

AI and Actuaries: Productivity Boost or Hype?

Southeastern Actuaries
Conference

November 21, 2024

Atlanta, GA



It's been two years
since ChatGPT
went online...

How is it going?

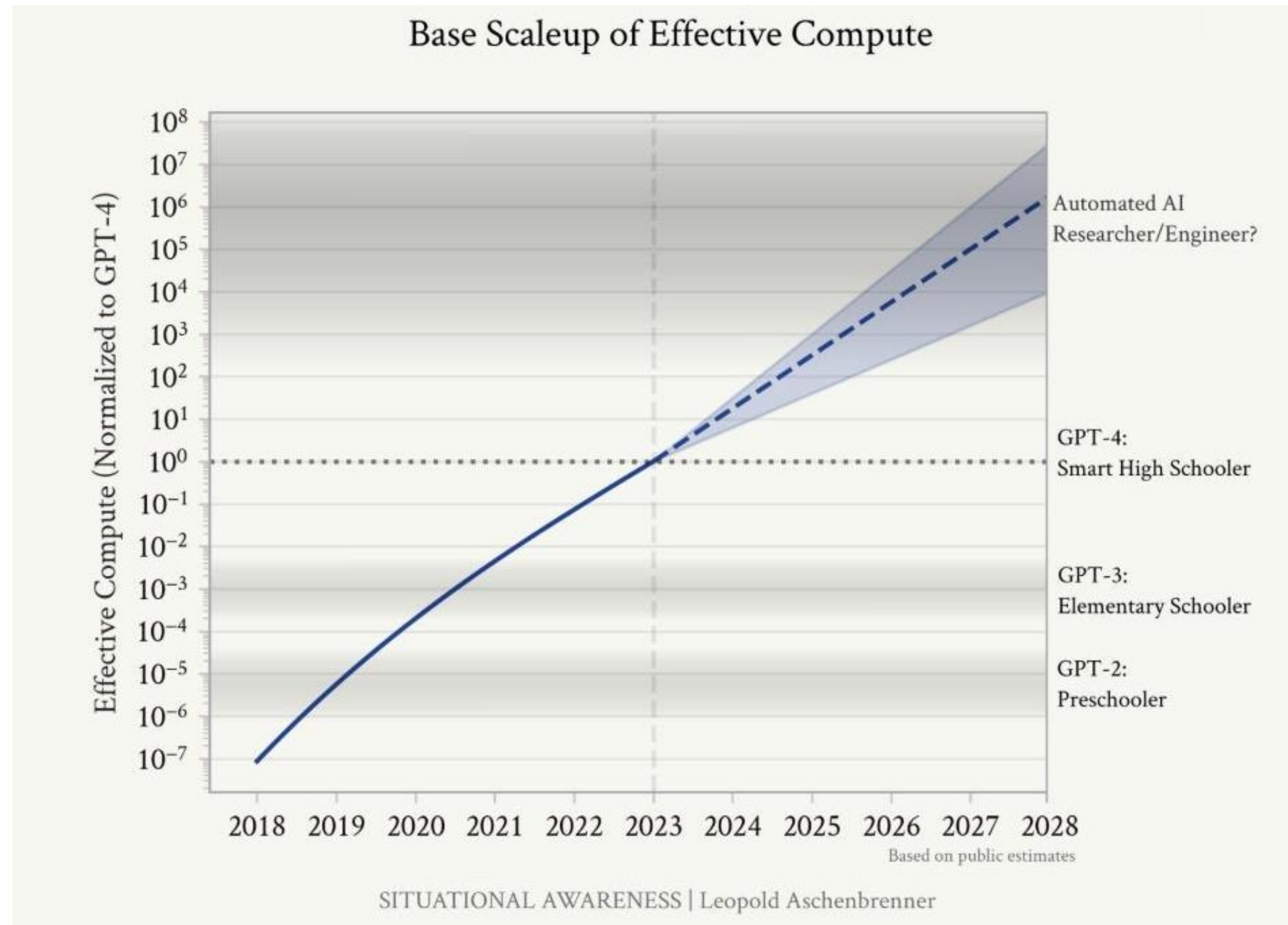
Producing ChatGPT

trained a model called ChatGPT which interacts
conversational way. The dialogue format makes it
possible for ChatGPT to answer followup questions,
correct its mistakes, challenge incorrect premises, and
refuse inappropriate requests.



