

1st Workshop on Adverse Impacts and Collateral Effects of Artificial Intelligence Technologies

Esma Aïmeur¹, Nicolás E. Díaz Ferreyra² and Hicham Hage³

¹*Department of Computer Science and Operations Research, University of Montréal, Canada*

²*Department of Computer Science and Applied Cognitive Science, University of Duisburg-Essen, Germany*

³*Computer Science Department, Notre Dame University - Louaize, Lebanon*

1. Preface

The rapid advancement and growth of Information and Communication Technologies in general and Artificial Intelligence (AI) in particular; has led to the seamless and yet indispensable integration of such technologies into our everyday activities.

Indeed, over the last decade, AI has infiltrated many aspects of our lives: people rely on it while driving or training; or when selecting which movie/song to play next, even when asking information about the weather or current traffic conditions. Moreover, individuals rely heavily on intelligent software applications across different domains including healthcare, logistics, agriculture, finance, education, defence, and governance. Particularly, AI systems facilitate decision-making processes across these domains through the automatic analysis and classification of large data sets and the subsequent identification of relevant patterns. To a large extent, such an approach has contributed to the sustainable development of modern societies and remains a powerful instrument for social and economic growth. However, recent events related to the discovery of biased AI, the massive spread of misinformation and deepfakes along with fears of AI powered autonomous weapons, have raised concerns among AI practitioners and researchers about the negative and detrimental impacts of these technologies. Indeed, like any other technology, AI can have some seriously negative consequences, whether intentionally or inadvertently.

Consequently, and due to the ubiquity of AI and the increasingly rapid rate of its development and adoption, there is an urgent call for guidelines, methods, and techniques to assess and mitigate the potentially adverse impacts and side effects of AI applications.

AIofAI'21: 1st Workshop on Adverse Impacts and Collateral Effects of Artificial Intelligence Technologies, Montreal, CA

✉ aimeur@iro.umontreal.ca (E. Aïmeur); nicolas.diaz-ferreyra@uni-due.de (N. E. Díaz Ferreyra);

hhage@ndu.edu.lb (H. Hage)


🌐 <http://www.iro.umontreal.ca/~aimeur/> (E. Aïmeur); <https://www.ndiaz-ferreyra.com/> (N. E. Díaz Ferreyra);

<https://www.ndu.edu.lb/> (H. Hage)

🆔 0000-0001-6304-771X (N. E. Díaz Ferreyra)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

Workshop Objectives

This 1st Workshop on Adverse Impacts and Collateral Effects of AI Technologies (AIofAI '21) is co-located with the 30th International Conference on Artificial Intelligence (IJCAI-21). The objective of the workshop is to bring together experts and practitioners to explore how and up to which extent AI technologies can serve deceptive and malicious purposes whether intentionally or not. Furthermore, it seeks to elaborate on guidelines, countermeasures and mitigation actions to prevent potential negative effects and collateral damages of AI systems. We therefore invited AI researchers and practitioners across different disciplines and knowledge backgrounds to submit contributions dealing with the following (or related) topics:

- Hazardous AI applications:
 - Deepfakes.
 - Fake news and misinformation.
 - Online deception.
 - Malicious personalization.
 - Social engineering.
- Adverse impacts of AI:
 - Privacy and security breaches.
 - Backfire effects.
 - Guidelines and mitigation actions.
 - Ethical conflicts and challenges.
 - Risk assessment methods.
- Responsible AI:
 - Case studies.
 - Best practices for trustworthy AI.

2. Accepted Papers

Nine papers were submitted and peer-reviewed by 3 members of the program committee in a single-blind process. Out of these, seven papers were accepted for this volume, three as long papers and four as short papers. In addition, we have four invited contributions from renowned AI researchers and experts.

2.1. Regular Papers

1. *Dataset Reconstruction Attack Against Language Models* (long)
Rrubaa Panchendrarajan and Suman Bhoi
2. *Using Mathematically-Grounded Metaphors to Teach AI-Related Cybersecurity* (short)
Bart Knijnenburg, Nicole Bannister and Kelly Caine
3. *The Presidential Deepfakes Dataset* (short)
Aruna Sankaranarayanan, Matt Groh, Rosalind Picard and Andrew Lippman

4. *ROBUST: Deep Learning for Malware Detection under Changing Environments* (long)
Adel Abusitta, Talal Halabi and Omar Abdel Wahab
5. *Mitigating Digital Mindlessness* (short)
Sushmita Khan, Mehtab Iqbal, Nushrat Humaira, Nina Hubig and Bart Knijnenburg
6. *Artificially Intelligent and Inclusive by Design: A Human-Centered Approach to Online Safety* (short)
Daricia Wilkinson
7. *SecuBot, a Teacher in Appearance: How Social Chatbots Can Influence People!* (long)
Jordi Saleilles and Esma Aïmeur

2.2. Invited Papers

1. *The Future of AI Ethics and the Role of Neurotechnology*
Sara Berger and Francesca Rossi
2. *Is My Model Biased? Exploring Unintended Bias in Misogyny Detection Tasks*
Daniela Godoy and Antonela Tommasel
3. *Fake News and AI: Fighting Fire With Fire?*
Kimiz Dalkir
4. *What Are You Afraid of? AI Doesn't Kill People, People Kill People*
Roger Schank and Ray Bareiss

3. Invited Talks and Panel

Three keynote talks and a discussion panel were included as part of the AIofAI's technical programme. Overall, eight internationally renowned experts have been invited to discuss and reflect on issues related to the adverse impacts and collateral effects of AI technologies.

