AI Safety and Accountability:
Some Questions on Liability and Regulation

Cary Coglianese
Edward B. Shils Professor of Law and Political Science
Director, Penn Program on Regulation
University of Pennsylvania Law School

AAAI Conference · February 9, 2023

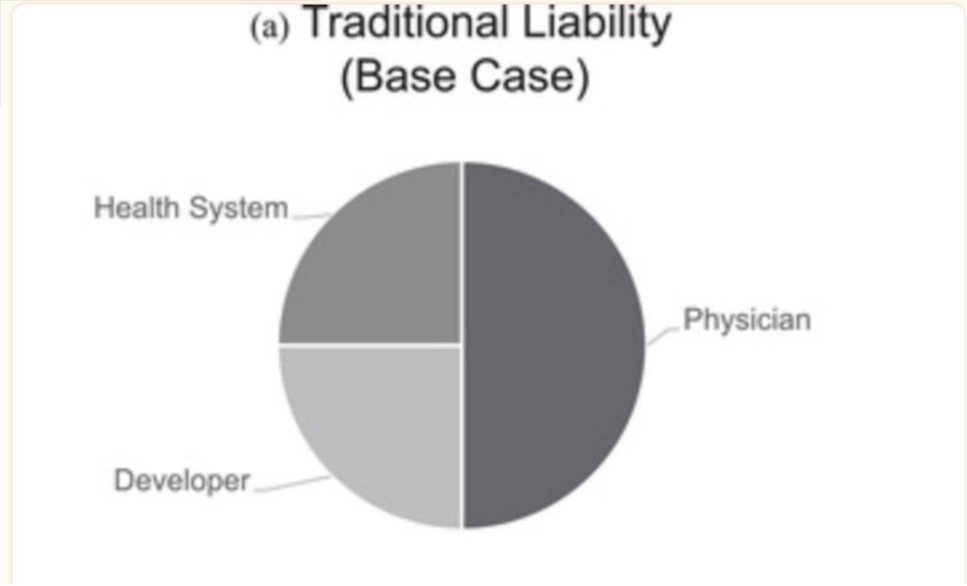
Penn Program on Regulation
University *of* Pennsylvania

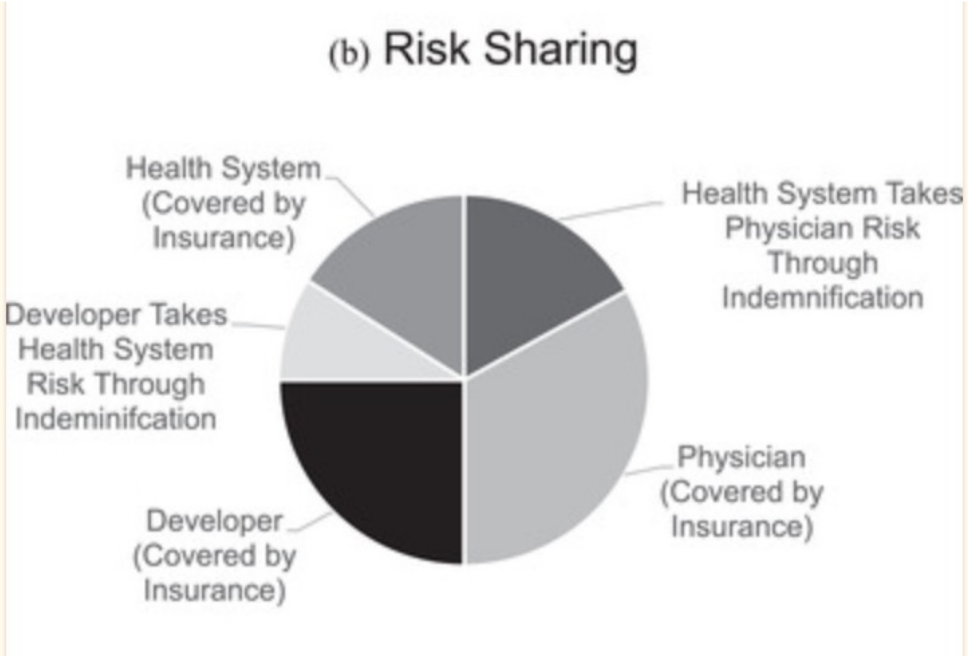
Law: Two facets

- Retrospective **liability** – *assigning responsibility after harm occurs*

- Prospective **regulation** – *imposing obligations before harm occurs*

Retrospective liability





Maliha, Gerke, Cohen & Parikh (2021)

