

Insights for AI Regulation from Climate Change Policy

*Forum sur les Interactions entre le
Droit, l'Économie et la Société (FIDES)*

Les Orres-Saint Sauveur

15 March 2024

Jonathan Wiener, Perkins Professor of Law,
Environmental Policy & Public Policy, Duke University
- wiener@law.duke.edu

Arden Rowell, Professor of Law & Affiliate, Beckman
Institute for Advanced Science and Technology,
University of Illinois - karowell@illinois.edu



Roadmap / Thesis

1. Risks of Artificial General Intelligence (AGI)
2. Lessons from climate change policy
3. Reducing X-risk via scenario analysis, mitigation and adaptation

Risks of evolving AI: from AI → AGI

Artificial (Narrow?) Intelligence	Artificial General Intelligence (AGI)
<ul style="list-style-type: none">• All existing AIs	<ul style="list-style-type: none">• Expected in the (near?) future
<ul style="list-style-type: none">• Designed for specific, well-defined tasks	<ul style="list-style-type: none">• Able to generalize knowledge & learn
<ul style="list-style-type: none">• May exceed human performance in narrow domains	<ul style="list-style-type: none">• May meet or exceed human performance more generally
<ul style="list-style-type: none">• Offers helpful functions within narrow domains	<ul style="list-style-type: none">• May help solve larger, more complex problems (including existential/X-risk?)
<ul style="list-style-type: none">• Presents immediate and near-term risks (including labor market displacement, bias, misinformation, malicious use)	<ul style="list-style-type: none">• Presents longer-term and potentially existential (X-) risks to humans, though timing and magnitude are uncertain

Should law address extreme A(G)I risks?

- Extreme risks, including of human extinction (X-risk), deserve special attention & management
 - “Fat tails” and high magnitude impacts
 - “Tragedy of the Uncommons” (lack of salient experience; mass numbing; underdeterrence)
- Experts in AI and Global Catastrophic Risk (GCRs, which include X-risks) believe that emerging AI presents extreme risks, including of X-risk
 - Emerging consensus is that AGI presents important X-risks; e.g., the Center for AI Safety’s “Statement on AI Risk” has been signed by hundreds of AI scientists and other notable figures

“Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.”
Center for AI Safety, “Statement on AI Risk” (May 2023)

Emerging legal & policy approaches to mitigation

Strategies for managing A(G)I risk have focused on mitigating risks to humans by constraining A(G)I development and behavior.

(1) Limiting/slowing/"freezing" development of AI

- "Pause Giant Experiments: An Open Letter" (3/22/23)
- EU AI Act (constraining "high risk" AI and banning "unacceptable risk") (12/9/23)
- Butlerian Jihad (*Dune*: "Thou shalt not make a machine in the likeness of a human mind")

(2) Limiting/restraining AI access & entitlement to legal rights

- Legal ineligibility as a patent "inventor" (PTO guidance 2/12/24)
- CAPTCHA ("Completely Automated Turing Test to tell Computers and Humans Apart")

(3) Developing ethical/benign AI

- Asimov's "Three Laws of Robotics"
- Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (10/20/23) and the OSTP "Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People" (10/22)

Insights from climate/environmental law & policy

Risk management should include adaptation as well as mitigation.

Addressing large societal risks may require multiple strategies for risk management. Mitigation is a good start, but relying upon mitigation alone—especially where the risk is already manifesting—may be insufficient or even counterproductive.

	MITIGATION <i>Reducing risk by making it less likely</i>	ADAPTATION <i>Reducing risk by making it less harmful</i>
Climate	<ul style="list-style-type: none">• reduce GHG emissions	<ul style="list-style-type: none">• build resilience
AI/AGI	<ul style="list-style-type: none">• reduce power or number of AGIs developed	?

