ExSS-ATEC: Explainable Smart Systems and Algorithmic Transparency in Emerging Technologies 2020

Alison Smith-Renner
Decisive Analytics Corporation
Arlington, VA, USA
alison.renner@dac.us

Styliani Kleanthous
OUC & Rise Research Centre
Nicosia, Cyrpus
styliani.kleanthous@gmail.com

Brian Lim
National University of Singapore
Singapore
brianlim@com.nus.edu.sg

Tsvi Kuflik
University of Haifa
Haifa, Israel
tsvikak@is.haifa.ac.il

Simone Stumpf
City, University of London
London, UK
simone.stumpf.1@city.ac.uk

Jahna Otterbacher
OUC & Rise Research Centre
Nicosia, Cyprus
jahna.otterbacher@me.com

Advait Sarkar
Microsoft Research
Cambridge, UK
advait@microsoft.com

Casey Dugan
IBM Research
Cambridge, MA, US
cadugan@us.ibm.com

Avital Shulner
University of Haifa
Haifa, Israel
avitalshulner@gmail.com

ABSTRACT

Smart systems that apply complex reasoning to make decisions and plan behavior, such as decision support systems and personalized recommendations, are difficult for users to understand. Algorithms allow the exploitation of rich and varied data sources, in order to support human decision-making and/or taking direct actions; however, there are increasing concerns surrounding their transparency and accountability, as these processes are typically opaque to the user. Transparency and accountability have attracted increasing interest to provide more effective system training, better reliability and improved usability. This workshop will provide a venue for exploring issues that arise in designing, developing and evaluating intelligent user interfaces that provide system transparency or explanations of their behavior. In addition, our goal is to focus on approaches to mitigate algorithmic biases that can be applied by researchers, even without access to a given system's inter-workings, such as awareness, data provenance, and validation.

CCS CONCEPTS

• **Human-centered computing~Human computer interaction (HCI)~Interactive systems and tools** • Computing methodologies~Machine learning • Computing methodologies~Artificial intelligence

KEYWORDS

Explanations; visualizations; machine learning; intelligent systems; intelligibility; transparency; fairness; accountability

1 Background

Smart systems that apply complex reasoning to make decisions and plan behavior, such as clinical decision support systems, personalized recommendations, home automation, machine learning classifiers, robots and autonomous vehicles, are difficult for a user to understand [2]. Fairness, accountability and transparency are currently hotly discussed aspects of machine learning systems, especially for deep learning systems that are seen to be very difficult to explain to users. Textual explanations and graphical visualizations are often provided by a system to give users insight into what the systems is doing and why it is doing it [3–6] and work is starting to investigate how to best engage in transparency design [1]. However, there are still numerous issues and problems regarding explanations and algorithm transparency that demand further attention, such as how can we build (better) explanations or transparent systems, what should be included in an explanation and how should they be presented, when should explanations be deployed, or when do they detract from the user experience, how can transparency expose biases in data or algorithmic processes, and how can we evaluate explanations or system transparency, especially from a user perspective.

The ExSS-ATEC 2020 workshop brings together academia and industry together to address these issues. This workshop is a follow-on from the ExSS 2018 and 2019 workshops in combination with the ATEC 2019 workshop previously held at IUI. This workshop includes a keynote, paper panels, and group activities, with the goal of developing concrete approaches to handling challenges related to the design and development of explanations and system transparency. ExSS-ATEC 2020 is supported by the Cyprus Center for Algorithm Transparency (CyCAT).

2 Workshop Overview

The workshop keynote is Dr. Carrie Cai, focusing on current challenges for explainable smart systems. Nine accepted papers are presented as three themed panel sessions. Accepted papers are:

- Jung et al. "Transparency of Data and Algorithms in a Persona System: Explaining Data-Driven Personas to End Users"
- Dodge and Burnett, "Position: We Can Measure XAI Explanations Better with 'Templates'"
- Hepenstal et al., "What Are You Thinking? Explaining Conversational Agent Responses for Criminal Investigations"
- Stockdill et al., "Cross-Domain Correspondences for Explainable Recommendations"
- Lindvall and Molin, "Verification Staircase: A Design Strategy for Actionable Explanations"
- Larasati et al., "The Effects of Explanation Styles on Users' Trust"
- Ferreira and Monteiro, "Do ML Experts Discuss Explainability for AI Systems? A Discussion Case in the Industry for a Domain-Specific Solution"
- Zürn et al., "What If? Interaction with Recommendations"
- Chromik and Schuessler, "A Taxonomy for Human Subject Evaluation of Black-Box Explanations in XAI"

The second part of the workshop is structured around hands-on activity sessions in small subgroups of 3-5 participants.

