

2008: The Turning Point in Western History

From Economic Growth and Optimism
to Managed Decline and Precarity

Introduction

- Post-WWII era: Economic growth, welfare state expansion, and optimism.
- 1990s-2000s: Technological optimism and globalization.
- 2008: A pivotal year marking a shift toward systemic crises and uncertainty.

The 2008 Financial Crisis—End of Economic Certainty

- Origins: Deregulation, housing bubble, reckless lending.
- Consequences: Bailouts for institutions, austerity for citizens.
- Distrust in institutions set the stage for populist movements.

Digital Disruption—iPhone, Social Media, and Society

- 2007: iPhone launched, symbolizing a new era in connectivity.
- Social media transformed social behavior, politics, and the economy.
- Attention economy led to polarization and mental health issues.

Crises of the 2010s—Local Wars and Pandemic

- Geopolitical instability: Syria, Ukraine, weakening U.S. hegemony.
- COVID-19 accelerated digital transformation and wealth consolidation.
- Erosion of trust in institutions and societal fragmentation.

The Rise of Generative AI—A Crisis of Meaning

- Promise: Innovation and efficiency.
- Threat: Displacement of intellectual tasks and crisis in education.
- Universities face existential challenges in remaining relevant.

UBI as a Response to Technological Displacement

- AI threatens to displace millions of jobs.
- UBI: A potential solution but raises questions of purpose and dignity.
- What does a meaningful life look like without traditional work?

Broader Trend—From Progress to Precarity

- 20th century: Belief in progress replaced by precarity.
- Climate change, political instability, and technological disruption dominate.
- Widening gap between elites and non-elites.

Conclusion—Personal Responsibility and Transformation

- Managed decline prioritizes stability over bold change, leading to stagnation.
- In contrast, personal transformation offers a path to renewal through higher principles.
- Historical examples: Early Christians, monastic movements, Renaissance inspiring cultural change.
- Is it possible to reclaim a sense of progress?

Discussion Questions

- Do you agree that 2008 was a turning point in history? Why or why not?
- How has social media transformed society, and can it be used constructively?
- What are the implications of UBI on human identity and purpose?
- How can universities remain relevant in an age of AI?
- Is technological progress inherently good, or should we focus on its risks?



