What are AI Agents? Why do they matter?

Prof. Fernando Perez-Cruz

This talk represent my views and not necessary those of the BIS

Game-Changer AI Agents

- Large Language Models (2018) …
  - have the capacity to (correctly?) respond to **any** input.

Cambrian Explosion!

Game-Changer AI Agents

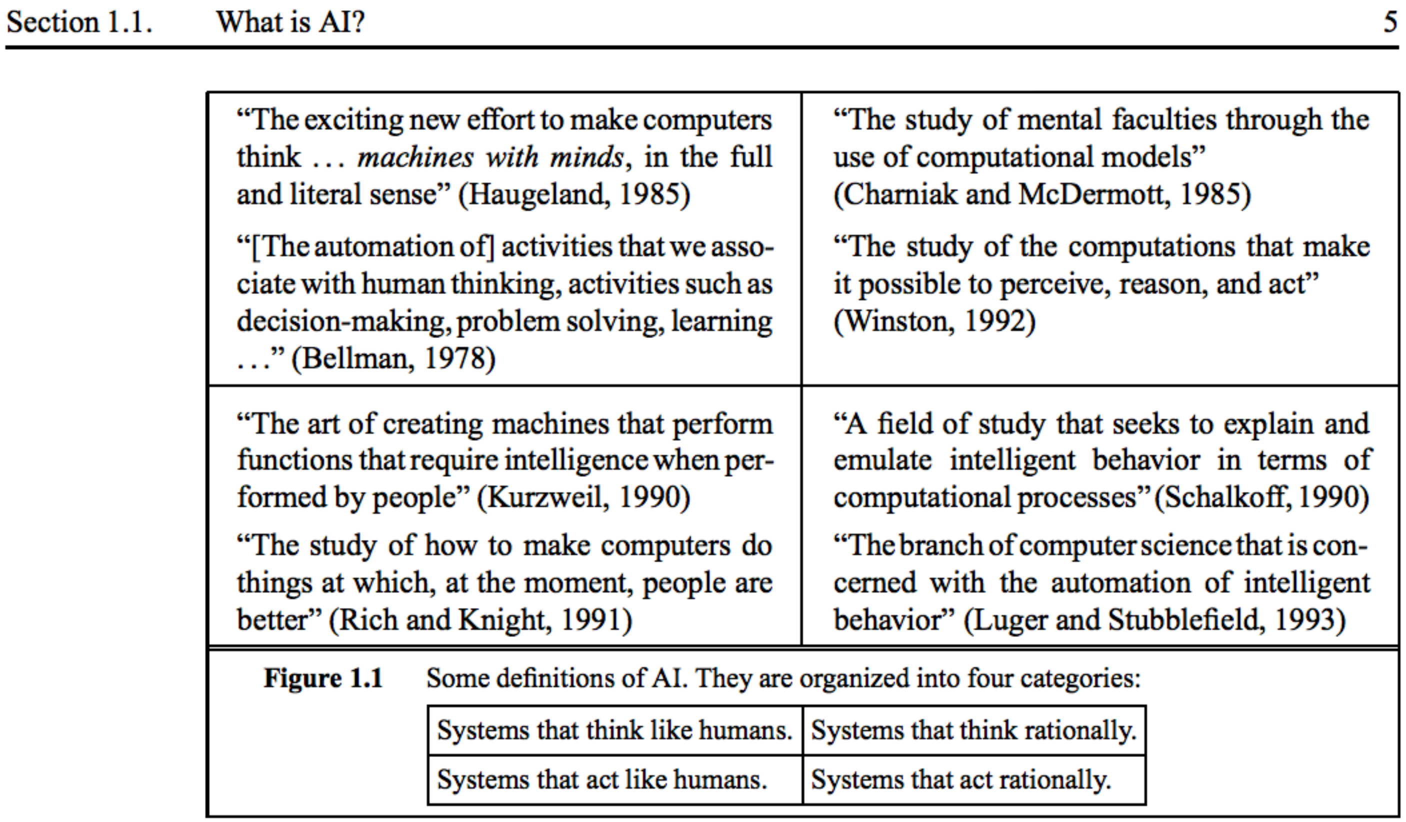
- Large Language Models (2018) …
  - have the capacity to (correctly?) respond to **any** question.

- Artificial General Intelligence (AGI):
  - A general-purpose AI system that can do almost all **cognitive** tasks a human can do.

- AI Agent …
  - is software that uses AI to pursue goals and complete tasks on behalf of users. (standard)
  - is a Multimodal LLM that can **control** a computer as a human. (useful)

- AI Agents would be the **gateway** to AGI (if there is such a thing).
  - Even if there were no AGI, they could still be extremely helpful today/tomorrow.

What is AI? What is Machine Learning?

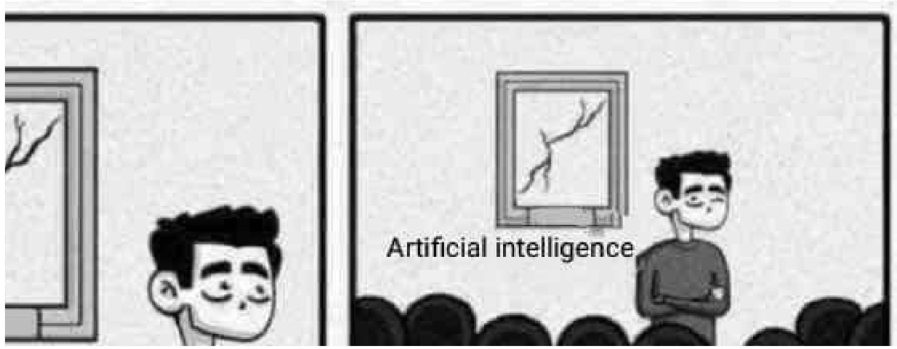


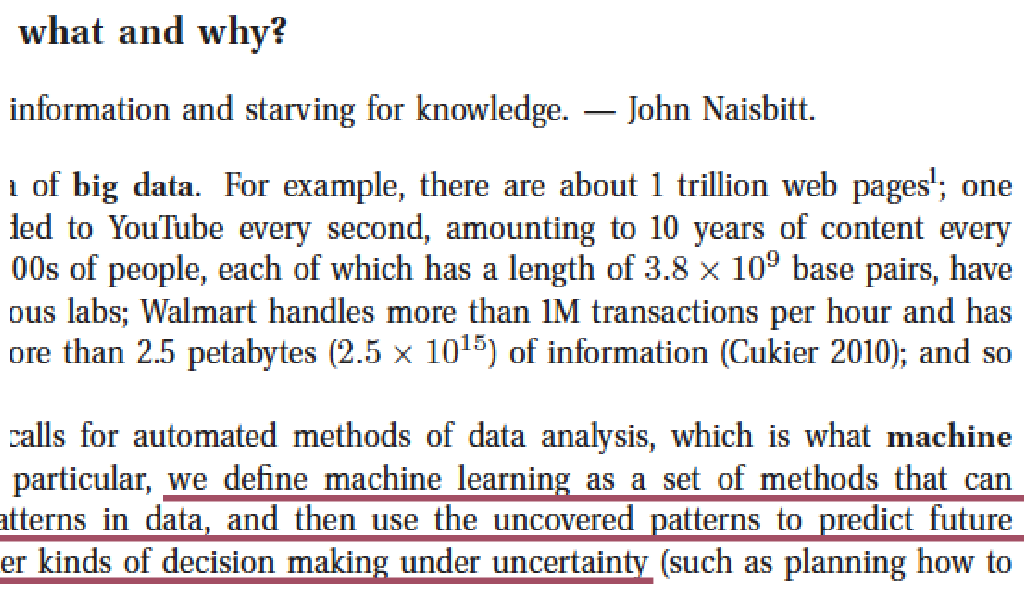
Section 1.1. What is AI? 5

"The exciting new effort to make computers think … *machines with minds*, in the full and literal sense" (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning …" (Bellman, 1978)	"The study of mental faculties through the use of computational models" (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act" (Winston, 1992)
"The art of creating machines that perform functions that require intelligence when performed by people" (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better" (Rich and Knight, 1991)	"A field of study that seeks to explain and emulate intelligent behavior in terms of computational processes" (Schalkoff, 1990) "The branch of computer science that is concerned with the automation of intelligent behavior" (Luger and Stubblefield, 1993)

Figure 1.1 Some definitions of AI. They are organized into four categories:

Systems that think like humans.	Systems that think rationally.
Systems that act like humans.	Systems that act rationally.





what and why?

information and starving for knowledge. — John Naisbitt.

of **big data**. For example, there are about 1 trillion web pages[1]; one led to YouTube every second, amounting to 10 years of content every 00s of people, each of which has a length of 3.8×10^9 base pairs, have ous labs; Walmart handles more than 1M transactions per hour and has ore than 2.5 petabytes (2.5×10^{15}) of information (Cukier 2010); and so

calls for automated methods of data analysis, which is what **machine learning** provides. In particular, we define machine learning as a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty (such as planning how to collect more data!).

Why ML is not statistics?

$$\min_{w} \frac{1}{n} \sum_{i=1}^{n} L\big(y_i, f_w(\boldsymbol{x}_i)\big)$$

$$\{(\boldsymbol{x}_i, y_i)\}_{i=1\ldots n} \underset{iid}{\sim} p_{X,Y}(\boldsymbol{x}, y)$$

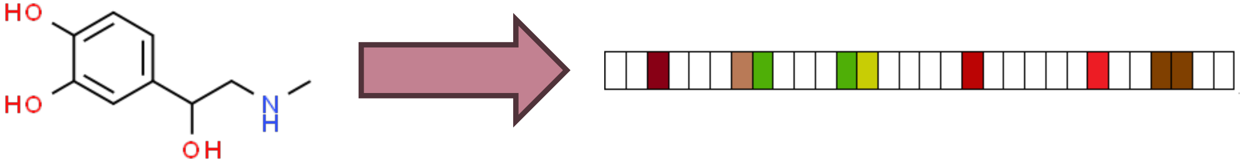
Machine Learners and Statisticians solved the **same** problem

Key aspects in Machine Learning

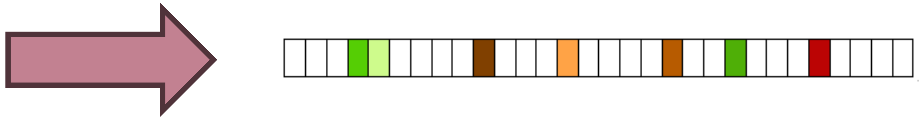
- Embeddings:
  - Transform non vectorial data into vectors.
- Over-parametrization:
  - x100 or x1000 more parameters than data.
- Data:
  - The more the merrier.
  - No diminishing return.
- Self-supervised Learning:
  - Labels are expensive.
  - Data is not.
- Zero-short learning:
  - From prediction to generation.

