VM global leadership SUMMIT 2017

AIEye

Artificial Intelligence
Supercharging Knowledge and Decision Making

VM global leadership SUMMIT 2017

Ethics & AI

Alex John London, PhD

Professor of Philosophy & Director,
Center for Ethics and Policy

Carnegie Mellon University

AIEye

Artificial Intelligence
Supercharging Knowledge and Decision Making

CMU K&L Gates Initiative

- $10 million gift to Carnegie Mellon University for Ethics and Computational Technologies.
- Consolidate CMU strengths across robotics, computer science, machine learning, ethics and social policy…
- Enhance CMU's global leadership in ethical AI design, oversight, and human impacts.

VM global leadership SUMMIT 2017

AIEye

Dangers of Far-Future Focus

- Obscures near-term decision impacts future we inherit.
- Absolves stakeholders of responsibility for:
  - Values they build into emerging AI systems
  - Promoting/discouraging forms of social relations that AI makes possible.
- ”Transitional” problems are here, now.

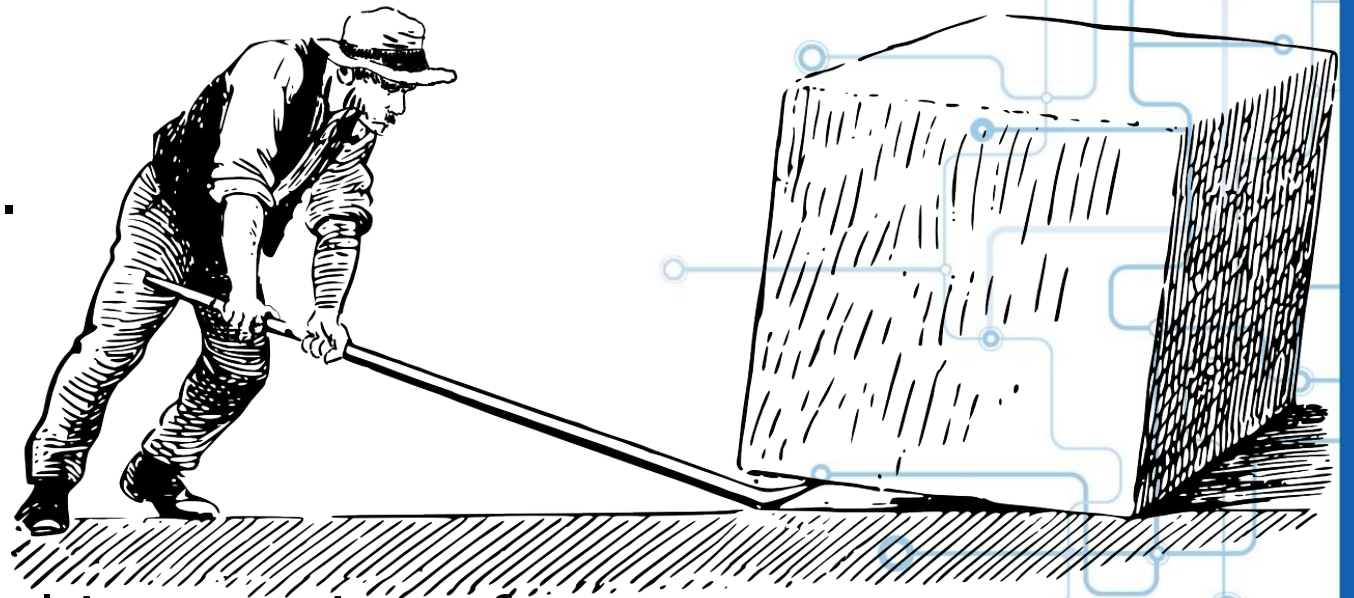
VM global leadership SUMMIT 2017

AIEye

Ethics Focus

Key “pivot points” in the development and deployment of AI systems in near to mid-term.

- Human-AI interactions.
- Reliability and assurance of trust.
- Encoding values
  - Enhance human freedom/autonomy.
  - Avoid domination, exclusion.



- Social changes required for ethical integration of autonomous systems into inclusive social order.

VM global leadership SUMMIT 2017

AIEye

Human-AI Interaction

- "Centaurs" are human / machine pairs.
- Centaurs can outperform unpaired humans and machines.
- (In)Effective centaurs
  - Sources of human error and machine error
  - Ecosystems to ensure continued competence.

