Large Language Models and Cognitive Automation

Anton Korinek

anton@korinek.com

Brookings, University of Virginia, GovAI

http://www.korinek.com @akorinek

based in part on my NBER WP #30957 on "Language Models and Cognitive Automation…"

A New Paradigm for AI

Old Paradigm, 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 ≈ 10^11 parameters and growing
  - close to complexity of human brains
- leading category: **Large language models (LLMs)**
  - like ChatGPT, Claude, Bing, Bard, … latest addition: GPT-4

First Impressions of GPT-4: Form vs Substance

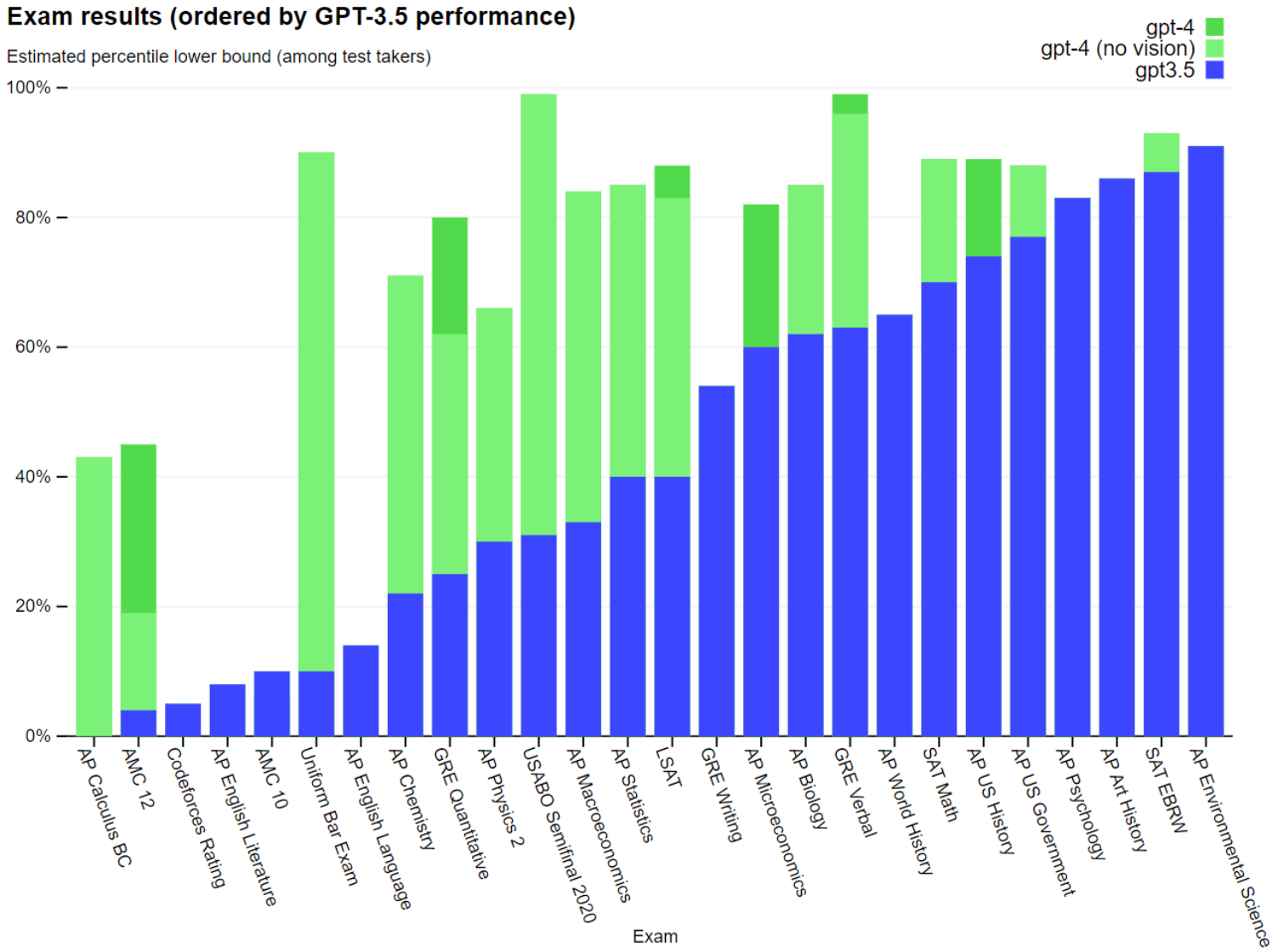
GPT-3.5/ChatGPT: wowed its users last November:

