SINTEF

The Future of Software Engineering in the light of LLMs

Erik Johannes Husom

24th May 2023

SCT – Department gathering

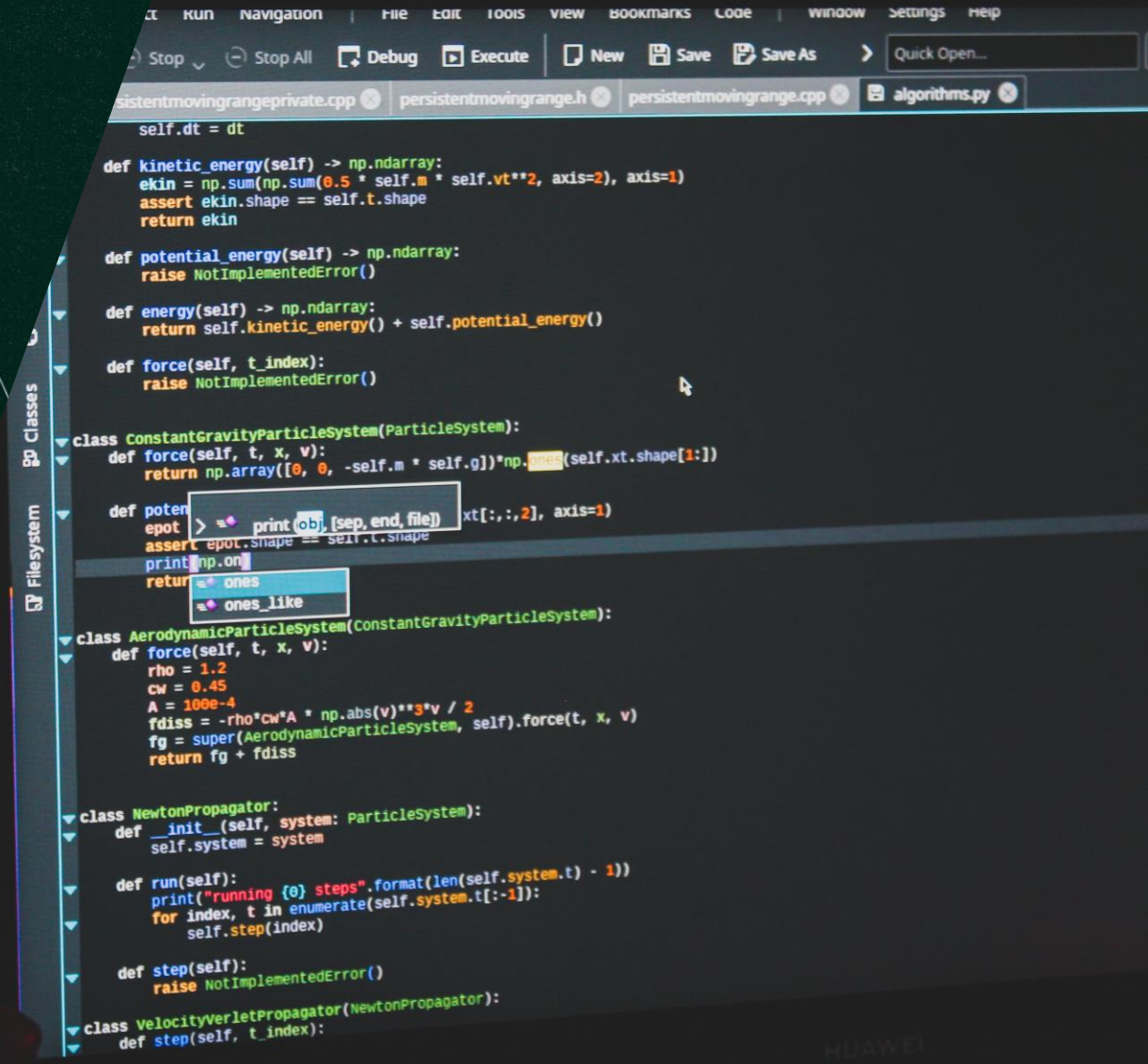


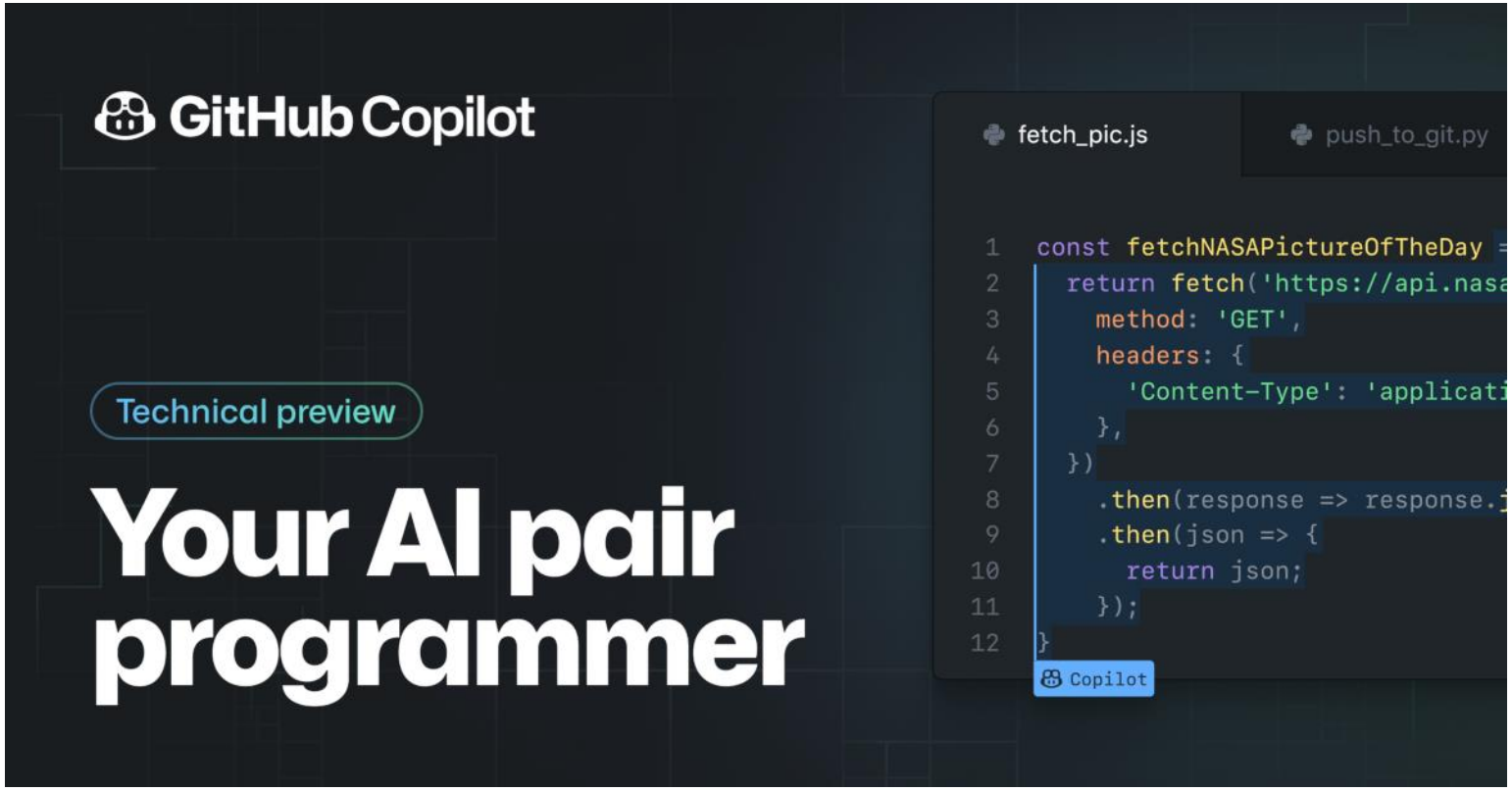
Photo by Árpád Czapp on Unsplash

Breakthroughs in the world of AI

Outline

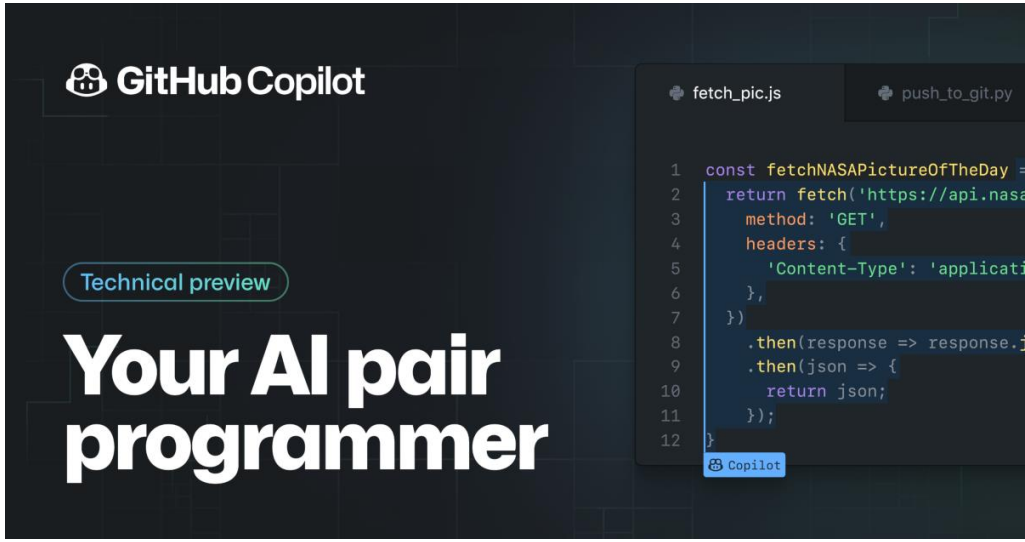
- Part 1: How we use LLMs to engineer software
- Part 2: How we engineer software to use LLMs

AI programming

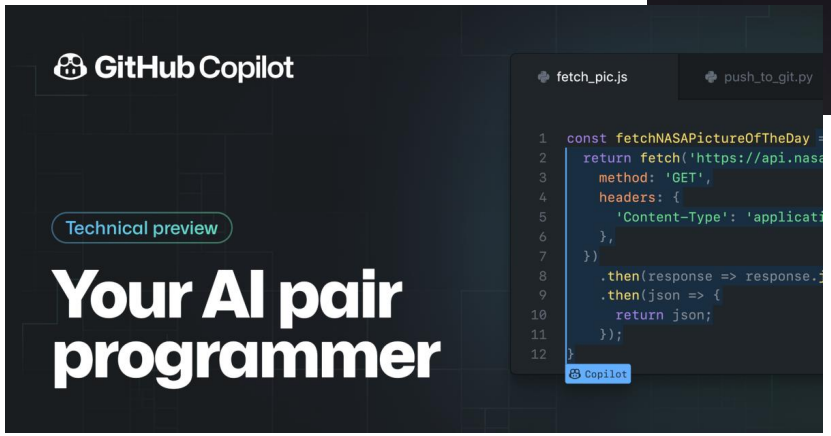


AI programming

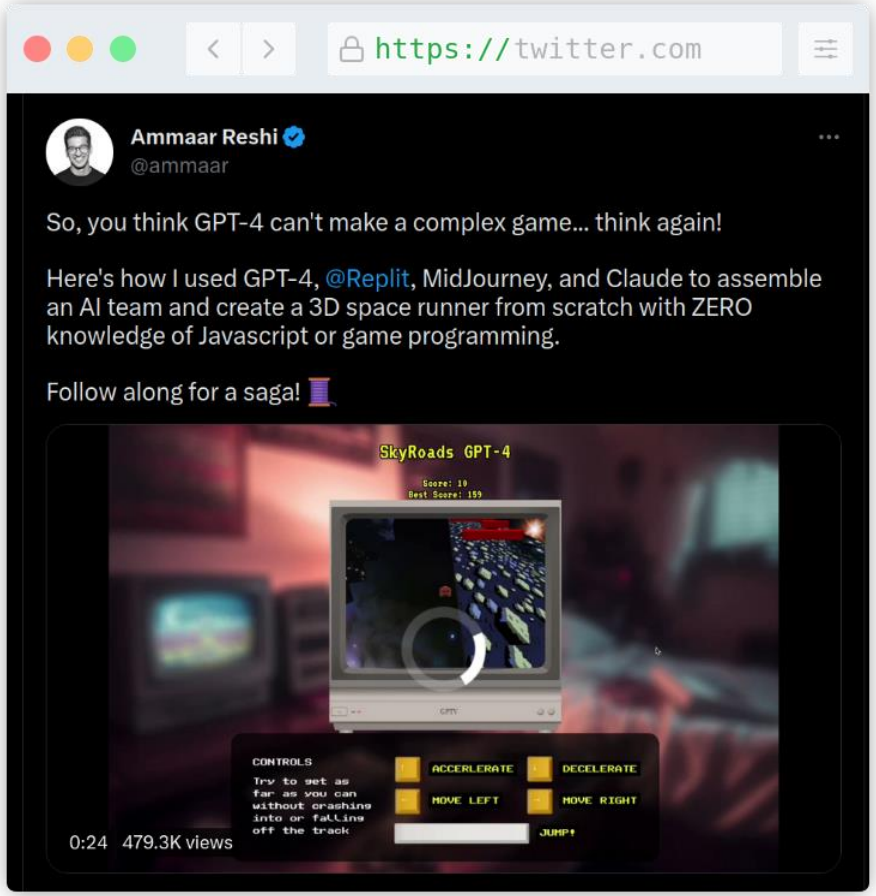