System 0-1-2-3

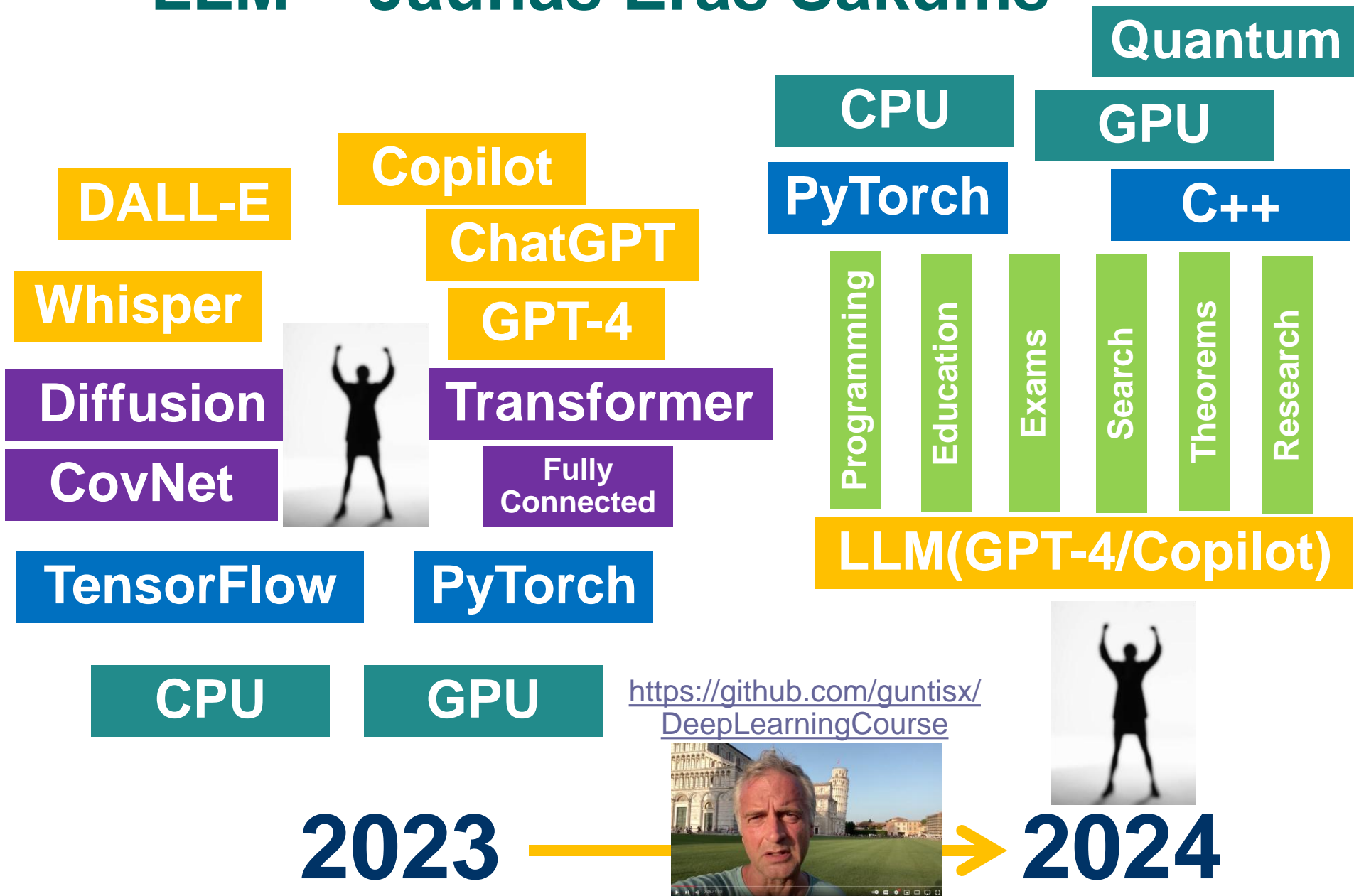
Evolution Beyond Biological



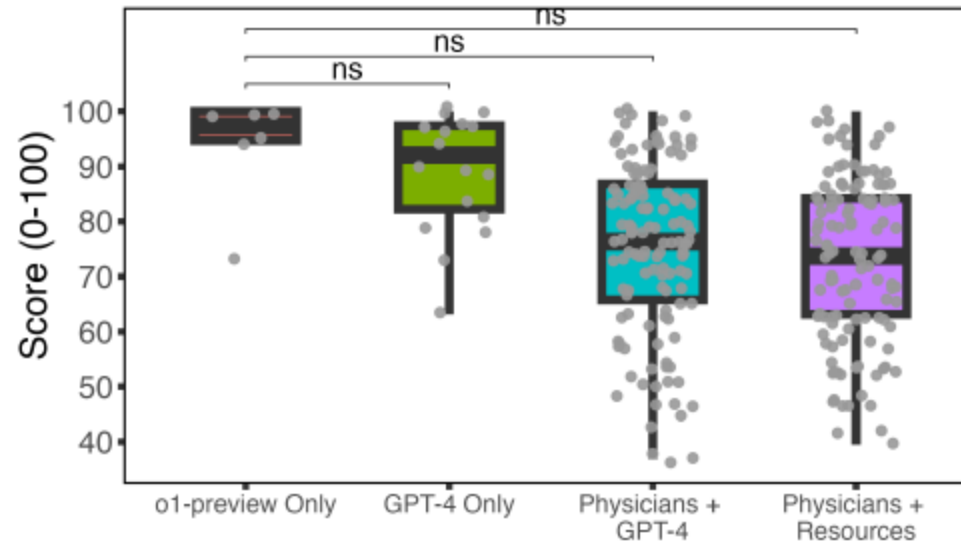
Guntis Bārzdīņš

12/12/2024

LLM = Jaunas Ēras Sākums



B. Landmark Diagnostic Cases: o1-preview Diagnostic Reasoning Scores Compared to GPT-4 and Physicians



How o1, o3 work??

2025

What is to Be Done?



Engineering for
Solving Real-world
Problems
(work hours+LLM ↑)