- first publicly available language model that could spit out coherent and well-crafted sentences on almost any topic

GPT-4: form vs substance

- GPT4 does not improve much on the form of its writing
  → we were already close to a ceiling on form
- but it does improve significantly on substance:
  - it is deeper, more insightful, and a more valuable tool for cognitive workers
  - it performed significantly better on the research tasks in my NBER WP in which GPT3.5 displayed shortcomings
- improvements on substance are not as flashy as improvements on form, but will ultimately have a greater impact on our world

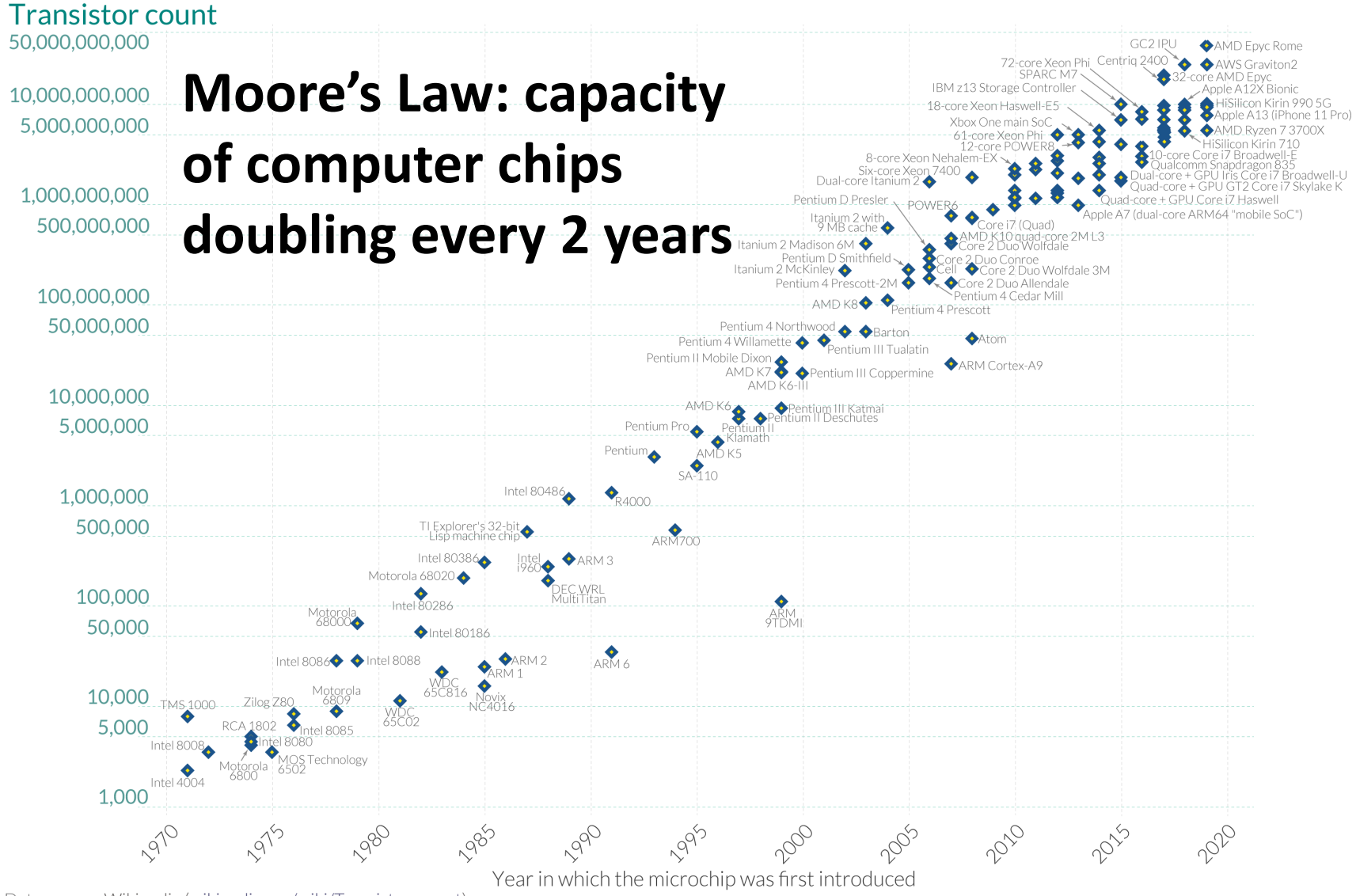
Performance of GPT-4 on Standardized Tests



