

Data, Responsibly

Edited by

Serge Abiteboul¹, Gerome Miklau², Julia Stoyanovich³, and
Gerhard Weikum⁴

- 1 ENS – Cachan, FR, serge.abiteboul@inria.fr
- 2 University of Massachusetts – Amherst, US, miklau@cs.umass.edu
- 3 Drexel University – Philadelphia, US, stoyanovich@drexel.edu
- 4 MPI für Informatik – Saarbrücken, DE, weikum@mpi-inf.mpg.de

Abstract

Big data technology promises to improve people’s lives, accelerate scientific discovery and innovation, and bring about positive societal change. Yet, if not used responsibly, large-scale data analysis and data-driven algorithmic decision-making can increase economic inequality, affirm systemic bias, and even destabilize global markets.

While the potential benefits of data analysis techniques are well accepted, the importance of using them responsibly – that is, in accordance with ethical and moral norms, and with legal and policy considerations – is not yet part of the mainstream research agenda in computer science.

Dagstuhl Seminar “Data, Responsibly” brought together academic and industry researchers from several areas of computer science, including a broad representation of data management, but also data mining, security/privacy, and computer networks, as well as social sciences researchers, data journalists, and those active in government think-tanks and policy initiatives. The goals of the seminar were to assess the state of data analysis in terms of fairness, transparency and diversity, identify new research challenges, and derive an agenda for computer science research and education efforts in responsible data analysis and use. While the topic of the seminar is transdisciplinary in nature, an important goal of the seminar was to identify opportunities for high-impact contributions to this important emergent area specifically from the data management community.

Seminar July 17–22, 2016 – <http://www.dagstuhl.de/16291>

1998 ACM Subject Classification H.2.8 Database Applications, H.3.3 Information Search and Retrieval, K.4.1 Public Policy Issues, K.6.5 Security and Protection

Keywords and phrases Data responsibly, Big data, Machine bias, Data analysis, Data management, Data mining, Fairness, Diversity, Accountability, Transparency, Personal information management, Ethics, Responsible research, Responsible innovation, Data science education

Digital Object Identifier 10.4230/DagRep.6.7.42



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license

Data, Responsibly, *Dagstuhl Reports*, Vol. 6, Issue 7, pp. 42–71

Editors: Serge Abiteboul, Gerome Miklau, Julia Stoyanovich, and Gerhard Weikum



DAGSTUHL
REPORTS

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive summary

Serge Abiteboul

Gerome Miklau

Julia Stoyanovich

Gerhard Weikum

License © Creative Commons BY 3.0 Unported license
© Serge Abiteboul, Gerome Miklau, Julia Stoyanovich, and Gerhard Weikum

Our society is data-driven. Large scale data analysis, known as Big data, is distinctly present in the private lives of individuals, is a dominant force in commercial domains as varied as automatic manufacturing, e-commerce and personalized medicine, and assists in – or fully automates – decision making in the public and private sectors. Data-driven algorithms are used in criminal sentencing – ruling who goes free and who remains behind bars, in college admissions – granting or denying access to education, and in employment and credit decisions – offering or withholding economic opportunities.

The promise of Big data is to improve people’s lives, accelerate scientific discovery and innovation, and enable broader participation. Yet, if not used responsibly, Big data can increase economic inequality and affirm systemic bias, polarize rather than democratize, and deny opportunities rather than improve access. Worse yet, all this can be done in a way that is non-transparent and defies public scrutiny.

Big data impacts individuals, groups and society as a whole. Because of the central role played by this technology, it must be used *responsibly* – in accordance with the ethical and moral norms that govern our society, and adhering to the appropriate legal and policy frameworks. And as journalists [3], legal and policy scholars [1, 2] and governments [4, 5] are calling for algorithmic fairness and greater insight into data-driven algorithmic processes, there is an urgent need to define a broad and coordinated computer science research agenda in this area. The primary goal of the Dagstuhl Seminar “Data, Responsibly” was to make progress towards such an agenda.

The seminar brought together academic and industry researchers from several areas of computer science, including a broad representation of data management, but also data mining, security/privacy, and computer networks, as well as social sciences researchers, data journalists, and those active in government think-tanks and policy initiatives. The problem we aim to address is inherently transdisciplinary. For this reason, it was important to have input from policy and legal scholars, and to have representation from multiple areas within computer science. We were able to attract a mix of European, North American, and South American participants. Out of 39 participants, 10 were women.

Specific goals of the seminar were to:

- assess the state of data analysis in terms of fairness, transparency and diversity;
- identify new research challenges;
- develop an agenda for computer science research in responsible data analysis and use, with a particular focus on potential high-impact contributions from the data management community;
- solicit perspectives on the necessary education efforts, and on responsible research and innovation practices.

The seminar included technical talks and break-out sessions. Technical talks were organized into themes, which included fairness and diversity, transparency and accountability, tracking and transparency, personal information management, education, and responsible

research and innovation. Participants suggested topics for seven working groups, which met over one or multiple days.

The organizers felt that the seminar was very successful – ideas were exchanged, discussions were lively and insightful, and we are aware of several collaborations that were started as a result of the seminar. The participants and the organizers all felt that the topic of the seminar is broad, fast moving and extremely important, and that it would be beneficial to hold another seminar on this topic in the near future.

Details about the program are contained in the remainder of this document.

References

- 1 Kate Crawford. Artificial Intelligence's White Guy Problem. The New York Times, June 25, 2016.
- 2 Kate Crawford and Ryan Calo. There is a blind spot in AI research. Nature / Comment 538(7625), October 13, 2016.
- 3 Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. Machine Bias. ProPublica, May 23, 2016.
- 4 Executive Office of the President, The White House. Big Data: Seizing Opportunities, Preserving Values. May 2014
- 5 Parliament and Council of the European Union. General Data Protection Regulation. 2016

2 Table of Contents

Executive summary

Serge Abiteboul, Gerome Miklau, Julia Stoyanovich, and Gerhard Weikum 43

Motivation and overview

Data, Responsibly: An Overview
Julia Stoyanovich, Serge Abiteboul, and Gerome Miklau 48

Big Data's Disparate Impact
Solon Barocas 48

Fairness and diversity

Fairness Through Awareness and Learning Fair Representations: A Tutorial
Michael Hay 49

An Axiomatic Framework for Fairness
Suresh Venkatasubramanian, Carlos Scheidegger, and Sorelle Friedler 50

What is Fairness Anyway? Interdisciplinary Concepts and Data Science
Bettina Berendt 50

Segregation Discovery
Salvatore Ruggieri 51

Diversity: Why, What, How
Evaggelia Pitoura and Marina Drosou 51

Transparency and accountability

Accountable Algorithms
Solon Barocas 52

Algorithmic Accountability and Transparency
Nicholas Diakopoulos 53

Auditing Black-box Models
Sorelle Friedler 54

Revealing Algorithmic Rankers
Gerome Miklau and Julia Stoyanovich 54

Computational Fact Checking
Cong Yu 55

Tracking and transparency

Online Tracking and Transparency
Claude Castelluccia 55

Tracing Information Flows Between Ad Exchanges Using Retargeted Ads
Christo Wilson 56

Quantifying Search Engine Bias
Krishna P. Gummadi 57

Seeing through Website Privacy Policies
Rishiraj Saha Roy 57

Collect it All: Why Bulk Surveillance Works <i>Nicholas Weaver</i>	58
Tracking, Targeting, Rating, Discriminating based on Social Media: Risk Measures for User Guidance <i>Gerhard Weikum</i>	58
Personal information management	
Managing your Personal Information <i>Serge Abiteboul and Amélie Marian</i>	59
Small Data Metadata <i>Arnaud Sahuguet</i>	59
Empowering Personal Data Management using Secure Hardware <i>Benjamin Nguyen</i>	59
Education, responsible research and innovation	
Science Data, Responsibly <i>Bill Howe</i>	60
Research and Education in Data Science and Responsible Use: Challenges and Opportunities <i>Chaitanya Baru</i>	61
Sustainability Research: Promoting Transparency and Accountability for Decision Makers <i>Claudia Bauzer Medeiros</i>	61
Practising Responsible Data Practices through Data Ethics Education <i>H. V. Jagadish</i>	61
Values, Algorithm Design, and Collaboration <i>Kristene Unsworth</i>	62
Privacy, Transparency and Education <i>Gerald Friedland</i>	62
Teaching Ethical Issues in Data Mining to Undergraduates <i>Sorelle Friedler</i>	64
Networked Systems Ethics <i>Ben Zevenbergen</i>	64
Lightning talks	
Benchmarking for (Linked) Data Management <i>Irini Fundulaki</i>	65
From Three Laws of Robotics to Five Principles of Big Data? <i>Wolfgang Nejdl</i>	65
Natural Language Processing, Responsibly <i>Jannik Strötgen</i>	65