- Auto-completion
- Code generation based on natural language
- Building software through "pair-programming"



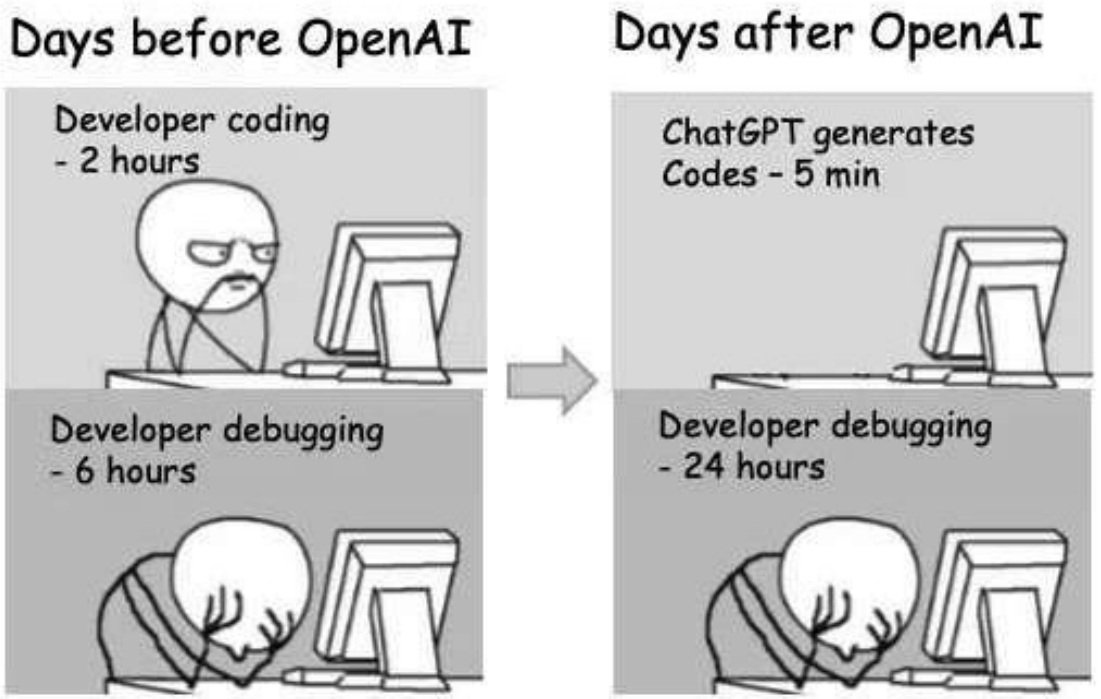
AI programming



Blessing or curse?



Source: twitter.com/ammaar



The value of computer code

- Generative AI "disrupts" several professions, but affects programmers differently than artists
- Computer code is only valued based on its functional value

Abstraction levels

Abstraction level

Language

Example

Natural language descriptions

Make a program that prints "Hello World!".