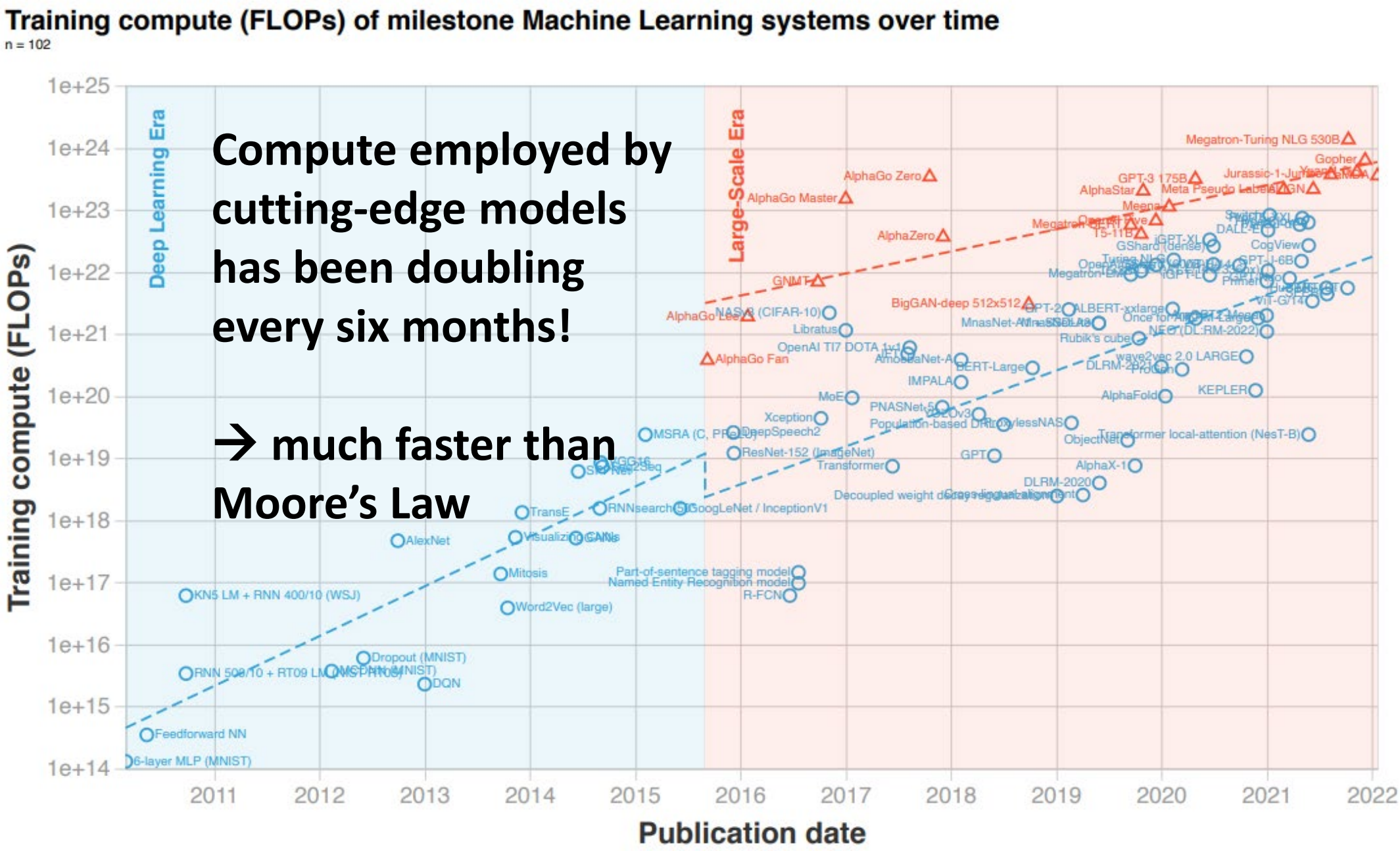
Moore's Law: The number of transistors on microchips has doubled every two years

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years.
This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

Our World in Data



Data source: Wikipedia (wikipedia.org/wiki/Transistor_count)
OurWorldinData.org – Research and data to make progress against the world's largest problems.