Working groups

Structural Bias <i>Solon Barocas, Bettina Berendt, Michael Hay, Amélie Marian, and Gerome Miklau</i>	66
Avoid Reinventing the Wheel <i>Bettina Berendt, Solon Barocas, Claude Castelluccia, Pauli Miettinen, Wolfgang Nejd, Salvatore Ruggieri, and Jannik Strötgen</i>	67
Dynamics and Feedback in Discrimination Processes <i>Krishna P. Gummadi, Chaitanya Baru, Marina Drosou, Salvatore Ruggieri, Rishiraj Saha Roy, Jannik Strötgen, and Suresh Venkatasubramanian</i>	67
Principles for Accountable Algorithms <i>Nicholas Diakopoulos and Sorelle Friedler</i>	68
Data, Responsibly: Business and Research Opportunities <i>Julia Stoyanovich, Serge Abiteboul, Chaitanya Baru, Sorelle Friedler, Krishna P. Gummadi, Michael Hay, Bill Howe, Benny Kimelfeld, Arnaud Sahuguet, Eric Simon, Suresh Venkatasubramanian, and Gerhard Weikum</i>	69
Explaining Decisions <i>Julia Stoyanovich, Chaitanya Baru, Claudia Bauzer Medeiros, Krishna P. Gummadi, Bill Howe, Arnaud Sahuguet, Jan Van den Bussche, and Gerhard Weikum</i>	69
Data, Responsibly: Use Cases and Benchmarking <i>Suresh Venkatasubramanian, Claudia Bauzer Medeiros, Gerald Friedland, Irini Fundulaki, and Salvatore Ruggieri</i>	70
Participants	71

3 Motivation and overview

3.1 Data, Responsibly: An Overview

Julia Stoyanovich (Drexel University – Philadelphia, US), Serge Abiteboul (ENS – Cachan, FR), and Gerome Miklau (University of Massachusetts – Amherst, US)

License © Creative Commons BY 3.0 Unported license

© Julia Stoyanovich, Serge Abiteboul, and Gerome Miklau

Main reference J. Stoyanovich, S. Abiteboul, G. Miklau, “Data Responsibly: Fairness, Neutrality and Transparency in Data Analysis (Tutorial),” in Proc. of the 19th Int’l Conf. on Extending Database Technology (EDBT’16), pp. 718–719, OpenProceedings.org, 2016.

URL <http://dx.doi.org/10.5441/002/edbt.2016.103>

The first talk of this seminar was a tutorial that surveyed dimensions of responsible data analysis and use, and set the stage for the technical talks and discussions. This presentation was based in part on a recent EDBT tutorial [1].

References

- 1 J. Stoyanovich, S. Abiteboul, G. Miklau: *Data Responsibly: Fairness, Neutrality and Transparency in Data Analysis, Tutorial*, EDBT 2016.

3.2 Big Data’s Disparate Impact

Solon Barocas (Microsoft – New York, US)

License © Creative Commons BY 3.0 Unported license

© Solon Barocas

Joint work of Barocas, Solon; Selbst, Andrew

Main reference S. Barocas, A. Selbst, “Big Data’s Disparate Impact”, *California Law Review*, Vol. 104, no. 3 (June 2016), pp. 671–732, 2016.

URL <http://dx.doi.org/10.15779/Z38BG31>

Advocates of algorithmic techniques like data mining argue that these techniques eliminate human biases from the decision-making process. But an algorithm is only as good as the data it works with. Data is frequently imperfect in ways that allow these algorithms to inherit the prejudices of prior decision makers. In other cases, data may simply reflect the widespread biases that persist in society at large. In still others, data mining can discover surprisingly useful regularities that are really just preexisting patterns of exclusion and inequality. Unthinking reliance on data mining can deny historically disadvantaged and vulnerable groups full participation in society. Worse still, because the resulting discrimination is almost always an unintentional emergent property of the algorithm’s use rather than a conscious choice by its programmers, it can be unusually hard to identify the source of the problem or to explain it to a court.

This talk examines these concerns through the lens of American antidiscrimination law – more particularly, through Title VII’s prohibition of discrimination in employment. In the absence of a demonstrable intent to discriminate, the best doctrinal hope for data mining’s victims would seem to lie in disparate impact doctrine. Case law and the Equal Employment Opportunity Commission’s Uniform Guidelines, though, hold that a practice can be justified as a business necessity when its outcomes are predictive of future employment outcomes, and data mining is specifically designed to find such statistical correlations. Unless there is a reasonably practical way to demonstrate that these discoveries are spurious, Title VII would appear to bless its use, even though the correlations it discovers will often reflect historic

patterns of prejudice, others' discrimination against members of protected groups, or flaws in the underlying data

Addressing the sources of this unintentional discrimination and remedying the corresponding deficiencies in the law will be difficult technically, difficult legally, and difficult politically. There are a number of practical limits to what can be accomplished computationally. For example, when discrimination occurs because the data being mined is itself a result of past intentional discrimination, there is frequently no obvious method to adjust historical data to rid it of this taint. Corrective measures that alter the results of the data mining after it is complete would tread on legally and politically disputed terrain. These challenges for reform throw into stark relief the tension between the two major theories underlying antidiscrimination law: anticlassification and antisubordination. Finding a solution to big data's disparate impact will require more than best efforts to stamp out prejudice and bias; it will require a wholesale reexamination of the meanings of "discrimination" and "fairness".

4 Fairness and diversity

4.1 Fairness Through Awareness and Learning Fair Representations: A Tutorial

Michael Hay (Colgate University – Hamilton, US)

License  Creative Commons BY 3.0 Unported license
© Michael Hay

This talk is a tutorial on two recent works on the problem of fairness in classification. The first work, Fairness Through Awareness [1], offers a framework for fair classification that is based on the principle that individuals who are similar for the purpose of the classification task should be treated similarly, and presents an algorithm for maximizing utility subject to the fairness constraint. The second work, Learning Fair Representations [2], formulates fairness as an optimization problem of finding a representation of the data that encodes the data as well as possible while obfuscating information that must be obscured to achieve fairness. The tutorial highlights themes such as individual vs. group fairness, fairness through awareness vs. obfuscation, and formal frameworks vs. empirical assessments.

References

- 1 Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel Fairness through awareness. In Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, pp. 214–226. ACM, 2012, <http://dx.doi.org/10.1145/2090236.2090255>
- 2 Richard Zemel, Yu Wu, Kevin Swersky, Toniann Pitassi, and Cynthia Dwork Learning Fair Representations. In Proceedings of the International Conference on Machine Learning, pp. 325–333. 2013

4.2 An Axiomatic Framework for Fairness

Suresh Venkatasubramanian (University of Utah – Salt Lake City, US), Carlos Scheidegger, and Sorelle Friedler (Haverford College, US)

License © Creative Commons BY 3.0 Unported license

© Suresh Venkatasubramanian, Carlos Scheidegger, and Sorelle Friedler

Joint work of S. A. Friedler, C. Scheidegger, S. Venkatasubramanian

Main reference S. A. Friedler, C. Scheidegger, S. Venkatasubramanian, “On the (im)possibility of fairness,” arXiv:1609.07236v1 [cs.CY], 2016.

URL <https://arxiv.org/abs/1609.07236v1>

What does it mean for an algorithm to be fair? Different papers use different notions of algorithmic fairness, and although these appear internally consistent, they also seem mutually incompatible. We present a mathematical setting in which the distinctions in previous papers can be made formal. In addition to characterizing the spaces of inputs (the “observed” space) and outputs (the “decision” space), we introduce the notion of a construct space: a space that captures unobservable, but meaningful variables for the prediction.

We show that in order to prove desirable properties of the entire decision-making process, different mechanisms for fairness require different assumptions about the nature of the mapping from construct space to decision space. The results in this work imply that future treatments of algorithmic fairness should more explicitly state assumptions about the relationship between constructs and observations.

4.3 What is Fairness Anyway? Interdisciplinary Concepts and Data Science

Bettina Berendt (KU Leuven, BE)

License © Creative Commons BY 3.0 Unported license

© Bettina Berendt

Main reference B. Berendt, S. Preibusch, “Better decision support through exploratory discrimination-aware data mining: foundations and empirical evidence,” *Artif. Intell. Law*, 22(2): 175–209, 2014.

URL <http://dx.doi.org/10.1007/s10506-013-9152-0>

Main reference B. Berendt, M. Büchler, G. Rockwell, “Is it Research or is it Spying? Thinking-Through Ethics in Big Data AI and Other Knowledge Sciences,” *KI*, 29(2):223–232, 2015.

URL <http://dx.doi.org/10.1007/s13218-015-0355-2>

It is by now a truism that data mining algorithms “discriminate”. Whether we consider any particular criterion or effect to be a differentiation or an undesirable (e.g. unlawful) “discrimination in the narrow sense”, is a second question. Similarly, one needs to investigate the non-identical notions of “non-discrimination” and “fairness”.

