

Transformative Advances in AI and Implications for Economic Policy

Anton Korinek

anton@korinek.com

Brookings, University of Virginia, GovAI

<http://www.korinek.com>

[@akorinek](#)

Technical Background

A New Paradigm for AI

Old paradigm of AI, 2010s: **Deep Learning**

- had a large impact on our world
- but still a category difference between human & artificial intelligence

New paradigm, 2020s: **Foundation models** underlying **Generative AI**

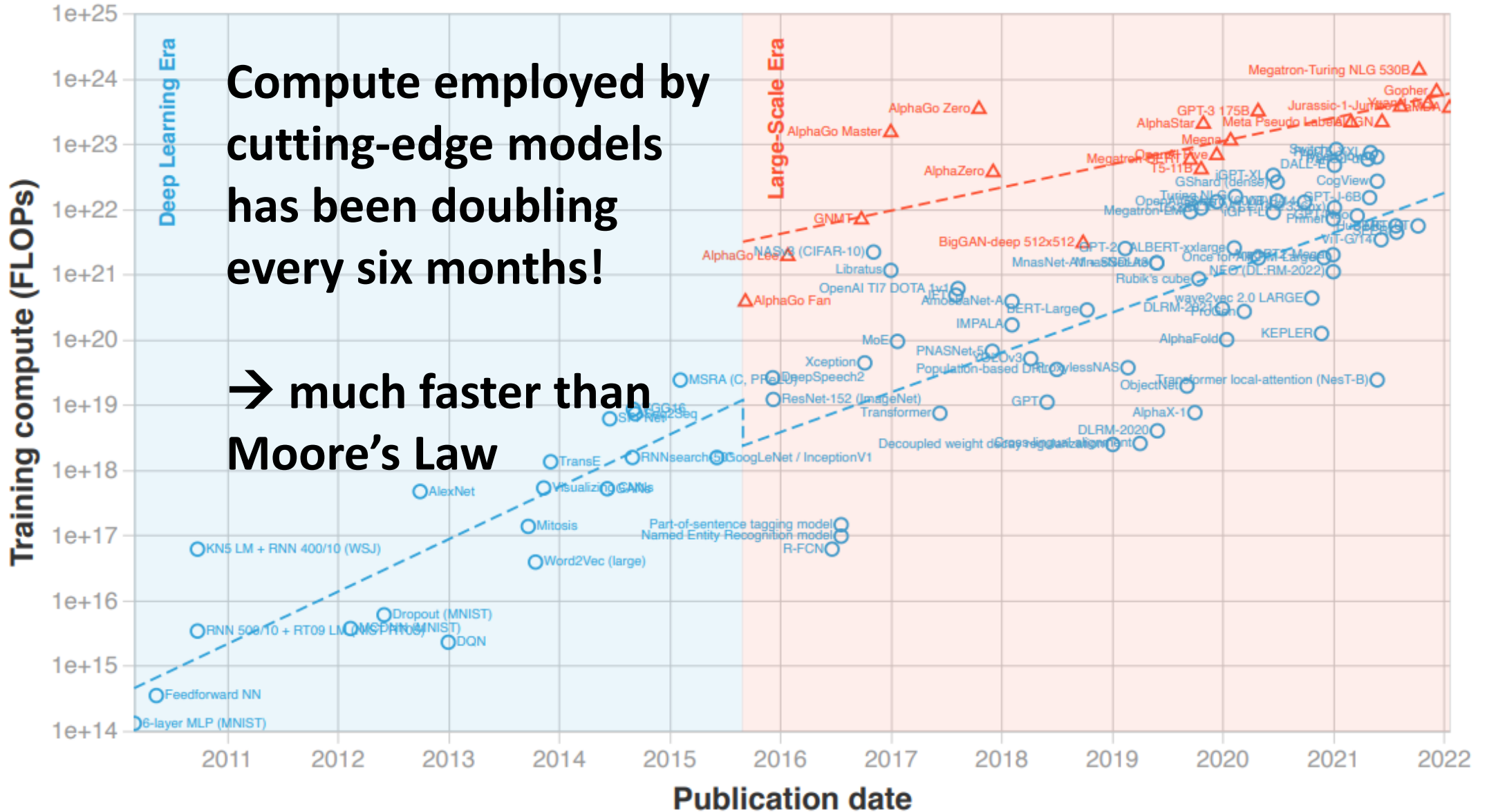
- builds on deep learning paradigm
- but qualitatively different – eerily human-like
- huge models with $\approx 10^{12}$ + parameters and growing
 - close to complexity of human brains

Leading category: **Large language models (LLMs)**

- like ChatGPT/Bing (GPT-4), Bard (PaLM-2, soon Gemini), Claude 2, Llama 2, ...

Training compute (FLOPs) of milestone Machine Learning systems over time

n = 102



Compute employed by cutting-edge models has been doubling every six months!