[Read about ChatGPT Plus](#)

“AI Today is the Worst it Will Ever Be”



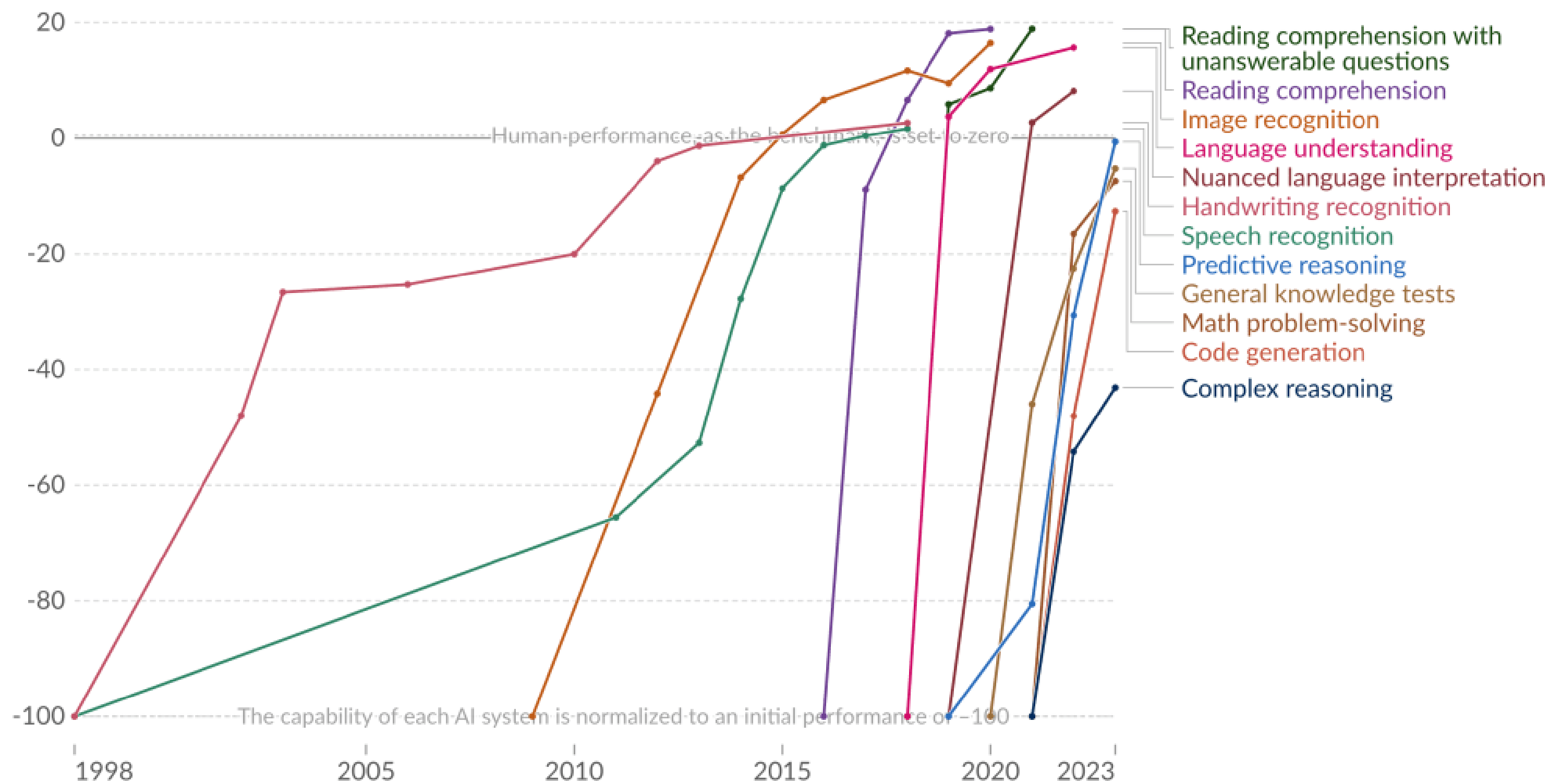
Source: [SITUATIONAL AWARENESS: The Decade Ahead \(situational-awareness.ai\)](https://situational-awareness.ai)

Rapidly Emerging Capabilities of AI

Test scores of AI systems on various capabilities relative to human performance



Within each domain, the initial performance of the AI is set to -100. Human performance is used as a baseline, set to zero. When the AI's performance crosses the zero line, it scored more points than humans.



Data source: Kiela et al. (2023)

OurWorldInData.org/artificial-intelligence | CC BY

Note: For each capability, the first year always shows a baseline of -100, even if better performance was recorded later that year.