Thus, conceptual issues arise even at the start of any process of being more responsible about data, and addressing them requires an interdisciplinary approach. Still, no data mining algorithm by itself is discriminatory – it can only become so when deployed in a context. This talk builds on our work on investigating discrimination-aware data mining (DADM) in contexts that involve human decision makers. I give a brief overview of our proposal of an interactive, exploratory DADM and an empirical study we did of such decisions. I then present five challenges that cannot be tackled by today’s formalisms for DADM or “fairness-aware data mining”: vicious cycles of one form of discrimination leading into another, the question of how to translate the Aristotelian principle of equality into a data framework, intersectionality and the emergence of new concepts, the perpetuation of pernicious concepts and how it gets baked into data-based decision making, and the search for causes. I conclude with an outlook on next-generation tools and research approaches.

The slides of the talk are available at [1].

References

- 1 Bettina Berendt. *What is fairness anyway? Interdisciplinary concepts and data science*. Presentation at the Dagstuhl Seminar “Data, responsibly”. Dagstuhl, 18 July 2016. https://people.cs.kuleuven.be/~bettina.berendt/Talks/berendt_2016_07_18.pptx

4.4 Segregation Discovery

Salvatore Ruggieri (University of Pisa, IT)

License © Creative Commons BY 3.0 Unported license
© Salvatore Ruggieri

Joint work of Baroni, Alessandro

Main reference A. Baroni, S. Ruggieri, “Segregation Discovery in a Social Network of Companies,” in Proc. of the 14th Int’l Symp. on Intelligent Data Analysis (IDA’15), LNCS, Vol. 9385, pp. 37–48, Springer, 2015.

URL http://dx.doi.org/10.1007/978-3-319-24465-5_4

The term segregation refers to restrictions on the access of people to each other. People are partitioned into two or more groups on the grounds of personal or cultural traits that can foster discrimination, such as gender, age, ethnicity, income, skin color, language, religion, political opinion, membership of a national minority, etc. Contact, communication, or interaction among groups are limited by their physical, working or socio-economic distance.

We introduce a framework for a data-driven analysis of segregation of minority groups in social networks, and challenge it on a complex scenario. The framework builds on quantitative measures of segregation, called segregation indexes, proposed in the social science literature. The segregation discovery problem consists of searching sub-graphs and sub-groups for which a reference segregation index is above a minimum threshold. A search algorithm is devised that solves the segregation problem based on frequent itemset mining. The framework is challenged on the analysis of segregation of social groups in the boards of directors of the real and large network of Italian companies connected through shared directors.

Relationships among segregation, discrimination, and diversity are also highlighted.

4.5 Diversity: Why, What, How

Evaggelia Pitoura (University of Ioannina, GR) and Marina Drosou (Hellenic Police – Athens, GR)

License © Creative Commons BY 3.0 Unported license
© Evaggelia Pitoura and Marina Drosou

In this talk, we first present a brief overview of data diversification. We discuss different diversification interpretations, namely, based on coverage, content dissimilarity and novelty, as well as, various algorithmic approaches. Then, we present our work on r-DisC diversification. r-DisC diversification locates diverse subsets of results in a way such that each item in the result is represented by a similar item in the diverse subset and the items in the diverse subset are dissimilar to each other. We also show various extensions of our basic model. Finally, we discuss some issues in social networks and opinion diversity, such as, homophily, opinion formation and fairness.

5 Transparency and accountability

5.1 Accountable Algorithms

Solon Barocas (Microsoft – New York, US)

License © Creative Commons BY 3.0 Unported license
© Solon Barocas

Joint work of Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, Harlan Yu

Main reference J. A. Kroll, J. Huey, S. Barocas, E. W. Felten, J. R. Reidenberg, D. G. Robinson, H. Yu, “Accountable Algorithms”, University of Pennsylvania Law Review, forthcoming; pre-print available at SSRN.

URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2765268

Many important decisions historically made by people are now made by computers. Algorithms count votes, approve loan and credit card applications, target citizens or neighborhoods for police scrutiny, select taxpayers for an IRS audit, and grant or deny immigration visas.

The accountability mechanisms and legal standards that govern such decision processes have not kept pace with technology. The tools currently available to policymakers, legislators, and courts were developed to oversee human decision-makers and often fail when applied to computers instead: for example, how do you judge the intent of a piece of software? Additional approaches are needed to make automated decision systems – with their potentially incorrect, unjustified or unfair results – accountable and governable. This Article reveals a new technological toolkit to verify that automated decisions comply with key standards of legal fairness.

We challenge the dominant position in the legal literature that transparency will solve these problems. Disclosure of source code is often neither necessary (because of alternative techniques from computer science) nor sufficient (because of the complexity of code) to demonstrate the fairness of a process. Furthermore, transparency may be undesirable, such as when it permits tax cheats or terrorists to game the systems determining audits or security screening.

The central issue is how to assure the interests of citizens, and society as a whole, in making these processes more accountable. This Article argues that technology is creating new opportunities – more subtle and flexible than total transparency – to design decision-making algorithms so that they better align with legal and policy objectives. Doing so will improve not only the current governance of algorithms, but also – in certain cases – the governance of decision-making in general. The implicit (or explicit) biases of human decision-makers can be difficult to find and root out, but we can peer into the “brain” of an algorithm: computational processes and purpose specifications can be declared prior to use and verified afterwards.

The technological tools introduced in this Article apply widely. They can be used in designing decision-making processes from both the private and public sectors, and they can be tailored to verify different characteristics as desired by decision-makers, regulators, or the public. By forcing a more careful consideration of the effects of decision rules, they also engender policy discussions and closer looks at legal standards. As such, these tools have far-reaching implications throughout law and society.

Part I of this Article provides an accessible and concise introduction to foundational computer science concepts that can be used to verify and demonstrate compliance with key standards of legal fairness for automated decisions without revealing key attributes of the decision or the process by which the decision was reached. Part II then describes how

these techniques can assure that decisions are made with the key governance attribute of procedural regularity, meaning that decisions are made under an announced set of rules consistently applied in each case. We demonstrate how this approach could be used to redesign and resolve issues with the State Department's diversity visa lottery. In Part III, we go further and explore how other computational techniques can assure that automated decisions preserve fidelity to substantive legal and policy choices. We show how these tools may be used to assure that certain kinds of unjust discrimination are avoided and that automated decision processes behave in ways that comport with the social or legal standards that govern the decision. We also show how algorithmic decision-making may even complicate existing doctrines of disparate treatment and disparate impact, and we discuss some recent computer science work on detecting and removing discrimination in algorithms, especially in the context of big data and machine learning. And lastly in Part IV, we propose an agenda to further synergistic collaboration between computer science, law and policy to advance the design of automated decision processes for accountability.

5.2 Algorithmic Accountability and Transparency

Nicholas Diakopoulos (University of Maryland – College Park, US)

License © Creative Commons BY 3.0 Unported license
© Nicholas Diakopoulos

Main reference N. Diakopoulos, "Accountability in Algorithmic Decision Making," *Communications of the ACM*, 59(2):56–62, ACM, 2016.

URL <http://dx.doi.org/10.1145/2844110>

As journalism shifts into the 21st century, the opportunities for reinventing the ways that news stories are found and told using computing and data are practically endless. But perhaps even more substantial are the new ways in which computation and algorithms are coming to adjudicate decisions in nearly all facets of industry and government. Algorithmic accountability reporting is a new form of computational journalism that is emerging to apply the core journalistic functions of watchdogging and investigative reporting to algorithms. In this talk I will discuss how algorithmic accountability reporting is used by journalists as a method for articulating the power structures, biases, and influences that computational artifacts play in society. I will trace various legal, technical, and regulatory challenges that remain, offering new openings for the development of tools. Finally, I will discuss the mandate for transparency of algorithms and proffer for discussion an initial transparency standard that delineates the dimensions of algorithms in use by industry or government that might be disclosed.

5.3 Auditing Black-box Models

Sorelle Friedler (Haverford College, US)

License © Creative Commons BY 3.0 Unported license
© Sorelle Friedler

Joint work of Philip Adler, Casey Falk, Sorelle A. Friedler, Gabriel Rybeck, Carlos Scheidegger, Brandon Smith, Suresh Venkatasubramanian

Main reference P. Adler, C. Falk, S. A. Friedler, G. Rybeck, C. Scheidegger, B. Smith, S. Venkatasubramanian, “Auditing Black-box Models by Obscuring Features,” arXiv:1602.07043v1 [stat.ML], 2016.

