multimodal foundation models.

3.8 Adapting to the Unknown: Novel Class Discovery, Open-Vocabulary Learning, & Beyond

Tyler Hayes (Georgia Institute of Technology – Atlanta, US)

In this talk, I argue that there are two complementary paradigms for enabling adaptation in neural networks: dynamic adaptation and static adaptation. Dynamic adaptation methods adapt the model on the fly by exploiting a stream of incoming data, as in novel class discovery. In contrast, static adaptation methods prepare the model in advance for unavailable semantic concepts, often by leveraging external knowledge sources such as natural language (e.g., open-vocabulary learning). I then present our PANDAS method that uses prototypes and a distance-based classifier to adapt to novel classes, before presenting our SHiNe method that fuses class taxonomy information into a classifier to improve open-vocabulary performance. I then conclude with my vision for the future of novel concept discovery and learning.

3.9 Is there a benefit in joint training?

Timm Felix Hess (KU Leuven, BE)

Joint work of Timm Felix Hess, Abhishek Jha, Gido M van de Ven, Tinne Tuytelaars

We analyze the challenges posed by partial observability when training on sequential data, examining its impact beyond catastrophic forgetting. Preliminary results indicate that the partial observability of sequential observations inhibits learning, causing a performance deficit relative to a jointly trained model. We position this perspective as essential to Continual Learning, where achieving joint-model performance typically is considered the ultimate goal. We also link it to the open problem of how (foundation) models can effectively integrate separate, sequential observations to build a generalized knowledge base.

3.10 The Death of "AGI" and the Rebirth of Continual Learning

Christopher Kanan (University of Rochester, US)

For years, continual learning was treated as a prerequisite for AGI. Today, many claim AGI is already here, at least under flexible, labor-centric definitions. Either way, the old continual learning agenda no longer fits. This talk proposes a reset that aligns with two concrete futures: (1) economic automation, where frontier models stay useful through fast, compute-aware updates; and (2) human-like synthetic minds, where systems form, consolidate, and reuse memories over long horizons. Minimizing forgetting is the wrong goal. The right objective is to maximize retained competence and forward transfer per unit of compute. I will share simple, high-leverage mechanisms that move the needle in practice: structured replay with wake and sleep, stability-gap mitigation for rapid adaptation, and selective rehearsal that progresses from easy to hard. These ideas improve production-style foundation models and lay the groundwork for memory-centric agents. The takeaway is clear: storage is cheap, data is abundant, compute is precious. Continual learning should be the default operating mode for models that matter.

3.11 Rethinking Efficiency in Continual Learning

Dhireesha Kudithipudi (University of Texas – San Antonio, US)

In this talk, we present the need for thinking across the stack to enable efficient continual learning that goes beyond performance improvements. We explore how integrating insights across the neuroscience, algorithm, architecture, and system layers offers a unique opportunity to design efficient systems. By distilling shared principles from hardware and software, we can develop more effective strategies for adaptation, replay, and regularization. This cross-stack co-design reveals how to build foundation models that are robust, compute, energy, and latency efficient.

3.12 Continual Learning on Vision-Language Pretrained Models

Xialei Liu (Nankai University – Tianjin, CN)

Joint work of Xusheng, Cao, Linglan Huang
Main reference Linlan Huang, Xusheng Cao, Haori Lu, Yifan Meng, Fei Yang, Xialei Liu: "Mind the Gap: Preserving and Compensating for the Modality Gap in CLIP-Based Continual Learning", CoRR, Vol. abs/2507.09118, 2025.
URL https://doi.org//10.48550/ARXIV.2507.09118
Main reference Xusheng Cao, Haori Lu, Linlan Huang, Xialei Liu, Ming-Ming Cheng: "Generative Multi-modal Models are Good Class-Incremental Learners", in Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024, pp. 28706–28717, IEEE, 2024.
URL https://doi.org//10.1109/CVPR52733.2024.02712
Main reference Linlan Huang, Xusheng Cao, Haori Lu, Xialei Liu: "Class-Incremental Learning with CLIP: Adaptive Representation Adjustment and Parameter Fusion", in Proc. of the Computer Vision – ECCV 2024 – 18th European Conference, Milan, Italy, September 29-October 4, 2024, Proceedings, Part LIV, Lecture Notes in Computer Science, Vol. 15112, pp. 214–231, Springer, 2024.
URL https://doi.org//10.1007/978-3-031-72949-2_13
Main reference Xusheng Cao, Haori Lu, Linlan Huang, Fei Yang, Xialei Liu, Ming-Ming Cheng: "Knowledge Graph Enhanced Generative Multi-modal Models for Class-Incremental Learning", CoRR, Vol. abs/2503.18403, 2025.
URL https://doi.org//10.48550/ARXIV.2503.18403

Continual learning is one of the key capabilities of next-generation artificial intelligence systems. It aims to enable systems to continuously acquire new knowledge in dynamically changing environments while avoiding catastrophic forgetting of previously learned knowledge, thereby mimicking the way humans learn. This talk explores continual learning in multimodal models based on vision-language pre-training frameworks, addressing current challenges from both the perspective of discriminative models (e.g., CLIP) and generative models (e.g., InternVL). Finally, it provides a summary and outlook on the advances in the field of continual learning.

