2nd Workshop on Adverse Impacts and Collateral Effects of Artificial Intelligence Technologies

Esma Aïmeur[1], Nicolás E. Díaz Ferreyra[2] and Hicham Hage[3]

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

[2]*Institute of Software Security, Hamburg University of Technology, 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'22: 2nd Workshop on Adverse Impacts and Collateral Effects of Artificial Intelligence Technologies, Vienna, AT

aimeur@iro.umontreal.ca (E. Aïmeur); nicolas.diaz-ferreyra@tuhh.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)

Workshop Objectives

This 1st Workshop on Adverse Impacts and Collateral Effects of AI Technologies (AIofAI '22) is co-located with the 31th International Conference on Artificial Intelligence (IJCAI-22). 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.

Special topics of interest:

1. **AI in the COVID-19 era**: Since its outbreak in 2020, the COVID-19 pandemic has plunged the world into a state of crisis. Far from over, the last two years have been characterized by a tsunami of misinformation, as well as the hasty deployment of inexact and biased AI models to detect COVID-19 and manage the pandemic, fostering mistrust in research and scientific evidence. AIofAI welcomes submissions elaborating on the detrimental impact of AI applications during the COVID-19 pandemic.
2. **AI Regulations**: The new regulatory framework for AI systems drafted by the European Commission seeks for instance to promote profound changes in the way such systems are developed and deployed. Still, many challenges are upfront particularly when it comes to the identification and assessment of risks potentially linked to AI solutions. We encourage the submission of papers elaborating on regulations, guidelines, methods and tools for assessing the risks of AI systems and their possible adverse impact on both individuals and societies at large.

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, six papers were accepted for this volume, three as long papers and three as short papers.

1. *Supposedly Fair Classification Systems and Their Impacts* (long)
   Mackenzie Jorgensen, Elizabeth Black, Natalia Criado, and Jose Such
2. *A Game for Crowdsourcing Adversarial Examples for False Information Detection* (long)
   Jan Cegin, Jakub Simko, and Peter Brusilovsky
3. *Utilising Assessment List for Trustworthy AI: Three Areas of Improvement* (short)
   Adrian Gavornik, Juraj Podrouzek, Matus Mesarcik, Sara Solarova, Stefan Oresko, and Maria Bielikova
4. *Good AI for Good: How AI Strategies of the Nordic Countries Address the Sustainable Development Goals* (short)
   Andreas Theodorou, Juan Carlos Nieves, and Virginia Dignum
5. *Towards Enhanced Privacy-Preserving Nudges* (long)
   Rim Ben Salem, Esma Aïmeur, and Hicham Hage
6. *Acquiring Knowledge Using Crowdsourcing and AI: Participatory Budget and Related Risks* (short)
   Lukasz Przysucha

3. Invited Talks

Two keynote talks were included as part of AIofAI's technical programme.

Keynote - "Collateral ethical effects of using facial recognition in the defense of Ukraine under attack"

Jean-Gabriel Ganascia

In the immediate aftermath of Russia's attack on Ukraine, a major technology company offered the Ukrainian government free access to its facial recognition techniques to once again combat its enemies and help identify Russian spies embedded in the population. At first glance, such a generous offer and fair use of technology should be welcomed. However, some possible uses of these AI-based devices may be ethically unacceptable. For instance, it has been said that by identifying Russian prisoners through facial recognition, they have been forced to say things publicly on social networks that they wouldn't have said freely and thus humiliate them, which is contrary to the laws of war set out in the Geneva Conventions. One mentions also that the identification by facial recognition of killed Russian soldier bodies allowed to dispatch their photos to their families and friends to undermine their national confidence; even if the defensive objective is legitimate, this desecration of the image of the dead is an inadmissible attempt to human dignity. While all these terrifying (and ethically questionable) usages of AI technologies have not been proven, it shows that, even with the better intents of the world, there are many potential bad usages that could cast opprobrium on AI and on the people who make use of them.

Keynote - "Robots that need to mislead: Biologically-inspired machine deception"

Ronald C. Arkin

Expanding our work in understanding the relationships maintained in teams of humans and robots, this talk overviews almost 15 years of research on deception and its application within robotic systems. Earlier we explored the use of psychology as the basis for producing deceit in robotic systems in order to evade capture. More recent work involves studying squirrel hoarding and bird mobbing behavior as it applies to deception, in the first case for misleading a predator, and in the second for feigning strength when none exists. Next, we review other-deception, where deceit is performed for the benefit of the mark. Finally, newly completed research on team deception where groups of agents using shills that serve to mislead others is presented. Results are presented in both simulation and simple robotic systems, as well as consideration of the ethical implications of this research.

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**
  - Hamburg University of Technology, Germany
  - Website: http://www.ndiaz-ferreyra.com/
  - Email: nicolas.diaz-ferreyra@tuhh.de
- **Hicham Hage**
  - Notre Dame University - Louaize, Lebanon
  - Website: https://www.ndu.edu.lb/
  - Email: hhage@ndu.edu.lb

Programme Committee

- Jeremy Clark (Concordia University, Canada)
- Steven Furnel (University of Nottingham, UK)
- Alison R. Panisson (Federal University of Santa Catarina, Brazil)
- Daniela Godoy (ISISTAN CONICET-UNICEN, Argentina)
- Antonela Tommasel (ISISTAN CONICET-UNICEN, Argentina)

- Pam Briggs (Northumbria University, UK)
- Lijie Guo (Clemson University, United States)
- Marios Belk (Cognitive UX GmbH, Cyprus)
- Jean-Gabriel-Ganascia (Paris-Sorbonne University, France)
- Josep Domingo-Ferrer (Universitat Rovira i Virgili, Spain)
- Damiano Spina (RMIT University, Australia)
- Kimiz Dalkir (McGill University, Canada)
- Timotheus Kampik (Umeå University | Singavio GmbH, Sweden)
- Juan Carlos Nieves (Umeå University, Sweden)
- Michael Floyd (Knexus Research Corporation, United States)
- Gabriel Pedroza (CEA-LIST, France)

Acknowledgments

This work was partially supported by Canada's Natural Sciences and Engineering Research Council (NSERC).