

What's Hot in AI & Ethics



Judy Goldsmith

University of Kentucky
College of Engineering
Computer Science Department
Lexington, KY, USA



- 1 Autonomous Vehicles
- 2 Autonomous Weapons
- 3 Reproducing Human Bias
- 4 Validity of News
- 5 The Future of Employment
- 6 The Singularity Cha Cha
- 7 Organizations and Links
 - AI Ethics-Related Groups
 - Writing Contests
 - Reports on the Future of AI



Trolley Problems

The trolley problem: should you pull the lever to divert the runaway trolley onto the side track? [Wikipedia] [*Do you take responsibility for killing one person to save several?*]

If a Pokémon Go player wanders in front of a self-driving car, should it swerve, putting rider(s) and others at risk?



Robot Soldiers and Smart Bombs

- If robots fight our wars, will fewer humans die?
- Can robot decisions be tempered by compassion?



Robot Soldiers and Smart Bombs

- If robots fight our wars, will fewer humans die?
- Can robot decisions be tempered by compassion?
- Can we program in the rules of war?
- Can we program ethics? (Which ones???)



Robot Soldiers and Smart Bombs

- If robots fight our wars, will fewer humans die?
- Can robot decisions be tempered by compassion?
- Can we program in the rules of war?
- Can we program ethics? (Which ones???)
- **Could autonomous agents of war be hacked?**



ML Applications Based on Historic Decision Data

- Mortgage loans
- Bail decisions
- ...



Recognizing and Generating Fake News

- Can we determine the provenance of a news item?



Recognizing and Generating Fake News

- Can we determine the provenance of a news item?
- Can we determine the validity of a news item?



Recognizing and Generating Fake News

- Can we determine the provenance of a news item?
- Can we determine the validity of a news item?
- Can we *generate* persuasive “news” items?



Post-Scarcity or Post-Employment?

If robots and software agents take over manufacturing, farming, service industries, education, etc., what will people do?



Post-Scarcity or Post-Employment?

If robots and software agents take over manufacturing, farming, service industries, education, etc., what will people do?

How will we share the prosperity?



Post-Scarcity or Post-Employment?

If robots and software agents take over manufacturing, farming, service industries, education, etc., what will people do?

How will we share the prosperity?

Will that come from governments or private sources?



Robot Sentience: Science Fiction ... Or Is It?

- Robot rights
- Inadvertent evil (Bostrom's paperclip scenario)
- Intentional harm to humans: Skynet
- ...



Why I Have Hyperlinks on My Slides

These slides will be available through the workshop site, and perhaps aaai.org



The Human Side

- **Never Again**

We refuse to participate in the creation of databases of identifying information for the United States government to target individuals based on race, religion, or national origin.

...

- **The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems**

The purpose of this Initiative is to ensure every technologist is educated, trained, and empowered to prioritize ethical considerations in the design and development of autonomous and intelligent systems.



Autonomous Agents

- **Future of Life Institute**

Mission: To catalyze and support research and initiatives for safeguarding life and developing optimistic visions of the future, including positive ways for humanity to steer its own course considering new technologies and challenges.

- **Campaign to Stop Killer Robots**



Industry Organizations

- Partnership on AI

Established to study and formulate best practices on AI technologies, to advance the public's understanding of AI, and to serve as an open platform for discussion and engagement about AI and its influences on people and society.

- Diversity in Artificial Intelligence

Inclusion | Balance | Neutrality

Preventing racial, age, gender, disability and other discrimination



Other Lists of Links

- [Ethics in Artificial Intelligence](#)

This is the website for a project based in the Computer Science Department, University of Oxford, Towards a Code of Ethics for Artificial Intelligence Research, which is being carried out by Professor Mike Wooldridge, Professor Peter Millican, and Dr. Paula Boddington.

- [Alphabetical List of Resources](#)

Here we [the Oxford group] list the resources briefly discussed on the parent Resources page, or elsewhere on this site, as well as links to relevant professional bodies and their statements on ethics.



Imagining More Ethical AI

- **ACM SIGAI Student Essay Contest on the Responsible Use of AI Technologies**

What do you see as the 1–2 most pressing ethical, social or regulatory issues with respect to AI technologies? What position or steps can governments, industries or organizations (including ACM SIGAI) take to address these issues or shape the discussions on them?

Deadline: 11:59pm on March 1, 2017 (UTC-12)

- **Sapiens Plurum: The Wisdom of the Many**

Earth Day Short Fiction Contest

Deadline: 12:59 PM [23:59], May 27, 2017

- [White House Report on the Future of AI 2016](#)
- [Stanford One Hundred Year Study on Artificial Intelligence \(AI100\)](#)
- [European Civil Law Rules in Robotics](#)



European Civil Law Rules in Robotics

4. ANALYSIS OF THE ETHICAL PRINCIPLES TO DEVELOP IN ROBOTICS

4.1. Devising a general ethical framework for robotics: proposal to establish the main roboethical principles for protecting humanity from robots

4.1.1. Protecting humans from harm caused by robots

4.1.2. Respecting the refusal of care by a robot

4.1.3. Protecting human liberty in the face of robots



European Civil Law Rules in Robotics

- 4.1.4. Protecting humanity against privacy breaches committed by a robot
- 4.1.5. Managing personal data processed by robots
- 4.1.6. Protecting humanity against the risk of manipulation by robots
- 4.1.7. Avoiding the dissolution of social ties
- 4.1.8. Equal access to progress in robotics
- 4.1.9. Restricting human access to enhancement technologies



Acknowledgments

- The Science Fiction & Computer Ethics group: Nicholas Mattei, Emanuelle Burton, Cory Siler, ...
- Marie desJardin
- Paula Boddington
- Toby Walsh
- Sven Koenig
- This material is based upon work supported by the National Science Foundation under Grant No. 1646887. *Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.*