→ much faster than Moore's Law

Training, Capabilities, and World Models

Training of modern LLMs proceeds in 3 steps:

Step 1: Pre-training (very costly: > \$100m for GPT-4, Gemini):

- LLM base model is pre-trained via “self-supervised” learning:
 - system is fed vast amounts of data
 - banal training objective: to predict the next word (token) in a sentence
- During this process, the model learns to represent the distribution of words in its training corpus
 - but to do so efficiently, the model develops an internal **representation or model of the world**

Training, Capabilities, and World Models

Training of modern LLMs proceeds in 3 steps:

Step 2: Instruction fine-tuning via supervised learning:

- LLM is taught to follow instructions rather than just continuing text:
 - fine-tuning to answer questions, summarize texts, brainstorm, etc.
 - but all this draws on the pre-trained foundation

Step 3: Reinforcement learning from human feedback (RLHF):

- Model learns how much humans like answer → teaches LLM to be friendly, sound authoritative, refrain from hateful text, etc.

Scaling and Emergent Capabilities

- **Scaling:** predictable performance gains from increase in training compute → “scaling laws” on how more parameters & training data reduce log-loss
- **Emergent capabilities:** arise unexpectedly at discrete points:
 - LLMs could suddenly translate, rhyme, do math, etc.
- **Multi-modal models:** LLMs combined with other modalities leading to rapid progress in other domains
e.g., robots with world models will lead to rapid advances in robotics

Productivity Effects

Productivity Effects of LLMs

Message 1: cognitive workers – including you – can derive enormous productivity benefits from LLMs

Steps required:

- a) safe access to cutting-edge systems
- b) training
- c) change in workflows
- d) focus on human comparative advantage:
 - generative AI models generate content → will be devalued
 - humans discriminate content and provide feedback → complementary

CEO of Coursera: “Anybody who doesn’t use this will shortly be at a severe disadvantage.”

Useful Capabilities for Cognitive Workers

Generative AI for Economic Research: Use Cases and Implications for Economists (for JEL, Dec 2023)

25 Capabilities split into six categories:

1. Ideation and feedback
2. Writing
3. Background research
4. Coding
5. Data analysis
6. Math

available at www.korinek.com

Impact on Macroeconomics & Growth

Message 2: Advances in AI will boost productivity and growth

→ implications for natural interest rates, inflation, monetary policy, etc.

a) Using only the current generation of models:

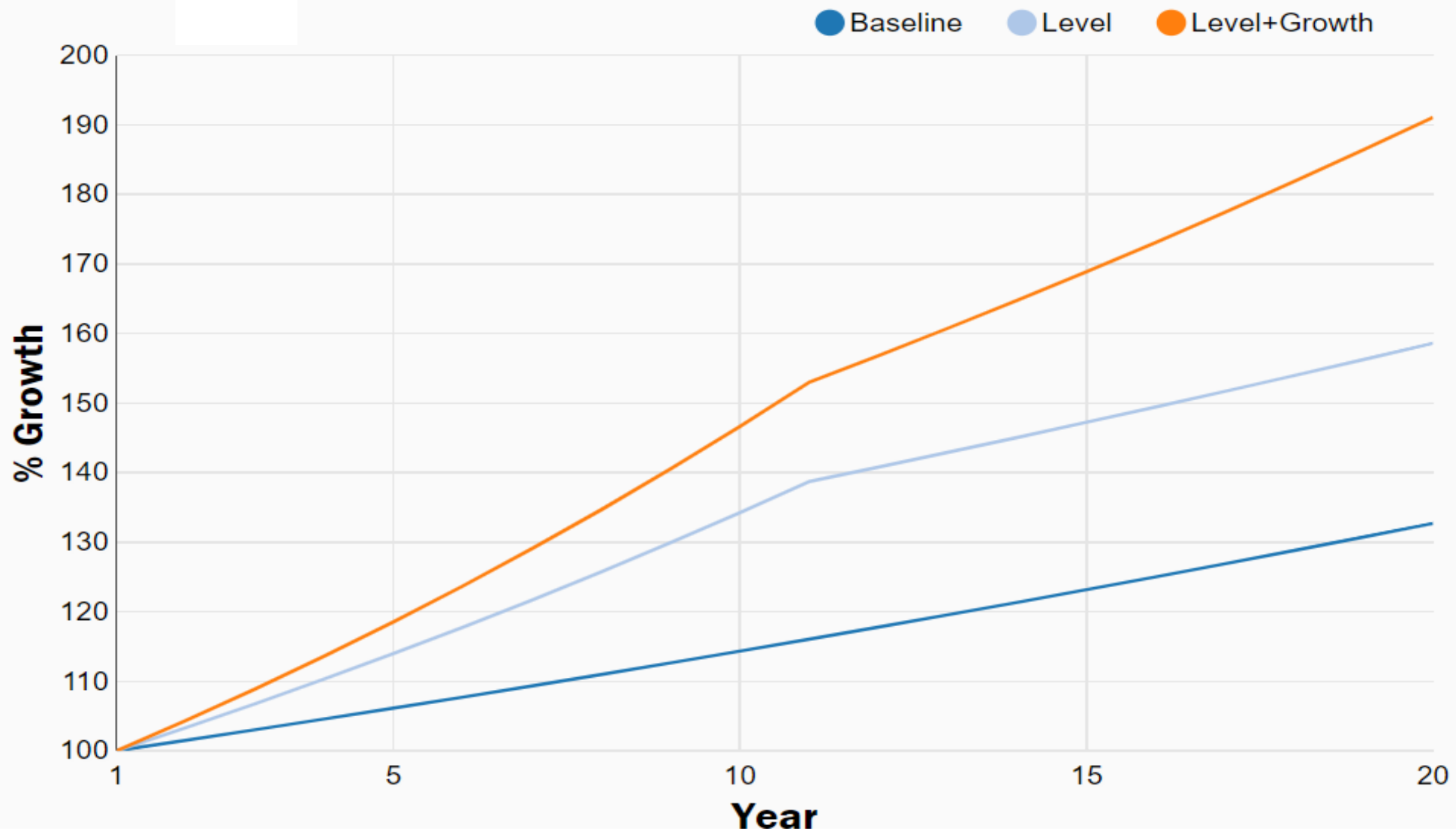
- Increase in level of productivity of cognitive workers > 10%
- Increase in rate of innovation and productivity growth
 - early estimates of 10%, 20%, 50% productivity gains for cognitive workers
 - rollout faster than in the past because it's cognitive automation