Performance on common exams
(percentile compared to human test-takers)

| | GPT-4 (2023) | GPT-3.5 (2022) |
|----------------------------|-----------------|-------------------|
| Uniform Bar Exam | 90th | 10th |
| LSAT | 88th | 40th |
| SAT | 97th | 87th |
| GRE (Verbal) | 99th | 63rd |
| GRE (Quantitative) | 80th | 25th |
| US Biology Olympiad | 99th | 32nd |
| AP Calculus BC | 51st | 3rd |
| AP Chemistry | 80th | 34th |
| AP Macroeconomics | 92nd | 40th |
| AP Statistics | 92nd | 51st |

SITUATIONAL AWARENESS | Leopold Aschenbrenner

Can AI Pass an Actuarial Exam?



Candidate Support

SOA Exam P Sample Exam

SOA Exam P Probability Sample Exam Results

Your Score:

25

Total Possible Score (points)

30

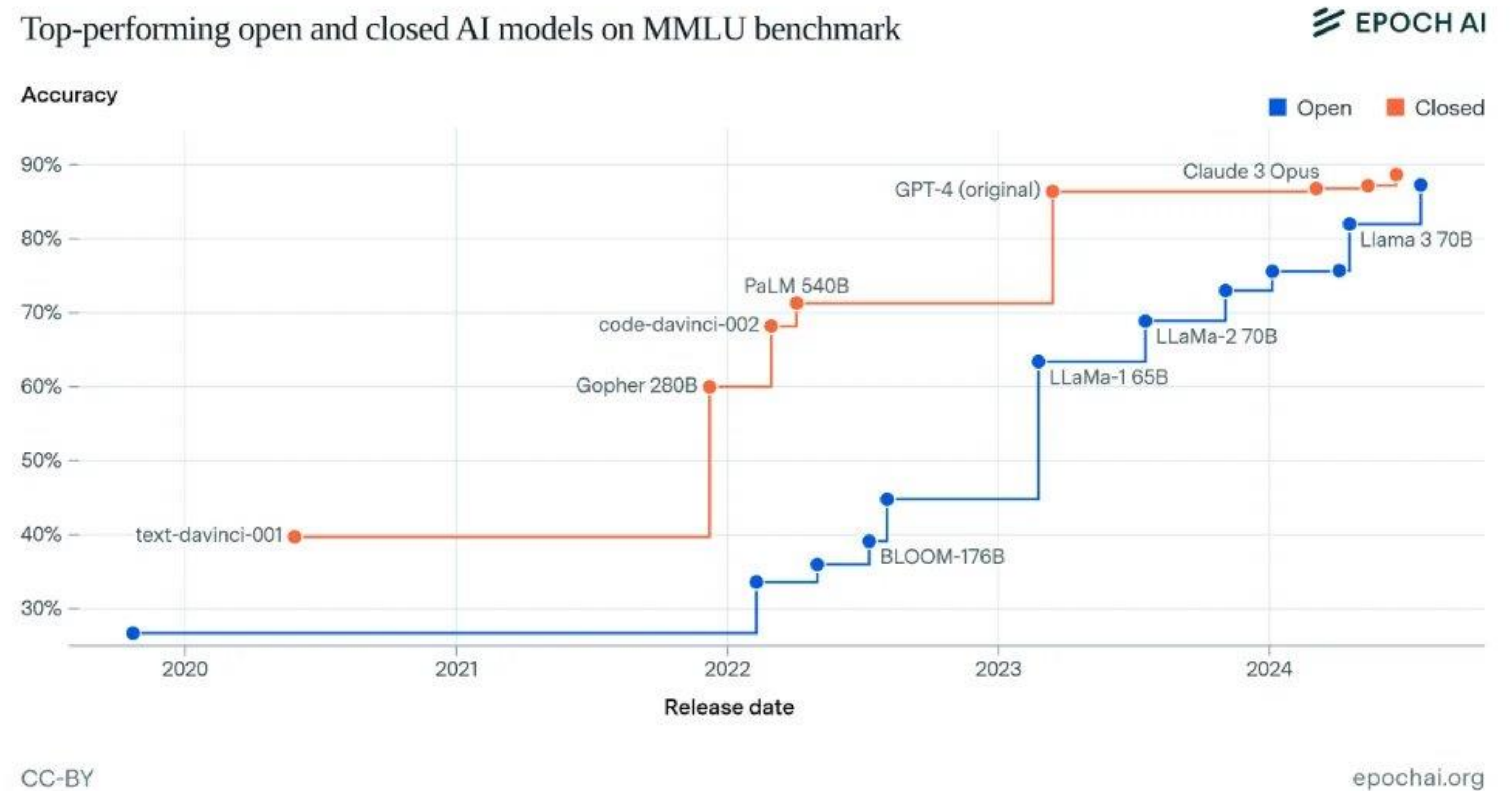
System Date and Time

11/17/2024, 12:13:57 PM

- 25 of 30 questions correct
(*Claude – 3.5 Sonnet Oct'24*)
- Completed a 3hr exam in ~20 mins
- Provided detailed work and explanations to each problem
- Parsed complex table input

Is Progress Plateauing?

- Two years ago, a single model (GPT3) was well ahead of competition
- Today (Late 2024) multiple models converging on similar performance
- Model Generational Improvements less significant
- Are we converging on Human Performance?



What does this mean for actuarial productivity?

Are we seeing the promised leap in productivity?

Producing ChatGPT

trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, correct its mistakes, challenge incorrect premises, and respond to inappropriate requests.

ChatGPT Plus

[Read about ChatGPT Plus](#)

Current State of AI: Consultants w/ GPT-4



Sources: Human-Generative AI Collaboration Experiment (May-June 2023); BCG analysis.

Note: Findings reflect results (on a 10-point scale) for the creative product innovation task only. Baseline task performance was used as a proxy for proficiency on this type of task. Both distributions reflect GPT-4-based performance grades rather than human grades for greater consistency of within-subject analysis.

[Source: How People Create and Destroy Value with Generative AI | BCG](#)

The Good Results:

- Group performance improved significantly
- Lowest performers benefited the most
- Consistency improved

Considerations:

- Performance was very task-dependent
- Humans exposed to training hurt AI performance; Humans modifying AI responses decrease quality
- Diversity of ideas decreased