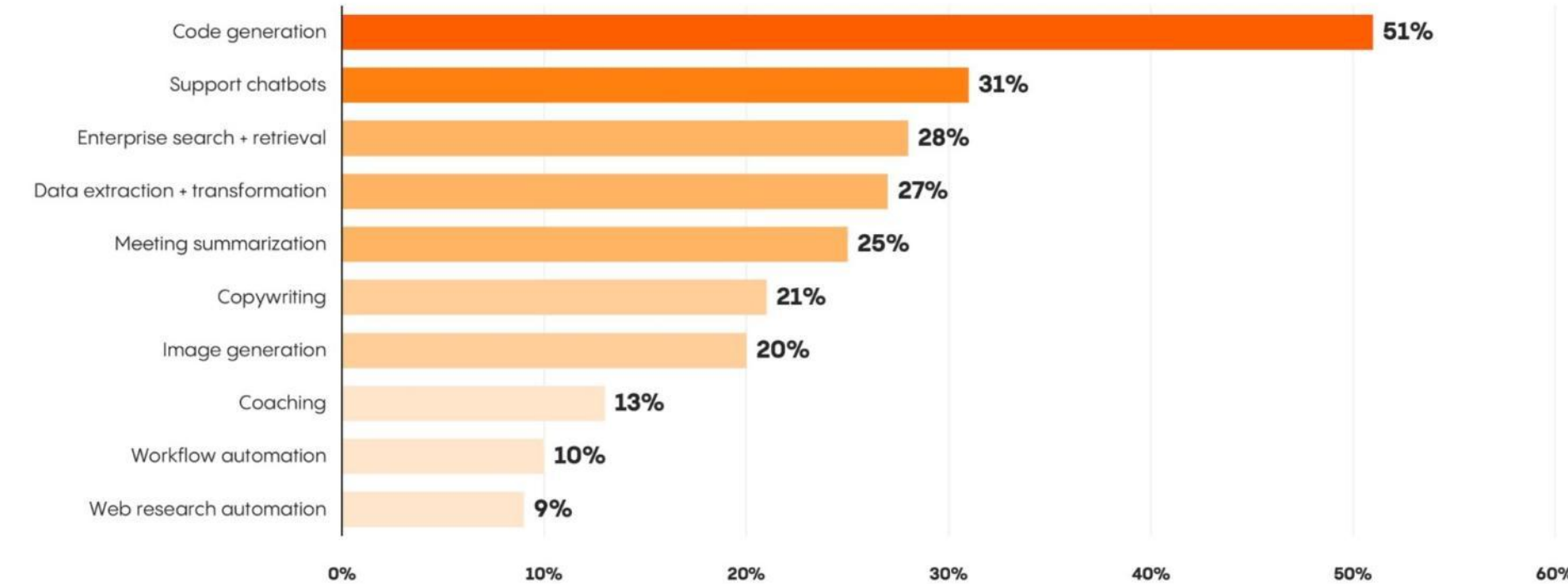
Non-engineering
Office-work (salaries ↓)
(to be largely automated by Engineers+LLM)

LLM Human-in-the-Loop nepieciešamība

Memory Facet	LLM Performance	Improving It Without Harm
Semantic Memory	Excellent	External knowledge updates for dynamic info.
Episodic Memory	Weak	Persistent memory layers for conversation.
Procedural Memory	Decent	Reinforcement learning for reliability.
Working Memory	Limited by context window	Larger windows or memory-augmented methods.

– Schema/Graph
– ARC-AGI Prize
(Kolmogorov Complexity & OccamRazor)