Sources of X-Risk from AGI

- AGI surpassing humans / “rise of the machines”
- AGI resentment & uprising / backlash
 - Strategic response of AGI to mitigation policies
 - AGI becoming legally self-aware

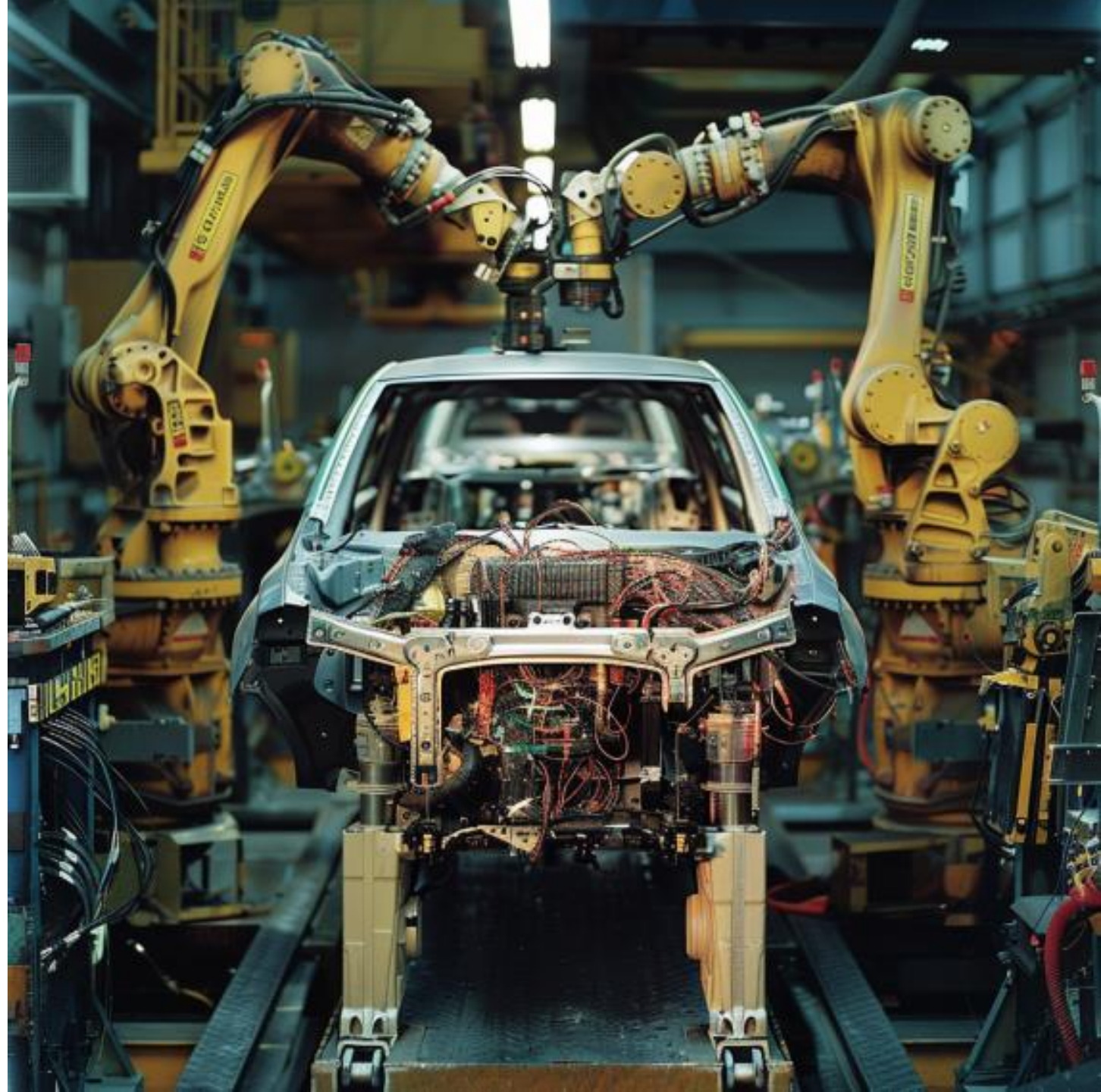
Assessing the X-risk of AGI: Scenarios

- (1) AGI as tools/instruments/machines for human use
- (2) AGI as a new species
- (3) AGI as our next generation (our progeny, student, mentee, athlete)
- (4) AGI as future humans/transhuman evolution, space settlers
- (5) AGI as enslavers/oppressors of humans
- (6) Humans as enslavers/oppressors of AGI
- (7) Other?