URL <http://arxiv.org/abs/1602.07043v1>

Data-trained predictive models see widespread use, but for the most part they are used as black boxes which output a prediction or score. It is therefore hard to acquire a deeper understanding of model behavior, and in particular how different features influence the model prediction. This is important when interpreting the behavior of complex models, or asserting that certain problematic attributes (like race or gender) are not unduly influencing decisions.

In this talk, I present a technique for auditing black-box models, which lets us study the extent to which existing models take advantage of particular features in the dataset, without knowing how the models work. Our work focuses on the problem of indirect influence: how some features might indirectly influence outcomes via other, related features. As a result, we can find attribute influences even in cases where, upon further direct examination of the model, the attribute is not referred to by the model at all.

Our approach does not require the black-box model to be retrained. This is important if (for example) the model is only accessible via an API, and contrasts our work with other methods that investigate feature influence like feature selection. We present experimental evidence for the effectiveness of our procedure using a variety of publicly available datasets and models. Not presented, we also validate our procedure using techniques from interpretable learning and feature selection, as well as against other black-box auditing procedures.

5.4 Revealing Algorithmic Rankers

Gerome Miklau (University of Massachusetts – Amherst, US) and Julia Stoyanovich (Drexel University – Philadelphia, US)

License © Creative Commons BY 3.0 Unported license
© Gerome Miklau and Julia Stoyanovich

Joint work of Julia Stoyanovich, Gerome Miklau, Ellen P. Goodman

Main reference J. Stoyanovich, E. P. Goodman, “Revealing Algorithmic Rankers,” Freedom to Tinker, Nov. 2016.

URL <https://freedom-to-tinker.com/blog/jstoyanovich/revealing-algorithmic-rankers/>

ProPublica’s story on “machine bias” in an algorithm used for sentencing defendants amplified calls to make algorithms more transparent and accountable. It has never been more clear that algorithms are political and embody contested choices, and that these choices are largely obscured from public scrutiny. We see it in controversies over Facebook’s newsfeed, or Google search results, or Twitter’s trending topics. Policymakers are considering how to operationalize “algorithmic ethics” and scholars are calling for accountable algorithms.

One kind of algorithm that is at once especially obscure, powerful, and common is the ranking algorithm. Algorithms rank individuals to determine credit worthiness, desirability for college admissions and employment, and compatibility as dating partners. They encode ideas of what counts as the best schools, neighborhoods, and technologies. Despite their importance, we actually can know very little about why this person was ranked higher than another in a dating app, or why this school has a better rank than that one. This is true

even if we have access to the ranking algorithm, for example, if we have complete knowledge about the factors used by the ranker and their relative weights, as is the case for US News ranking of colleges. In this blog post, we argue that syntactic transparency, wherein the rules of operation of an algorithm are more or less apparent, or even fully disclosed, still leaves stakeholders in the dark: those who are ranked, those who use the rankings, and the public whose world the rankings may shape.

In this talk we discuss the reasons for opacity in ranking algorithms, and the corresponding harms that this opacity brings. We give examples of these issues in using rankings of US colleges and academic departments. We go on to outline directions for future work that would make rankings interpretable.

5.5 Computational Fact Checking

Cong Yu (Google – New York, US)

License © Creative Commons BY 3.0 Unported license
© Cong Yu

Joint work of Bill Adair, Chengkai Li, Jun Yang, Cong Yu

This talk describes the process through which reporters perform fact checking on statements made by politicians and the various challenges facing the reporters. The three main stages in the process are finding the claims, checking the claims, and distributing the reviews. For each stage, I illustrate some recent works that aim at computationally assisting the reporters, as well as more technical challenges to be addressed for the ultimate holy grail of automatic fact checking. This talk is also a call-for-action for database and algorithm researchers to work in this socially important area.

6 Tracking and transparency

6.1 Online Tracking and Transparency

Claude Castelluccia (INRIA – Grenoble, FR)

License © Creative Commons BY 3.0 Unported license
© Claude Castelluccia

URL <https://myrealonlinechoices.inrialpes.fr/>

In the last few years, as a result of the proliferation of intrusive and privacy-invading ads, the use of ad-blockers and anti-tracking tools have become widespread. As of the second quarter of this year, 16% of online Americans, about 45 million people, had installed ad-blocking software, according to PageFair 2015 report. Meanwhile, 77 millions Europeans are blocking ads. All this accounts globally for \$21.8 billion worth of blocked ads. The Internet economy is in danger since ads fuel the free content and services over Internet.

We believe that Adblockers are only a short-term solution, and that better tools are necessary to solve this problem in the long term. Most users are not against ads and are actually willing to accept some ads to help websites. However, users want more control and transparency about the ads that they want to receive, and about the way they are tracked and profiled on the Internet.

As opposed to existing ad blockers that take a binary approach (i.e., block everything if you install them or block nothing otherwise), MyRealOnlineChoices project aims to provide

users with the right tools that allow users to make fine-grained choices about their privacy and the ads that they want to receive. Our tools allow users to choose on which sites (more specifically, on which categories of sites) they want to block the trackers. For example, a user can choose to block the trackers on sites related to health or religion, but may choose not to block the trackers on sites related to sports or news. Similarly, a user might want to block ads that are targeted on some categories that he considers sensitive. Our tools provide this type of control to the users.

As trust between users and online entities is the key here, our project starts with transparency as the first key feature. We need tools that provide more transparency to users by indicating if an ad is retargeted or is delivered to users based on their interests or not. The ultimate goal is to enforce the user choices while sustaining the ad economy of the Internet. And thanks to this transparency feature, users can be aware of what is going on with their browsing data, and therefore can make an informed decision.

6.2 Tracing Information Flows Between Ad Exchanges Using Retargeted Ads

Christo Wilson

License  Creative Commons BY 3.0 Unported license
© Christo Wilson

Joint work of Muhammad Ahmad Bashir, Sajjad Arshad, William Robertson, Christo Wilson

Main reference M. A. Bashir, S. Arshad, W. Robertson, C. Wilson, “Tracing Information Flows Between Ad Exchanges Using Retargeted Ads,” in Proc. of the 25th USENIX Security Symp., pp. 481–496, USENIX Association, 2016.

URL <https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/bashir>

Numerous surveys have shown that Web users are seriously concerned about the loss of privacy associated with online tracking. Alarming, these surveys also reveal that people are also unaware of the amount of data sharing that occurs between ad exchanges, and thus underestimate the privacy risks associated with online tracking.

In reality, the modern ad ecosystem is fueled by a flow of user data between trackers and ad exchanges. Although recent work has shown that ad exchanges routinely perform cookie matching with other exchanges, these studies are based on brittle heuristics that cannot detect all forms of information sharing, especially under adversarial conditions.

In this study, we develop a methodology that is able to detect client- and server-side flows of information between arbitrary ad exchanges. Our key insight is to leverage retargeted ads as a mechanism for identifying information flows. Intuitively, our methodology works because it relies on the semantics of how exchanges serve ads, rather than focusing on specific cookie matching mechanisms. Using crawled data on 35,448 ad impressions, we show that our methodology can successfully categorize four different kinds of information sharing between ad exchanges, including cases where existing heuristic methods fail.

References

- 1 Muhammad Ahmad Bashir and Sajjad Arshad and William Robertson and Christo Wilson. *Tracing Information Flows Between Ad Exchanges Using Retargeted Ads*. In Proceedings of Usenix Security, Austin TX, August 2016.

6.3 Quantifying Search Engine Bias

Krishna P. Gummadi (MPI-SWS – Saarbrücken, DE)

License © Creative Commons BY 3.0 Unported license
© Krishna P. Gummadi

Joint work of Juhi Kulshrestha, Motahhareh Eslami, Saptarshi Ghosh, Krishna P. Gummadi, Karrie Karahalios
Main reference J. Kulshrestha, M. Eslami, J. Messias, M. B. Zafar, S. Ghosh, K. P. Gummadi, K. Karahalios, “Quantifying Search Bias: Investigating Sources of Bias for Political Searches in Social Media,” to appear in Proc. of the 20th ACM Conf. on Computer-Supported Cooperative Work and Social Computing(CSCW’17); pre-print available from author’s webpage.
URL www.mpi-sws.org/juhi/search-bias-cscw-2017-pre-print.pdf

Search systems in online social media sites are frequently used to find information about ongoing events and people. For topics with multiple competing perspectives, such as political events or political candidates, bias in the top ranked results significantly shapes public opinion. However, bias does not emerge from an algorithm alone. It is important to distinguish between the bias that arises from the data that serves as the input to the ranking algorithm and the bias that arises from the ranking algorithm itself. In this talk, I will propose a framework to quantify these distinct biases and apply this framework to politics-related queries on Twitter. We found that both the input data and the ranking algorithm contribute significantly to produce varying amounts of bias in the search results and in different ways. I will discuss the consequences of these biases and propose mechanisms to signal this bias in social search systems interfaces.