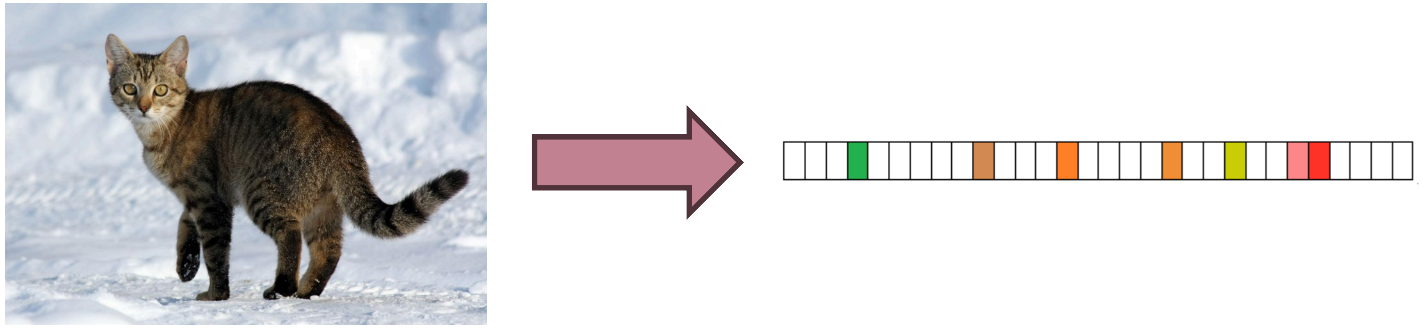
Embeddings

- What is an embeddings?
  - Transformation of an image, word, chemical structure, protein, game position … any data into a vector.
- Why we need embeddings?
  - We can compute similarities between vectors.