3 Key People

3.1 Keynote Speaker

Dr. Carrie Cai is a senior research scientist at Google Brain and PAIR (Google's People+AI Research Initiative). Her research aims to make human-AI interactions more productive and enjoyable to end-users, ranging from novel tools to help doctors steer AI cancer-diagnostic systems in real-time, to frameworks for effectively onboarding end-users to AI assistants. Her work has been published in HCI venues such as CHI, IUI, CSCW, and VL/HCC, receiving 4 best paper / honorable mention awards and profiled on TechCrunch and the Boston Globe. Before joining Google, Carrie completed her PhD in computer science at MIT, where she created intelligent wait-learning systems to help people accomplish long-term goals in short chunks while waiting. Carrie first learned to program at age 24, after having completed undergraduate degrees in human biology and education at Stanford. She feels that it's never too late to learn machine learning, and that some of the world's best AI innovations come from the humanities powered by computing.

3.1 Workshop Committee

The workshop committee includes Gagan Bansal (UW), Veronika Bogina (Haifa University), Robin Burke (UC Boulder), Jonathan Dodge (OSU), Fan Du (Adobe), Malin Eiband (LMU), Michael Ekstrand (Boise State), Melinda Gervasio (SRI), Fausto Giunchiglia (U Toronto), Alan Hartman (Afeka), Judy Kay (U Sydney), Bran Knowles (U Lancaster), Todd Kulesza (Google), Tak Lee (Adobe), Loizos Michael (Cyprus), Shabnam Najafan (Delft), Alicja Piotrkowicz (U Leeds), Forough Poursabzi-Sangdeh (Microsoft), Gonazalo Ramos (MSR), Stephanie Rosenthal (CMU), Martin Schuessler (TU Berlin), Ramya Srinivasan (Fujitsu), Mike Terry (Google), Sarah Völkel (U Munich), and Jürgen Ziegler (U Duisburg).

3.1 Workshop Organizers

The workshop organizing committee includes: **Alison Smith-Renner,** Director of the Machine Learning Visualization Lab for Decisive Analytics Corporation and PhD Candidate at University of Maryland, College Park; **Dr. Styliani Kelanthous**, senior researcher in the Faculty of Pure and Applied Sciences at Open University of Cyprus and RISE Research Centre, Cyprus; **Dr. Brian Lim**, Assistant Professor in the Department of Computer Science at the National University of Singapore, **Prof. Tsvi Kuflik**, professor and former head of the Information Systems Department at the University of Haifa, Israel; **Dr. Simone Stumpf**, Senior Lecturer at City, University of London, **Jahna Otterbacher**, founder of the Behavioral & Language Traces research lab, which is housed in the Faculty of Pure and Applied Sciences, Open University of Cyprus; **Dr. Advait Sarkar**, senior researcher at Microsoft Research in Cambridge (UK); **Casey Dugan**, manager of the AI Experience Team at IBM Research Cambridge (MA, USA); and **Avital Shulner**, PhD student in the Information Systems Department at the University of Haifa, Israel.

REFERENCES

[1] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing Transparency Design into Practice. In *Proceedings of the 2018 Conference on Human InformationInteraction&Retrieval.* https://doi.org/10.1145/3172944.3172961

[2] Alyssa Glass, Deborah L. McGuinness, and Michael Wolverton. 2008. Toward establishing trust in adaptive agents. In *Proceedings of the 13th international conference on Intelligent user interfaces - IUI '08*, 227. https://doi.org/10.1145/1378773.1378804

[3] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *ACM Conference on Computer Supported Cooperative Work (CSCW)*, 241–250. https://doi.org/10.1145/358916.358995

[4] Carmen Lacave and Francisco J. Díez. 2002. A review of explanation methods for Bayesian networks. *Knowledge Engineering Review 17*, 107–127. https://doi.org/10.1017/S026988890200019X

[5] Pearl Pu and Li Chen. 2006. Trust building with explanation interfaces. In *International conference on Intelligent User Interfaces (IUI)*, 93. https://doi.org/10.1145/1111449.1111475

[6] William Swartout, Cecile Paris, and Johanna Moore. 1991. Explanations in knowledge systems: Design for Explainable Expert Systems. *IEEE Expert 6*, 3: 58–64. https://doi.org/10.1109/64.87686