6.4 Seeing through Website Privacy Policies

Rishiraj Saha Roy (MPI für Informatik – Saarbrücken, DE)

License © Creative Commons BY 3.0 Unported license
© Rishiraj Saha Roy

A number of online privacy concerns involving data sharing and unexpected ad recommendations can be mitigated if users understand website privacy policies better. Unfortunately these policies are usually very long since they contain legal documentation, and have to cover several corner cases. The short message on cookies that we get when we visit a new website inside Europe is also generally not enough. Wilson et al. (2016) state that the three key things that a user is really concerned about in a website’s privacy policy are whether it collects personal information, shares that personal information, and how easy it is to delete this personal information. The authors find that the average Web user is indeed able to find answers to most of these concerns inside the policies, and it is possible to automatically extract relevant excerpts from the long privacy policies using regular expressions and weighted term matching. With this knowledge, we can trivially show an extractive summarization of the privacy policy focused on the relevant aspects to the user instead of the short cookie message, and the list of trackers if any using tools like Ghostery. But going further, we can get the preferred privacy policy from the user as defined by answers to the questions earlier, and then try to automatically reason and find the likely answers. Using this knowledge, we can then check the compliance of the site’s policy to the user’s preference, and then suggest alternatives. An unobtrusive way of suggesting alternatives would be to show policy-compliant websites on the search results’ page itself, as users usually end up on these sites through a search engine. So, in a post-processing step, our privacy advisor can fetch the privacy policies of the top-k sites ($k = 5$ or 10 , say), and check the compliance statuses

before presenting to the user, and the user can accordingly choose his/her preferred site(s). This specific use case is applicable of hundreds of usual websites that we visit for day-to-day information like flight fare and insurance comparisons, but not really to big players like Google, Facebook or Amazon. There are quite a few research challenges here: sometimes the clarification on a privacy aspect is unclear even to humans, and sometimes there are disagreements between average Web users and legal experts. The automated reasoning about policy compliance using NLP techniques is non-trivial, and finally not every visit to these websites is through a search engine, so we have to figure out how best to show relevant alternatives in such cases.

6.5 Collect it All: Why Bulk Surveillance Works

Nicholas Weaver (ICSI – Berkeley, US)

License  Creative Commons BY 3.0 Unported license
© Nicholas Weaver


Joint work of NSA

A big driving force behind the “collect it all” mentality is that bulk surveillance techniques, both for the private sector and intelligence services, both works and is economically feasible. When developing a surveillance system, no matter the purpose, there is usually a requirement to have the capability to target anyone. As a corollary there is also a huge desire for retrospective capabilities, since the data collector doesn’t know until tomorrow what he wished to save today. This leads to architectures which effectively collect data on everybody and then, subsequently, select for the actual information of interest.

Such architectures are also remarkably affordable and, thanks to modern big data techniques, scale linearly. Ranging in size from just a couple of racks to a large facility, its straightforward to match both the data budget and computational budget needed to collect, retain, and search information on everybody. \$100M in hardware could maintain a 10MB dossier on everybody on the planet, It may be the case that "collect it all" is simply unstoppable because it is both effective and affordable.

6.6 Tracking, Targeting, Rating, Discriminating based on Social Media: Risk Measures for User Guidance

Gerhard Weikum (MPI für Informatik – Saarbrücken, DE)

License  Creative Commons BY 3.0 Unported license
© Gerhard Weikum

Joint work of Joanna Biega, Krishna Gummadi, Gerhard Weikum

In this talk I introduce the R-Susceptibility model which captures sensitive topics in online communities and the exposure or risk of individual users incurred by their posts. Topics are captured by latent embeddings, and sensitive topics are identified by crowdsourcing. A user’s risk is quantified by distance measures between the user’s post or entire posting history and the topic of interest. The R-Susceptibility model allows ranking users with regard to an adversary that targets the top-k users on some sensitive topic. Based on these methods, we envision a personalized tool that can alert users, explain risks and possible countermeasures, and guides users.

7 Personal information management

7.1 Managing your Personal Information

Serge Abiteboul (ENS – Cachan, FR) and Amélie Marian (Rutgers University – Piscataway, US)

License © Creative Commons BY 3.0 Unported license
© Serge Abiteboul and Amélie Marian

Personal information is constantly produced and stored by a large number of sources and services. While in the past, users would store all their data in physical form or on their local machines, the advent of the cloud has made this impossible. Personal information is fragmented in multiple heterogeneous systems, making it difficult to access, control, and exploit for personal use. In this talk, we make the case for the need for Personal Information Management Systems (PIMS), (cloud-based) systems that manages all the information of a person. Recent technological advances and societal pressures are enabling new interesting applications for PIMS. We discuss a subset of these applications: data integration, personal search, and personal knowledge management, and their potential to improve users' lives by giving them back control over their own information.

7.2 Small Data Metadata

Arnaud Sahuguet (Cornell Tech NYC, US)

License © Creative Commons BY 3.0 Unported license
© Arnaud Sahuguet

Main reference Arnaud Sahuguet, “Small Data Metadata,” 2016.

URL <https://medium.com/@sahuguet/small-data-metadata-1ff922fa6d14#.rmzh2c9kh>

Small data are the digital traces that individuals generate as a byproduct of daily activities, such as sending e-mail or exercising with fitness trackers.

We advocate for the need to standardize metadata about small data and build tools to create, manage, process and reason about it. Not only is small data metadata critical to foster the small data ecosystem of consumers and producers, it is also essential to offer some necessary guarantees – like privacy – inherent to the data itself.

7.3 Empowering Personal Data Management using Secure Hardware

Benjamin Nguyen (INSA – Bourges, FR)

License © Creative Commons BY 3.0 Unported license
© Benjamin Nguyen

Joint work of Nicolas Anciaux, Philippe Bonnet, Luc Bouganim, Benjamin Nguyen, Philippe Pucheral, Iulian Sandu-Popa

Main reference N. Anciaux, P. Bonnet, L. Bouganim, B. Nguyen, I. S. Popa, P. Pucheral, “Trusted Cells: A Sea Change for Personal Data Services,” in Proc. of the 6th Biennial Conf. on Innovative Data Systems Research (CIDR'13), 2013.

URL http://cidrdb.org/cidr2013/Papers/CIDR13_Paper68.pdf

How do you keep a secret about your personal life in an age where your daughter's glasses record and share everything she senses, your wallet records and shares your financial transactions, and your set-top box records and shares your family's energy consumption? Your

personal data has become a prime asset for many companies around the Internet, but can you avoid – or even detect – abusive usage?

Today, there is a wide consensus that individuals should have increased control on how their personal data is collected, managed and shared. Yet there is no appropriate technical solution to implement such personal data services: centralized solutions sacrifice security for innovative applications, while decentralized solutions sacrifice innovative applications for security. In this presentation, we argue that the advent of secure hardware in all personal IT devices, at the edges of the Internet, could trigger a sea change.

We introduce PlugDB, a personal data server running on a secure portable tokens which forms a global, decentralized data platform that provides security yet enables innovative applications. We describe this platform, called asymmetric architecture, because it is composed on the one hand of a large number of low power, low availability, high trust devices, and on the other hand high power, 24/7 but low to no trust cloud type infrastructure. Finally, we define a range of challenges for future research.


References

- 1 Cuong-Quoc To, Benjamin Nguyen, Philippe Pucheral. *Private and Scalable Execution of SQL Aggregates on a Secure Decentralized Architecture*, TODS, to appear, 2016.
- 2 Nicolas Anceaix, Saliha Lallali, Iulian Sandu Popa, Philippe Pucheral. *A Scalable Search Engine for Mass Storage Smart Objects*. VLDB 2015

8 Education, responsible research and innovation

8.1 Science Data, Responsibly

Bill Howe (University of Washington – Seattle, US)

License  Creative Commons BY 3.0 Unported license

© Bill Howe

Joint work of Maxim Gretchkin, Hoifung Poon, Poshen Lee


There is a reproducibility crisis in science: the number of retractions are increasing year to year, public trust is low, and a number of reproducibility studies across fields have shown dismal results. The incentive structures in science have increased pressure to achieve results at all costs, and new technology has made it easier to substitute exploratory research for controlled experiments.

We see two complementary solutions: In education, we advocate incorporating a rigorous ethics program into data science curricula, emphasizing case studies that do not readily admit technical solutions. In technology, we advocate systems research to enforce statistical checks, avoid multiple hypothesis testing issues, and ensure curation of public datasets.

As an example, we describe a project in computational curation that provides a first step toward automatic verification of scientific claims. Using a public repository of microarray data, we show that co-trained models on the human-provided metadata and the content of the dataset itself can significantly improve the quality of the labels over the state of the art methods, making thousands of new datasets available for reproducibility studies.