Three Estimates of AI Productivity Gains

Daron Acemoglu: 0.07% GDP boost per year

VS

Goldman Sachs: 1.5% GDP per year

McKinsey: 1.5-3% ++ GDP per year boost

The Simple Macroeconomics of AI*

Daron Acemoglu
Massachusetts Institute of Technology

April 5, 2024

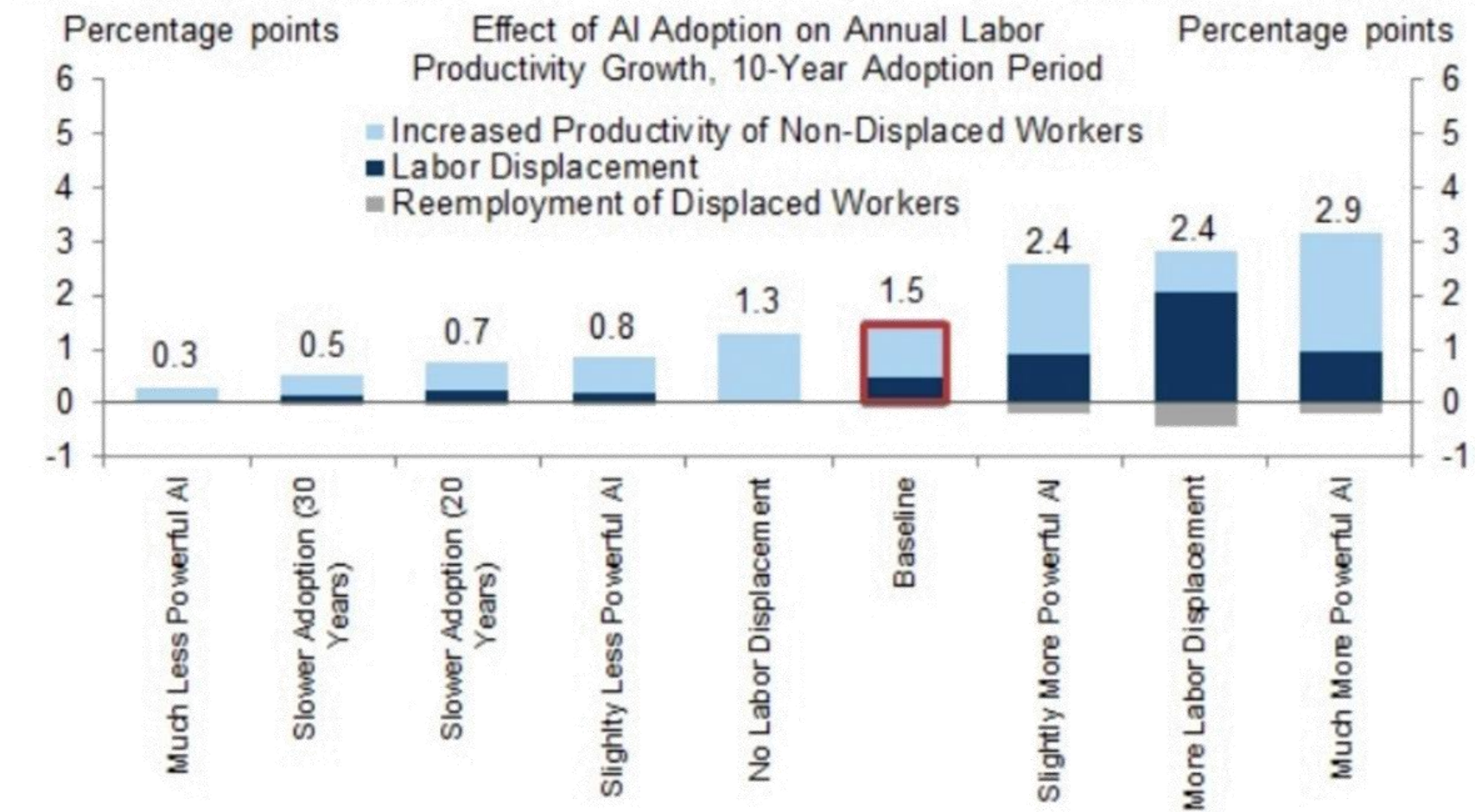
Abstract

This paper evaluates claims about the large macroeconomic implications of new advances in AI. It starts from a task-based model of AI's effects, working through automation and task complementarities. It establishes that, so long as AI's microeconomic effects are driven by cost savings/productivity improvements at the task level, its macroeconomic consequences will be given by a version of Hulten's theorem: GDP and aggregate productivity gains can be estimated by what fraction of tasks are impacted and average task-level cost savings. Using existing estimates on exposure to AI and productivity improvements at the task level, these macroeconomic effects appear nontrivial but modest—no more than a 0.71% increase in total factor productivity over 10 years. The paper then argues that even these estimates could be exaggerated, because early evidence is from easy-to-learn tasks, whereas some of the future effects will come from hard-to-learn tasks, where there are many context-dependent factors affecting decision-making and no objective outcome measures from which to learn successful performance. Consequently, predicted TFP gains over the next 10 years are even more modest and are predicted to be less than 0.55%. I also explore AI's wage and inequality effects. I show theoretically that even when AI improves the productivity of low-skill workers in certain tasks (without creating new tasks for them), this may increase rather than reduce inequality. Empirically, I find that AI advances are unlikely to increase inequality as much as previous automation technologies because their impact is more equally distributed across demographic groups, but there is also no evidence that AI will reduce labor income inequality. AI is also predicted to widen the gap between capital and labor income. Finally, some of the new tasks created by AI may have negative social value (such as design of algorithms for online manipulation), and I discuss how to incorporate the macroeconomic effects of new tasks that may have negative social value.