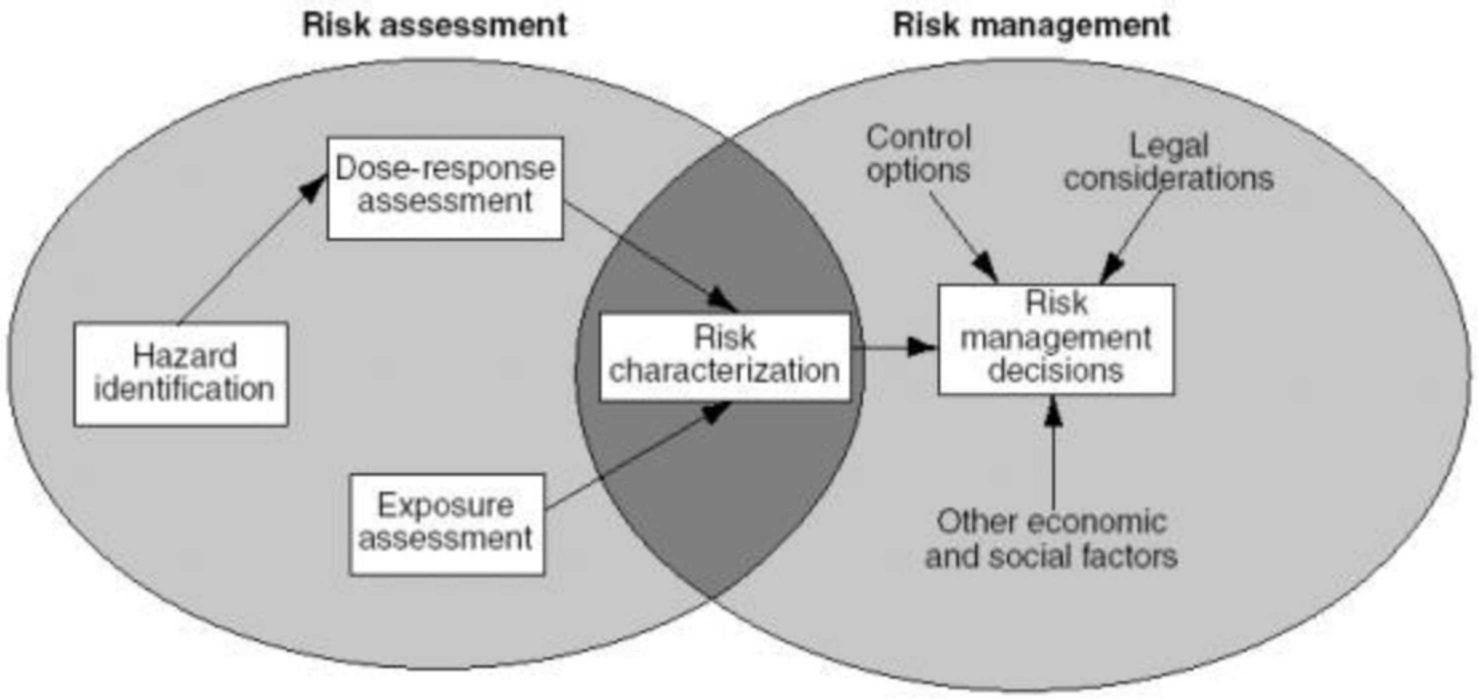
Prospective regulation

- Expect legislative and regulatory action in the coming years

- A prevailing regulatory trend is to be “risk-based”
  - Proposed EU legislation would place different AI uses into different “risk” categories and subject them to differential levels of regulatory scrutiny
    - 3 categories: “unacceptable risk,” “high risk,” and “low or minimal risk”
    - Risks are to “health and safety or fundamental rights of persons”
  - U.S. NIST has recently issued a voluntary “risk management framework” for AI

Risk assessment vs. risk management

Figure 1. Diagram of NRC risk assessment/risk management paradigm.



Source: EPA Office of Research and Development.