3.13 Continual, Decentralized Compositionality for Sustainable Artificial Intelligence

Vincenzo Lomonaco (University of Pisa, IT)

URL https://www.vincenzolomonaco.com/collage-lab

Sustainable Artificial Intelligence envisions a shift from centralized, resource-intensive models toward decentralized, reusable systems that are both environmentally efficient and socially inclusive. By distributing intelligence across networks rather than concentrating it in isolated silos, decentralization enables collective innovation, transparency, and equitable access to AI capabilities. This approach aligns ecological sustainability with democratization: reducing the carbon and computational footprint of AI while empowering diverse communities to contribute, adapt, and benefit from shared knowledge. Through continual adaptation and compositional design, AI can evolve as a green, open, and participatory infrastructure – towards the definition of a circular economy, building new AI solutions from existing ones.

3.14 Modularity for Continual Robot Learning

Jorge Mendez-Mendez (Stony Brook University, US)

Standard monolithic models struggle with continual learning, as they store knowledge in an unstructured manner that is difficult to update with new information. This talk argues that for domains with well-understood structures, like robotics, modular solutions offer a significantly more powerful framework for lifelong learning. Task and motion planning (TAMP) decomposes problems both temporally and functionally, providing a concrete exploration of this concept. I present a lifelong learning algorithm for TAMP's parameter samplers, showing how a robot can continually improve at household chores. I then conclude by discussing how these modular principles relate to the future of continual learning in the era of foundation models.

3.15 What does it mean to accumulate knowledge in lifelong (machine) learning?

Martin Mundt (Universität Bremen, DE)

When we speak of lifelong or continual (machine learning), we often refer to the ability to accumulate knowledge over long periods of time from possibly changing non-stationary data or experiences. This seems to have implicitly resulted in a tendency to believe the ability to "add on" knowledge without forgetting is the main ingredient to success. In this talk, I argue that this is not necessarily realistic. To this end, I show two examples, tackling a controlled small scale and a large-scale real-world scenario. In the first, "adding on" knowledge is problematic, because confounders, i.e. spurious correlations, change over time due to partial observability – leading to contradictions and the necessity to revise knowledge. In the second example, a dataset spanning 250 years of books, changes in language and society lead to knowledge becoming outdated, obsolete, or possibly changing substantially. In an analysis of training LLMs on this kind of data over time and applying content moderation tools it becomes clear that our current (continual learning) tools are not yet set up to tackle such scenarios adequately.

3.16 CoPeP: Continual Pretraining for Protein Language Models

Darshan Patil (MILA – Montreal, CA)

In recent years, protein language models (pLMs) have gained significant attention for their ability to capture the structure and function of proteins, accelerating the discovery of new therapeutic drugs. These models are typically trained on large, evolving corpora of proteins that are continuously updated by the biology community. motivating the need for continual

learning. As a result, we introduce the Continual Pretraining of Protein Language Models (CoPeP) benchmark, a novel benchmark for evaluating continual learning approaches on pLMs. We evaluate methods from the continual learning literature spanning replay, unlearning, and plasticity-based methods, some of which have never been applied to models and data of this scale. Our findings reveal that continual pretraining can improve the perplexity of the model over the traditional, non-continual training and several continual learning methods do outperform naive continual pretraining.

3.17 A Metacognition Lens on Continual Foundation Model Learning: Progress & Challenges

Ameya Prabhu (Universität Tübingen, DE)

Joint work of Ameya Prabhu, Mathias Bethge

I broadly compare human metacognition with language models. Humans build a coherent map of the world and monitor their own thinking while "reading" and "writing" throughout their life. The key question: Can LLMs do the same – continually learn from mistakes, backtrack, correct, check, and plan? I conjecture there are three pillars: metacognitive knowledge (learning of facts, procedures, strategies as internal knowledge offline i.e. across large text corpora), metacognitive regulation (given a problem at test-time, trying to continually planning, monitoring, verification, and self-correction but online), and metacognitive experience (combining the both after each task – reflecting on past to continually improve future choices via world modeling). Open problems include how to self-improve with better data, algorithms, and pipelines; how to measure retention and transfer at the sample level; how to trade off facts, tool use, and strategy; how to detect and recover from errors using language feedback; how to evaluate and refine planning; and how to formalize continual world modeling while assessing benchmarks, metrics, and methods. The goal overall pitched is to chart progress and surface open problems across these three metacognitive axes in continual foundation-model learning.