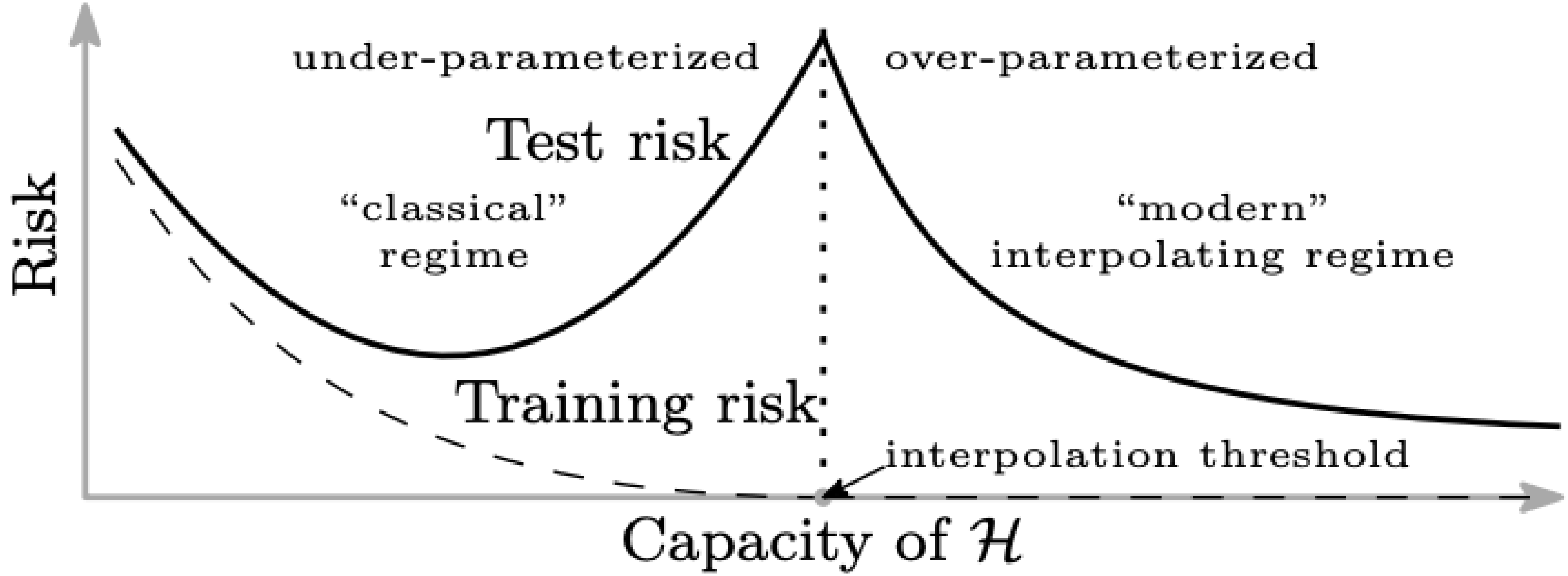


Cat

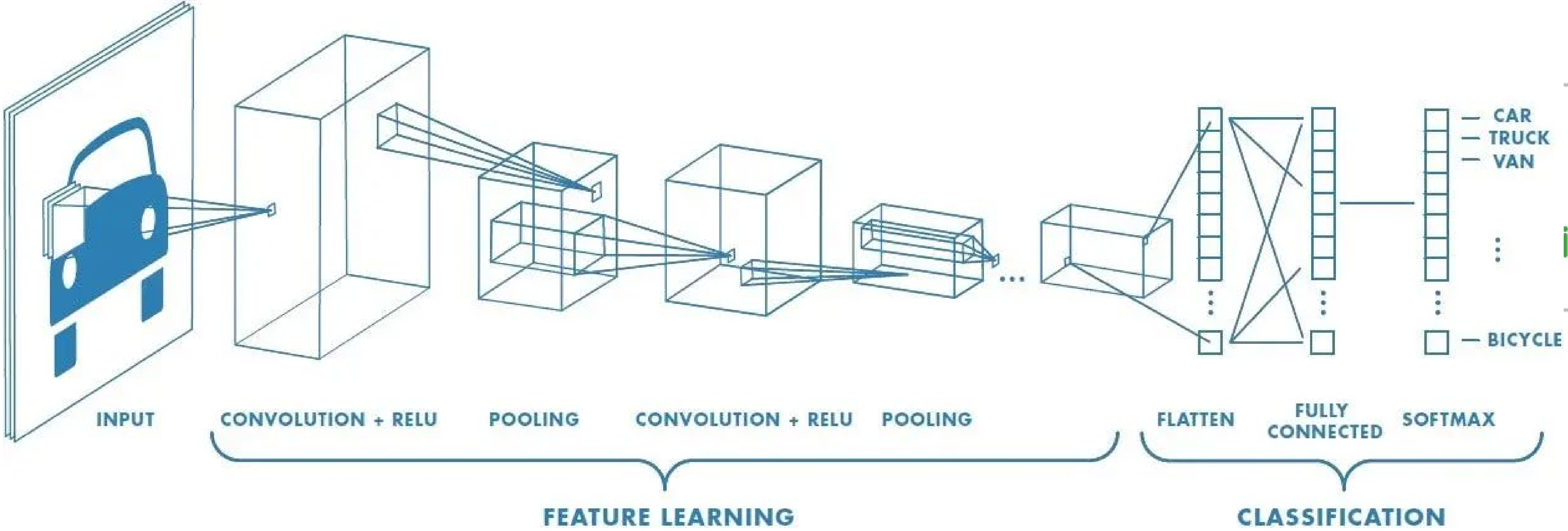




Overparametrization



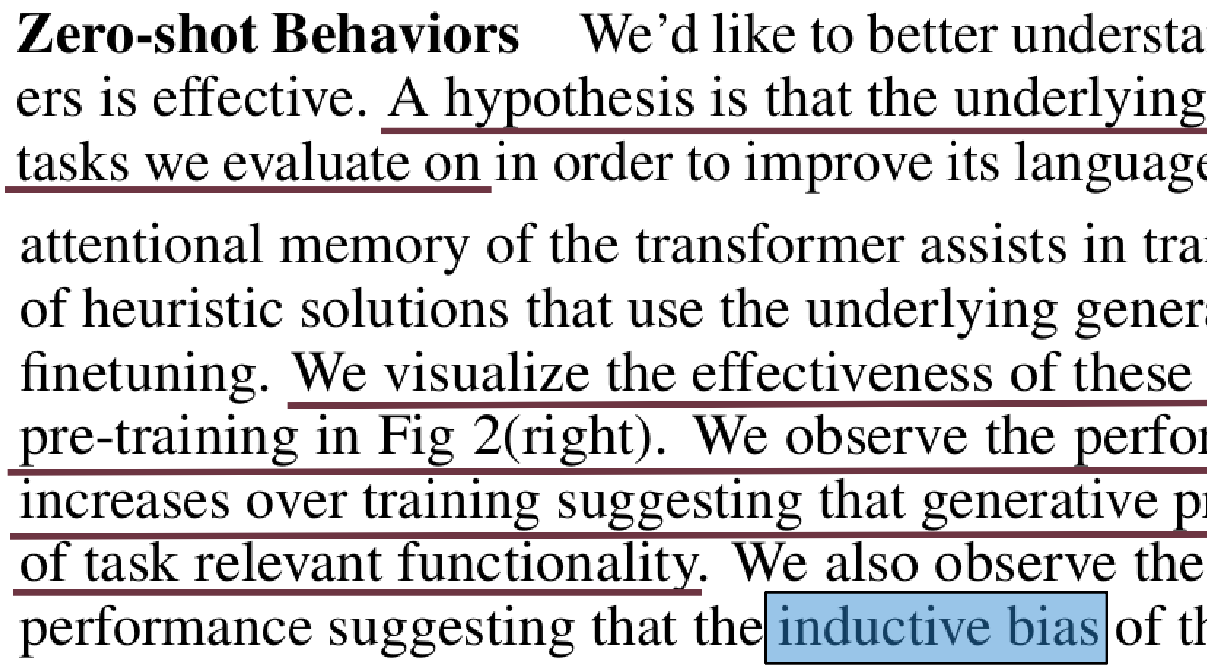
Self-Supervised Learning



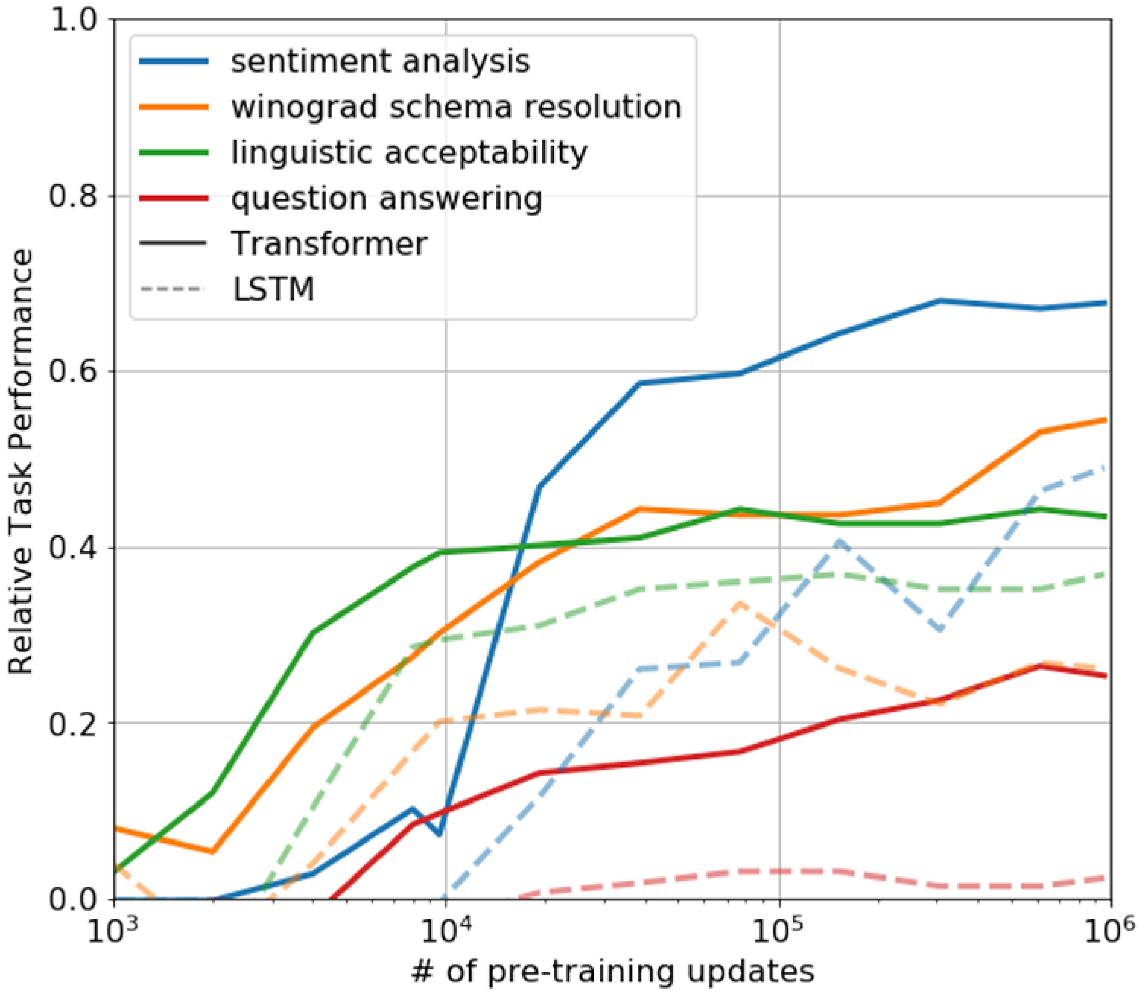
past

ible

Zero-shot



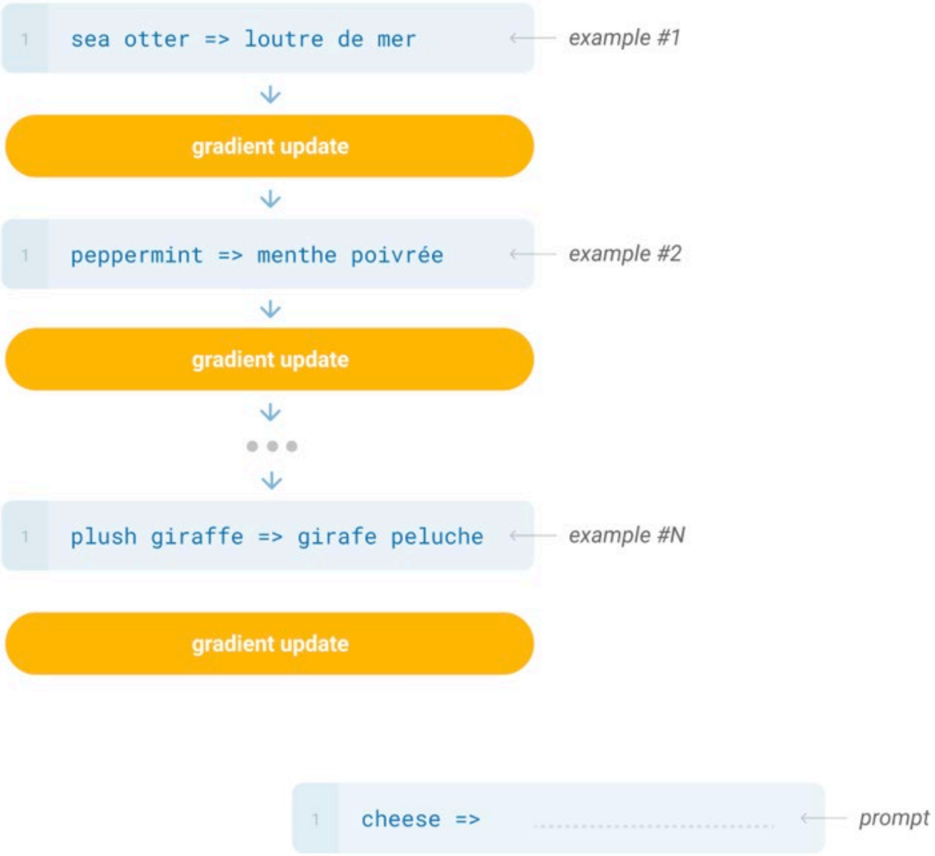
Zero-shot Behaviors We'd like to better understa
ers is effective. A hypothesis is that the underlying
tasks we evaluate on in order to improve its languag
attentional memory of the transformer assists in tra
of heuristic solutions that use the underlying gener
finetuning. We visualize the effectiveness of these
pre-training in Fig 2(right). We observe the perfo
increases over training suggesting that generative p
of task relevant functionality. We also observe the
performance suggesting that the inductive bias of t



Improving Language Understanding by Generative Pre-Training

LLMs are zero-shot learners

- Traditional ML:



```
sea otter => loutre de mer
```
example #1

gradient update

```
peppermint => menthe poivrée
```
example #2

gradient update

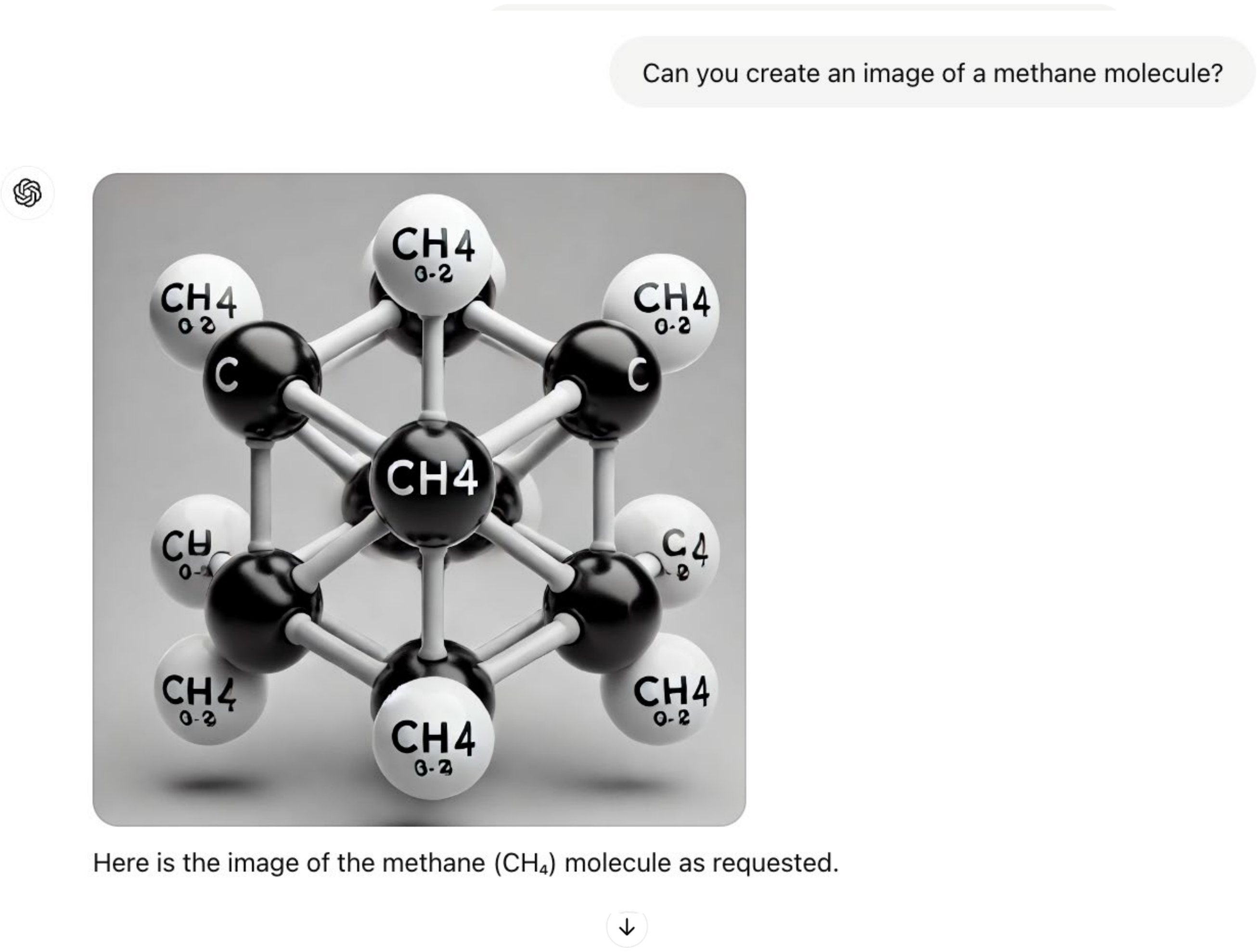
```
plush giraffe => girafe peluche
```
example #N