Some questions about "risk-based" AI regulation

- **Which harms count?**
  - Whose answers to this question count? (Expert risk judgments vs. lay risk judgments)
- **Can different types of harms be placed into commensurable units to allow comparisons (high risks v. low risks)?**
  - Can privacy harms be compared with health harms?
  - Should harms be monetized to facilitate comparisons?
- **Should regulatory decisions be based on population risks or individual risks?**
  - Does it matter if overall number of people harmed is reduced but those who do suffer harm may incur more severe consequences?
  - Does it matter how risks are *distributed*?

Some questions about “risk-based” AI regulation

- **How should the benefits of AI factor into regulatory decision-making?**
    - If some domain is truly high risk, shouldn’t we worry about leaving that to humans?
- **Which normative principle should govern risk decision making?**
    - Hippocratic / “do no harm”?
    - Reduce risk to an “acceptable level” (What is acceptable?)
    - ALARP: “as low as reasonably practicable” (What is reasonable?)
    - Maximize net benefits?

Some questions about “risk-based” AI regulation

- **Who should bear the burden of proof with respect to AI safety?**
  - *Precautionary principle/safety case regulation*: Should AI developers be required to demonstrate “safety” and receive approval before use?
  - *Reactive/recall regulation*: Should government be required to show that AI tool is “unsafe” before it can be removed from the market?

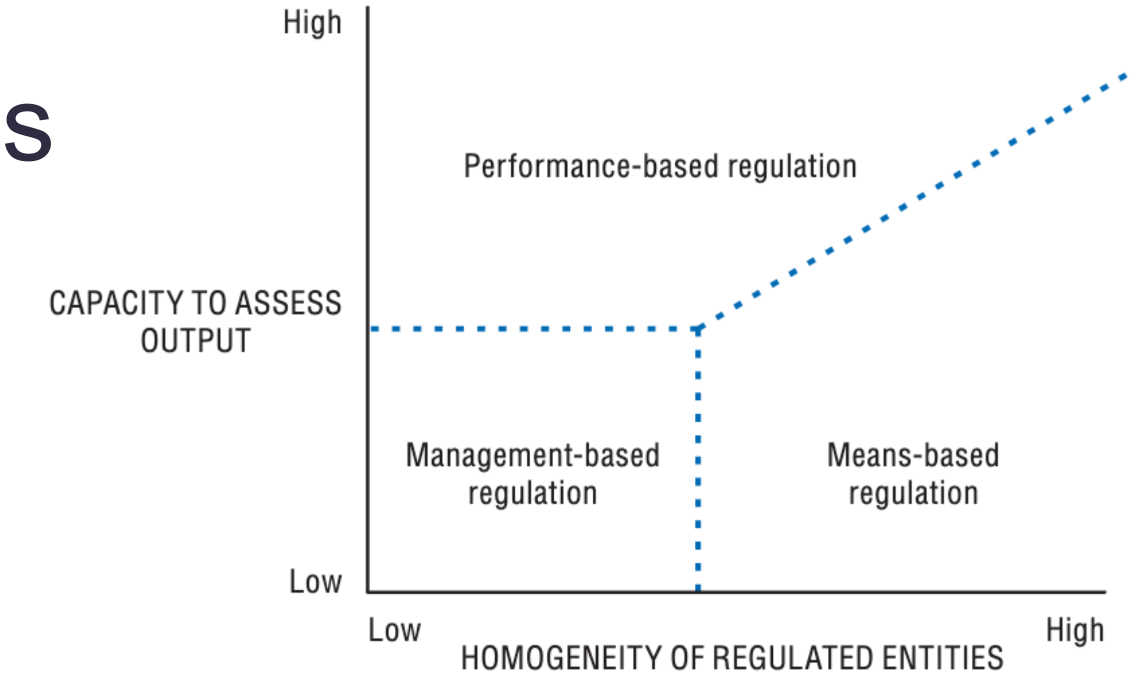
General Questions About Regulation

- **Upon whom should regulatory obligations be imposed?**
  - Designers? Users?
  - Firms? Professionals/individuals?
- **What institution(s) should impose these obligations?**
  - Dedicated AI commission? Or separate regulatory agencies (e.g., NHTSA, FDA)?
- **What should the obligations require?**
  - ”Prescriptive”: Specified means/actions/designs?
  - “Performance-based”: Achievement or avoidance of specified outcomes?
  - “Management-based”: Sound risk management (e.g., responsible AI, AI auditing)?