3.18 Beyond Zero-Shot Generalization: Strengthening Vision-Language Models with Adaptation & Personalization

Elisa Ricci (University of Trento, IT)

Groundbreaking achievements in Vision-Language pretraining have increased the interest in crafting Vision-Language Models (VLMs) that can understand visual content alongside natural language, enabling a new definition of zero-shot classification. Despite huge pretraining databases, VLMs still face limitations, suffering from performance degradation in case of large train-test dissimilarity and requiring the design of highly generalizing textual templates. Test-Time Adaptation (TTA) and Few-Shot Adaptation (FSA) can effectively improve the robustness of VLMs by adapting a given model to online inputs or to a specific downstream

task given a few annotated examples. In this talk, I will present an overview of a few recent works from my research group on TTA, FSA and other strategies aimed at improving the robustness of VLMs across various downstream tasks.

3.19 Some reflections on Deep Continal Learning in the foundation model era

Tinne Tuytelaars (KU Leuven, BE)

In this talk, I briefly discuss two recent works from our lab. The first one investigates whether it is possible to predict which samples will be forgotten when learning continually. We find that what is learned slowest is forgotten first. Adapting the sampling accordingly when constructing a replay buffer can reduce forgetting. In the second work, we investigate continual learning with access to all data from previous tasks, with the goal to make the training more compute efficient. We show that careful initialization, regularization and batch sampling can make training twice as fast with similar or better performance than joint incremental learning.

3.20 Do we need test-time adaptation for VLMs? The side-effects of unlearning

Bartlomiej Twardowski (Computer Vision Center – Barcelona, ES)

This presentation investigates the efficacy of test-time adaptation (TTA) for Vision Language Models (VLMs), particularly in more realistic scenarios. Our preliminary findings indicate a consistent degradation in VLM performance with increasing levels of visual corruption that cannot be overcome by improved prompting techniques. Furthermore, the study explores the unintended consequences of unlearning, demonstrating how the removal of a single concept can significantly impact the representation and behavior of other, seemingly unrelated concepts. Mechanistic Interpretability (MI) offers a promising framework for a more nuanced understanding of learning and unlearning processes within continual learning settings.

3.21 How to look at forgetting?

Gido van de Ven (University of Groningen, NL)

Joint work of Timm Hess, Eli Verwimp, Gido M. van de Ven, Tinne Tuytelaars
Main reference Timm Hess, Eli Verwimp, Gido M. van de Ven, Tinne Tuytelaars: "Knowledge Accumulation in Continually Learned Representations and the Issue of Feature Forgetting", Trans. Mach. Learn. Res., Vol. 2024, 2024.
URL https://openreview.net/forum?id=aHtZuZfHcf

Continual learning research has shown that neural networks suffer from catastrophic forgetting "at the output level", but it is debated whether this is also the case at the level of learned representations. We show that, even though forgetting in the representation (i.e., feature forgetting) can be small in absolute terms, when measuring relative to how much was learned during a task, forgetting in the representation tends to be just as catastrophic as forgetting at the output level. We also show that this feature forgetting is problematic as it substantially slows down the incremental learning of good general representations (i.e., knowledge accumulation).

3.22 The Butterfly Effect of Minimal Updates on Generative Visual Models

Joost van de Weijer (Computer Vision Center – Barcelona, ES)

Main reference Héctor Laria Mantecon, Alex Gomez-Villa, Imad Eddine Marouf, Kai Wang, Bogdan Raducanu, Joost van de Weijer: "Assessing Open-world Forgetting in Generative Image Model Customization", CoRR, Vol. abs/2410.14159, 2024.
URL https://doi.org//10.48550/ARXIV.2410.14159

In this talk, I investigate forgetting in visual foundation models for image generation. We introduce the term open-world forgetting to refer to the fact that forgetting in foundation models can occur throughout its vast latent space. We look at the special case of minimal updates, which refer to tasks that only require small parameter updates to the model. We propose three methods to measure catastrophic forgetting in DMs, and show that for all three these methods we observe catastrophic forgetting even for the case of minimal model updates.

3.23 Continual Learning for Multi-modal Human-centric Applications

Liyuan Wang (Tsinghua University – Beijing, CN)

Continual learning is a fundamental mechanism for enabling long-term adaptation and knowledge accumulation in intelligent systems. However, in the era of large-scale pretraining, it remains a critical challenge to extend continual learning to dynamic, heterogeneous real-world environments. This talk presents our recent efforts on continual learning in pretrained models, with a focus on two key directions. First, we propose a modality-heterogeneous

continual pretraining framework for multi-modal physiological signal generation, enabling real-time and robust monitoring of human health conditions. Second, inspired by the spatial cognition mechanisms of the biological brain, we develop embodied agents that construct and refine cognitive maps through continuous collection of spatial knowledge, thus equipping multi-modal language models with strong long-horizon generalization in complex environments. Together, these advances point toward the development of brain-inspired embodied intelligence with lifelong adaptability.