→ large positive supply shock

b) But the next generation of models will be even more powerful

→ progress accelerating

Figure 2. Possible Growth Trajectories



Brookings report “[Machines of mind: The case for an AI-powered productivity boom](#),” with Martin Baily and Erik Brynjolfsson

Integrating AI into our Economy

Main barriers to adoption:

- institutional inertia / new processes required
- regulation (eg in medicine, law, education, etc)
- false narratives (“AI is not creative”, “AI can’t do xyz”, etc...)

Drivers of adoption:

- rollout via software rather than new physical capital
- integration into existing software already happening (email, office, ...)
- interaction in natural language is easy to learn

Scenarios for the Future

Where is this all going?

- There is no obvious ceiling to the capabilities of AI
 - Just scaling current systems will deliver very powerful capabilities
- Significant uncertainty about:
 - speed of progress
 - whether there is a ceiling to what AI can do
 - ratio of benefits & harms

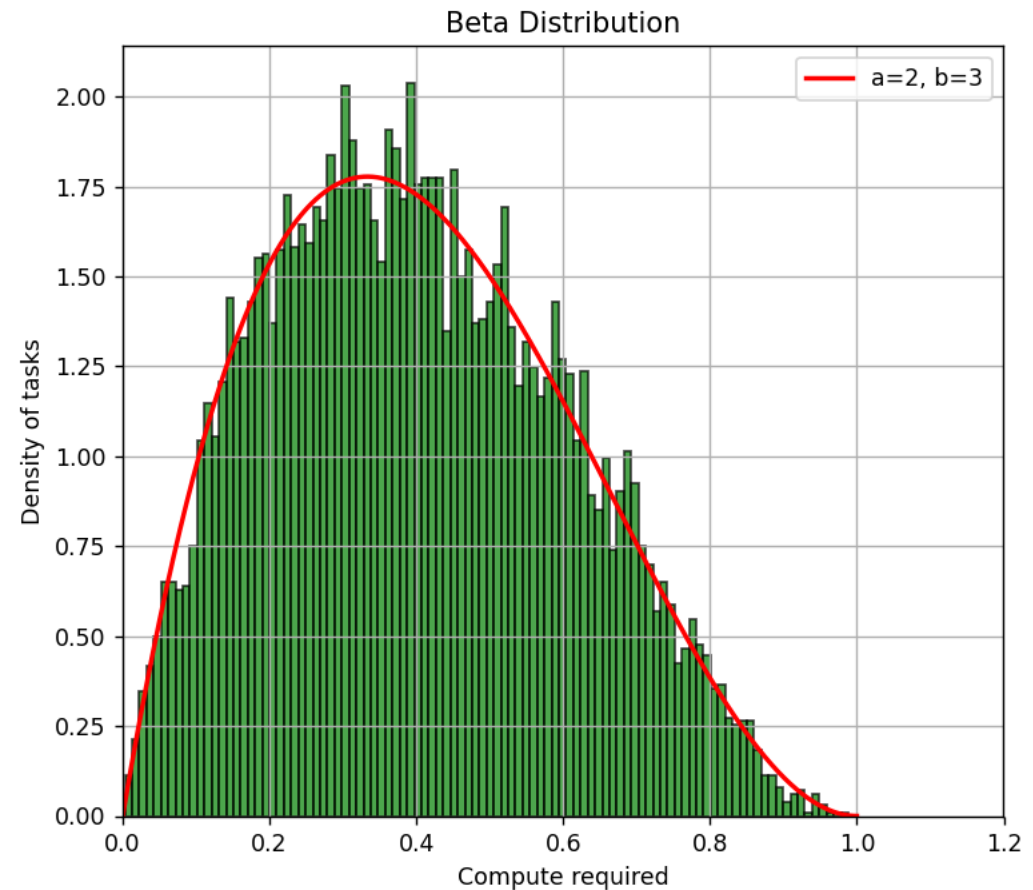
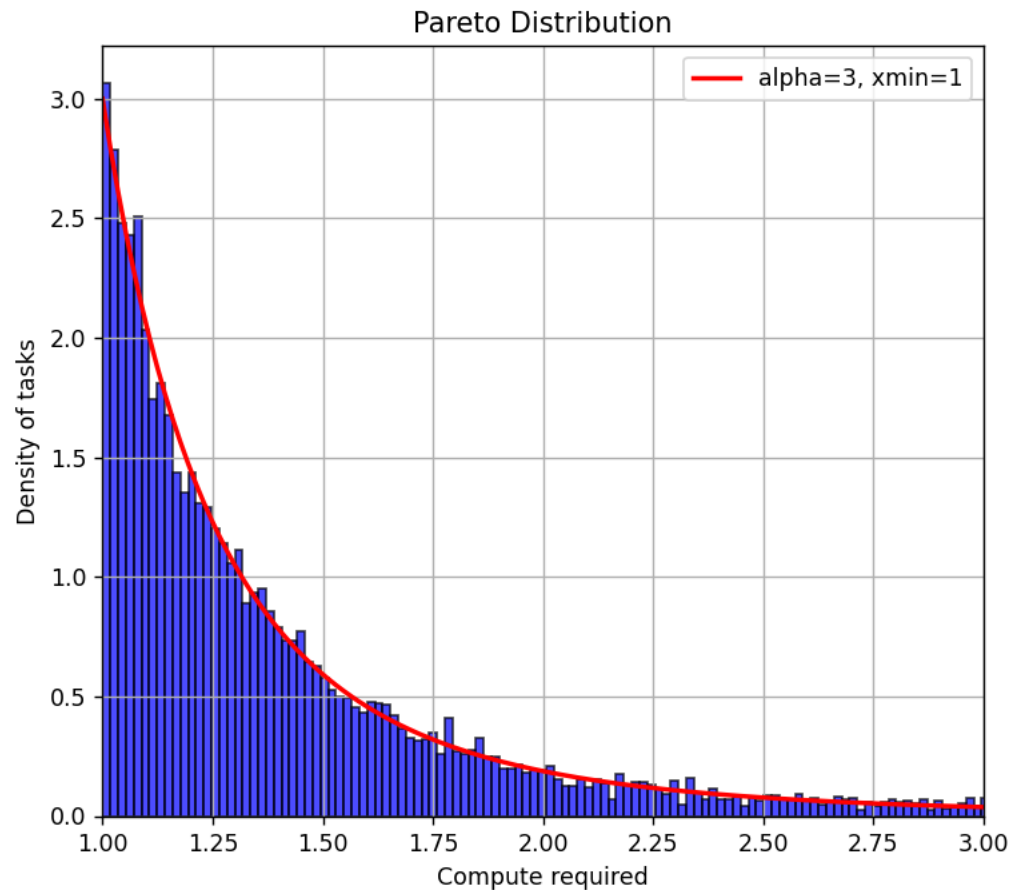
Preparing for AGI

- We need to prepare for a range of possible scenarios, including for the potential of **Artificial General Intelligence (AGI)**
 - AGI = ability of AI to perform all cognitive tasks that can be performed by humans
- AGI would be transformative for our world
 - and raises many important economic and governance questions

Thinking about AGI Scenarios

Computational complexity of tasks performed by humans (from ongoing work):

Are the capabilities of human brains infinite or bounded?



Disruption of Labor Markets

Message 3: Prepare for the disruption of labor markets*

Story of the industrial revolution: we have retreated into the unautomated tail of the task distribution → will continue in short run

BUT: AGI would simply make humans an obsolete technology

→ see my 2022 Brookings report on “[Preparing for the \(Non-Existent?\) Future of Work](#),” with Megan Juelfs

