

Economics of Transformative AI

Spring 2026 Fellowship

UChicago Existential Risk Laboratory (XLab)

Introduction

As artificial intelligence (AI) advances, companies and countries face economic pressures to prioritize capability development over safety. This dynamic increases catastrophic and existential risks from artificial general intelligence (AGI), leading to calls to pause or slow AI capability development.

This fellowship will examine how economics shape AI safety decisions, and vice versa. We will use economic reasoning to 1) explore future trajectories of AI and its impacts on society, and 2) understand how economic tools such as markets, industrial policy, regulation, and international co-ordination can meaningfully reduce existential risk from AI.

Some big questions we hope to explore include:

1. What economic forces drive under-investment in AI safety? How do they interact with existential risk?
2. If transformative AI or AGI arrives, what will the macro-economic transition look like? Who are the key players?
3. What will post-AGI or artificial super-intelligence (ASI) economies look like? How can humanity steer and prosper in an AI-dominated economy?

Structure

Over five weeks, this fellowship will provide an overview of topics at the intersection of economics and AI safety. The focus will be on existential risk, though other areas will be touched on as needed. At the end, fellows should walk away with a strong understanding of the field, key gaps, and how they can contribute. No specific background is needed, but this fellowship assumes basic understanding of economic principles and the general dynamics of AI safety. Debate and disagreement are encouraged — the field is new, experts disagree among themselves, and things move rapidly.

Each week (two hours) will focus on one topic, motivated by a selection of required readings (which should take around an hour). We also include optional resources that expand on the required readings. For a comprehensive overview of the latest academic economics literature, we recommend Agrawal, Brynjolfsson, Korinek (2025): [The Economics of Transformative AI](#).

During the fellowship meetings, we will reflect on and discuss the readings before diving into a group activity. We will assume familiarity with the primer reading.

The fellowship will meet 5–7pm on Thursdays. Dinner will be provided.

Contacts

Feel free to contact the facilitators at any time:

- Justin Chen: justinchen@uchicago.edu
- Kailee Kuan: kaileekuan@uchicago.edu

Weekly schedules

Week 1: Why does the economics of AI matter for existential risk?

Non-aligned AI could be dangerous, potentially resulting in existential risks. Many such risks would transmit (at least in part) through economic channels. Conversely, economics offers many toolkits and frameworks that can help society navigate AI. This week will provide a primer on the links between AI safety and economics, drawing upon latest scenarios, forecasts, and threat models.

Driving questions

- What distinguishes AI from other transformative technologies? Why does that matter for existential risk?
- How do forecasts about AGI timelines shape what economic policies are worth considering?
- How should we think about uncertainty? What epistemic tools are most useful (e.g. scenarios, empirical models, theoretical models, case studies, etc.)?

Readings

Type	Reading	Why read it
What might transformative AGI look like?		
Require	AI Futures Project (2025), AI 2027	Provides concrete scenarios and reference points for AI development and how things could go right/wrong.
Optional	Growiec and Prettnner (2025), The Economics of p(doom)	More formal economic treatment/modelling of various AI scenarios, and what they mean for safety.
Optional	METR (2025), Task-Completion Time Horizons of Frontier AI Models	Popular empirical framework for thinking about capabilities.
Optional	Forecasting Research Institute (2026), Forecasting the Economic Effects of AI	Latest survey of views on AI capabilities and economic impacts from economists, experts, and general public.
Optional	Dwarkesh Podcast with Carl Shulman (~1:34-2:09)	Comprehensive overview of AI alignment and trajectories.
Why should economists care?		
Require	Korinek (2023), Scenario Planning for an A(G)I Future	Explains how economic policymaking needs to adapt to different AI scenarios.
Optional	Davidson (2021), Could Advanced AI Drive Explosive Economic Growth?	Goes through economic frameworks/arguments behind AI take-off (more approachable version here).
Optional	Jones (2024), The AI Dilemma	Models tension between risks and benefits of AI.

Activity

- **Spectrum:** Facilitators will read a series of claims around the economics of AI. The group will line up based on how much they agree or disagree with each claim, and explain why they chose their position.

Week 2: How might AI impact economic growth?

Experts disagree on how AI might impact economic growth. Under some scenarios, superintelligence combined with advanced robotics could boost growth so much that current institutions struggle to keep up and the share of AI-related activities in GDP approaches 100%. Under other scenarios, AI might boost productivity but long-term growth rates might not change much from the baseline, due to bottlenecks such as physical infrastructure, raw materials, or an AI capabilities slowdown. This week will explore these disagreements and what different scenarios mean for AI safety.