gradient update

```
cheese =>
```
prompt

Language Models are Few Shot Learners

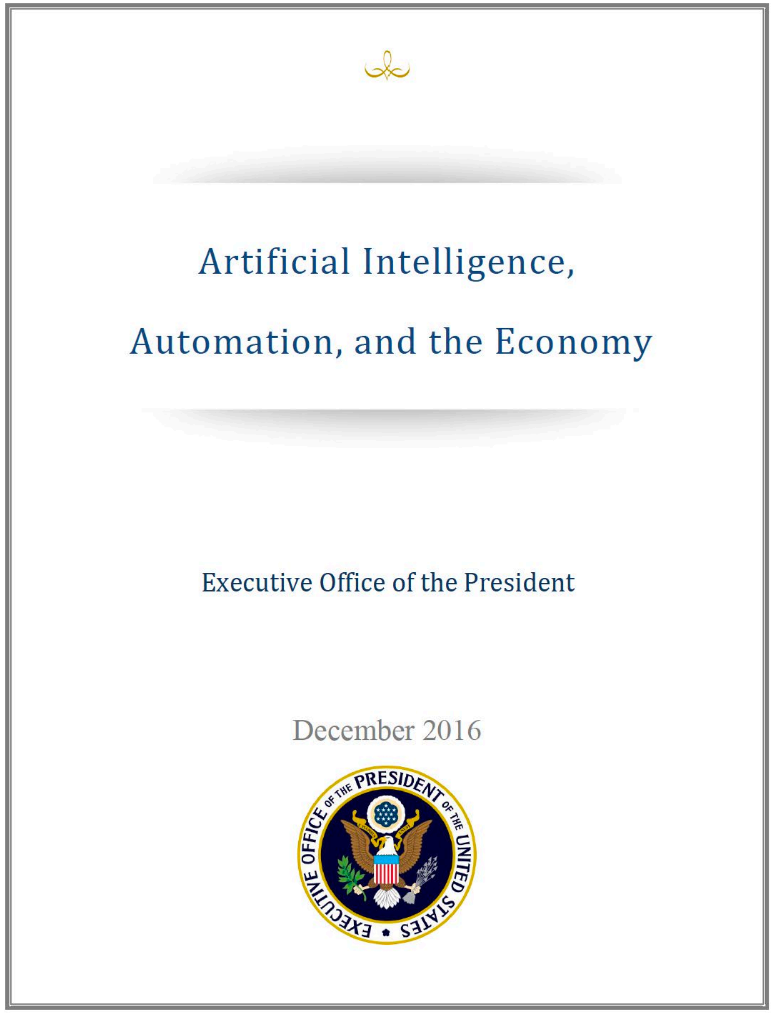
Universal



Can you create an image of a methane molecule?

Here is the image of the methane (CH_4) molecule as requested.

Have we been here before?