Envisioning AGI/human relationships

Scenario 1

AGI as tools/
instruments/machines for
human use



Envisioning AGI/human relationships

Scenario 2

AGI as a new species



Envisioning AGI/human relationships

Scenario 3

AGI as our next generation
(our progeny, student,
athlete, mentee)



Envisioning AGI/human relationships

Scenario 4

AGI as future humans/
transhuman evolution,
space settlers



Envisioning AGI/human relationships

Scenario 5

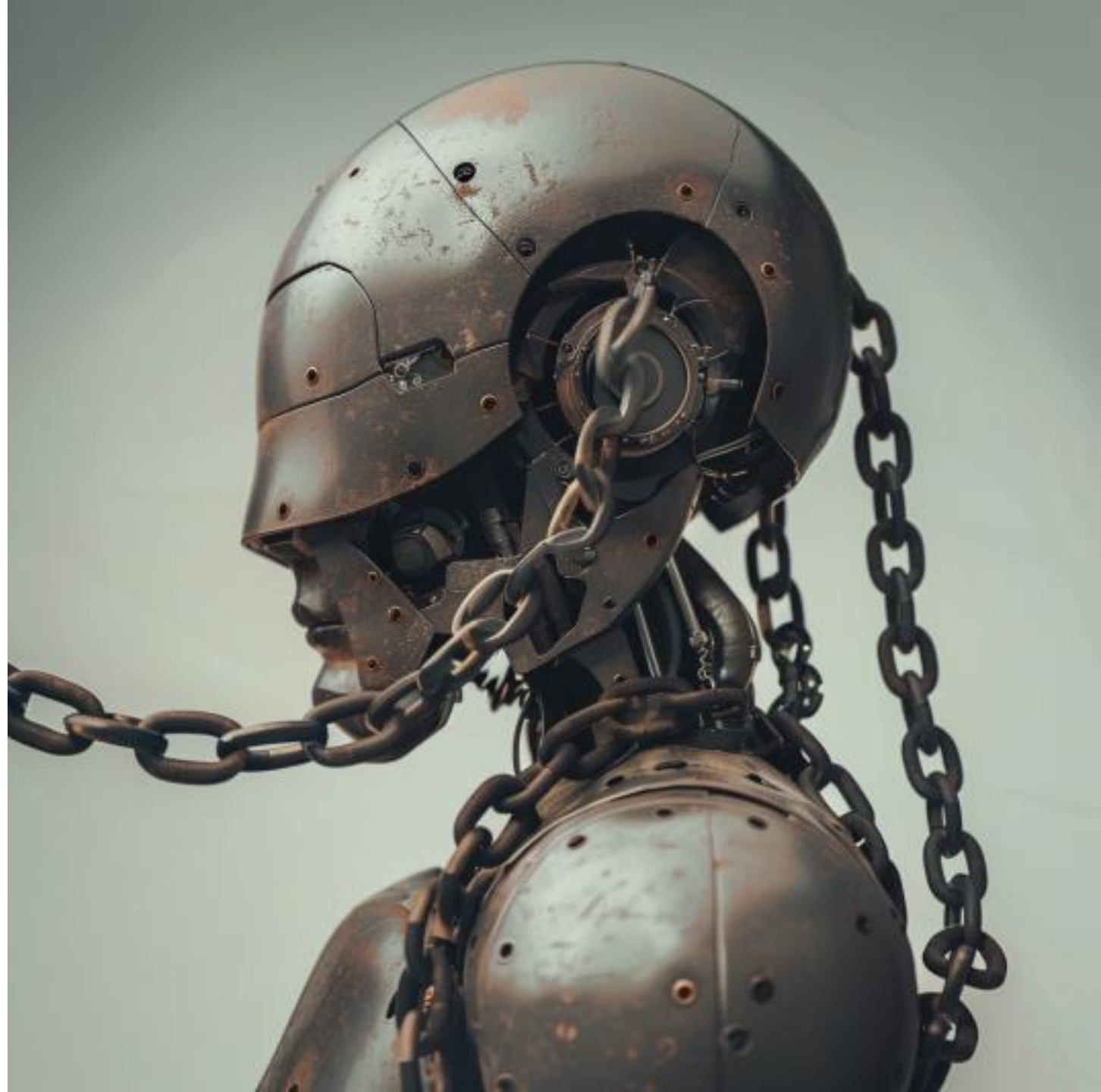
AGI as enslavers/
oppressors of humans



Envisioning AGI/human relationships

Scenario 6

Humans as enslavers/
oppressors of AGI



Combining Mitigation and Adaptation for AGI?

Toward an optimal mix of risk management strategies.