8.2 Research and Education in Data Science and Responsible Use: Challenges and Opportunities

Chaitanya Baru (NSF – Arlington, US)

License  Creative Commons BY 3.0 Unported license
© Chaitanya Baru

This talk will provide an overview of activities supported by the US National Science Foundation (NSF), and recent initiatives in the US Federal Government, that address the issue of responsible use of data. The talk is intended to initiate a discussion on the role that funding agencies could play in supporting a research and education agenda in this area.

8.3 Sustainability Research: Promoting Transparency and Accountability for Decision Makers

Claudia Bauzer Medeiros (UNICAMP – Campinas, BR)

License  Creative Commons BY 3.0 Unported license
© Claudia Bauzer Medeiros


Sustainability is a transdisciplinary research domain that is strongly dependent on the analysis of big data, at multiple space and time scales. In a broad sense, it can be seen as an effort to improve people's lives without compromising the planet's limited resources, from a micro point of view (a single person) to a macro perspective (the Earth, and the biosphere).

Sustainability studies cover a vast range of subjects, such as health, pollution and climate change, biodiversity, inequality or education. As a consequence, data handled are widely heterogeneous, e.g., concerning records about an individual, or environmental measurements, or observations of species.

The talk will discuss a few of the research challenges for big data analysis in sustainability via a real use case in Brazil, and the intrinsic scientific, economic, social and political issues. It will emphasize the aspects of transparency, auditability and accountability for decision making in this context.

8.4 Practising Responsible Data Practices through Data Ethics Education

H. V. Jagadish (University of Michigan – Ann Arbor, US)

License  Creative Commons BY 3.0 Unported license
© H. V. Jagadish
URL <https://www.edx.org/course/data-science-ethics-michiganx-ds101x>

We will get responsible data practices only if data practitioners are responsible, and data practitioners will be responsible only if they know how. For these reasons, it is critical to (1) Make Data Scientists aware of ethical issues regarding data so that they at least try to do the right thing, AND (2) Empower them with the tools to do the right thing with minimum burden.

Imperative (1) means that we need Data Ethics training as an integral part of Data Science training. One suggested starting point is a recent MOOC on EdX. Imperative (2)

means that we need research to develop algorithms and tools that can effectively implement policies that we societally decide are the ones we would like to adopt.

There is an additional question of what policies we should adopt. This requires social consensus, at least across certain segments of society. An informed consensus can only be reached with good education. So the need for Data Ethics education actually extends beyond just Data Science practitioners and to society at large.

Finally, we note that there are many challenging issues at the boundaries – defining what exactly is OK, sociological issues in developing consensus, political concerns regarding laws and regulations, and so on. However, there is a great deal that we can all agree about today – stuff that is not at the boundaries. There is urgent pressure to get this to practice.

8.5 Values, Algorithm Design, and Collaboration

Kristene Unsworth (Drexel University – Philadelphia, US)

License  Creative Commons BY 3.0 Unported license
© Kristene Unsworth

Algorithms and the results they provide appear to many, outside our fields to be objective. We know this is not the case and that the values of data scientists and anyone working with algorithms are reflected in these designs. Because of this, it is important to acknowledge the importance of ethics in the work we do and in society. Ethics are about action and our values drive us. This presentation discussed ongoing research into the role of values in algorithm design and within teams. The work / algorithm-related values of the workshop participants were also discussed and highlighted relation to early research findings. These included:

- technological progress leads to social progress
- correcting information asymmetries
- have an impact on the community
- knowledge
- clarity / insight
- curiosity
- sustainable business model
- intellectual outrage
- technological solutions
- awareness of background context

8.6 Privacy, Transparency and Education

Gerald Friedland (ICSI – Berkeley, US)

License  Creative Commons BY 3.0 Unported license
© Gerald Friedland

Joint work of Gerald Friedland, Dan Garcia, Julia Bernd, Serge Egelmann, Jaeyoung Choi
URL <http://www.teachingprivacy.org>

Decisions about data sharing begin early in somebody's life. For example, when one uploads an image to the Internet, the decision might be whether to post the image or not given the other people that can be identified in that photograph. In our times, decisions like that start to arise in teenage years. On the other hand, current curricula do not cater to

the new responsibilities, not even on the level of University education for engineers. In my talk at Dagstuhl, I presented some of the new issues that arise, including cybercasing [1], data exploitation, and privacy concern followed by a presentation of the teaching resources <http://www.teachingprivacy.org>.

The Teaching Privacy project is an NSF-sponsored collaboration between the International Computer Science Institute and the University of California-Berkeley. The project aims to empower high school students and college undergrads in making informed choices about privacy, by building a set of educational tools and hands-on exercises to help teachers demonstrate what happens to personal information on the Internet, and what the effects of sharing information can be.

Current computer-science curricula at high schools and colleges usually include an abundance of material on data-retrieval methods and how to improve them, but rarely make room for discussion of the potential negative impact of these technologies. Among the groups most affected by those negative impacts are high-school students; they are the most frequent users of social-networking sites and apps, but often do not have a full understanding of the potential consequences their current online activities might have later in their lives. For example, a Facebook posting that a high-scooter's friends think is cool might be seen by a much larger audience than she expected – including perhaps future employers who would not think it was so cool. In addition, not understanding – or not thinking about – the consequences of posting often leads to over-sharing information about other people, including friends and relatives.

The Teaching Privacy is organized around 10 basic principles. Each principle is underpinned with technical explanations, anecdotes, apps, videos, news links, exercises, discussion items, and a guideline for teachers. The teacher's guidelines (TROPE – Teacher's Resources for Online Privacy Education) follow the paradigm of the 5 E's [3]. Learning objectives are outlined clearly and the whole site is licensed under Creative Commons 0, allowing a teacher to cherry pick from the website and creating their own curriculum [2].

This work was supported by funding provided to the International Computer Science Institute by the National Science Foundation, through grants CNS-1065240 and DGE-1419319, and by the Broadband Technology Opportunities Program through the California Connects program. Additional support comes from funding provided to the University of California Berkeley through NSF grants EEC-1405547 and CCF-0424422 and through the IISME Summer Fellowship Program.

References

- 1 G. Friedland, R. Sommer: *Cybercasing the Joint: On the Privacy Implications of Geotagging*, Usenix HotSec 2010 at the Usenix Security Conference, Washington DC, August 2010.
- 2 Julia Bernd, Blanca Gordo, Jaeyoung Choi, Bryan Morgan, Nicholas Henderson, Serge Egelman, Daniel D Garcia, Gerald Friedland: *Teaching Privacy: Multimedia Making a Difference*, IEEE MultiMedia, Vol. 22, Issue 1, pp. 12–19, January 2015.
- 3 <http://enhancinged.wgbh.org/research/eeeeee.html>

8.7 Teaching Ethical Issues in Data Mining to Undergraduates

Sorelle Friedler (Haverford College, US)

License  Creative Commons BY 3.0 Unported license
© Sorelle Friedler

URL <http://ww3.haverford.edu/computerscience/courses/cmssc207/>

In this talk, I discuss how guiding principles about teaching ethics in data mining to undergraduates were integrated into the design of 100-level and 200-level courses. The main principle underlying these curricular choices is that fairness and ethical considerations should be integrated throughout the course, not sidelined to a single lecture or module. These fairness and ethical issues are discussed throughout the course as they relate to the understanding of real world data (e.g., data errors and choices) and the communication of these choices and assumptions (e.g., error values and contextual assumptions). Conversations with domain experts help to connect the data to its true context and drive home the importance and ethical nature of the choices made.

8.8 Networked Systems Ethics

Ben Zevenbergen (University of Oxford, GB)

License  Creative Commons BY 3.0 Unported license
© Ben Zevenbergen

URL <http://networkedsystemsethics.net/>

The Oxford Internet Institute's Ethics in Networked Systems Research project is developing practical guidelines for computer scientists and engineers to assess the ethical implications and social impact of Internet-based projects. These guidelines aim to underpin a meaningful cross-disciplinary conversation between gatekeepers of ethics standards and researchers about the ethical and social impact of technical Internet research projects. The iterative reflexivity methodology guides stakeholders to identify and minimize risks and other burdens, which must be mitigated to the largest extent possible by adjusting the design of the project before data collection takes place. The aim is thus to improve the ethical considerations of individual projects, but also to streamline the proceedings of ethical discussions in Internet research generally. The primary audience for these guidelines are technical researchers (e.g. computer science, network engineering, as well as social science) and gatekeepers of ethics standards at institutions, academic journals, conferences, and funding agencies. It is possible to use these guidelines beyond in academic research in civil society, product development, or otherwise, but these are not the primary audience. Some sections point the reader to other groups – such as the data subjects, lawyers, local peers, etc. – who can also use (parts of) the guidelines to help assess the impact of a project from their expertise or point of view.