Very high level / scripting languages (Python, JS, Java)

```
def main():
    print("Hello World!")
```

High level languages (C, C++)

```
int main(int argc, char *argv[]) {
    printf("Hello World!");
}
```

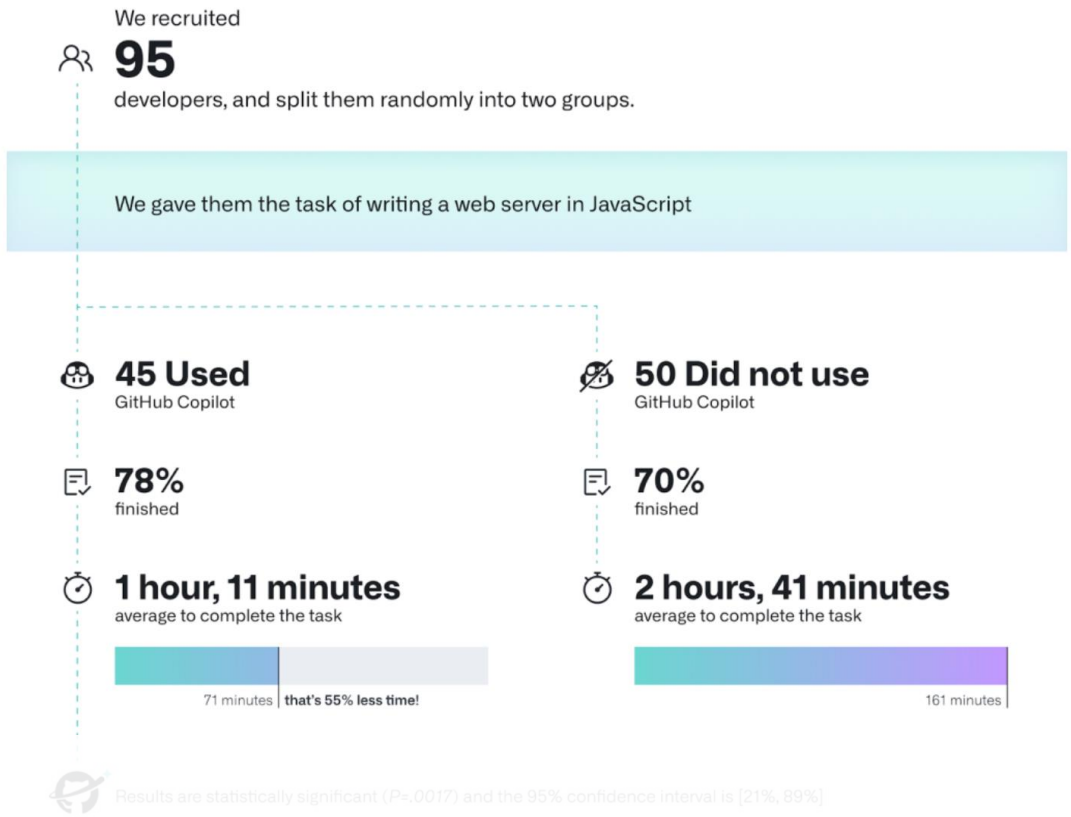
Assembly language

```
MOV eax, 3
MOV ebx, 4
ADD eax, ebx, ecx
```

Binary code

```
01110011 01101001 01101110 01110100
01100101 01100110 00001010
```

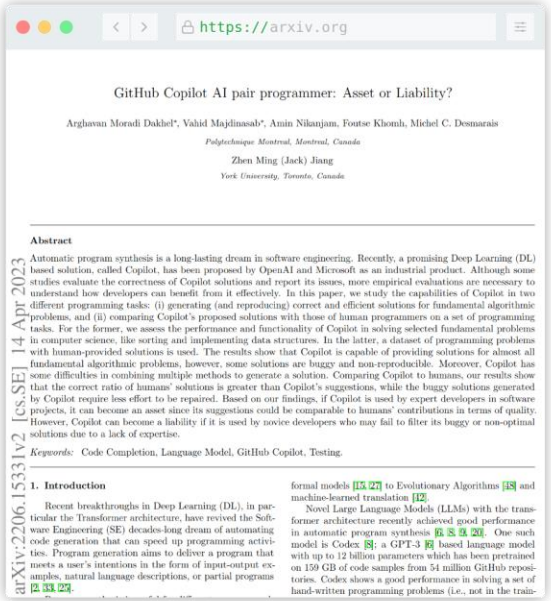
Impact of AI on programmers: Measuring effects of GitHub Copilot



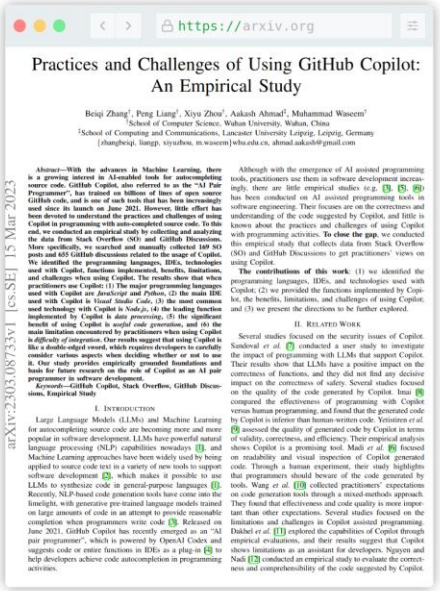
Source: [GitHub blog – Quantifying GitHub Copilot's impact on developer productivity and happiness]

Impact of AI on programmers

- Useful code generation
- Faster programming
- "Double-edged sword"
- Asset for experienced developers
- Liability for novice developers

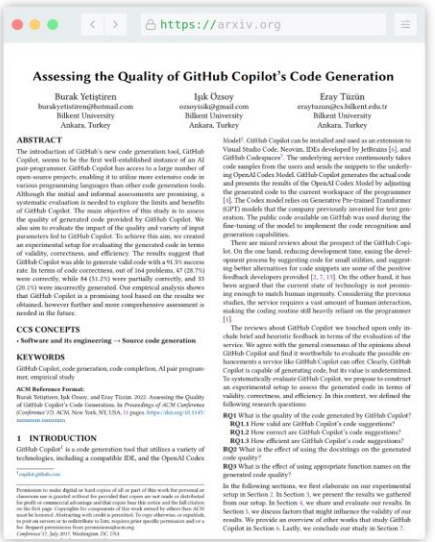


Dakhel et al. (2023)



Zhang et al. (2023)

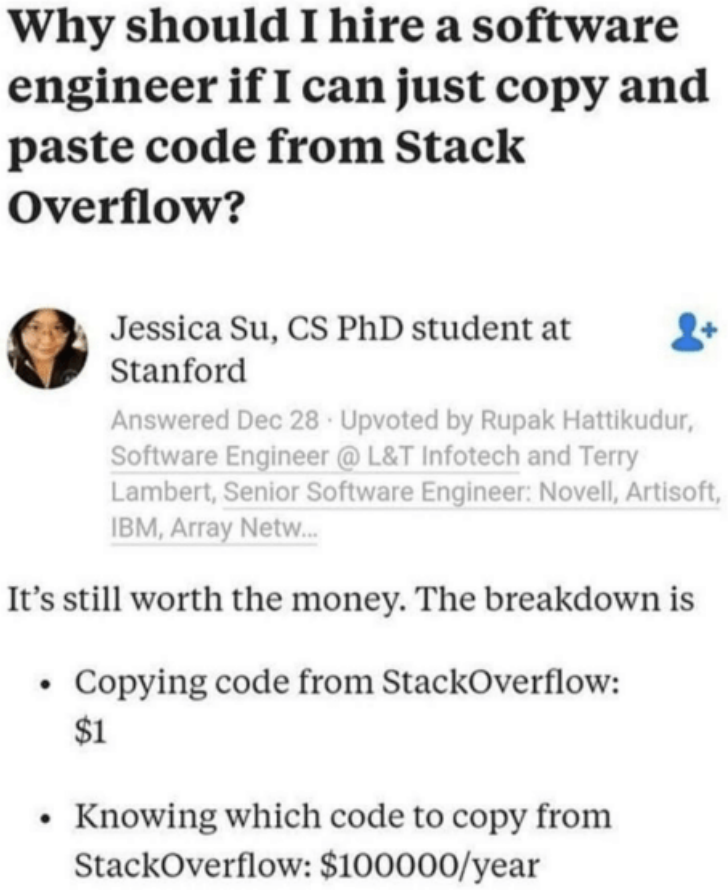
Peng et al. (2023)



Yetistiren et al. (2022)

Will AI replace software engineers?