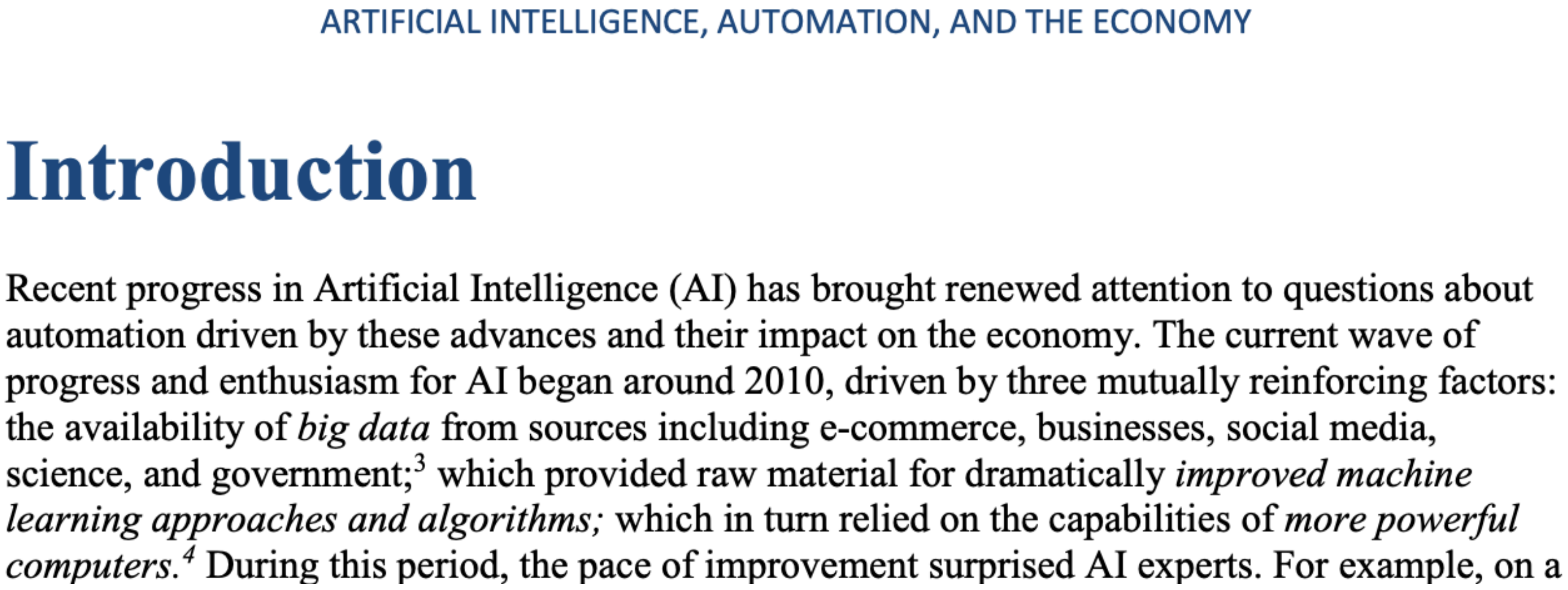


Artificial Intelligence,

Automation, and the Economy

Executive Office of the President

December 2016

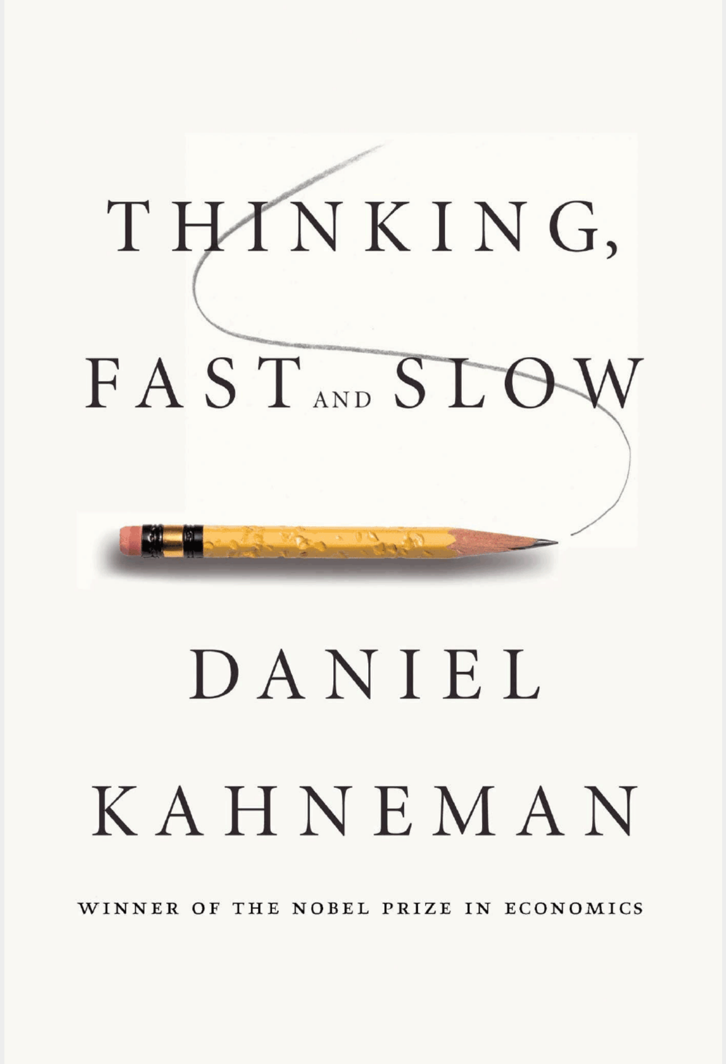


ARTIFICIAL INTELLIGENCE, AUTOMATION, AND THE ECONOMY

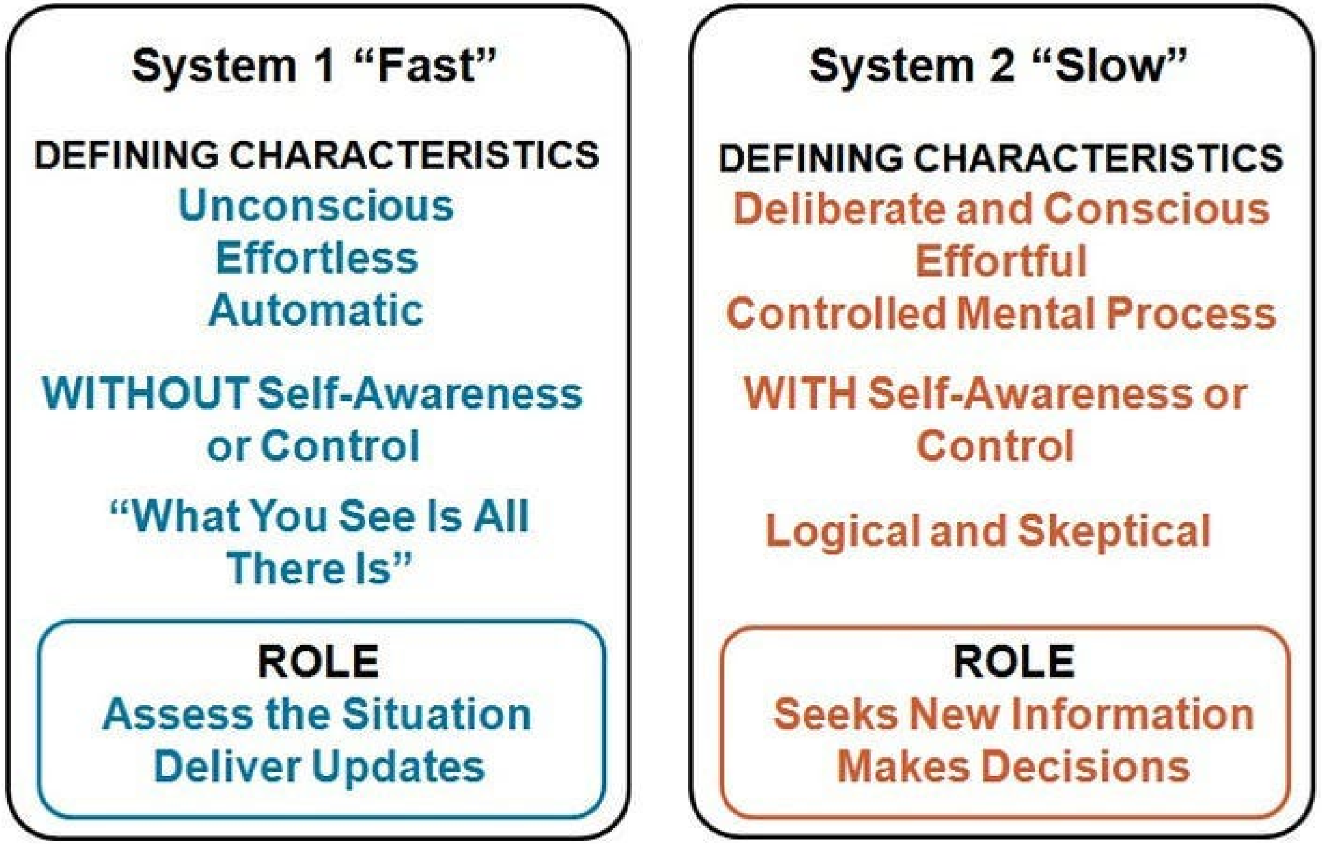
Introduction

Recent progress in Artificial Intelligence (AI) has brought renewed attention to questions about automation driven by these advances and their impact on the economy. The current wave of progress and enthusiasm for AI began around 2010, driven by three mutually reinforcing factors: the availability of *big data* from sources including e-commerce, businesses, social media, science, and government;[3] which provided raw material for dramatically *improved machine learning approaches and algorithms;* which in turn relied on the capabilities of *more powerful computers.*[4] During this period, the pace of improvement surprised AI experts. For example, on a

Framework



System 1 and 2 thinking



System 1 "Fast"

DEFINING CHARACTERISTICS

Unconscious
Effortless
Automatic

WITHOUT Self-Awareness or Control

"What You See Is All There Is"

ROLE

Assess the Situation
Deliver Updates

System 2 "Slow"

DEFINING CHARACTERISTICS

Deliberate and Conscious
Effortful
Controlled Mental Process

WITH Self-Awareness or Control

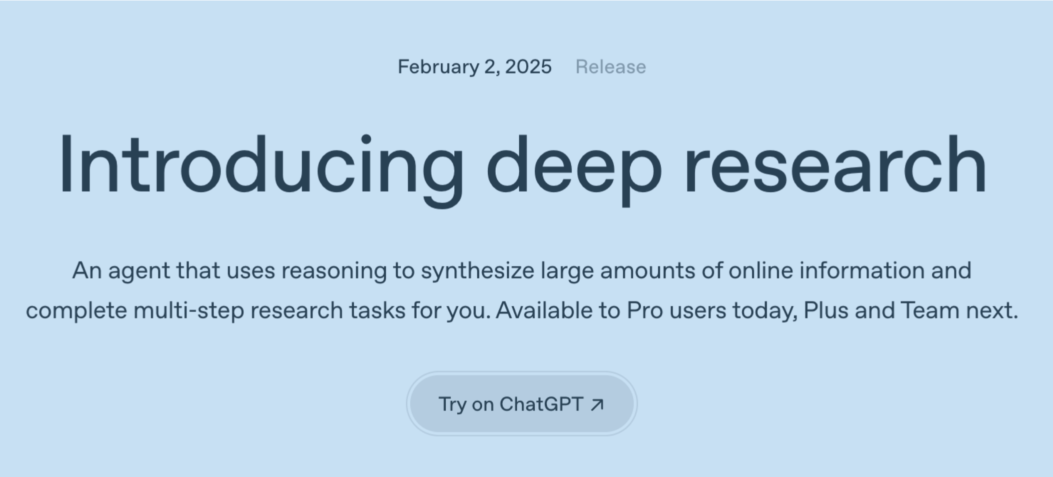
Logical and Skeptical

ROLE

Seeks New Information
Makes Decisions

Two Types of Agents

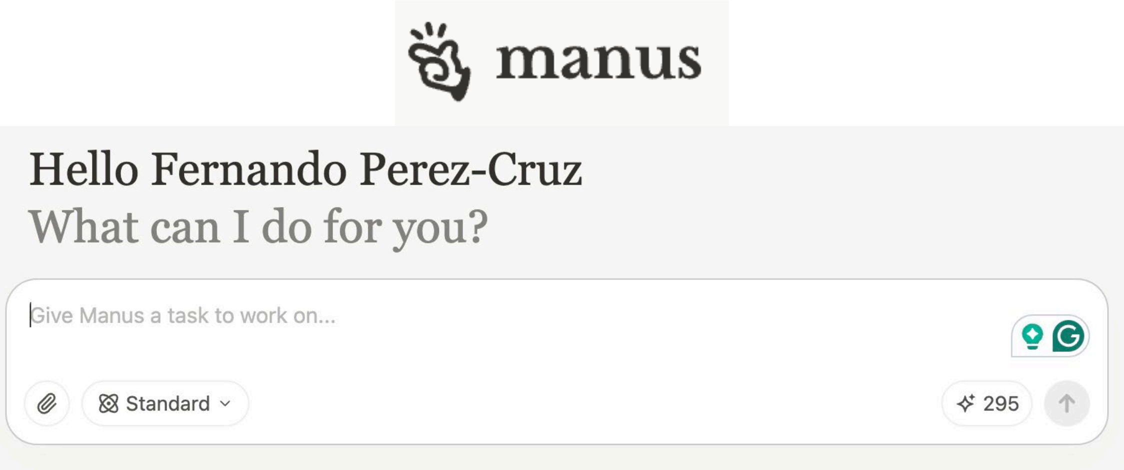
- Narrow Agents: Deep Research and Reasoning
  - Emulate System 2 (not AGI).
  - They are as good as the person using it.
  - Examples:
    - OpenAI, xAI and DeepMind.
- General Purpose: LLMs that use computers like a human.
  - Engage System 1 and emulate System 2.
  - Many cognitive functions at the same time.
  - Errors are too high.
  - Available today:
    - Anthropic (Claude with CU),
    - Open AI (Operator), Manus,
    - DeepMind (Mariner).



February 2, 2025 Release

Introducing deep research

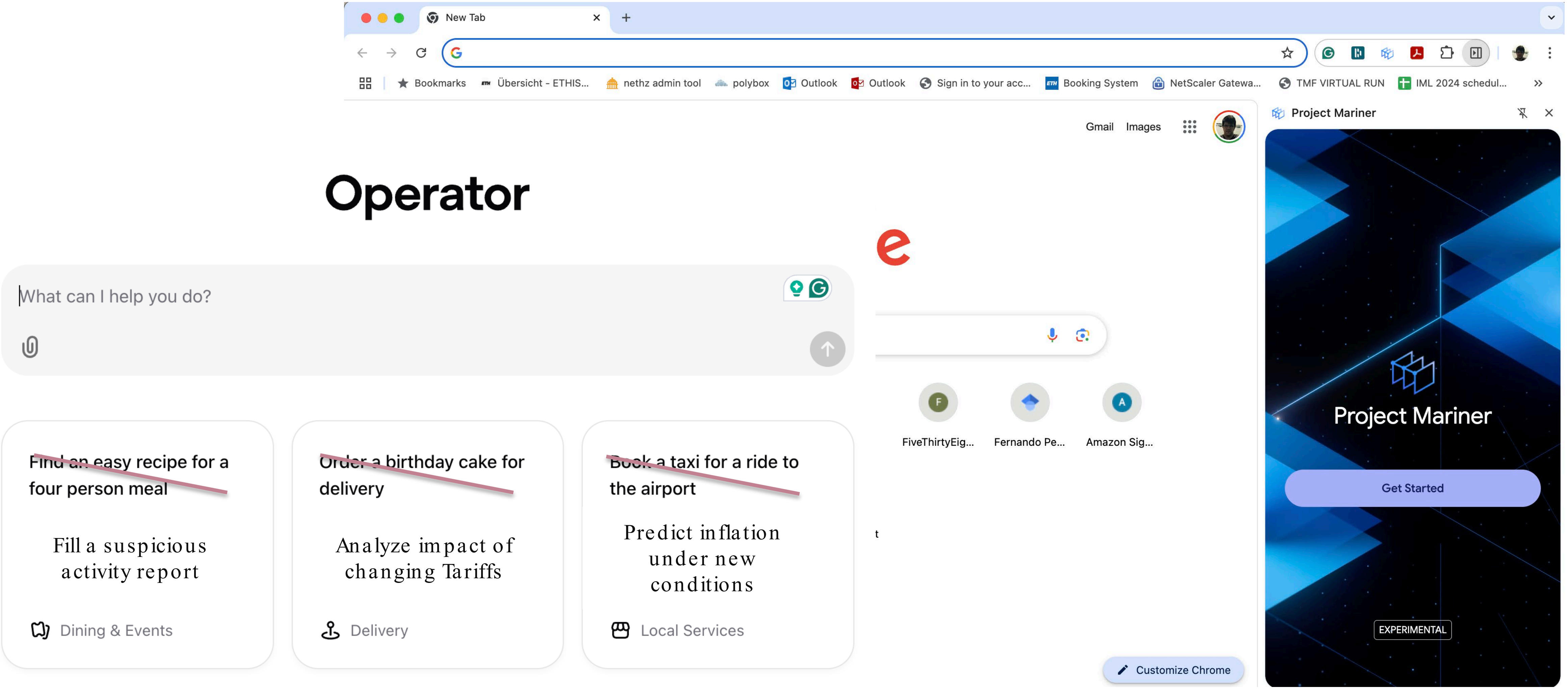
An agent that uses reasoning to synthesize large amounts of online information and complete multi-step research tasks for you. Available to Pro users today, Plus and Team next.



Hello Fernando Perez-Cruz
What can I do for you?

Video

An AI agent is a Multimodal LLM that can **control** a computer as a human



Operator

What can I help you do?