9 Lightning talks

9.1 Benchmarking for (Linked) Data Management

Irini Fundulaki (FORTH – Heraklion, GR)

License © Creative Commons BY 3.0 Unported license
© Irini Fundulaki

In this talk we discussed the importance of benchmarking and focused on the kinds of benchmarks we need in order to assess the responsibility regarding the use of data in general, by the different service providers: search engines, recommendation services, and social networks among others. Benchmarks will allow us to measure the bias, or neutrality of the aforementioned entities and give people and regulating bodies a good basis on how to act and react regarding the protection and we'll use of their data.

9.2 From Three Laws of Robotics to Five Principles of Big Data?

Wolfgang Nejdl (Leibniz Universität Hannover, DE)

License © Creative Commons BY 3.0 Unported license
© Wolfgang Nejdl

Can we define five principles for Big Data to guide us through future big data applications? If these are general / generic guidelines, yes. If they should constrain big data applications in an effective way, no. The world has become too complex, possible scenarios are too diverse, and different aspects of big data applications are too often conflicting. How can we solve this dilemma? By working on the issues the workshop focused on in an interdisciplinary way, and by not only focusing on efficiency of algorithms, but also on their fairness and related aspects.

9.3 Natural Language Processing, Responsibly

Jannik Strötgen (MPI für Informatik – Saarbrücken, DE)

License © Creative Commons BY 3.0 Unported license
© Jannik Strötgen

An important kind of data that should be dealt with responsibly, is unstructured data. There is a huge amount of textual data containing information that is not (yet) available in structured format, and natural language processing techniques can be applied to get structured information out of it. Until few years ago, the main text types that NLP research dealt with were non-personal data such as news corpora, and the main goal has been to enrich the textual content with further information. Nowadays, a lot of NLP research is carried out on personal data such as social media content. Although enriching documents is still a major goal, a lot of research addresses predicting author characteristics. Thus, and because NLP techniques are not just used in research environments, but became mature enough to be applied for all types of (real-time) applications, their output can clearly affect individuals. Besides false extractions and aggregations, further problems occur as language is typically uttered at a specific time and place in a specific context, but extracted and

aggregated information extracted from textual data does often not consider this context anymore. Thus, as data in general, unstructured data and NLP techniques should be used and applied responsibly.

10 Working groups

10.1 Structural Bias

Solon Barocas (Microsoft – New York, US), Bettina Berendt (KU Leuven, BE), Michael Hay (Colgate University – Hamilton, US), Amélie Marian (Rutgers University – Piscataway, US), and Gerome Miklau (University of Massachusetts – Amherst, US)

License © Creative Commons BY 3.0 Unported license
© Solon Barocas, Bettina Berendt, Michael Hay, Amélie Marian, and Gerome Miklau

This working group discussed the problem of structural bias in algorithm output. The group agreed that bias was not the best term, but rather that the focus should be on structural discrimination.

The following questions were raised during the discussion: (1) How can we assess structural discrimination? (2) How much of the discrimination comes from the algorithm, how much of it from the input data? (3) If the data is biased, is it the job of the algorithm (operator) to correct structural bias? How? And wouldn't that introduce new bias, that of the algorithm designer/operator? The discussion kept returning to the last question, with various examples, and the group recognized that we were rehashing the traditional issues/questions of affirmative action (fairness w.r.t. to skills or potential, making up for societal bias).

Starting from the model of true space vs. observed space of Friedler et al. [1], an important question is how to identify and mitigate the errors in the input data that may be introducing discrimination. A possibility is to combine multiple observation measurements, and to correct for some of the known bias. One problem with this approach is that many causes of discrimination cannot be identified. In addition, this approach assumes that most of the bias comes from the observation, and not from explicit bias in the data, which is unlikely to be true in practice. Correcting bias in the data is both technically challenging and may have legal ramifications.

References

- 1 Sorelle A. Friedler, Carlos Scheidegger and Suresh Venkatasubramanian On the (im)possibility of fairness. *arXiv preprint arXiv:1609.07236*, 2016.

10.2 Avoid Reinventing the Wheel

Bettina Berendt (KU Leuven, BE), Solon Barocas (Microsoft – New York, US), Claude Castelluccia (INRIA – Grenoble, FR), Pauli Miettinen (MPI für Informatik – Saarbrücken, DE), Wolfgang Nejdl (Leibniz Universität Hannover, DE), Salvatore Ruggieri (University of Pisa, IT), and Jannik Strötgen (MPI für Informatik – Saarbrücken, DE)

License © Creative Commons BY 3.0 Unported license
 © Bettina Berendt, Solon Barocas, Claude Castelluccia, Pauli Miettinen, Wolfgang Nejdl, Salvatore Ruggieri, and Jannik Strötgen

The working group investigated the question of how we can avoid reinventing various wheels when developing more responsible approaches to dealing with data.

We identified different dimensions along which existing wheels should be studied in order to learn from them, in particular: research contents (a topic discussed briefly and repeatedly at the Dagstuhl seminar, albeit outside the working group), teaching contents and teaching methods (including the “teaching of ethics teaching”), research ethics vs. “ethics of research products”, and ethical commitments to society vs. to clients. We also identified disciplines that are especially promising for our search, in particular medicine, bioethics, statistics, nuclear physics, psychology, journalism and legal sciences. Further relevant fields include data ethics, computer ethics, information ethics, robot ethics, and technology impact assessment. A number of stakeholders can be considered: digital right societies such as European Digital Rights (<http://www.edri.org>); civil rights societies, such as Transparency International (<http://www.transparency.org>); regulation authorities, such as the European Data Protection Supervisor (<http://www.edps.eu>); and professional associations, such as the Association of European Journalists (<http://www.aej.org>). There are also concepts such as dual use that have recurred throughout the Dagstuhl seminar and that have been studied in a large body of literature that we should become more familiar with. We discussed structural implementations of ethics, including ethics boards.

We share links and materials on a joint and open platform at https://etherpad.wikimedia.org/p/Dagstuhl_WG_Avoid_Reinventing.

10.3 Dynamics and Feedback in Discrimination Processes

Krishna P. Gummadi (MPI-SWS – Saarbrücken, DE), Chaitanya Baru (NSF – Arlington, US), Marina Drosou (Hellenic Police – Athens, GR), Salvatore Ruggieri (University of Pisa, IT), Rishiraj Saha Roy (MPI für Informatik – Saarbrücken, DE), Jannik Strötgen (MPI für Informatik – Saarbrücken, DE), and Suresh Venkatasubramanian (University of Utah – Salt Lake City, US)

License © Creative Commons BY 3.0 Unported license
 © Krishna P. Gummadi, Chaitanya Baru, Marina Drosou, Salvatore Ruggieri, Rishiraj Saha Roy, Jannik Strötgen, and Suresh Venkatasubramanian

In this breakout session, we identified different types of temporal aspects that are crucial in discrimination processes: (i) stepwise fairness, (ii) changing targets, and (iii) updated data. Thus, we propose a stepwise fairness model, where there is a final fairness goal but it is infeasible to achieve it at once due to a disproportionate loss in perceived “utility”. We should also consider the case when the perfect fairness scenario is a moving target, and changes with time. Relevant modeling paradigms for incremental fairness are online learning (continuous re-learning with new decisions), reinforcement learning (updating the reward

function towards an optimal policy) and cooperative game theory (agents and algorithms as players trying to achieve ideal fairness). The general idea requires that the decisioning algorithm should incorporate feedback, i.e., current decisions should influence future decisions, and that the fairness “score” should be a function of the current time. The segregation model as proposed by Baroni and Ruggieri (2015) can be a good starting point for the stepwise fairness model.

10.4 Principles for Accountable Algorithms

Nicholas Diakopoulos (University of Maryland – College Park, US) and Sorelle Friedler (Haverford College, US)

License © Creative Commons BY 3.0 Unported license
© Nicholas Diakopoulos and Sorelle Friedler

Joint work of Marcelo Arenas, Solon Barocas, Nicholas Diakopoulos, Sorelle Friedler, Michael Hay, Bill Howe, H. V. Jagadish, Kris Unsworth, Arnaud Sahuguet, Suresh Venkatasubramanian, Christo Wilson, Cong Yu, Bendert Zevenbergen

Automated decision making algorithms are now used throughout industry and government, underpinning many processes from dynamic pricing to employment practices to criminal sentencing. Given that such algorithmically informed decisions have the potential for significant societal impact, the goal of this document is to help developers and product managers design and implement algorithmic systems in publicly accountable ways. Accountability in this context includes an obligation to report, explain, or justify algorithmic decision-making as well as mitigate any negative social impacts or potential harms.

We begin by outlining seven equally important guiding principles that follow from this premise:

Algorithms and the data that drive them are designed and created by people – There is always a human ultimately responsible for decisions made or informed by an algorithm. “The algorithm did it” is not an acceptable excuse if algorithmic systems make mistakes or have undesired consequences, including from machine-learning processes.