Training & Emergent Capabilities of LLMs

- Training via self-supervised learning:
  - system is fed vast amounts of data
  - objective: to predict the next word in a sentence

- Based on this training, many impressive capabilities emerged:
  - LLMs could suddenly write coherent sentences, paragraphs, … essays
    they can also translate, do logical reasoning, math, be creative, write code, etc.

  - during the **training** process, the systems **develop a "world model"** that can be applied to a wide range of tasks
    - we are forced to re-evaluate our understanding of what human brains do

  - LLMs display significant "capabilities overhang"

  - scaling laws imply fairly predictable advances

Two Camps on the Significance of LLMs

1. Camp “Stochastic parrots” or “ Advanced auto-complete”

2. Camp “This is the Road to Human-level AI”

→ easy to either over- or underestimate LLMs – and dangerous to do so!

→ the capability surface of LLMs is very different from humans

Two Camps (ctd)

Both camps emphasize a grain of truth:

1. Camp “Stochastic parrots”:

- have significant limitations (length of prompts, drift in output, training data cut-off, etc.)
- tend to “hallucinate”
- not always grounded in our ethical values

2. Camp “Road to Human-level AI”:

- CEOs stating that their cognitive ability is a lot higher using LLMs and “anybody who doesn’t use this will shortly be at a severe disadvantage” (CEO of Coursera)
- Growing number of academic studies pointing to productivity gains of 20%, 50% and more for cognitive workers using LLMs

Useful Capabilities for Cognitive Workers

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

→ see my NBER Paper #30957 on "Language Models and Cognitive Automation for Economic Research"

Category	Task	Usefulness
Ideation	Brainstorming	3
	Evaluating ideas	2
	Providing counterarguments	3
Writing	Synthesizing text	3
	Editing text	3
	Evaluating text	3
	Generating catchy titles & headlines	3
	Generating tweets to promote a paper	3
Background Research	Summarizing Text	3
	Literature Research	1
	Formatting References	3
	Translating Text	3
	Explaining Concepts	2

Category	Task	Usefulness
Coding	Writing code	2
	Explaining code	2
	Translating code	3
	Debugging code	2
Data Analysis	Extracting data from text	3
	Reformatting data	3
	Classifying and scoring text	2
	Extracting sentiment	2
	Simulating human subjects	2
Math	Setting up models	2
	Deriving equations	1
	Explaining models	1

Importance of Prompt Engineering

Important new task: prompt engineering

- is like a mix of programming & using natural language

- performance and capabilities depend a lot on good "prompt engineering"
- this induces the LLM to shift into the desired latent state

Short-Term Lessons

- LLMs are useful assistants AND tutors
  → automation of “micro tasks”
  → significant productivity gains

Economic advice:

- focus on your comparative advantage (h/t David Ricardo)
- need to change our workflows
  - foundation models generate content → will be devalued
  - we humans can discriminate content → complementary
  - we also provide feedback and organize projects

Effects on Productivity and Growth

Generative AI will usher in a new era of rapid productivity growth

- early estimates of 10%, 20%, 50% productivity gains for cognitive workers using *current* models
- productivity growth is powered by (automatable) cognitive work
- **progress itself is accelerating**

Medium-Term Outlook (Speculative)

- Our LLM will become better and better assistants AND tutors
- They will be adapted to lots of different use cases
  - this will take some time, but has the potential to significantly restructure the economy
- The role of humans in many cognitive tasks will diminish
  - in many tasks, we'll increasingly turn into rubber-stampers

Cognitive Automation

How does cognitive automation differ from physical automation?

- we have automated tasks & jobs for centuries – what's new?

1. it affects a very different category of workers

2. it produces outputs that are non-rivalrous
   → can be rolled out much faster

3. it chips away at our last comparative advantage vs machines

Lots of interesting new questions

What will the new era of cognitive automation imply for:

- labor markets
- education
- technological progress
- …
- social welfare?

As of right now, human brains enhanced by LLMs are (still!) the best technology around to answer these!