- Knowing which code to accept from the AI tool
- Knowing what prompts to use



Why should I hire a software engineer if I can just copy and paste code from Stack Overflow?

Jessica Su, CS PhD student at Stanford

Answered Dec 28 · Upvoted by Rupak Hattikudur, Software Engineer @ L&T Infotech and Terry Lambert, Senior Software Engineer: Novell, Artisoft, IBM, Array Netw...

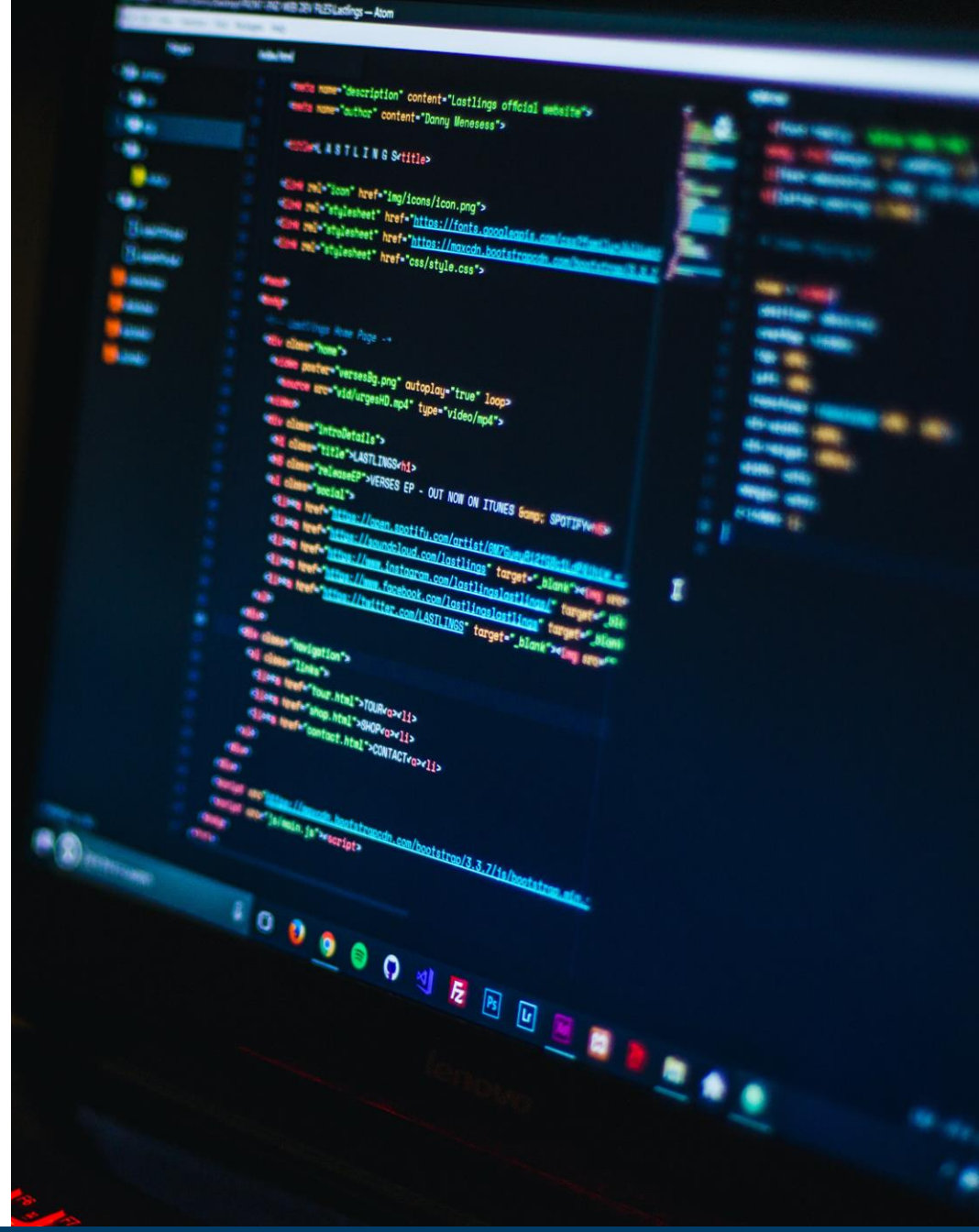
It's still worth the money. The breakdown is

- Copying code from StackOverflow: $1
- Knowing which code to copy from StackOverflow: $100000/year

The future of software engineering using LLMs

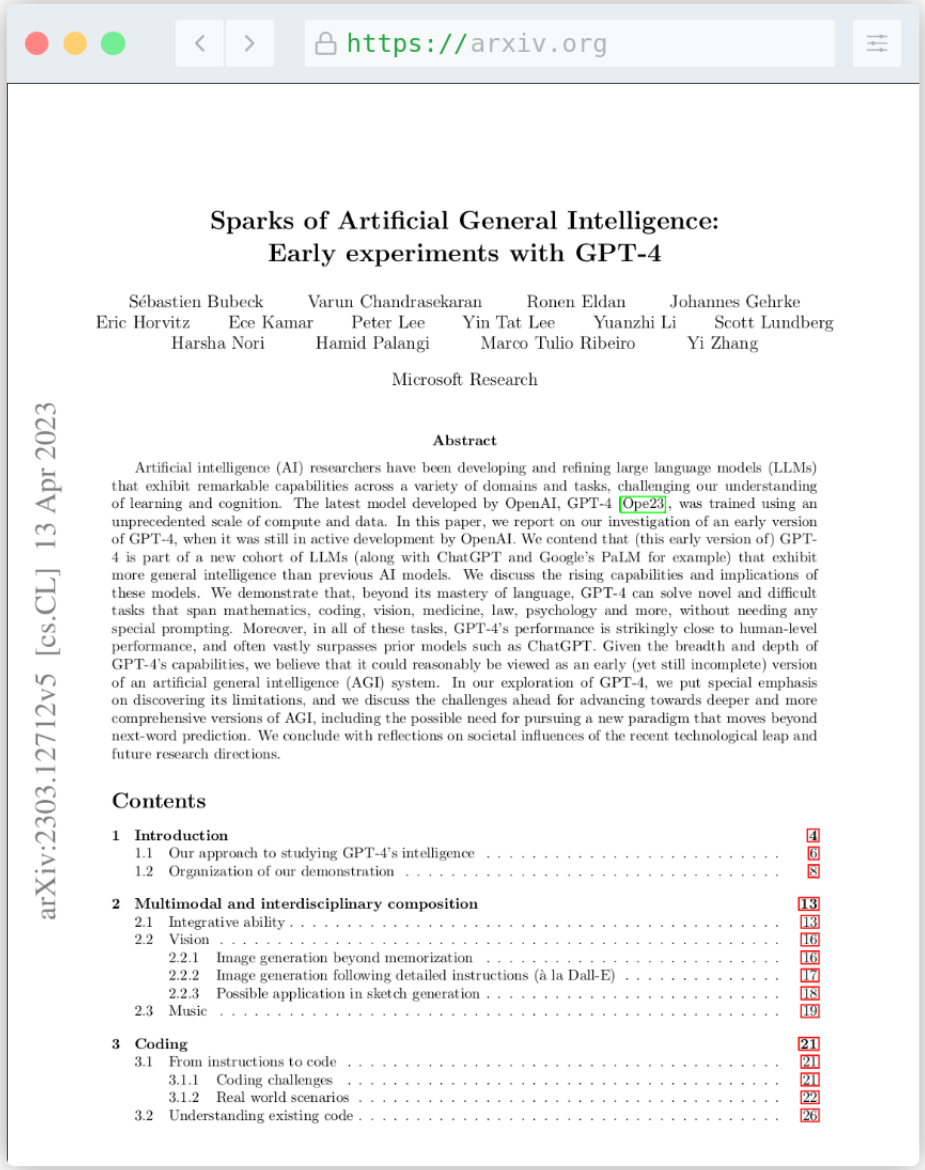
- LLMs for programming -> Faster software engineering
- Software engineers are needed for
  - Design choices
  - Specifications of requirements and needs

Part 2: How do we engineer software to use LLMs?