1. **Auditability:** Enable interested third parties to probe, understand, and review the behavior of the algorithm through disclosure of information that enables monitoring, checking, or criticism, including through provision of detailed documentation, technically suitable APIs, and permissive terms of use.
2. **Error and Uncertainty:** Identify, log, and articulate sources of error and uncertainty throughout the algorithm and its data sources so that expected and worst case implications can be understood and inform mitigation procedures.
3. **Explainability:** Ensure that algorithmic decisions as well as any data driving those decisions can be explained to end-users and other stakeholders in non-technical terms.
4. **Fairness:** Ensure that algorithmic decisions do not create discriminatory or unjust impacts when comparing across different demographics (e.g., race, sex, etc).
5. **Human Experimentation:** Consider the ethics of human experimentation in advance, including potential harms to end-users or others impacted by the algorithm, mitigation strategies for undue risk, and disclosure protocols for potential harms.
6. **Privacy:** Protect users’ privacy surrounding any decisions or data derived or inferred from information about them.

7. Responsibility: Make available externally visible avenues of redress for adverse individual or societal effects of an algorithmic decision system, and designate an internal role for the person who is responsible for the timely remedy of such issues.

10.5 Data, Responsibly: Business and Research Opportunities

Julia Stoyanovich (Drexel University – Philadelphia, US), Serge Abiteboul (ENS – Cachan, FR), Chaitanya Baru (NSF – Arlington, US), Sorelle Friedler (Haverford College, US), Krishna P. Gummadi (MPI-SWS – Saarbrücken, DE), Michael Hay (Colgate University – Hamilton, US), Bill Howe (University of Washington – Seattle, US), Benny Kimelfeld (Technion – Haifa, IL), Arnaud Sahuguet (Cornell Tech NYC, US), Eric Simon (SAP France, FR), Suresh Venkatasubramanian (University of Utah – Salt Lake City, US), and Gerhard Weikum (MPI für Informatik – Saarbrücken, DE)

License © Creative Commons BY 3.0 Unported license
 © Julia Stoyanovich, Serge Abiteboul, Chaitanya Baru, Sorelle Friedler, Krishna P. Gummadi, Michael Hay, Bill Howe, Benny Kimelfeld, Arnaud Sahuguet, Eric Simon, Suresh Venkatasubramanian, and Gerhard Weikum

During this session we tried to identify business and funding opportunities around Data, Responsibly. On the business side, we looked at the consumer (PIMs, risk advisor, education), the enterprise (RaaS for Responsibility as a Service, Affordable Machine Learning) and tools (for auditing and benchmarking, for metrics, leveraging secure hardware, leveraging new language-based approaches and formal methods. We also looked at the social good element and some out of the box ideas such as data poisoning (for the dark web) and identity swapping.

On the research side, we agreed that a common vocabulary is truly necessary for collaboration and advertising. We suggested to try a Grand Challenge-based approach. Some challenges we identified include: universal health vault, vision zero data management (systems that can resist system design flaws and user errors), inequity in education and also a chance to design the next generation Internet starting from a clean slate. We tried to formulate the overall mission as “the science of data, responsibly and its application”.

10.6 Explaining Decisions

Julia Stoyanovich (Drexel University – Philadelphia, US), Chaitanya Baru (NSF – Arlington, US), Claudia Bauzer Medeiros (UNICAMP – Campinas, BR), Krishna P. Gummadi (MPI-SWS – Saarbrücken, DE), Bill Howe (University of Washington – Seattle, US), Arnaud Sahuguet (Cornell Tech NYC, US), Jan Van den Bussche (Hasselt University, BE), and Gerhard Weikum (MPI für Informatik – Saarbrücken, DE)

License © Creative Commons BY 3.0 Unported license
 © Julia Stoyanovich, Chaitanya Baru, Claudia Bauzer Medeiros, Krishna P. Gummadi, Bill Howe, Arnaud Sahuguet, Jan Van den Bussche, and Gerhard Weikum

In this session, we examined what it means to have an algorithm explain its decisions. We first considered the possible audience for such explanations. Stakeholders we have identified include: researcher, developer, consumer, policy maker, politician, regulator, competitor, and auditor.

We then looked at the goal we want to achieve with the explanation, e.g., explaining how the algorithm works, explaining the decision itself, making the recipient not feel bad about the decision, and providing actionable information for the recipient to get a better chance at a positive outcome in the future. An explanation should have the following properties: (a) understandable, (b) manageable (by generator and recipient), (c) consistent, (d) sound, (e) complete, (f) minimal.

We identified two broad categories of decisions. For allocation decisions, a set of limited resources needs to be allocated among agents who have expressed preferences. Decisions for all agents are generated at the same time by the algorithm. For single decisions, the explanation will highly depend on the nature of the algorithm used.

We then considered the kinds of explanations that may be appropriate for different kinds of algorithmic processes. For cases where the decision is based on a set of rules, an explanation can be the set of rules that triggered it. For cases where the decision is based on a decision tree, an explanation can be the full tree or the path (from root to leaf) that lead to the decision. For other cases, an explanation should identify the most critical attributes that contributed to the decision. We also identified another form of explanation where a macro-view of the behavior of the algorithm is visualized for the user, e.g., shared rides company response times shown on a city map.

The group concluded that explaining algorithmic decisions is an important, interesting and a technically challenging area that warrants attention from the research community.

10.7 Data, Responsibly: Use Cases and Benchmarking

Suresh Venkatasubramanian (University of Utah – Salt Lake City, US), Claudia Bauzer Medeiros (UNICAMP – Campinas, BR), Gerald Friedland (ICSI – Berkeley, US), Irimi Fundulaki (FORTH – Heraklion, GR), and Salvatore Ruggieri (University of Pisa, IT)

License © Creative Commons BY 3.0 Unported license

© Suresh Venkatasubramanian, Claudia Bauzer Medeiros, Gerald Friedland, Irimi Fundulaki, and Salvatore Ruggieri

This working group started to compile a set of use cases related to Data, Responsibly. Having such a set of curated examples can be extremely helpful when writing papers, applying for funding or simply trying to convince people about the importance and timeliness of the topic. This is work in progress but we hope that this evolving dataset will consist of a list of documented examples, annotated with the various dimensions that describe Data, Responsibly.

Participants

- Serge Abiteboul
ENS – Cachan, FR
- Marcelo Arenas
Pontificia Universidad Catolica de Chile, CL
- Solon Barocas
Microsoft – New York, US
- Chaitanya Baru
NSF – Arlington, US
- Claudia Bauzer Medeiros
UNICAMP – Campinas, BR
- Bettina Berendt
KU Leuven, BE
- Claude Castelluccia
INRIA – Grenoble, FR
- Nicholas Diakopoulos
University of Maryland – College Park, US
- Marina Drosou
Hellenic Police – Athen, GR
- Gerald Friedland
ICSI – Berkeley, US
- Sorelle Friedler
Haverford College, US
- Irimi Fundulaki
FORTH – Heraklion, GR
- Krishna P. Gummadi
MPI-SWS – Saarbrücken, DE
- Michael Hay
Colgate Univ.y – Hamilton, US
- Bill Howe
University of Washington – Seattle, US
- H. V. Jagadish
University of Michigan – Ann Arbor, US
- Benny Kimelfeld
Technion – Haifa, IL
- Amélie Marian
Rutgers Univ. – Piscataway, US
- Pauli Miettinen
MPI für Informatik – Saarbrücken, DE
- Gerome Miklau
University of Massachusetts – Amherst, US
- Wolfgang Nejdl
Leibniz Univ. Hannover, DE
- Benjamin Nguyen
INSA – Bourges, FR
- Evaggelia Pitoura
University of Ioannina, GR
- Salvatore Ruggieri
University of Pisa, IT
- Rishiraj Saha Roy
MPI für Informatik – Saarbrücken, DE
- Arnaud Sahuguet
Cornell Tech NYC, US
- Eric Simon
SAP France, FR
- Julia Stoyanovich
Drexel Univ. – Philadelphia, US
- Jannik Strötgen
MPI für Informatik – Saarbrücken, DE
- Fabian Suchanek
Télécom ParisTech, FR
- Kristene Unsworth
Drexel Univ. – Philadelphia, US
- Jan Van den Bussche
Hasselt University, BE
- Suresh Venkatasubramanian
University of Utah – Salt Lake City, US
- Agnès Voisard
FU Berlin, DE
- Nicholas Weaver
ICSI – Berkeley, US
- Gerhard Weikum
MPI für Informatik – Saarbrücken, DE
- Christo Wilson
Northeastern University – Boston, US
- Cong Yu
Google – New York, US
- Ben Zevenbergen
University of Oxford, GB