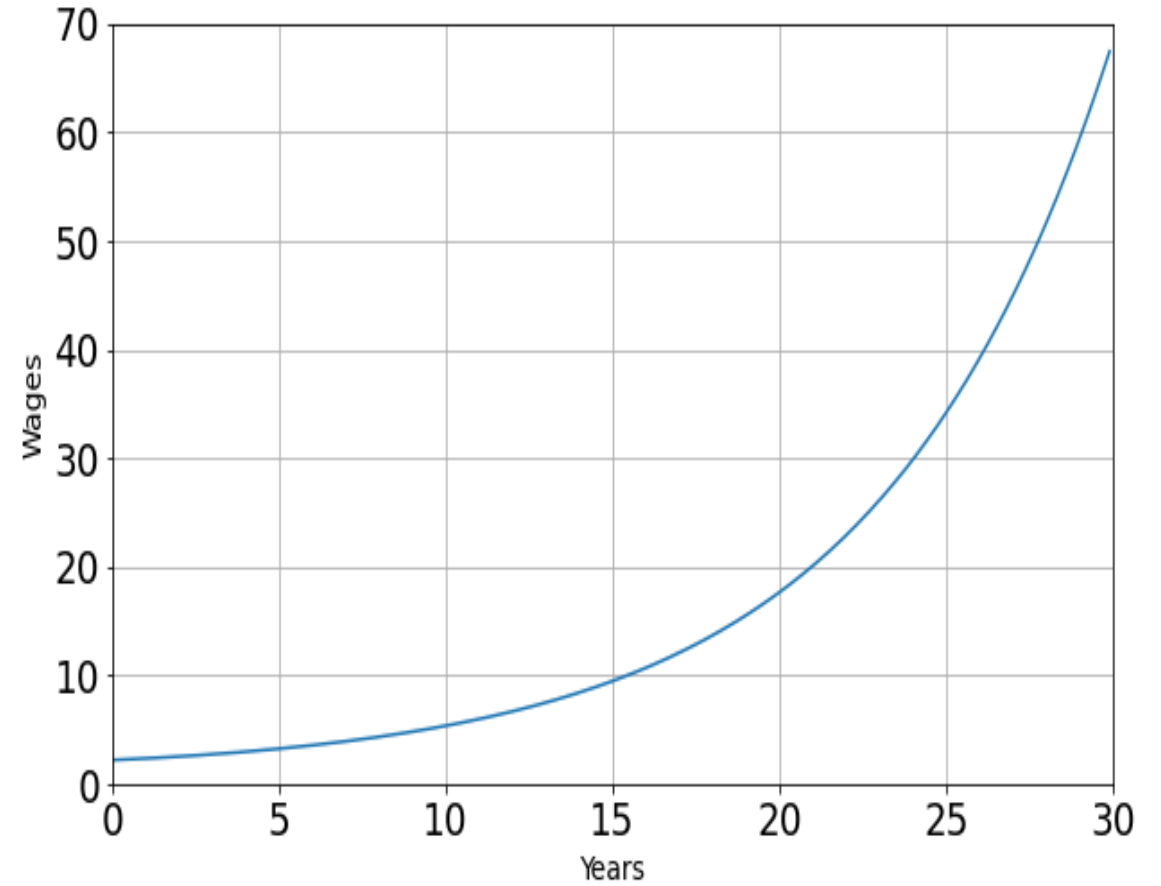
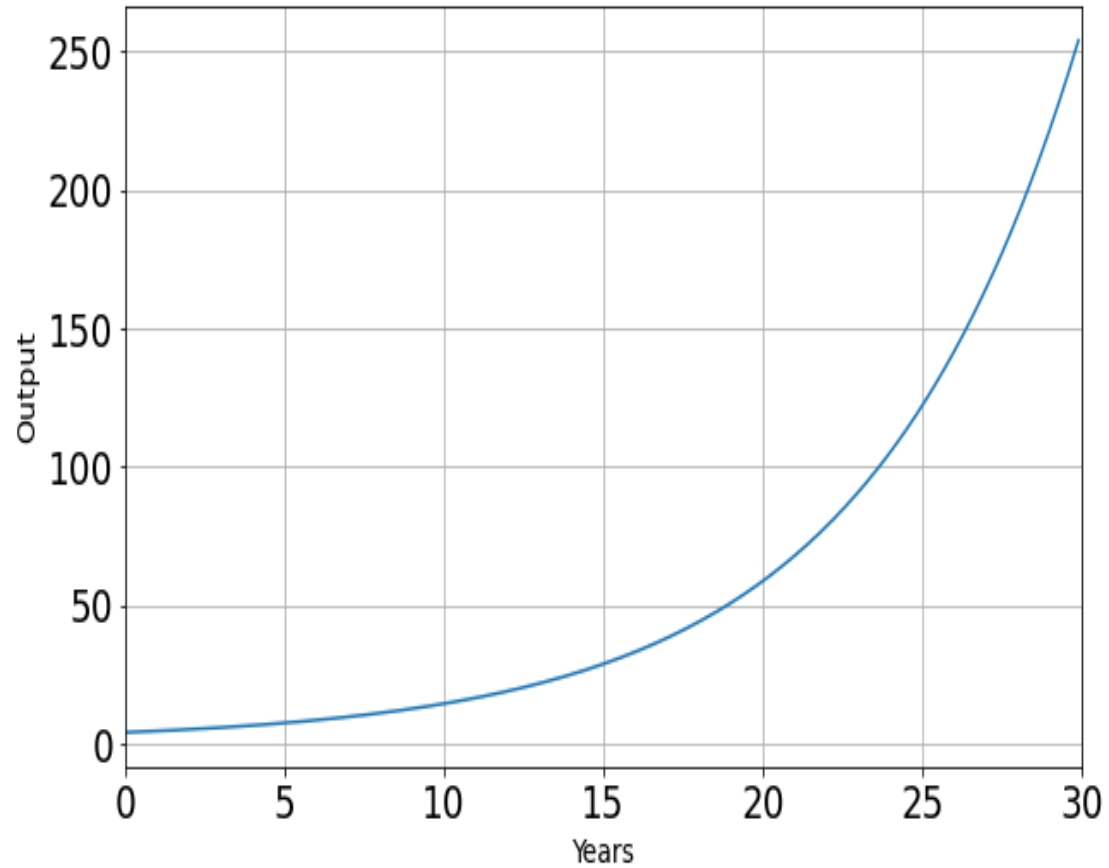
*and much more