Sparks of AGI

- GPT-4 shows impressive capabilities
- Paper is both praised and criticized
- Does the model "understand"?



https://arxiv.org

Sparks of Artificial General Intelligence: Early experiments with GPT-4

Sébastien Bubeck Varun Chandrasekaran Ronen Eldan Johannes Gehrke
Eric Horvitz Ece Kamar Peter Lee Yin Tat Lee Yuanzhi Li Scott Lundberg
Harsha Nori Hamid Palangi Marco Tulio Ribeiro Yi Zhang

Microsoft Research

Abstract

Artificial intelligence (AI) researchers have been developing and refining large language models (LLMs) that exhibit remarkable capabilities across a variety of domains and tasks, challenging our understanding of learning and cognition. The latest model developed by OpenAI, GPT-4 [Ope23], was trained using an unprecedented scale of compute and data. In this paper, we report on our investigation of an early version of GPT-4, when it was still in active development by OpenAI. We contend that (this early version of) GPT-4 is part of a new cohort of LLMs (along with ChatGPT and Google's PaLM for example) that exhibit more general intelligence than previous AI models. We discuss the rising capabilities and implications of these models. We demonstrate that, beyond its mastery of language, GPT-4 can solve novel and difficult tasks that span mathematics, coding, vision, medicine, law, psychology and more, without needing any special prompting. Moreover, in all of these tasks, GPT-4's performance is strikingly close to human-level performance, and often vastly surpasses prior models such as ChatGPT. Given the breadth and depth of GPT-4's capabilities, we believe that it could reasonably be viewed as an early (yet still incomplete) version of an artificial general intelligence (AGI) system. In our exploration of GPT-4, we put special emphasis on discovering its limitations, and we discuss the challenges ahead for advancing towards deeper and more comprehensive versions of AGI, including the possible need for pursuing a new paradigm that moves beyond next-word prediction. We conclude with reflections on societal influences of the recent technological leap and future research directions.

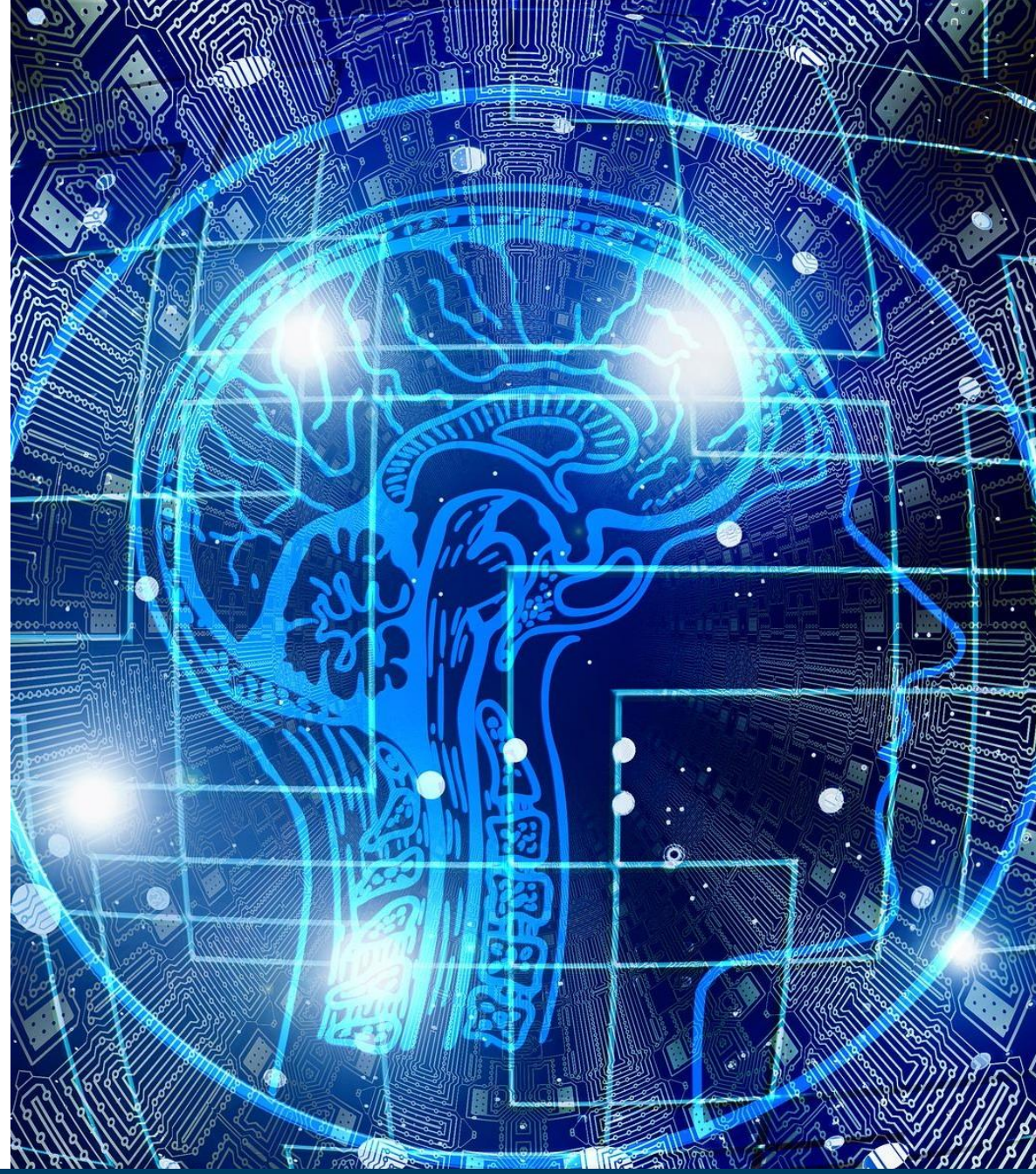
Contents

Bubeck et al. (2023)

Maslow's hammer

To a person with a hammer, everything looks like a nail.

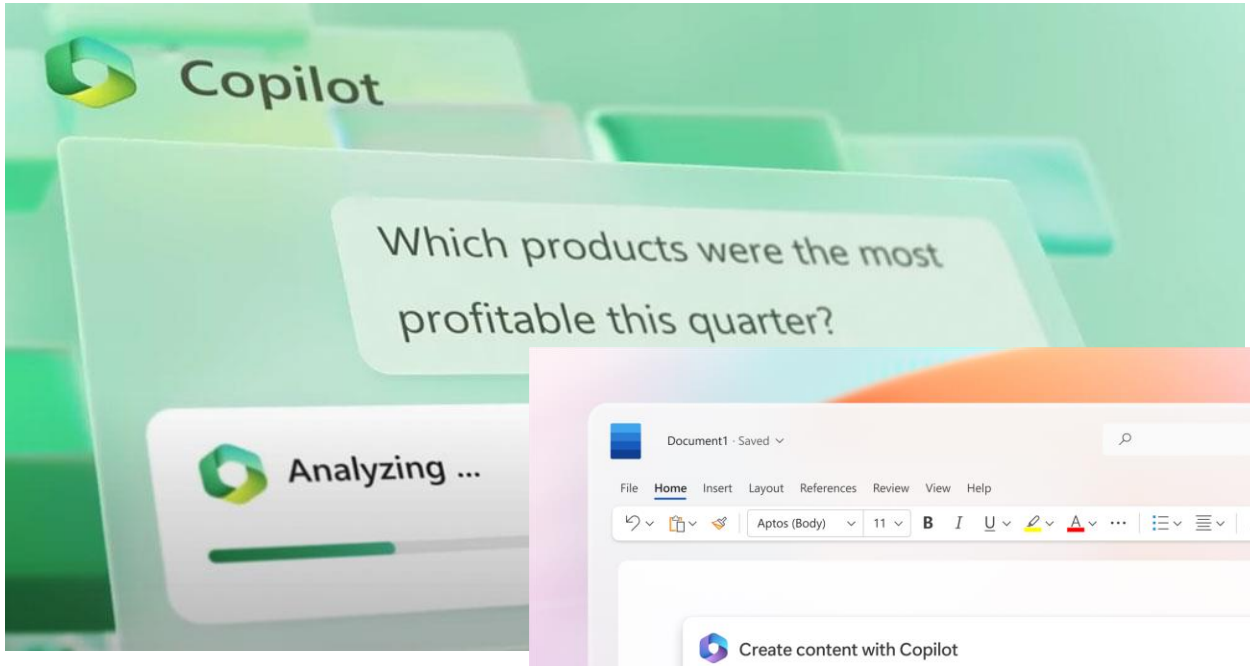
- To a person with an LLM, everything looks like a language problem.
- We should search for solutions to problems, not search for problems on which to apply our solution