Keynote - "How Not to Destroy the World With AI"

Stuart Russell

I will briefly survey recent and expected developments in AI and their implications. Some are enormously positive, while others, such as the development of autonomous weapons and the replacement of humans in economic roles, may be negative. Beyond these, one must expect that AI capabilities will eventually exceed those of humans across a range of real-world-decision making scenarios. Should this be a cause for concern, as Elon Musk, Stephen Hawking, and others have suggested? And, if so, what can we do about it? While some in the mainstream AI community dismiss the issue, I will argue that the problem is real and that the technical aspects of it are solvable if we replace current definitions of AI with a version based on provable benefit to humans.

Keynote - "Ethical Issues of Personalized Persuasive Technology"

Julita Vassileva

Personalization / user-adaptation is essentially an optimization of a technological system with a particular purpose. Using persuasive technology as an example, this talk will discuss two

purposes of personalization: for the benefit of individual users and for the sustainability of the community. It will show how persuasive technology for behaviour change can be optimized by personalizing it to the individual user, and how adaptive incentive mechanisms (a form of persuasive technology) can encourage behaviours contributing to a sustainable community. The talk will also address some ethical issues related to personalized persuasive technology and how we as technology developers can contribute to emerging social (regulatory) solutions.

Keynote - “Privacy is Power!”

Carissa Véliz

One of the most misleading myths about privacy is that it is something of a personal preference. I will argue that people should protect their privacy first and foremost privacy is a kind of power. If we give too much of our data to corporations, the wealthy will rule. If we give too much personal data to governments, we risk sliding into authoritarianism. For democracy to be strong, the bulk of power needs to be with the citizenry, and whoever has the data will have the power. Privacy is not a personal preference; it is a political concern. Personal data is a toxic asset, and should be regulated as such. The trade in personal data has to end. Personal data is not the kind of thing that should be bought or sold.

Have you ever been denied insurance, a loan, or a job? Have you had your credit card number stolen? Do you have to wait too long when you call customer service? Have you paid more for a product than one of your friends? Have you been harassed online? Have you noticed politics becoming more divisive in your country? You might have the data economy to thank for all that and more.

The moment you check your phone in the morning you are giving away your data. Before you've even switched off your alarm, a whole host of organisations have been alerted to when you woke up, where you slept, and with whom. Our phones, our TVs, even our washing machines are spies in our own homes.

Without your permission, or even your awareness, tech companies are harvesting your location, your likes, your habits, your relationships, your fears, your medical issues, and sharing it amongst themselves, as well as with governments and a multitude of data vultures. They're not just selling your data. They're selling the power to influence you and decide for you. Even when you've explicitly asked them not to. And it's not just you. It's all your contacts too, all your fellow citizens. Privacy is as collective as it is personal.

Digital technology is stealing our personal data and with it our power to make free choices. To reclaim that power, and our democracy, we must take back control of our personal data. Surveillance is undermining equality. We are being treated differently on the basis of our data. Privacy is the blindfold of justice.

What can we do? The stakes are high. We need to understand the power of data better. We need to start protecting our privacy. And we need regulation. We need to pressure our representatives. It is time to pull the plug on the surveillance economy.

Panel Discussion - “Adverse Impacts and Collateral Effects of AI Technologies”

Yoshua Bengio, Gilles Brassard, Jean-Gabriel Ganascia, Francesca Rossi, Stuart Russell, Roger Schank

The role of AI in people’s everyday life has grown exponentially over the last decade. From making bank transactions to finding a life partner, AI technologies have acquired an increasingly important role in the dynamics of modern societies. However, while AI is providing several benefits and can serve noble purposes, it can also be used to deceive or even inflict intentional harm. Far away from fiction, the use of AI for malicious purposes is nowadays a reality demanding not only a profound debate but also urgent solutions. Hence, this panel will elaborate on the challenges surging from unethical and malicious AI applications from a holistic perspective. For this, it will bring together a group of renowned experts to reflect on such challenges and share their views.

4. Organization and Committees

Workshop Organizers

- **Esma Aïmeur**
 - University of Montréal, Canada
 - Website: <http://www.iro.umontreal.ca/~aimeur/>
 - Email: aimeur@IRO.UMontreal.CA
- **Nicolás E. Díaz Ferreyra**
 - University of Duisburg-Essen, Germany
 - Website: <http://www.ndiaz-ferreyra.com/>
 - Email: nicolas.diaz-ferreyra@uni-due.de
- **Hicham Hage**
 - Notre Dame University - Louaize, Lebanon
 - Website: <https://www.ndu.edu.lb/>
 - Email: hhage@ndu.edu.lb

Programme Committee

- Ludovico Boratto (Centre Tecnològic de Catalunya, Spain)
- Jean-Gabriel-Ganascia (Paris-Sorbonne University, France)
- Damiano Spina (RMIT University, Australia)
- Stefan Stieglitz (University of Duisburg-Essen, Germany)
- Christos Fidas (University of Patras, Greece)
- Leon Derczynski (ITU Copenhagen, Denmark)
- Daniela Godoy (ISISTAN CONICET-UNICEN, Argentina)
- Julita Vassileva (University of Saskatchewan, Canada)
- Bart Knijnenburg (Clemson University, United States)
- Antonela Tommasel (ISISTAN CONICET-UNICEN, Argentina)

- Jeremy Clark (Concordia University, Canada)
- Steven Furnel (University of Nottingham, UK)
- Pam Briggs (Northumbria University, UK)
- Marios Belk (Cognitive UX GmbH, Cyprus)
- Norman Sadeh-Konieczpol (Carnegie Mellon University, United States)
- Alison R. Panisson (Federal University of Santa Catarina, Brazil)

Acknowledgments

This work was partially supported by Canada's Natural Sciences and Engineering Research Council (NSERC) and the Deutsche Forschungsgemeinschaft (DFG) under grant No. GRK 2167, Research Training Group "User-Centred Social Media".