Disruption of Labor Markets

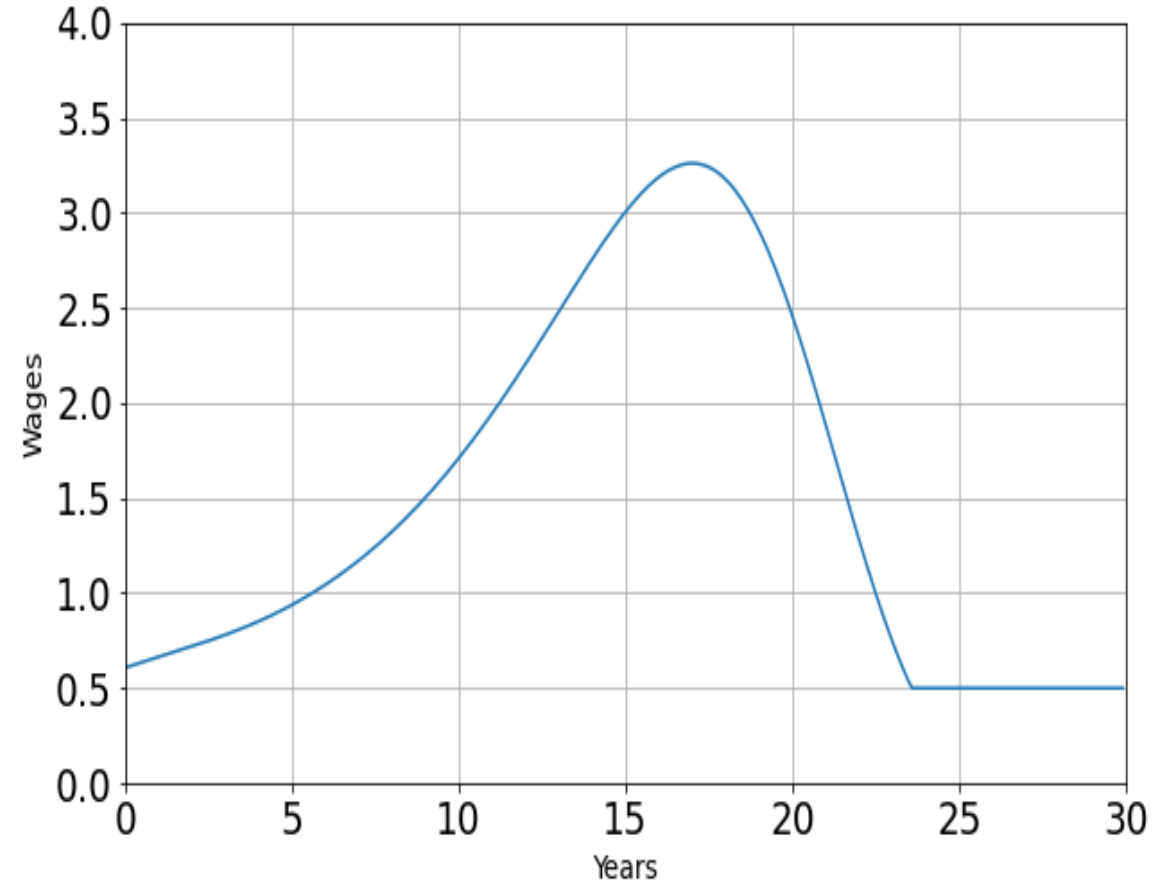
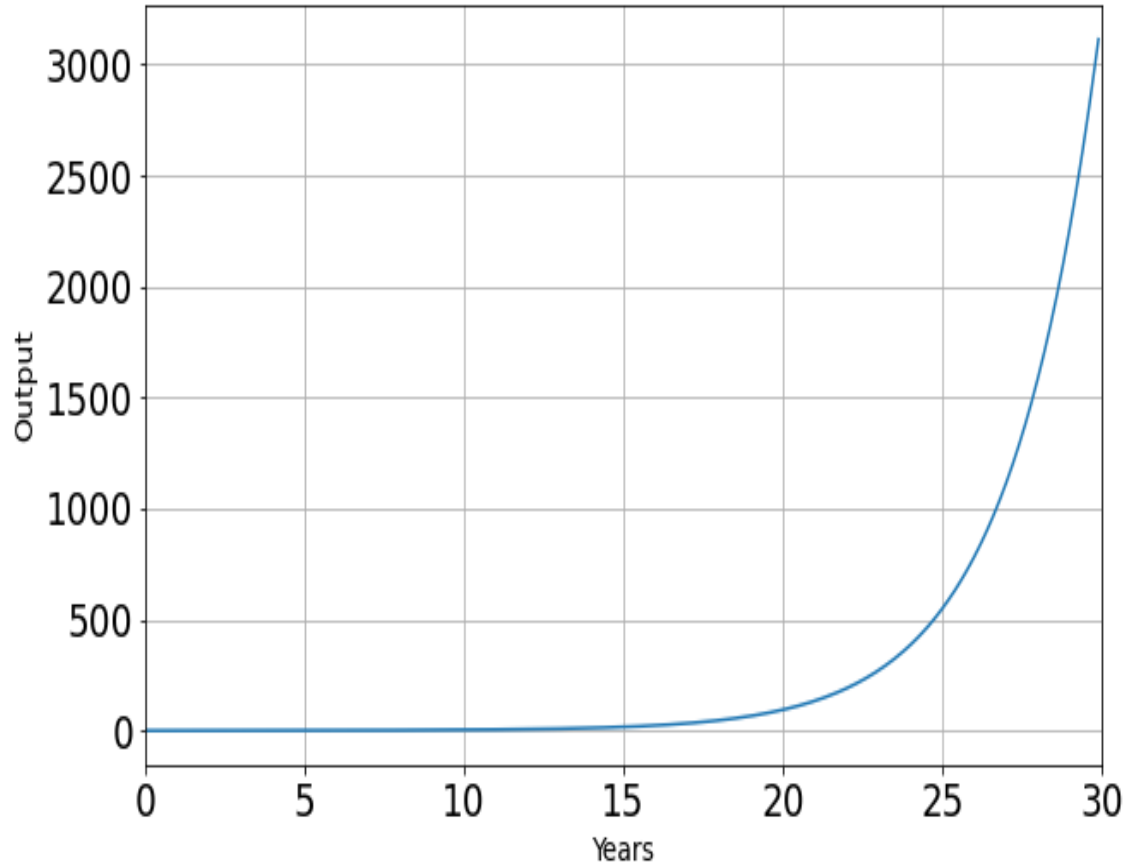
Broad implications:

- income distribution
- financial stability ←
- systems of taxation ←
- social safety nets ←
- education
- political stability

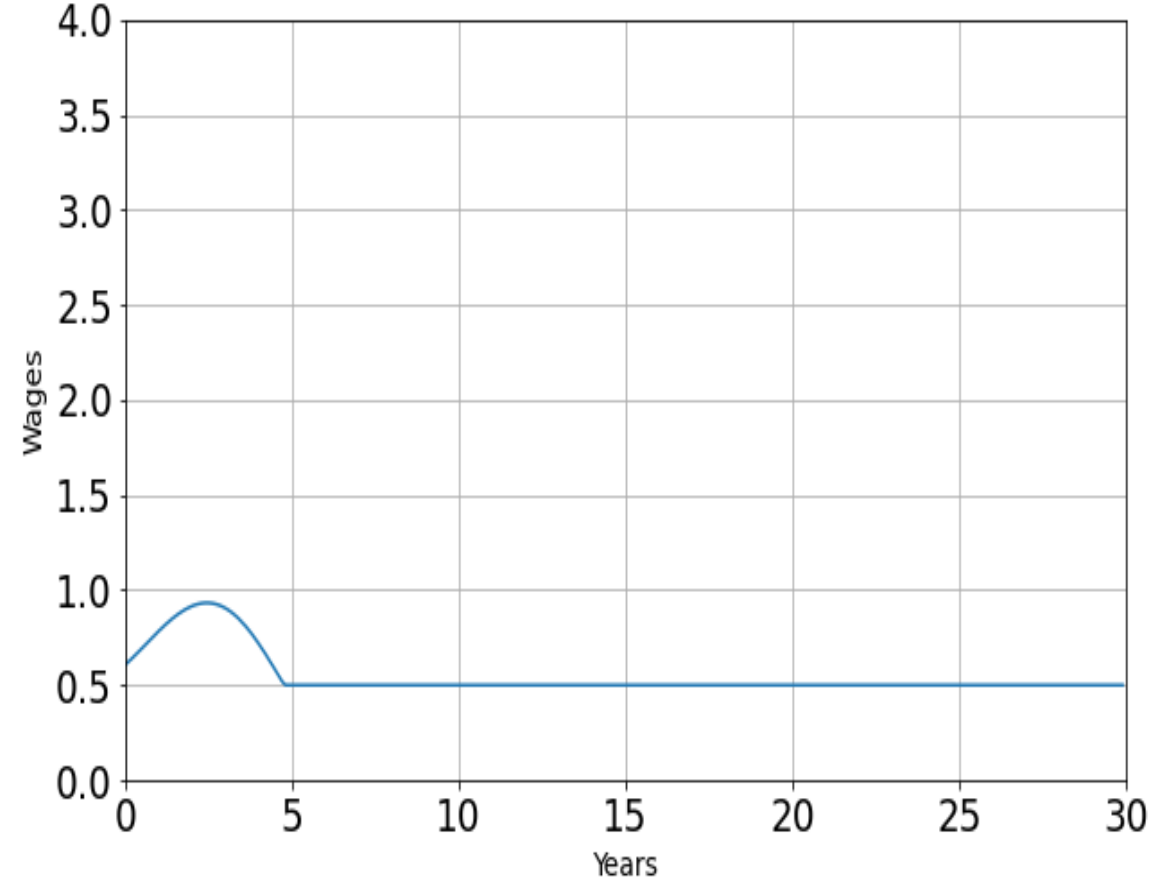
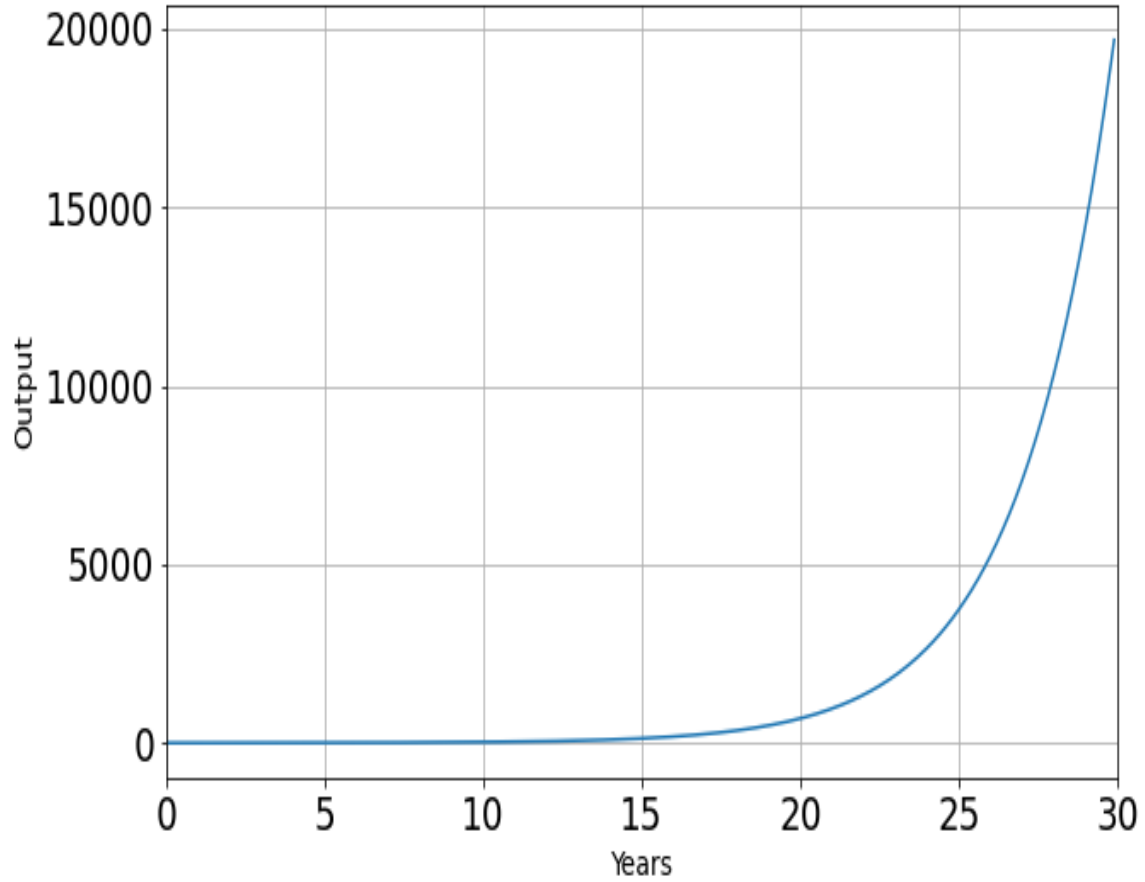
Growth as Usual in Output and Wages



AGI within 20 years: Growth Takeoff, Plummeting MPL



AGI within 5 years: Growth takes off even faster



Risks for Financial Markets

Financial sector is the “brain” of the economy

- Financial sector has always been a **magnet for intelligence**
 - ever more sophisticated regulatory arbitrage → requires reg-tech solutions
 - ever greater systemic complexity → risk of disruptions, flash crashes, etc
 - cybersecurity challenges
- **Disruptions in the real economy** will translate into financial risks
 - last significant economic disruption: tractors in the 1920s
 - we do not want to repeat 1929