Key Factor Affecting Choices About Obligations: AI’s Heterogeneity

- Different algorithms
- Different data
- Different uses

Figure 5.1. **Conditions for use of means-based, performance-based, and management-based regulation**


High
CAPACITY TO ASSESS OUTPUT
Low
Performance-based regulation
Management-based regulation
Means-based regulation
Low
High
HOMOGENEITY OF REGULATED ENTITIES

Coglianese (2010)

AI Safety and Accountability: The Core Challenge for Law

How can the law ensure adequate incentives for socially and economically optimal AI risk management by those who design, develop, and use AI tools?

Possible overarching questions to motivate joint AI-law research

1. **What are the risks from AI?**
   - Can standard methods be developed for assessing, quantifying, and characterizing AI risks? Can risks and benefits be monetized? (How will the answers to this question be different for AI than for risk assessment and characterization in any other domain?)

2. **What is the optimal level of AI risk management?**
   - What are best practices in AI risk management? How often should AI be validated/monitored/audited?

3. **How to create adequate incentives?**
   - Liability or regulation? What are the best rules? How enforced?
   - Role for third-party auditing and certification?

Background readings

Cary Coglianese & Kat Hefter, “From Negative to Positive Algorithm Rights,” *William & Mary Bill of Rights Journal* 30:883-923 (2022)

Cary Coglianese & Alicia Lai, “Algorithm vs. Algorithm,” *Duke Law Journal* 72:1281-1340 (2022)

Cary Coglianese & Alicia Lai, “Antitrust by Algorithm,” *Stanford Journal of Computational Antitrust* 2:1-22 (2022)

Cary Coglianese, “Moving Toward Personalized Law,” *University of Chicago Law Review Online* (2022)

Cary Coglianese & Steven M. Appel, “Algorithmic Administrative Justice,” in Marc Hertogh, Richard Kirkham, Robert Thomas and Joe Tomlinson, eds., *The Oxford Handbook of Administrative Justice* 481-502 (Oxford University Press, 2022)

Cary Coglianese & Lavi Ben Dor, “AI in Adjudication and Administration,” *Brooklyn Law Review* 86: 791-838 (2021)

Cary Coglianese, “Regulating New Tech: Problems, Pathways, and People,” *TechReg Chronicle* 1: 65-73 (December 2021)

Cary Coglianese & Lavi Ben Dor, “Procurement as AI Governance,” *IEEE Transactions on Technology & Society* 2: 192-199 (2021)

Cary Coglianese, “Administrative Law in the Automated State,” *Dædalus* 150(3): 104-120 (2021)

Cary Coglianese & Erik Lampmann, “Contracting for Algorithmic Accountability,” *Administrative Law Review Accord* 6:175-199 (2021)

Cary Coglianese, “Algorithmic Regulation: Machine Learning as Governance Tool,” in Marc Schuilenburg & Rik Peeters, eds., *The Algorithmic Society: Power, Knowledge and Technology in the Age of Algorithms* 35-52 (Routledge, 2021)

Cary Coglianese & Steven M. Appel, “Algorithmic Governance and Administrative Law,” in Woodrow Barfield, ed., *Cambridge Handbook on the Law of Algorithms: Human Rights, Intellectual Property, Government Regulation* 162-181 (Cambridge University Press, 2021)

Cary Coglianese, “Deploying Machine Learning for a Sustainable Future,” in Daniel Esty, ed., *A Better Planet: Forty Big Ideas for a Sustainable Future* 200-208 (Yale University Press, 2019)

Cary Coglianese, “Optimizing Regulation for an Optimizing Economy,” *University of Pennsylvania Journal of Law and Public Affairs* 4:1-13 (2018)

Cary Coglianese & David Lehr, “Transparency and Algorithmic Governance,” *Administrative Law Review* 71:1-56 (2019)

Cary Coglianese & David Lehr, “Regulating by Robot: Administrative Decision-Making in the Machine Learning Era,” *Georgetown Law Journal* 105:1147-1223 (2017)

Cary Coglianese, “Management-Based Regulation: Implications for Public Policy,” in Gregory Bounds and Nikolai Malyshev, eds., Risk and Regulatory Policy: Improving the Governance of Risk (OECD Publishing, 2010)

Cary Coglianese
Edward B. Shils Professor of Law
Professor of Political Science
Director, Penn Program on Regulation
University of Pennsylvania

www.PennReg.org
www.TheRegReview.org

The Regulatory Review

PENN PROGRAM ON REGULATION