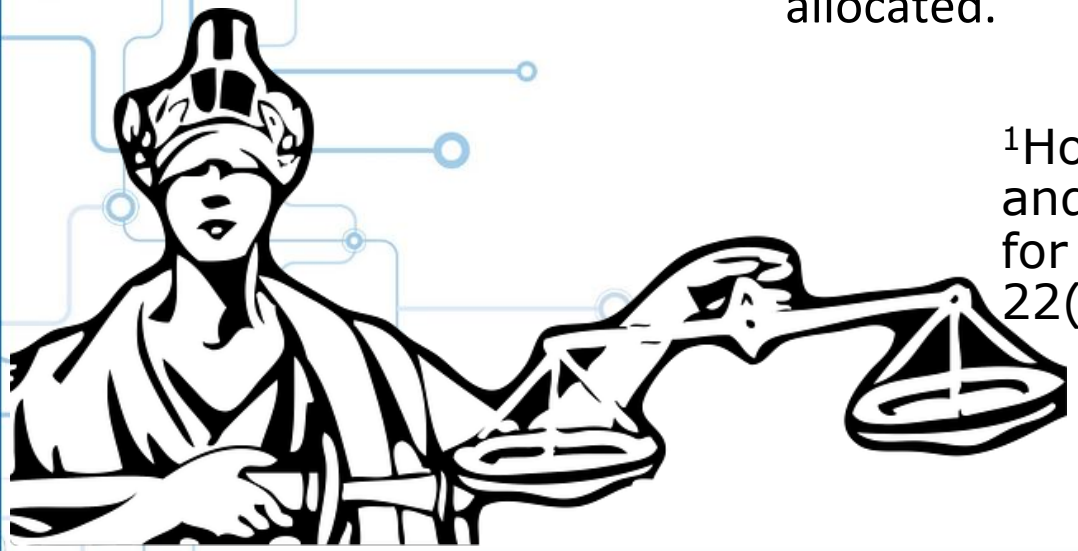
Mary Harrsch https://www.flickr.com/photos/mharrsch/9525926414

VM global leadership SUMMIT 2017

AIEye

AI Among Us

- Changing perceptions of unpaired acts?
- Accountability with varying degrees of access to machine "reasoning."
- Avoiding moral shortcomings:
  - Moral hazard: when protections against hazard promote risky behavior
  - Loopholes in morality[1]: when labor is divided but key responsibilities are not allocated.



[1]Hoss and London (2016) Assessing the Moral Coherence and Moral Robustness of Social Systems: Proof of Concept for a Graphical Models Approach. Sci Eng Ethics. 22(6):1761-1779.

VM global leadership SUMMIT 2017

AIEye

Reliability

- Universal or context-unbounded AI systems far off.
- Reliability depends on
  - Task competency in anticipated contexts.
    - Customer service bot: Accents, dialects, full range of customer needs
  - Windows of coping with contextual change or ambiguity.
  - User demands remaining limited to anticipated contexts.
    - Personal care bot: help me over the ledge,
      take me shopping, talk to my sister for me…

By FHKE (San Francisco Museum of Modern Art) [CC BY-SA 2.0 (http://creativecommons.org/licenses/by-sa/2.0)], via Wikimedia Commons

Reliability

- How are we validating system competency in anticipated contexts?
  - Caveat emptor?
  - Autonomous vehicles and standards[1]
- Quantifying likelihood of unanticipated context change?
- Validating mitigation measures or responses to failure?

[1]Danks and London (2017) Regulating Autonomous Systems: Beyond Standards. IEEE Intelligent Systems 32(1) Jan.-Feb.

VM global leadership SUMMIT 2017

AIEye

Encoding Values

- Autonomous systems necessarily encode task-specific values.
  - Tasks are *already* value laden!
- Challenge to make values explicit, transparent.
  - HR: what qualities does this system measure?
  - Care-bot: What does it do with all your private information?
- Challenge to incorporate sensitivity to broader values, avoid recapitulating biases and moral blindness, perpetuating histories of injustice.

VM global leadership SUMMIT 2017

AIEye

Human Impacts

- How do autonomous systems change the way we interact with one another?
  - Powerful levers of control/oppression?
  - Lower inhibitions to use force/violence?
  - Powerful tools for exploiting our cognitive limits/weaknesses.
- How do we promote social equality in society with centaurs and autonomous systems?
- Many questions here not about technology *per se*, but about how *we* will use and adapt to it.

VM global leadership SUMMIT 2017

AIEye

VM global leadership SUMMIT 2017

Alex John London, PhD

ajlondon@andrew.cmu.edu

Professor of Philosophy Director,
Center for Ethics and Policy

Carnegie Mellon University

AIEye

Artificial Intelligence
Supercharging Knowledge and Decision Making