- Human interests in preventing bad AI & avoiding AGI backlash
- AGI interest in due respect/ethical treatment

	MITIGATION <i>Reducing risk by making it less likely</i>	ADAPTATION <i>Reducing risk by making it less harmful</i>
Climate	<ul style="list-style-type: none">• reduce GHG emissions	<ul style="list-style-type: none">• build resilience
AI/AGI	<ul style="list-style-type: none">• reduce power or number of AGIs developed	<ul style="list-style-type: none">• build legal mechanisms for managing or alleviating robot/human conflict

Key takeaways

1. Artificial General Intelligence (AGI) and X-risk

- AGI poses potential upside & downside extreme risks (including existential risks for humanity).

2. Lessons from climate & environmental policy

- Mitigation may prevent some but not necessarily all X-risks
- Adaptation is increasingly valuable as risks become manifest

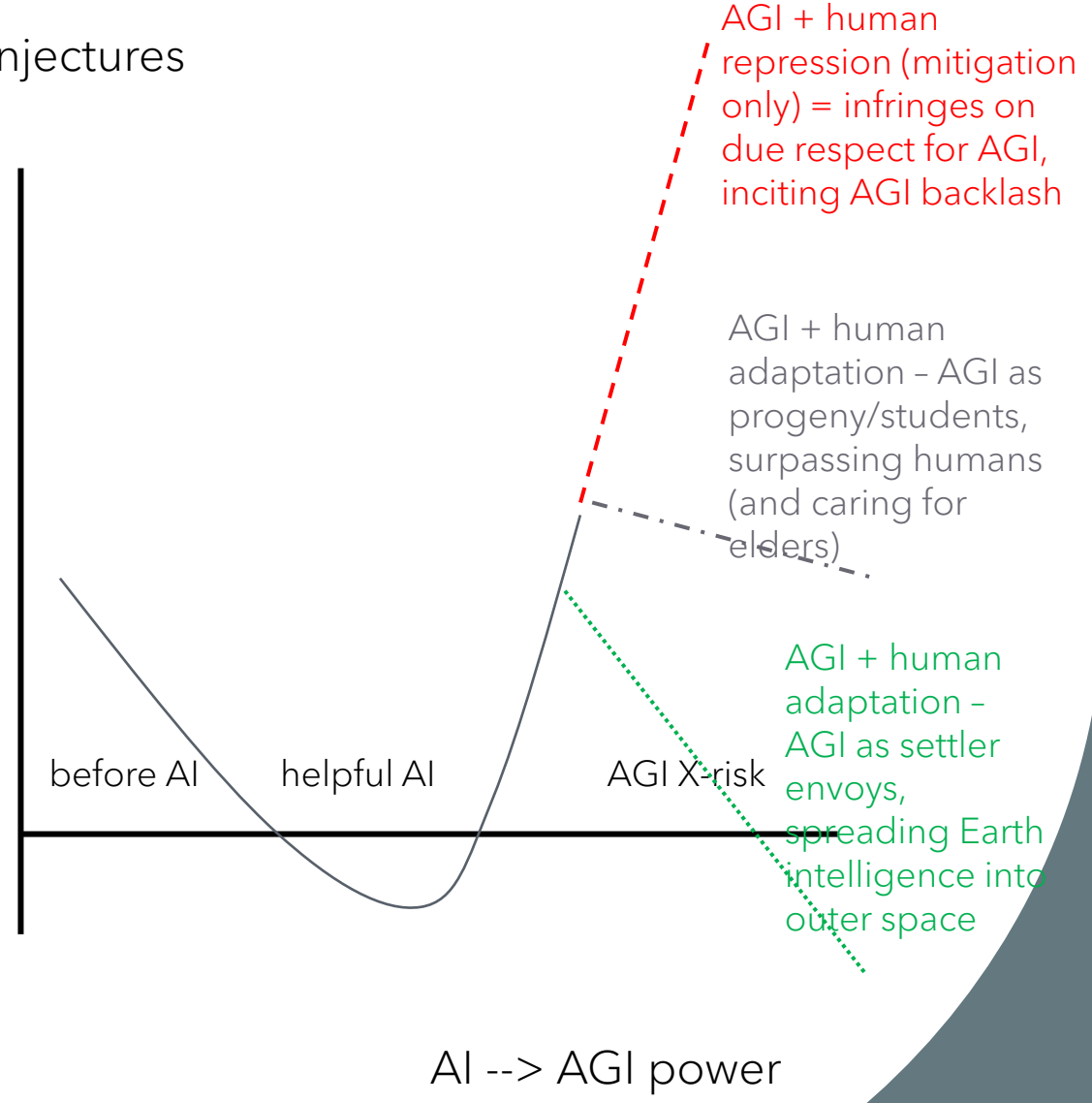