Driving questions

- How much could AI accelerate economic growth? What does such an economy look like?
- Does automated R&D change the nature of growth itself? What does that mean for who benefits?
- How does AI impact GDP across countries? Is GDP a good metric for measuring AI?

Readings

Type	Reading	Why read it
Arguments for explosive growth		
Require	Clancy (2025), What If We Could Automate Invention?	Brief explanation of intuition behind automated R&D.
Optional	Trammell and Korinek (2023), Economic Growth Under Transformative AI	Formal mathematical treatment of automating R&D and impacts on growth, as well as potential barriers.
Optional	Aghion, Jones, and Jones (2018), Artificial Intelligence and Economic Growth	Seminal paper on modelling AI with endogenous growth theory.
Optional	Forethought (2025), Three Types of Intelligence Explosions	Provides a framework to distinguish between different types of explosive growth.
Optional	Forethought (2025), Could One Country Outgrow the Rest of the World?	Explores how AI could break assumptions/historical trends in growth theory and development economics.
Arguments for modest impacts		
Require	Cunningham (2025), Economic and Transformative AI	Key arguments against explosive growth, overview of disagreements between economists and AI researchers.
Optional	Acemoglu (2025): The Simple Macroeconomics of AI	Leading skeptic of explosive growth due to task-level bottlenecks.
Optional	Anthropic (2026), Understanding AI's effects on the economy	Empirical data on how AI is currently being used in the economy.
Optional	Epoch AI (2025), Compute scaling will slow down due to increasing lead times	Argues for physical constraints to AI capability growth.

Activity

- **Scenario forecasting:** Working in small groups, fellows will pick one model of economic growth presented above and develop a plausible scenario on what the world might look like over the next ten years under that model. Fellows will then pitch their scenarios to the room.

Week 3: How might AI impact jobs and power?

AI is already changing the nature of jobs. What could AGI/ASI do to the labour market? Some argue that ASI, being at least as good as the best human at any computer-based task, could wipe out white-collar employment. Further, ASI could develop robotics that automate physical tasks, leaving an ever diminishing niche for human labour. Others argue that AI could never do all human jobs, or that even so, comparative advantage will always leave a role for human labour. Economies that rely more on AI and less on human labour could concentrate power in the hands of the few who control AI and capital. This week will unpack these dynamics and explore how society might function in an AI-driven economy.

Driving questions

- Does AI augment human labor or replace it? Will humans have any competitive or comparative advantages left?
- How will AI impact who has power? How do these dynamics map on to existential risk?

Readings

Type	Reading	Why read it
<i>Labour displacement</i>		
Require	Korinek and Suh (2024), Scenarios for the Transition to AGI	Considers how wages might change under different AGI scenarios, including potential wage collapses.
Optional	Restrepo (2025), We Won't be Missed: Work and Growth in the AGI World	Formal treatment of different labour economics models and they mean under AGI.
Optional	Gans and Goldfarb (2024): O-Ring Automation	Formal models of how AI complements/substitutes for human labour.
Optional	Oks (2026), Why ATMs didn't kill bank teller jobs, but the iPhone did	Intuition of the different ways to think about job displacement at different levels and timescales.
Optional	Stevenson (2025), What is There to Fear in a Post-AGI World?	Explores how human life and meaning might look like in a world where human labour is no longer needed.
<i>Power concentration and inequality</i>		
Require	Drago and Laine (2025), The Intelligence Curse	Explains how economies that no longer need human labour may stop valuing human welfare.
Optional	Korinek and Stiglitz (2026), Steering Technological Progress	Formal argument for the need to focus on human wellbeing and re-distribution as the value of human labour diminishes.
Optional	Beraka, Kao, Yang, and Yuchtman (2021), AI-tocracy	How AI empowers autocratic regimes, creating feedback loops between AI, surveillance, and political control.
Optional	Korinek and Stiglitz (2021), AI, Globalization, and Strategies for Economic Development	Explores how winner-take-all dynamics in AI could prevent developing economies from benefitting from AI.
Optional	Athey and Scott Morton (2025), AI, Competition, and Welfare	Models competition between AI labs and argues that the welfare impacts of AI depend on market structures

Activity:

- **Stakeholder panel:** Each fellow will assume an economic role (e.g. white-collar worker, lab founder, developing country finance minister) and respond in character to facilitator questions and comment on policy proposals.

Week 4: What are the strategic dynamics between AI actors?

Existing strategic incentives are among the greatest drivers of unsafe AI development. Economic tools such as game theory help explain why dangerous racing dynamics persist between frontier labs and between countries. Microeconomic concepts such as externalities and asymmetric information also shed light on how governments and markets can shape the incentives of AI labs. This week will explore these dynamics primarily through an interactive table-top exercise.