~~Find an easy recipe for a four person meal~~

Fill a suspicious activity report

Dining & Events

~~Order a birthday cake for delivery~~

Analyze impact of changing Tariffs

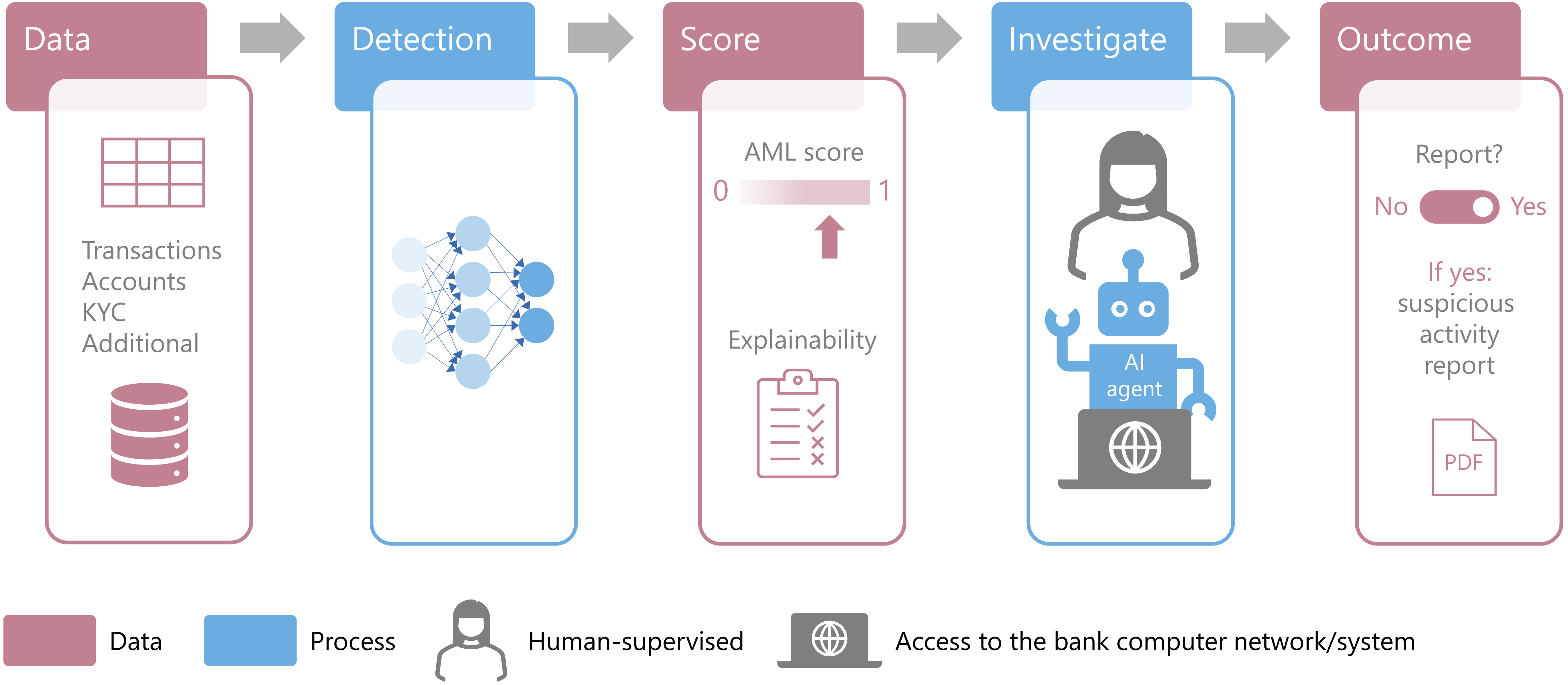
Delivery

~~Book a taxi for a ride to the airport~~

Predict inflation under new conditions

Local Services

The AI future of Anti Money Laundry compliance



Data

Transactions
Accounts
KYC
Additional

Detection

Score

AML score
0 1

Explainability

Investigate

AI agent

Outcome

Report?
No Yes

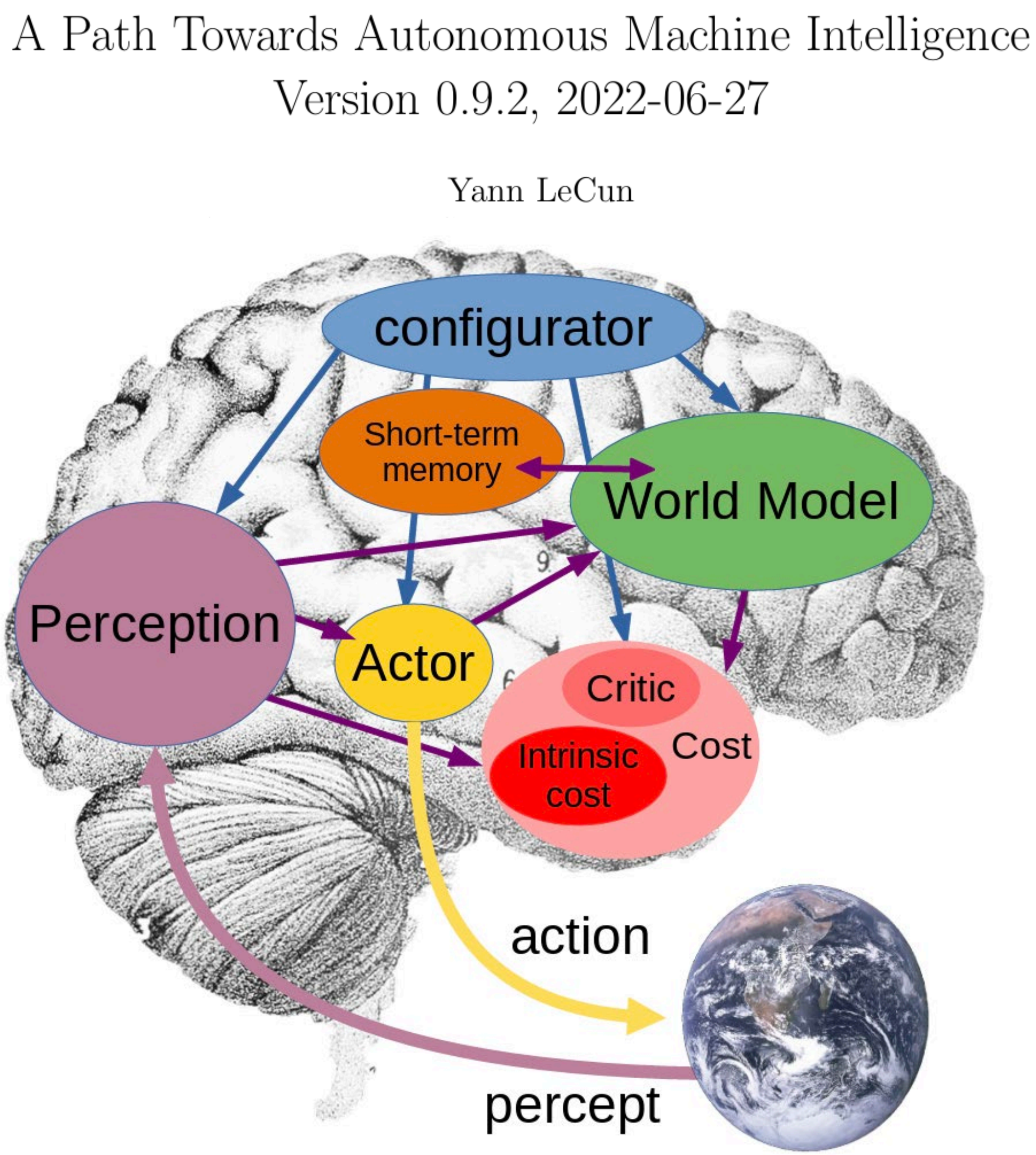
If yes: suspicious activity report

PDF

🟥 Data
🟦 Process
Human-supervised
Access to the bank computer network/system

Self-driving car experience

- AI is where self-driving cars were 10 years ago:
  - The first 90% **was** “easy”.
  - The next 5-9.9% **is** tougher.
- Two main problems:
  - There are 31.5 million seconds in a year.
  - Low probability events.
- LLMs are still a unique engineering solution.
- Modular solution:
  - Gatekeeper.
  - Many Systems 2.
  - Memory.
  - I/O modules.