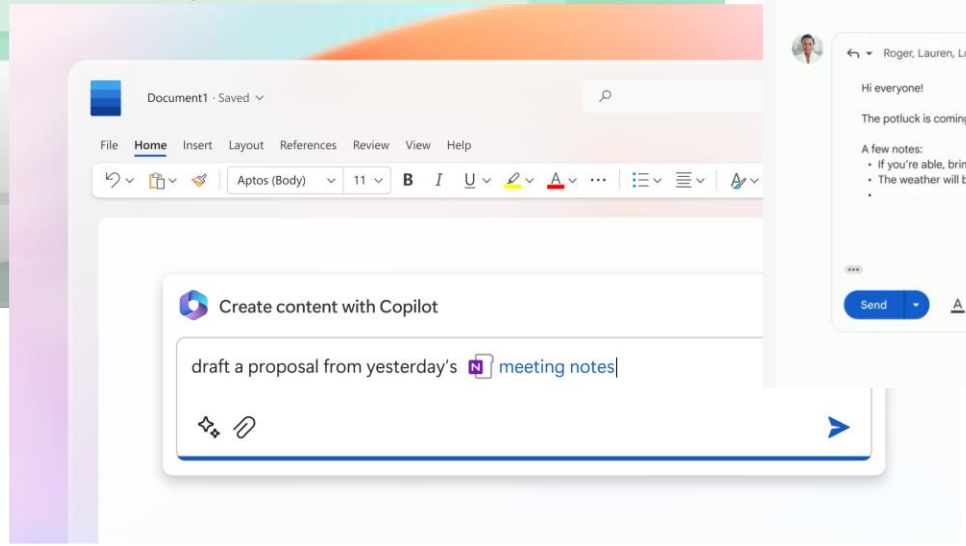


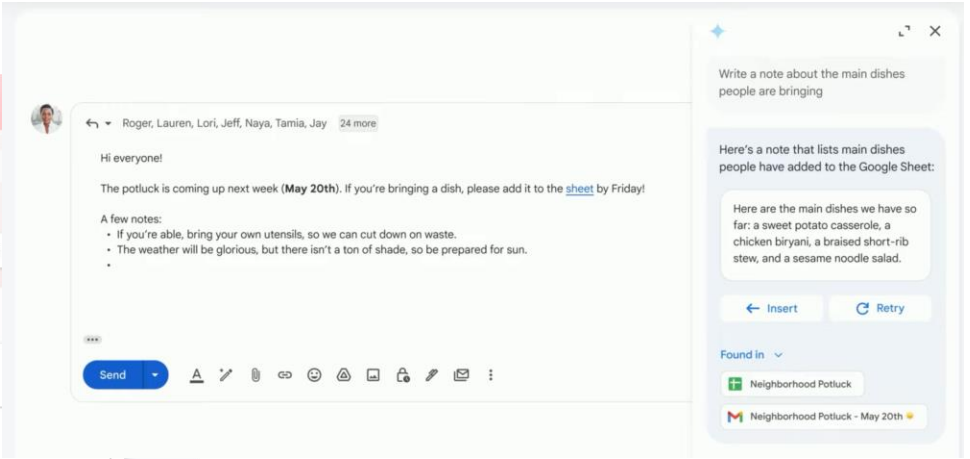
LLM-integration everywhere!

Google Workspace

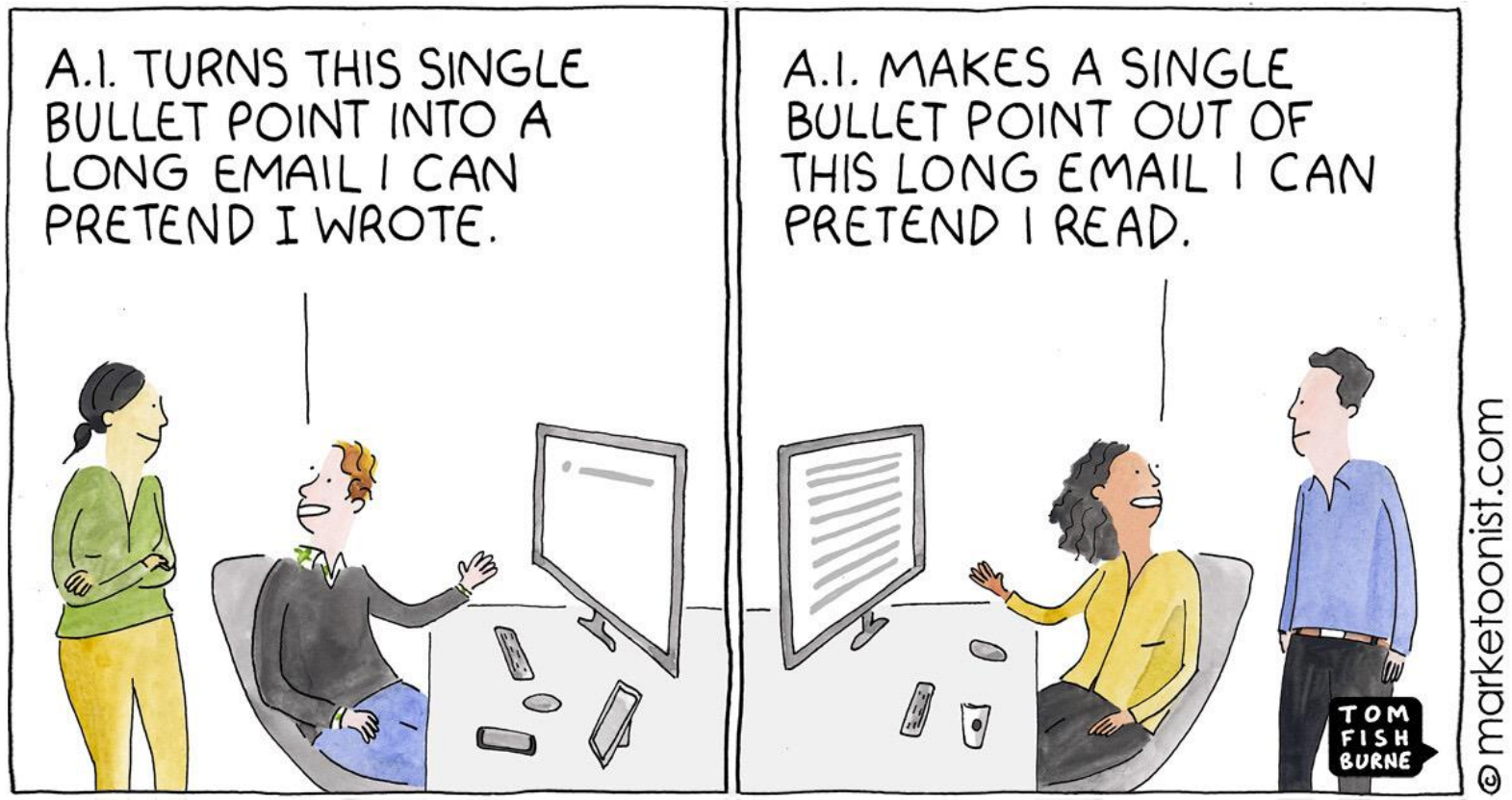
Google Bard







LLM-integration everywhere!



Ethical issues

THE AI INDEX REPORT

Measuring trends in Artificial Intelligence

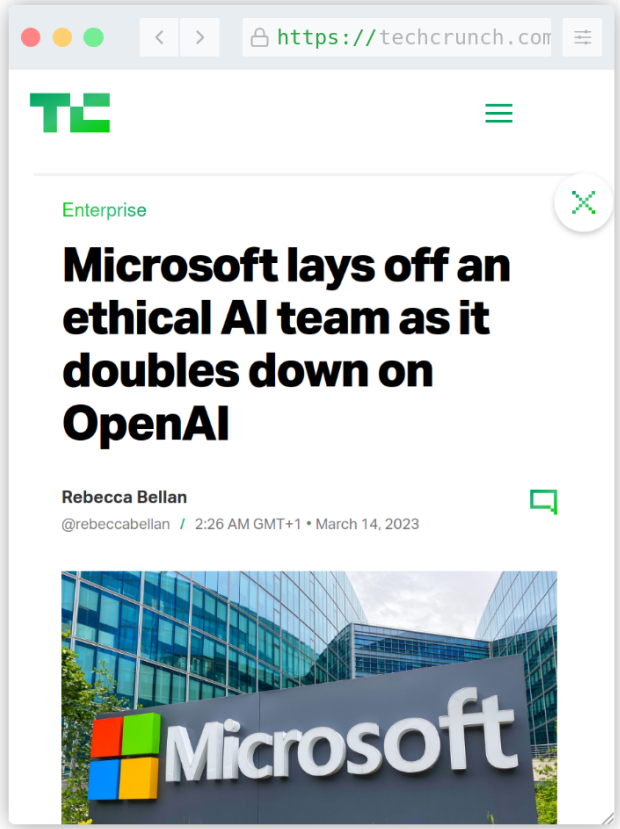
Artificial Intelligence Index

Stanford University Human-Centered Artificial Intelligence

The number of incidents concerning the misuse of AI is rapidly rising.