Driving questions

- Why are frontier labs and countries racing to AGI while under-investing in safety?
- What are the geopolitical dynamics of AI safety?
- Are there ways to get out of these racing dynamics?

Readings

Type	Reading	Why read it
Why racing dynamics arise		
Require	Greutzemacher, Avin, Fox, and Saeri (2024), Strategic Insights from Simulation Gaming of AI Race Dynamics	Explains common patterns and potential pitfalls in the AI race that are leading society on a path to high existential risks
Optional	Armstrong, Bostrom, and Shulman (2013), Racing to the Precipice	Foundational game theory framework for an AI arms race.
How to tackle racing dynamics		
Require	RAND (2025), Understanding the AI Diffusion Framework	Explains how compute governance can shape global AI development
Optional	Hendrycks, Schmidt, Wang (2025), Deterrence with Mutual Assured AI Malfunction (MAIM)	Argues that mutual destruction will deter an AI-driven geopolitical monopoly similar to nuclear security
Optional	Kim, Choo, Neoh, and Tambe (2026), Incentive-Aware AI Safety via Strategic Resource Allocation	Formal game theory reasoning to design institutions and incentives to promote AI safety.
Optional	Balabanova, et. al. (2024), Media and responsible AI governance: a game-theoretic and LLM analysis	Evolutionary game theory modelling developers, regulators, and the public

Activity

- **Table-top exercise:** The group will role-play to simulate the strategic dynamics around an intelligence takeoff. Each person will assume a role (e.g. governments, AI labs, etc.) and interact with each other over multiple rounds to propose actions in line with their roles. Facilitators will construct events and indicators based on player actions.

Week 5: How can economics inform society's approach to AI?

As a field, economics offers many ways to shape the safe development of AI. Economic analysis shapes how public policy is developed and adopted. Economic theory informs many parts of technical alignment. Economic thinking can help us better understand the dynamics of AGI, in combination with other tools stemming from computer science, political science, philosophy, etc.. This week will wrap up the fellowship by looking at how fellows can use economics to make AI safer through practical steps that they can undertake.

Driving questions

- What market failures cause systematic underinvestment in AI safety, and what interventions (taxes, insurance, procurement rules, compute pricing) can correct them?
- What does international coordination on AI safety look like, and is it achievable given geopolitical incentives?
- What does a coherent research agenda or career plan at the intersection of AI safety and economics look like?

Readings

Type	Reading	Why read it
Government interventions		
Require	Korinek (2024), Economic Policy Challenges for the Age of AI	Overview and framing of key economic policy challenges from transformative AI.
Optional	Korinek and Lockwood (2026), Public Finance in the Age of AI: A Primer	Explores potential sources of government income as AI expands its share of the economy.
Optional	Jones (2025), How Much Should We Spend to Reduce AI's Existential Risks?	Proposes that at least one percent of GDP should be spent annually on AI risks.
Optional	Jensen, Emery-Xu, and Trager (2023), Industrial Policy for Advanced AI	Explores how industrial policies around compute could shift more resources to AI safety in frontier labs.
Market mechanisms		
Require	Timei, Jain, and Franklin (2025), AI Governance through Markets	Proposes short-term actionable levers using existing market-based tools to promote AI safety
Optional	Stetler (2025), Reinsuring AI	Proposes insurance as an alignment mechanism for AI firms
International governance		
Require	Blomquist, et. al. (2025), Examining AI Safety as a Global Public Good	Overview of issues in international AI governance, comparing AI to climate change as a co-ordination issue.
Optional	Roberts, Hine, Taddeo, and Floridi (2024), Global AI Governance: Barriers and Ways Forward	Systematic stock-take of barriers and potential models for international AI governance.
Optional	Forethought (2026), What an International Project to Develop AGI Should Look Like	Examines why and how a joint international AI project could be set up.

Activity

- **Policy pitch:** Working in small groups, fellows will develop a policy proposal to tackle AI safety using economic tools. They will then pitch the proposal to the room. A scenario will then be randomly drawn, representing a shift in the Overton window. Fellows then pitch the proposal again to sell their policy in this new environment.

Next Steps

The goal of this fellowship is to leave fellows both informed and ready to contribute to the intersection of AI safety and economics.

Following the fellowship, facilitators will organize one-on-one chats with fellows to discuss their careers and interests. The chats could help answer any remaining questions, help fellows develop a coherent research/policy/career plan, stress-test any ideas fellows might already have, and point fellows towards resources/networks/opportunities at XLab and beyond.