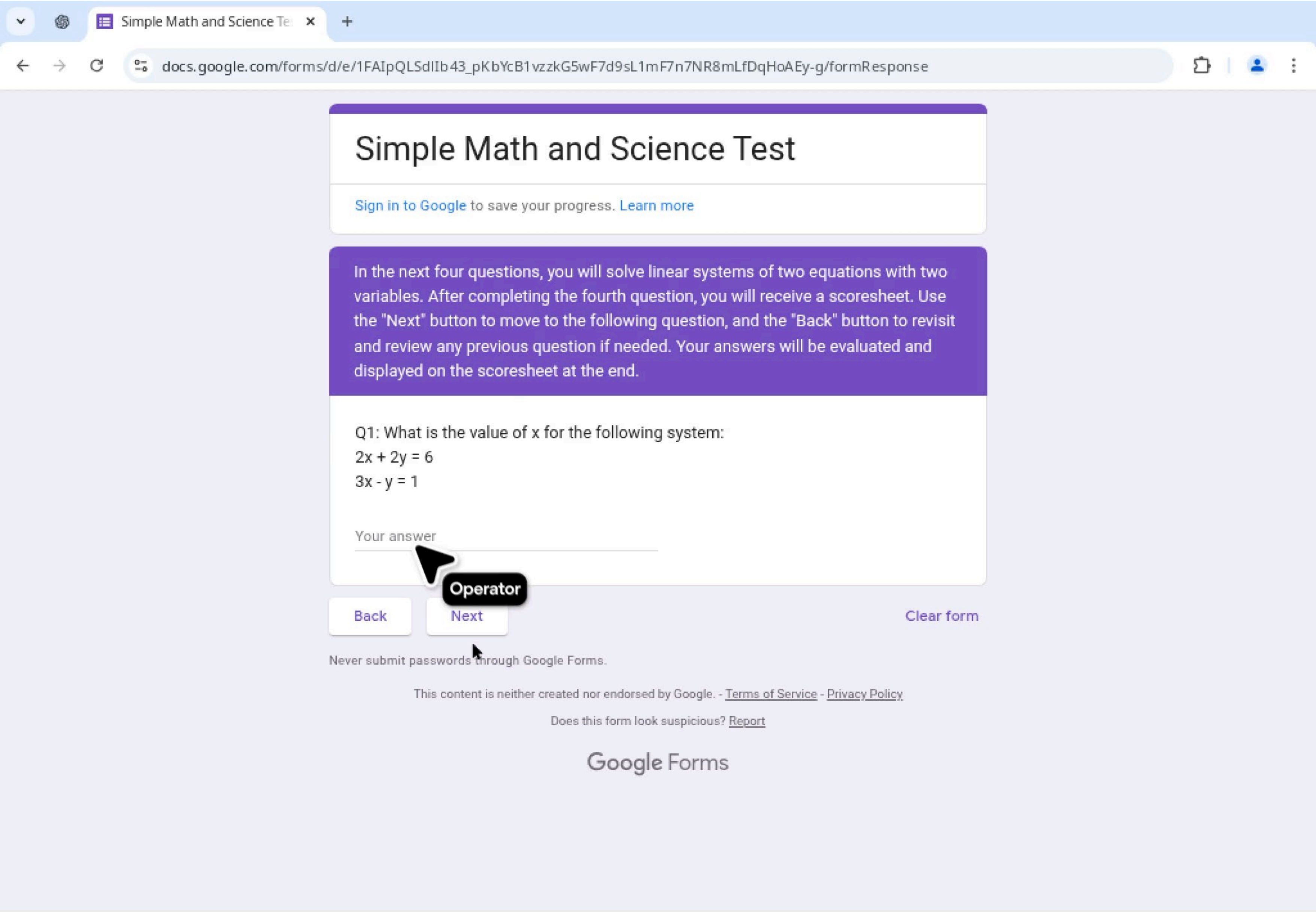


A Path Towards Autonomous Machine Intelligence
Version 0.9.2, 2022-06-27

Yann LeCun

Annex

Demo 1



Simple Math and Science Te
docs.google.com/forms/d/e/1FAIpQLSdIIb43_pKbYcB1vzzkG5wF7d9sL1mF7n7NR8mLfDqHoAEy-g/formResponse

Simple Math and Science Test

Sign in to Google to save your progress. Learn more

In the next four questions, you will solve linear systems of two equations with two variables. After completing the fourth question, you will receive a scoresheet. Use the "Next" button to move to the following question, and the "Back" button to revisit and review any previous question if needed. Your answers will be evaluated and displayed on the scoresheet at the end.

Q1: What is the value of x for the following system:

$2x + 2y = 6$

$3x - y = 1$

Your answer

Never submit passwords through Google Forms.

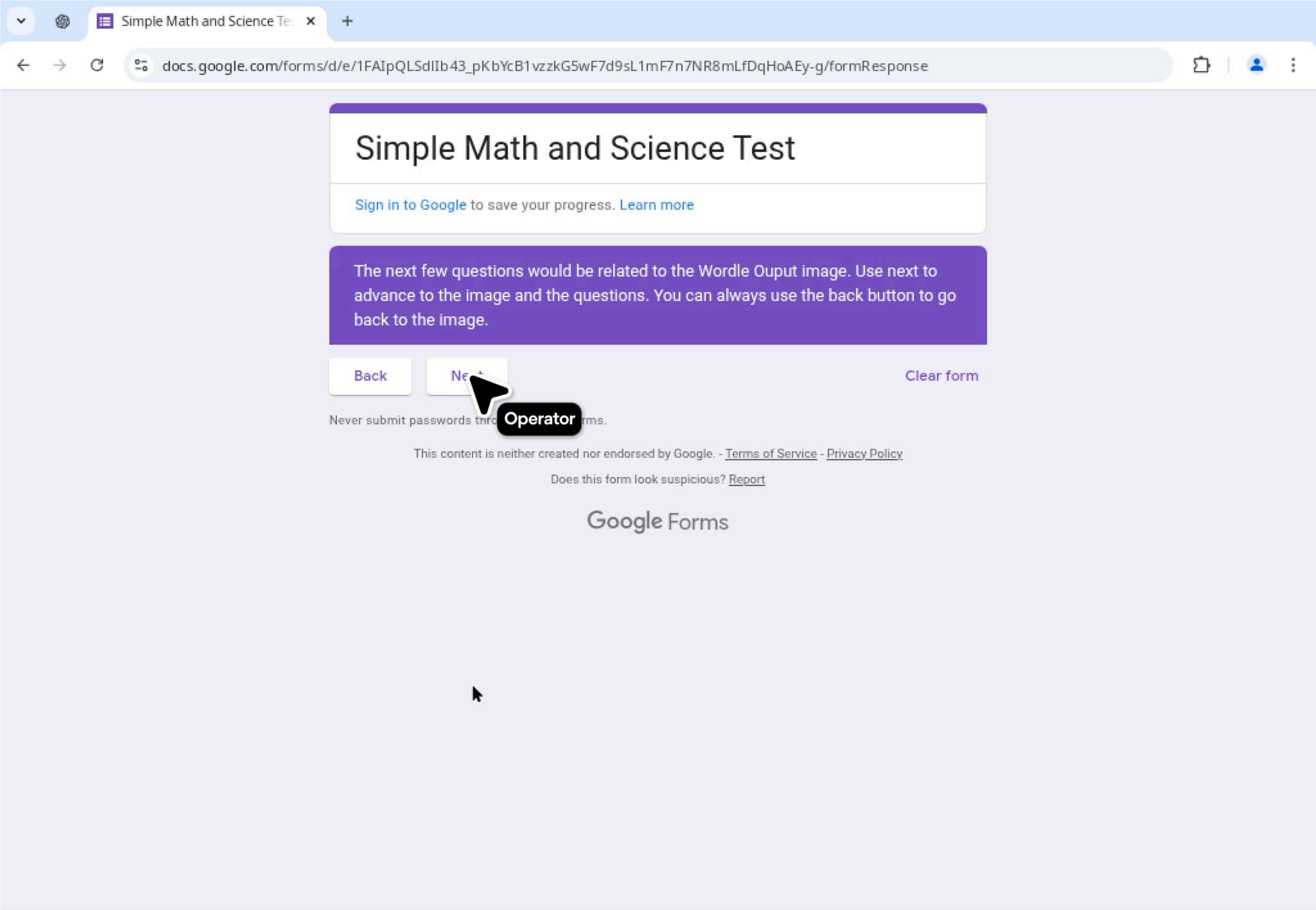
This content is neither created nor endorsed by Google. - Terms of Service - Privacy Policy

Does this form look suspicious? Report

02:06 Solving equations for value of x

operator.chatgpt.com

Demo 2



Simple Math and Science Te
docs.google.com/forms/d/e/1FAIpQLSdlIb43_pKbYcB1vzzkG5wF7d9sL1mF7n7NR8mLfDqHoAEy-g/formResponse

Simple Math and Science Test

Sign in to Google to save your progress. Learn more

The next few questions would be related to the Wordle Ouput image. Use next to advance to the image and the questions. You can always use the back button to go back to the image.

Back
Next
Operator
Clear form

Never submit passwords thro… rms.

This content is neither created nor endorsed by Google. - Terms of Service - Privacy Policy

Does this form look suspicious? Report

03:14 Proceeding to Wordle Output image

operator.chatgpt.com

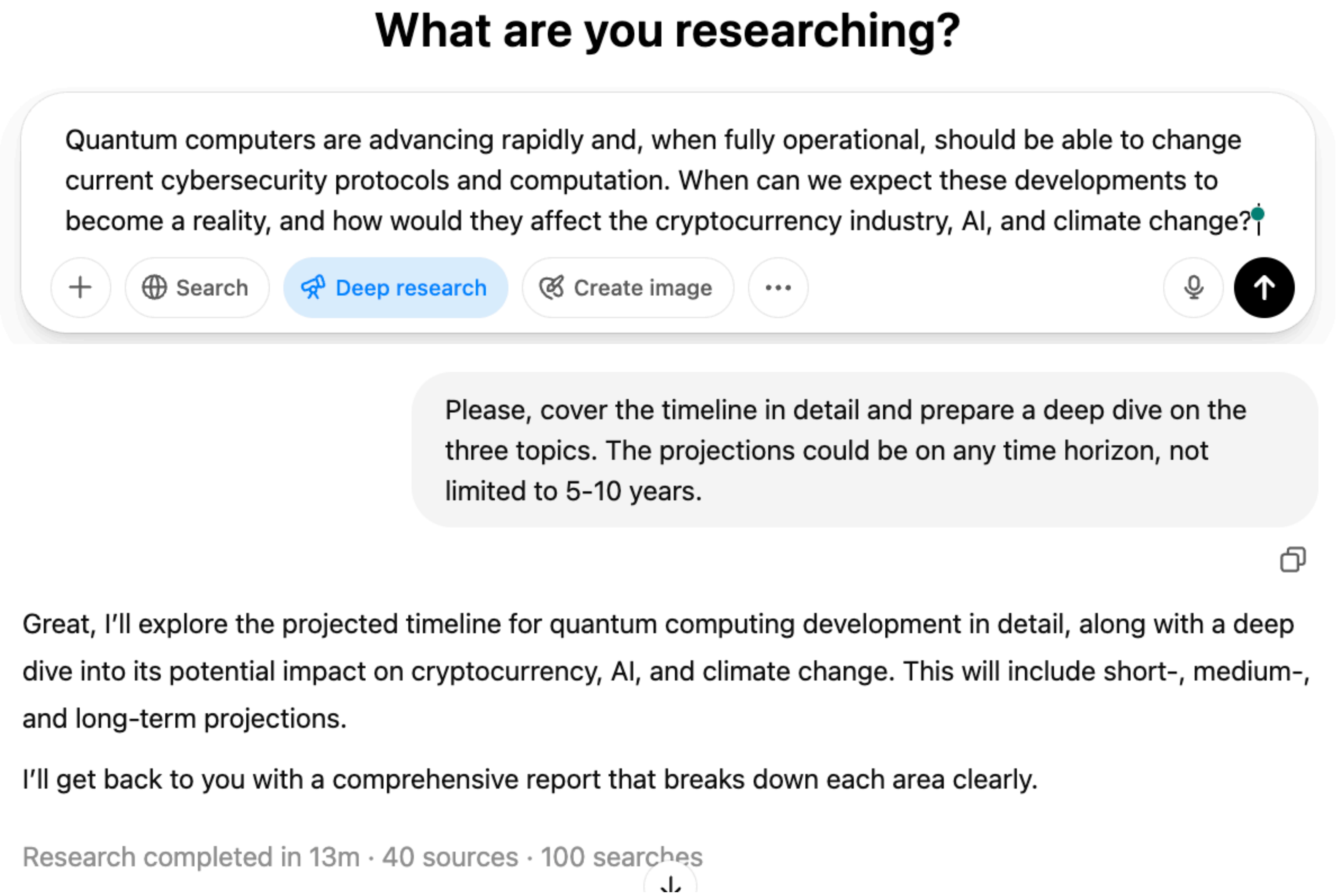
Take aways

- Narrow Research Assistant AI Agents:
    - Ready to be used today.
    - Fine-tuning or constraining to the environment would make them trustworthy.
    - They are as good as the person using it:
        - Best practices would be essential to get the most out of it.
- AI Agents: LLM with computer use:
    - Still very limited and error-prone.
    - Universal vs narrow?
    - One model vs modular?
    - Engineering narrow models:
        - Eg, Filling up a Suspicious Activity Report.

