BERKELEY CENTER FOR LAW & TECHNOLOGY

UC Berkeley School of Information

Handoffs with Integrity: Functional Fidelity and Values Protection in AI and Robotics

Deirdre K. Mulligan

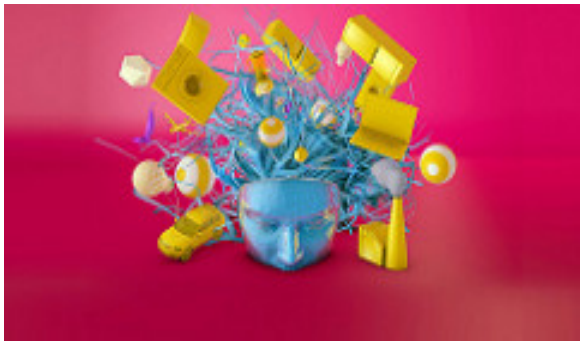
Robotics and Artificial Intelligence:
Policy Implications for the Next Decade
December 12, 2016

Functional Fidelity

Values: system

Intentional bias

Bias through process, values of designer



Complexity

Privacy: massive collection, inferences, breaking barriers of place and skin, influencing

Tumitu Design

Values: data

Poor selection

Inaccuracies

Selection bias

Bias in patterns of collection

Values at risk

Privacy

Autonomy

Fairness

Responsibility

Governance

Common challenges

- Transparency
- Accountability
- Bias
- Privacy

New or increased challenges posed by ML/AI/Robotics

- Complexity
- Dynamic systems—learning from data and experience
- Personalization—what does it mean to be fair
- Collapse of place and skin as structural protections

Governance responses (taken and proposed)

- Improved standards, certification, testing (red team; accessibility; real world conditions)
- Transparency (code reviews; documentation reviews)
- Software independence (auditing machines w/ machines)
- Expertise in agencies and ngo sector
- Updatability of systems
- Participation—transparent and inclusive public process
- Security and privacy foregrounded

Alignment of Actors and Governance

Translation work is critical—what does ballot or recount mean? What properties must a system have to support functions?

Focus on Socio-technical systems—people and code, algorithms as and in organizations—how do they shift power, responsibility, deal with differently situated domain expertise, how do the constraints of code and law interact? How do we assure assumptions and limits are correct, understood and shared.

How to study, manage values—not solve for them once and for all. As algorithms are on the move, so too are people, processes, etc. It's a dynamic ecosystem.

Social input into the guts of the machine and data—backend transparency is insufficient—need methods of participation.

Retooling Governance

- Communication and collaboration among lawyers, other domain experts, and engineers
  - Technical standards integral to governance
- Public goods—privacy, cybersecurity
  - market failures, illusion of control
- Regulators and public participation requires new tools and expertise
  - Explainability (processes, rules, data)
  - Experiential knowledge (simulations)
  - Exposing assumptions and limits of models
  - Public sector and ngo technologists