4 Working Groups

4.1 Working Group on Main Changes for Continual Learning in the Foundation Model Era

Participants identified several key ways in which continual learning in the foundation model era differs from classical settings. First, starting from large pretrained models changes the plasticity–stability trade-off: the initial model already encodes vast knowledge. Second, high-quality generative models enable synthetic data and self-improvement loops reminiscent of generative replay. In addition, parameter-efficient methods (e.g., LoRA) redefine how adaptation is performed. Third, evaluation becomes significantly harder: forgetting is less visible, benchmarks are costly and incomplete, and new metrics are needed. Finally, the scientific context has changed: scale and financial constraints limit experimentation within the academic context, emergent properties and in-context learning alter adaptation dynamics, and real-world scenarios increasingly motivate CL.

4.2 Working Group on Memory in Continual Learning in the Foundation Model Era

In this workgroup, the discussion was about how to optimally employ memory during continual learning of foundation models. Especially the newly introduced technique of prompt learning provides efficient ways for fast adaptation. The discussion stressed the parallels with memory in neuroscience: an agent receives a stream of data and must decide what to retain, compress, or forget over time. In the age of foundation and agentic models, memory extends beyond stored data to include the neural network parameters, external mechanisms like retrieval-augmented generation (RAG), and token-based context handled via attention, each solving complementary problems. Key challenges include handling long and redundant contexts, deciding which information truly matters, and updating memory without supervision, possibly via compression or separate memory-management models.

This topic has been chosen by the particpants to aim to write a perspectives paper on in the near future.

4.3 Working Group on Forgetting in the Foundation Model Era

This working group focused on how to measure forgetting of foundation models during continual updates. Measuring catastrophic forgetting becomes especially difficult when updating foundation models due to the scale of their knowledge and the cost of evaluation.

In classical continual learning, forgetting is typically quantified by re-evaluating performance on previously seen tasks, but for foundation models this approach is often infeasible: comprehensive benchmarks are extremely expensive to run and only sparsely cover the breadth of capabilities such models encode. This motivates research on efficient evaluation strategies for continual adaptation of foundation models. Promising directions include reduced evaluation protocols that preserve broad capability coverage, as well as methods that infer forgetting from changes in model parameters or internal representations, enabling targeted assessment of vulnerable knowledge without requiring full-scale benchmark evaluation.

4.4 Working Group on Benchmarks, Setups, and Evaluation in the Foundation Model Era

Participants highlighted several open challenges around benchmarks, setups, and evaluation for continual learning with foundation models. A first issue is cost: evaluation can easily exceed fine-tuning itself, motivating a rethinking of benchmark design toward compute-efficient protocols, such as reduced task sequences, proxy or synthetic datasets. Second, while forgetting and adaptability remain core challenges, their relative importance is application-dependent, and different forms of forgetting (e.g., factual knowledge, reasoning ability, or core behaviors) may matter unequally; however, systematic large-scale evaluations comparing methods are still largely missing. Third, participants emphasized grounding continual learning in concrete applications – such as unlearning, personalization, deepfake detection, or tool use – to clarify objectives and expose real-world failure modes. Overall, the group argued that current benchmarks are often artificial and insufficient and new benchmarks need to be designed.

Participants

- Rahaf Aljundi
Toyota Motor Europe – Zaventem, BE
- Andrew D. Bagdanov
University of Florence, IT
- Lucas Caccia
Microsoft Research – Montréal, CA
- Antonio Carta
University of Pisa, IT
- Laurent Charlin
HEC Montréal, CA
- Jonghyun Choi
Seoul National University, KR
- Barbara Hammer
Universität Bielefeld, DE
- Tyler Hayes
Georgia Institute of Technology – Atlanta, US
- Timm Felix Hess
KU Leuven, BE
- Christopher Kanan
University of Rochester, US
- Dhireesha Kudithipudi
University of Texas – San Antonio, US
- Xialei Liu
Nankai University – Tianjin, CN
- Vincenzo Lomonaco
University of Pisa, IT
- Jorge Mendez-Mendez
Stony Brook University, US
- Martin Mundt
Universität Bremen, DE
- Darshan Patil
MILA – Montreal, CA
- Ameya Prabhu
Universität Tübingen, DE
- Elisa Ricci
University of Trento, IT
- Tinne Tuytelaars
KU Leuven, BE
- Bartlomiej Twardowski
Computer Vision Center – Barcelona, ES
- Gido van de Ven
University of Groningen, NL
- Joost van de Weijer
Computer Vision Center – Barcelona, ES
- Liyuan Wang
Tsinghua University – Beijing, CN