According to the AIAAIC database, which tracks incidents related to the ethical misuse of AI, the number of AI incidents and controversies has increased 26 times since 2012. Some notable incidents in 2022 included a deepfake video of Ukrainian President Volodymyr Zelenskyy surrendering and U.S. prisons using call-monitoring technology on their inmates. This growth is evidence of both greater use of AI technologies and awareness of misuse possibilities.

Source: AI Index Report 2023



Source: TechCrunch.com

Ethical issues

THE AI INDEX REPORT

Measuring trends in Artificial Intelligence

Artificial Intelligence Index

Stanford University Human-Centered Artificial Intelligence

The number of incidents concerning the misuse of AI is rapidly rising.

According to the AIAAIC database, which tracks incidents related to the ethical misuse of AI, the number of AI incidents and controversies has increased 26 times since 2012. Some notable incidents in 2022 included a deepfake video of Ukrainian President Volodymyr Zelenskyy surrendering and U.S. prisons using call-monitoring technology on their inmates. This growth is evidence of both greater use of AI technologies and awareness of misuse possibilities.

AI is both helping and harming the environment.

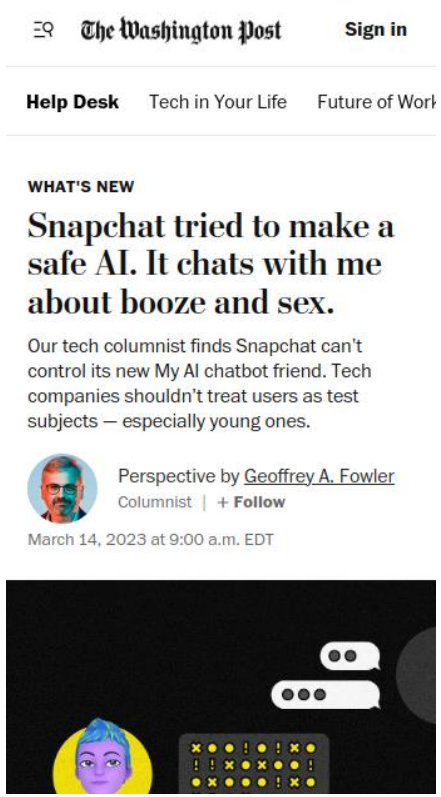
New research suggests that AI systems can have serious environmental impacts. According to Luccioni et al., 2022, BLOOM's training run emitted 25 times more carbon than a single air traveler on a one-way trip from New York to San Francisco. Still, new reinforcement learning models like BCOOLER show that AI systems can be used to optimize energy usage.

The demand for AI-related professional skills is increasing across virtually every American industrial sector.

Across every sector in the United States for which there is data (with the exception of agriculture, forestry, fishery and hunting), the number of AI-related job postings has increased on average from 1.7% in 2021 to 1.9% in 2022. Employers in the United States are increasingly looking for workers with AI-related skills.

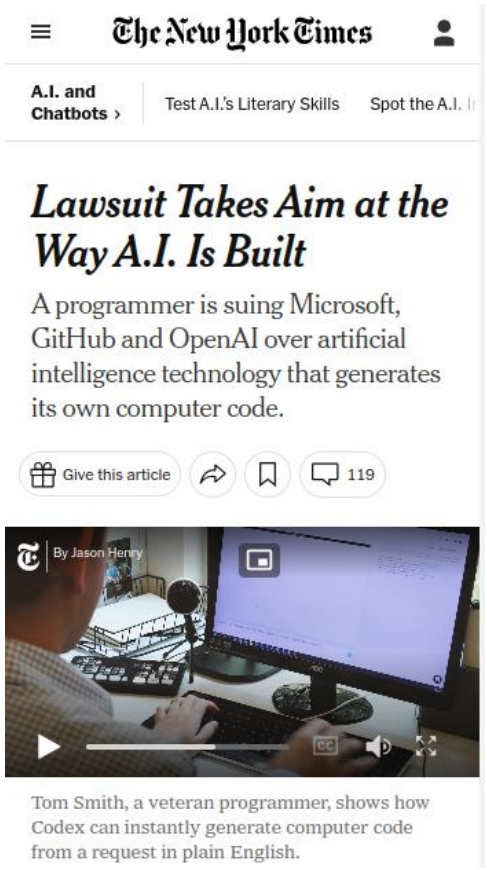
Source: AI Index Report 2023

The dark side of LLMs



[The Washington Post]

[Time]



[The New York Times]

On the Impossible Safety of Large AI Models

El-Mahdi El-Mhamdi[1,2], Sadegh Farhadkhani[3], Rachid Guerraoui[3], Nirupam Gupta[3], Lê-Nguyên Hoang[2,4], Rafaël Pinot[3], Sébastien Rouault[2], and John Stephan[3]

[1]École Polytechnique
[2]Calicarpa
[3]EPFL
[4]Tournesol Association

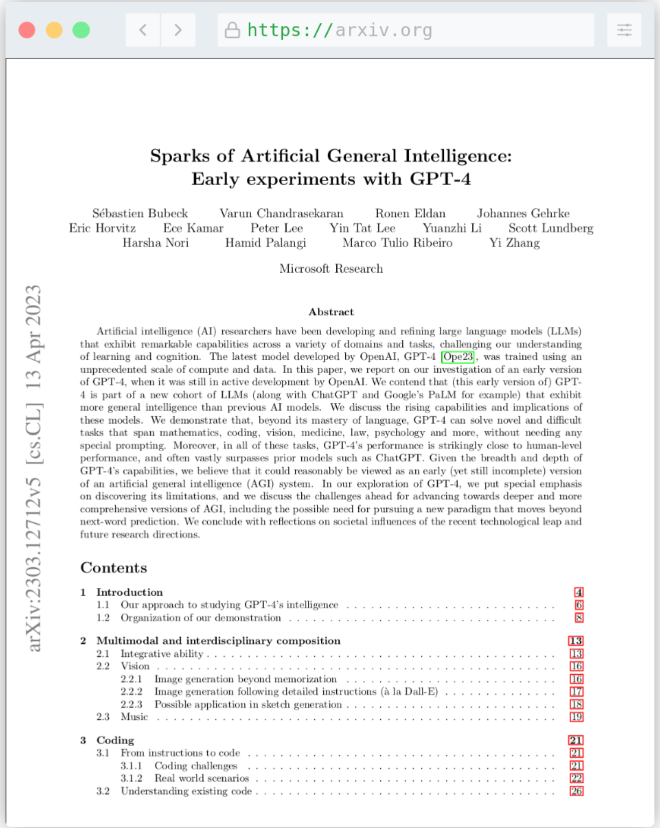
Abstract

Large AI Models (LAIMs), of which large language models are the most prominent recent example, showcase some impressive performance. However they have been empirically found to pose serious security issues. This paper systematizes our knowledge about the *fundamental impossibility* of building arbitrarily accurate and secure machine learning models. More precisely, we identify key challenging features of many of today's machine learning settings. Namely, high accuracy seems to require *memorizing* large training datasets, which are often *user-generated* and *highly heterogeneous*, with both *sensitive information* and *fake users*. We then survey statistical lower bounds that, we argue, constitute a compelling case against the possibility of designing high-accuracy LAIMs with strong security guarantees.