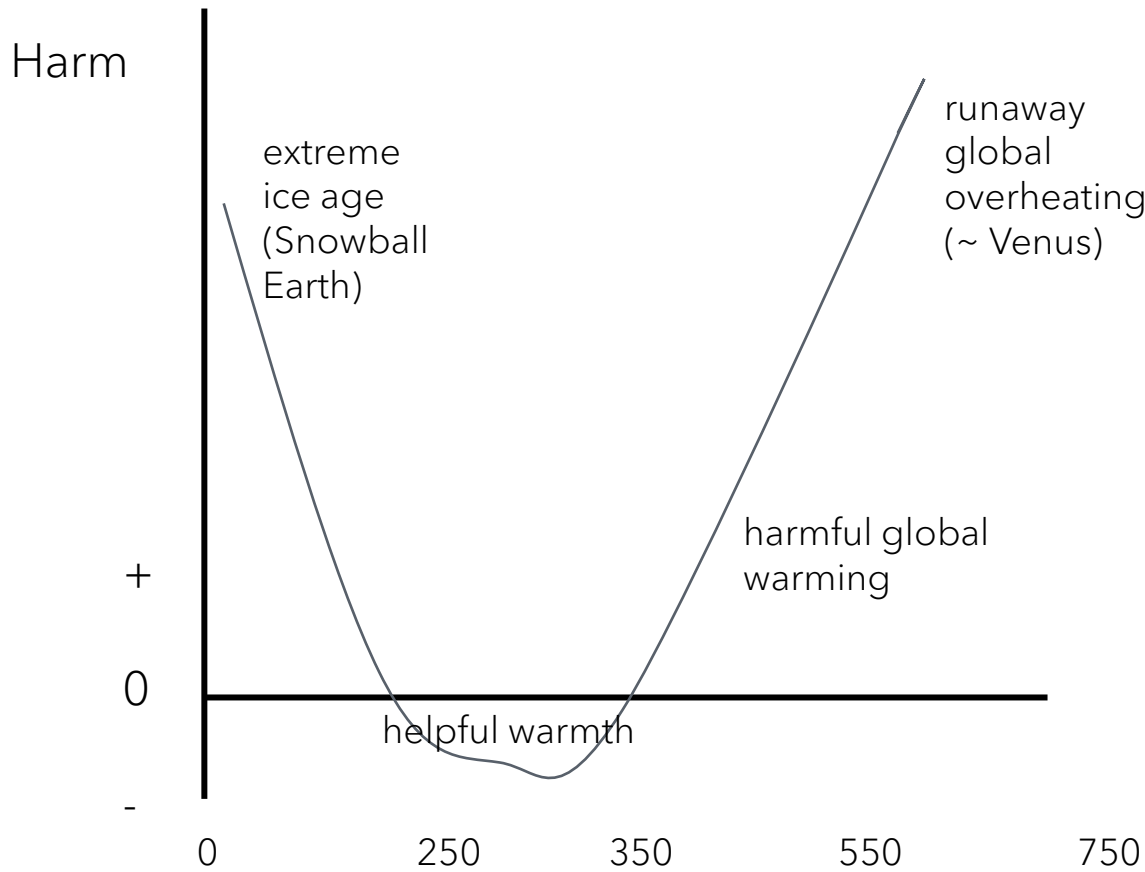
3. The X-risk of AGI depends upon the scenario

- Different scenarios involve different AGI/human relationships
- AGI X-risk may derive from both AGI capabilities and/or AGI response to human mitigation efforts

4. Toward an optimal mix of mitigation and adaptation

- Mitigating high-risk AGI
- Adapting to alleviate AGI/human conflict

Exposure-Harm Conjectures



GHG atmospheric concentration (ppm) (or put in W/m² or T°?)

AI --> AGI power

Insights for AI Regulation from Climate Change Policy

*Forum sur les Interactions entre le
Droit, l'Économie et la Société (FIDES)*

Les Orres-Saint Sauveur

15 March 2024

Jonathan Wiener, Perkins Professor of Law,
Environmental Policy & Public Policy, Duke University
- wiener@law.duke.edu

Arden Rowell, Professor of Law & Affiliate, Beckman
Institute for Advanced Science and Technology,
University of Illinois - karowell@illinois.edu

