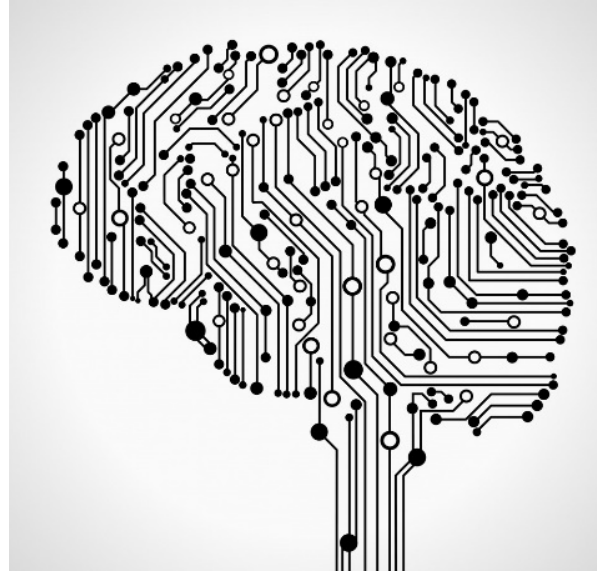


Public Risk Management for A.I.: The Path Forward



MATTHEW U. SCHERER, J.D.

BUCHANAN ANGELI ALTSCHUL &
SULLIVAN LLP

BENEFICIAL A.I. 2017

JANUARY 8, 2017

What is a “public risk?”

A potential source of harm that is:

- 1) Centrally or mass-produced or widely distributed
and
- 2) Outside the control of the individual risk bearer

Plain English: A public risk is something that could harm a lot of people, and individual potential victims have no way of stopping the harm from happening.

Examples

Nuclear technology
Environmental threats
Mass-produced consumer goods
Mechanized transportation
...Autonomous Systems

History of Public Risk Management

Human-generated public risks did not really exist prior to industrialization.

Modern forms of public risk management were developed because pre-industrial institutions were fundamentally incapable of managing the risks and harms generated by industrialization.

Industrial-Era Methods of Public Risk Management

	Formal	Informal
Preemptive	Legislation Agency rulemaking Subsidies	Industry standards
Reactive	Common law	Free market (consumer choice)

The Big Question for A.I. Risk Management

Will these industrial-era methods of risk management be capable of managing the risks associated with artificial intelligence and autonomous machines?

Shortcomings of Traditional Formal Regulation in Managing AI Risk

- Machines are not people
 - Existing legal systems operate by assigning and allocating legal rights and responsibilities to “persons”
- Discreetness
 - Risky AI development might be done in locations and using methods that escape detection by regulators
- Discreteness
 - Risks might stem from the interaction of components created at different places and times, without conscious coordination
- Diffuseness
 - Designers and manufacturers of components may be in different jurisdictions (and operators in yet other jurisdictions)
- Opacity
 - Regulators may not be able to discover or understand the underlying mechanisms that create risks

Shortcomings of the free market

- Information asymmetry
 - Producers have more information about risk than consumers
 - Particularly acute with emerging technologies
 - Failure of free market in managing industrial era risks is what led to rise of regulatory state
- Insurance?
 - Difficult to estimate risks with new technologies
 - Difficult to insure against large-scale public risks

Shortcomings of industry standards and self-regulation

- **Fox guarding the henhouse**
 - Industry effectively decides acceptable level of risk for public
 - Only works if industry's interests are very closely aligned with public at large.
 - Rarely the case for large companies, which traditionally generate most public risks.
- **Enforcement**
 - Market participants can avoid restrictions by simply leaving (or never joining)

Shortcomings of industry standards and self-regulation

Usually results in one of two outcomes

Weak and ineffective standards

or

Strong but never-enforced standards

Could it be different with AI? Reasons to be optimistic about industry standards

- **Early buy-in from key industry players**
 - Industry is not waiting for harm to occur
 - Reflected in private sector initiatives encouraging transparency and beneficial AI
 - Partnership on AI
 - OpenAI
- **Broad stakeholder participation**
 - AI community has been welcoming of input from policy, ethics, and legal experts
 - Stakeholders have direct role in many current initiatives
 - Means that industry standards will align better with interests of society at large
- **Emerging consensus of goals**

Could it be different with AI? Reasons to be Optimistic About Industry Standards

- → Could be amplified by common law
 - System where legal rules do not come from fixed legislation, but instead develop organically from precedents derived from resolving individual disputes
 - Dominant system in English-speaking world
 - Elements becoming increasingly common in other countries (esp. in Asia and Africa)
 - Mechanism: Tort law
 - People have “duty of care” to avoid causing harm
 - When true industry consensus emerges and is reflected in written standards, those standards can effectively become law through adoption by common law courts as part of the duty of care

Current Developments: IEEE

- Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems
 - Divided into committees focused on different aspects of ensuring ethically aligned design (law, economics, privacy, etc)
 - Dec 2016: Published draft version of *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems (AI/AS)*
 - In public comment period
 - In other words, this is the time to engage

Current Developments: IEEE

- Sample standard: “Transparency in autonomous systems”
 - Will set standards so that autonomous systems will “be transparent to a wide range of stakeholders,” including users, regulatory entities, accident investigators, attorneys and expert witnesses, and society at large
 - Recognizes that “transparency” will mean different things for (and to) different groups of stakeholders
- General consensus that openness is good, but...
 - Intellectual property concerns
 - FRAND is not so useful if tech can easily be pirated
 - Security concerns
 - Might give bad actors blueprint to do harm

Other Current Developments

- **Governance Coordinating Committees**
 - Proposed by Wendell Wallach and Gary Marchant
 - Institutional mechanism for coordinating risk management efforts by both governmental and non-governmental actors
- **Building Global Infrastructure to Ensure AI and Robotics Are Beneficial**
 - Spearheaded by Wendell Wallach
 - Joint effort by The Hastings Center and The Carnegie Council for Ethics in International Affairs
 - Will seek to act as an “issues manager” for different institutional stakeholders worldwide
- **Sector-specific**
 - Autonomous weapons
 - Autonomous vehicles

In Closing

- The momentum generated by this conference and others like it must not only be sustained, but intensified: we cannot stop with general statements of principles
- There is a rare opportunity here to shape policy without going to a government for approval
- Don't fear law, policy, and ethics people: engage them (and each other!)