Prototypical AI Agent Definition

- McKinsey: **AI agents are the tools we use to interact with AI.**
- IBM: An AI agent refers to a system or program that is capable of autonomously performing tasks on behalf of a user by designing its workflow and utilizing available tools.
- Google: AI agents are software systems that use AI to pursue goals and complete tasks on behalf of users.
- Microsoft: An agent takes the power of generative AI a step further, because instead of just assisting you, agents can work alongside you or even on your behalf.
- BCG: AI agents are artificial intelligence that use tools to accomplish goals.
- AWS: An AI agent is a software program that can interact with its environment, collect data, and use the data to perform self-determined tasks to meet predetermined goals.
- Salesforce: An AI agent is an intelligent system that can understand and respond to customer enquiries without human intervention.

Deep Research: Only System 2



What are you researching?

Quantum computers are advancing rapidly and, when fully operational, should be able to change current cybersecurity protocols and computation. When can we expect these developments to become a reality, and how would they affect the cryptocurrency industry, AI, and climate change?

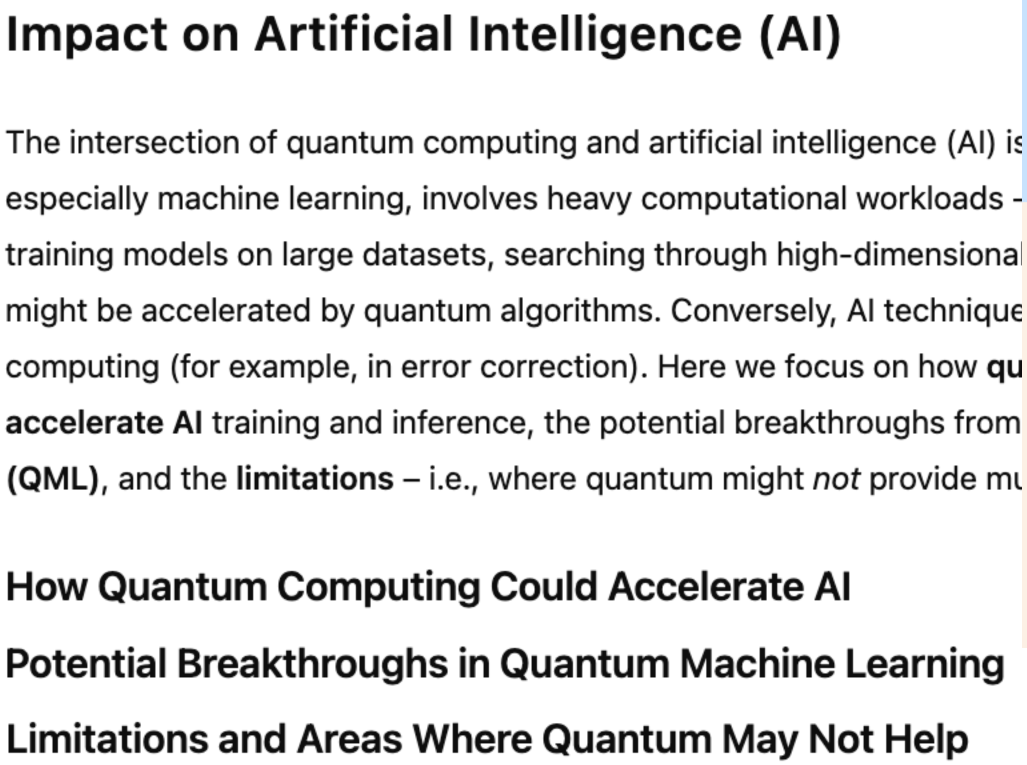
Please, cover the timeline in detail and prepare a deep dive on the three topics. The projections could be on any time horizon, not limited to 5-10 years.

Great, I'll explore the projected timeline for quantum computing development in detail, along with a deep dive into its potential impact on cryptocurrency, AI, and climate change. This will include short-, medium-, and long-term projections.

I'll get back to you with a comprehensive report that breaks down each area clearly.

Research completed in 13m · 40 sources · 100 searches

Deep Research: Only System 2



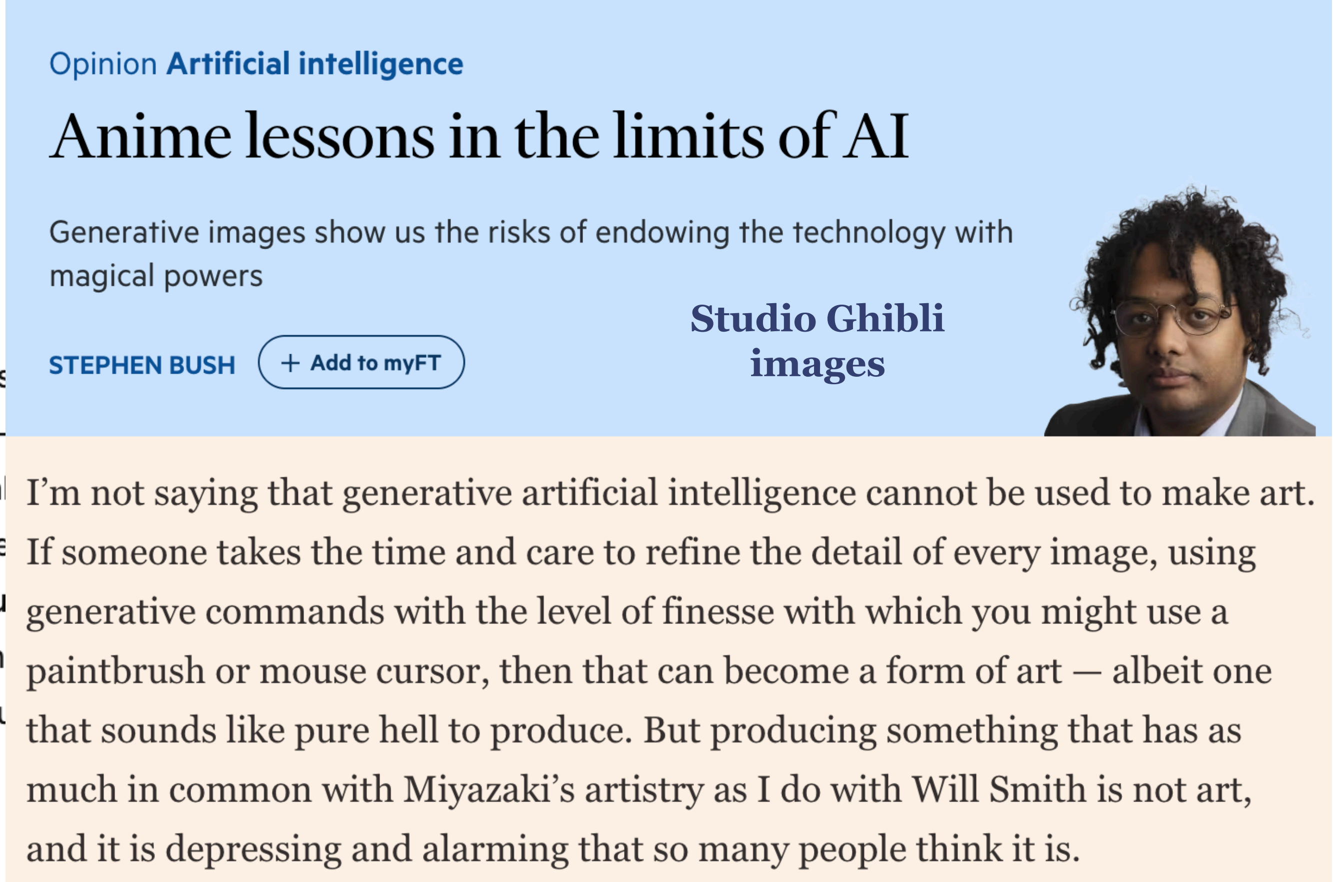
Impact on Artificial Intelligence (AI)

The intersection of quantum computing and artificial intelligence (AI) is
especially machine learning, involves heavy computational workloads -
training models on large datasets, searching through high-dimensional
might be accelerated by quantum algorithms. Conversely, AI technique
computing (for example, in error correction). Here we focus on how **qu**
accelerate AI training and inference, the potential breakthroughs from
(QML), and the **limitations** – i.e., where quantum might *not* provide mu

How Quantum Computing Could Accelerate AI

Potential Breakthroughs in Quantum Machine Learning

Limitations and Areas Where Quantum May Not Help



Opinion **Artificial intelligence**

Anime lessons in the limits of AI

Generative images show us the risks of endowing the technology with magical powers

STEPHEN BUSH

Studio Ghibli images

I'm not saying that generative artificial intelligence cannot be used to make art. If someone takes the time and care to refine the detail of every image, using generative commands with the level of finesse with which you might use a paintbrush or mouse cursor, then that can become a form of art — albeit one that sounds like pure hell to produce. But producing something that has as much in common with Miyazaki's artistry as I do with Will Smith is not art, and it is depressing and alarming that so many people think it is.

Replace 'art' by 'research'

- They are as good as the person using it.

Prototypical AI Agent Definition

- McKinsey: AI agents are the tools we use to interact with AI.
- IBM: An AI agent refers to a system or program that is capable of autonomously performing tasks on behalf of a user by designing its workflow and utilizing available tools.
- Google: AI agents are software systems that use AI to pursue goals and complete tasks on behalf of users.
- Microsoft: An agent takes the power of generative AI a step further, because instead of just assisting you, agents can work alongside you or even on your behalf.
- BCG: AI agents are artificial intelligence that use tools to accomplish goals.
- AWS: An AI agent is a software program that can interact with its environment, collect data, and use the data to perform self-determined tasks to meet predetermined goals.
- Salesforce: An AI agent is an intelligent system that can understand and respond to customer enquiries without human intervention.

These definitions fall flat. They're outdated and **uninspired automation** at best. the bottleneck is not AI: it's **poor data accessibility** and current inefficient processes.