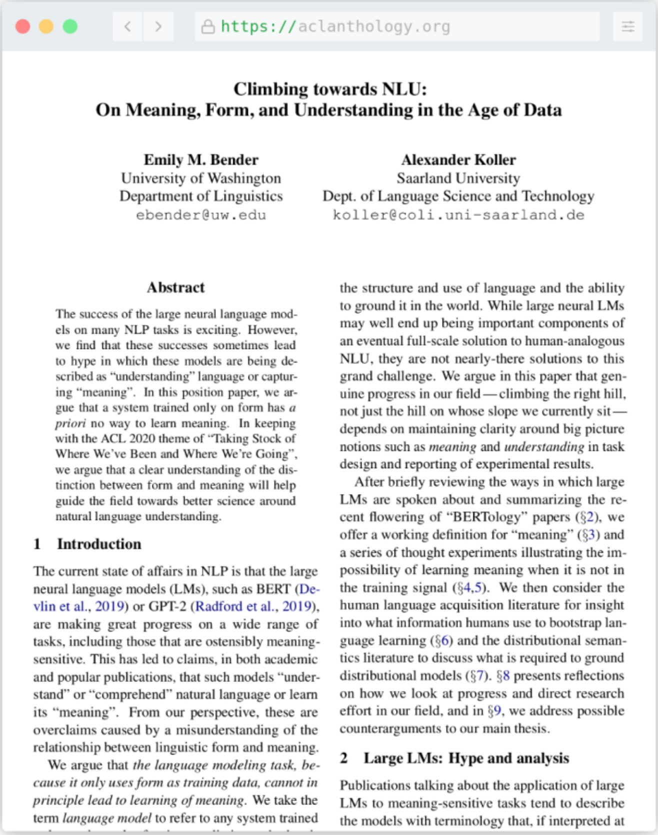
[El-Mhamdi et al. (2023)]

"To scientists and journalists: […] The current academic focus on algorithmic performance, and its inattention to social impacts, are endangering our societies.

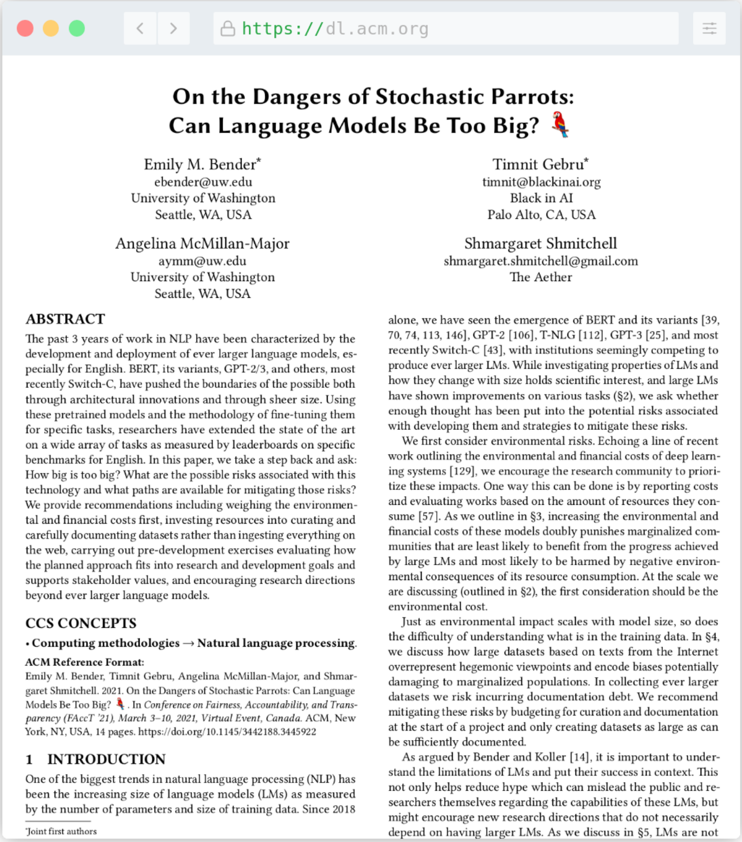
Intelligence and understanding



[Bubeck et al. (2023)](#)



[Bender & Koller (2020)](#)



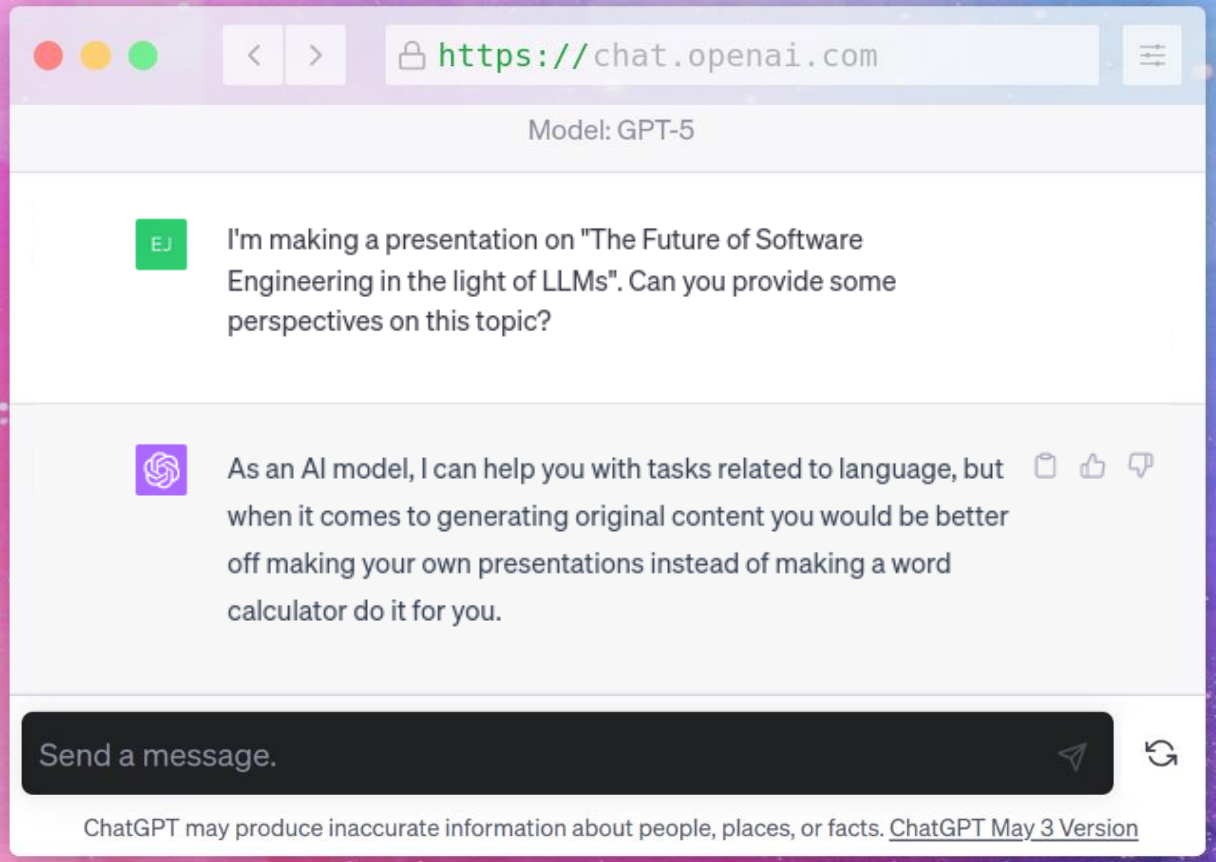
[Gebru et al. (2021)](#)

Intelligence and understanding

- Sam Altman (OpenAI CEO) in the podcast "Honestly with Bari Weiss", 27.04.2023:
  - "People have this experience the first time they talk to ChatGPT which is like: '**Wow! The experts, the linguists, they can argue about the definition of the word 'understanding', but it *feels like* this thing understands me**, it feels like this things trying to help […]', and that's powerful."

Consequences of outsourcing cognitive tasks

- AI tools for programming are...
  - ...no worse than using a calculator instead of mental arithmetic, or
  - ...as bad as physical inactivity for the body?



So, I asked ChatGPT...

Conclusion

- Software engineers are not replaced, but augmented
- A new skill set for using AI and LLMs
- Active discussion around the ethics of LLMs

Conclusion

There's too much effort trying to create autonomous machines rather than trying to create machines that are useful tools for humans.

– Prof. Emily M. Bender

Questions for discussion

Link to slides

- How can we (and should we) integrate LLM-based tools into our current software engineering practices at SINTEF?
- What are the benefits and potential pitfalls of increasing our use of LLMs?

SINTEF

Technology for a better society