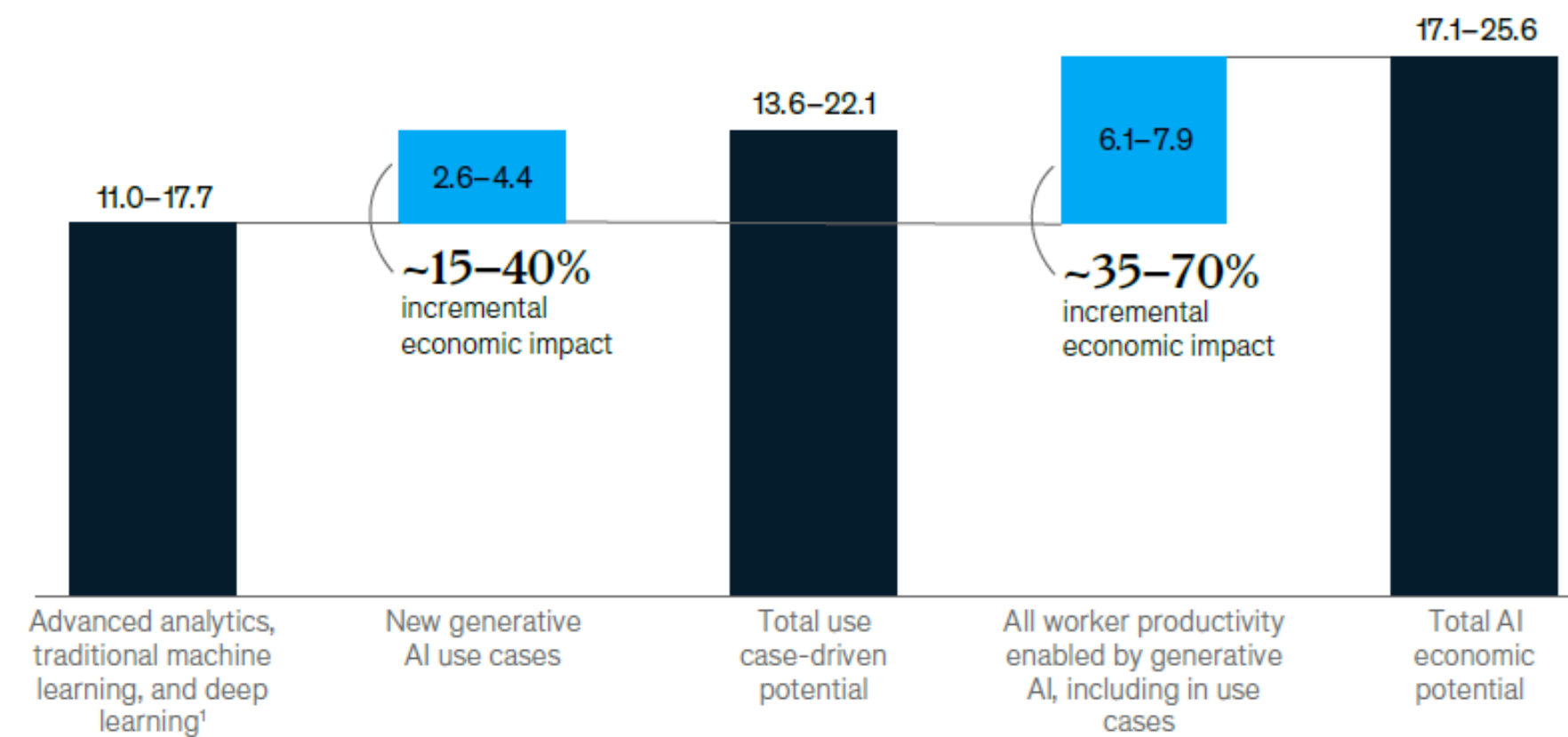
JEL Classification: E24, J24, O30, O33.

Keywords: Artificial Intelligence, automation, ChatGPT, inequality, productivity, technology adoption, wage.

Exhibit 13: We Estimate That Generative AI Could Boost Aggregate Labor Productivity Growth by 1.5pp in the US, Although the Size of the Boost Will Depend on AI's Capability and Adoption Timeline



Source: Goldman Sachs Global Investment Research



¹Updated use case estimates from "Notes from the AI frontier: Applications and value of deep learning," McKinsey Global Institute, April 17, 2018.

What about Actuarial Applications?

- **AI will benefit “knowledge workers”**
(including MDs, PhD, JDs and ASA/FSAs!)
- Actuaries are constantly challenged to produce work faster, to incorporate more data into their work, and improve quality and consistency
- Improve existing analytics and automation use cases
- Boost productivity of knowledge workers; Improve collaboration
- Enable small teams to develop incredibly complex solutions



AI Boosted Tasks in Actuarial Work Today

| General Business Productivity | Advanced Pattern Recognition | Translation and Code Generation | Data Automation |
|---|---------------------------------|--|--------------------------|
| Assistants and Copilots | Risk Segmentation | Querying large databases with plain language (Text-to-SQL) | Document Data Extraction |
| Document Search | Pricing & Reserving | Initial Code Generation | Data classification |
| Meeting Transcripts | Trend Forecasting | | Initial data analysis |
| Summarization / Executive Summary Writing | Seasonality / Anomaly Detection | | |
| Text / Email Generation | | | |

A.I. TURNS THIS SINGLE BULLET POINT INTO A LONG EMAIL I CAN PRETEND I WROTE.



A.I. MAKES A SINGLE BULLET POINT OUT OF THIS LONG EMAIL I CAN PRETEND I READ.



©marketoonist.com

AI Agents: Upcoming Applications

TBW / Slide coming

Reflection, Tool Use, Planning, Multi-Agent Collaboration

Amara's Law



We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run.

- Bill Gates

Thank You