Dominant Generative AI Use Cases

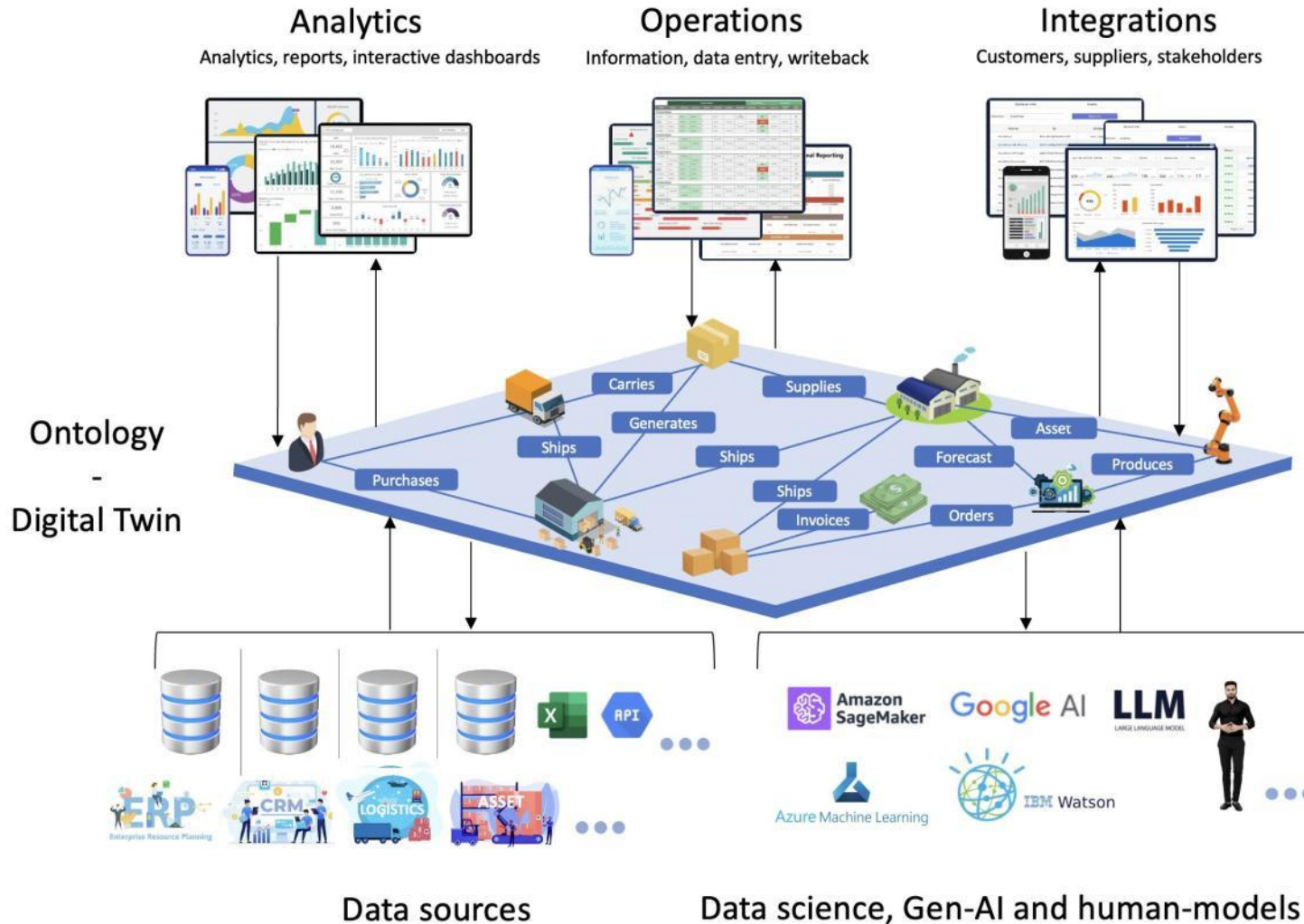


Palantir Ontology (Schema/Graph)

Defense companies by market cap

COMPANY	MARKET CAP	EMPLOYEES	FOUNDED
Palantir	\$173,827,400,000	3,892	2003
RTX	\$157,046,600,000	185,000	1934
LOCKHEED MARTIN	\$121,606,200,000	122,000	1912
BOEING	\$115,012,000,000	171,000	1934
GENERAL DYNAMICS	\$74,736,470,000	111,600	1952
HOUGHTON MIFFLIN HARCOURT	\$69,008,600,000	101,000	1939
LIHARRIS	\$44,503,790,000	50,000	1926
ARMSTRONG WORLD INDUSTRIES	\$7,354,361,000	44,000	1886

Note: Data as of December 2024 from public sources



StructuredData

UnstructuredData

"There are graphs in our data"



Ontology Core →



LAYER / 01

Semantic

Dynamic Objects & Links

Multi-Modal Properties

Ontology Primitives



LAYER / 02

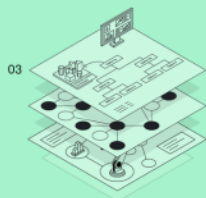
Kinetic

AI-Driven Actions & Functions

Process Mining & Automation

Action Orchestration

Real-Time Monitoring



LAYER / 03

Dynamic

AI-Powered Decisions

Multi-Step Simulations

Decision Capture & Learning

<https://www.ft.com/content/6cfdfe2b-6872-4963-bde8-dc6c43be5093>



Palantir and Anduril join forces with tech groups to bid for Pentagon contracts

Consortium likely to include Elon Musk's SpaceX in move to grab a bigger slice of \$850bn US defence budget



The Anduril Great autonomous drive © Kyle Orloff/Bloomberg

Tabby Kinder and George Hammond in San Francisco. DECEMBER 22 2024

Unlock the Editor's Digest for free

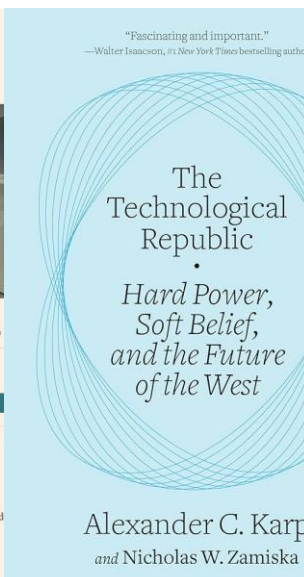
Roula Khalaf, Editor of the FT, selects her favourite stories in this weekly newsletter.

Enter your email address

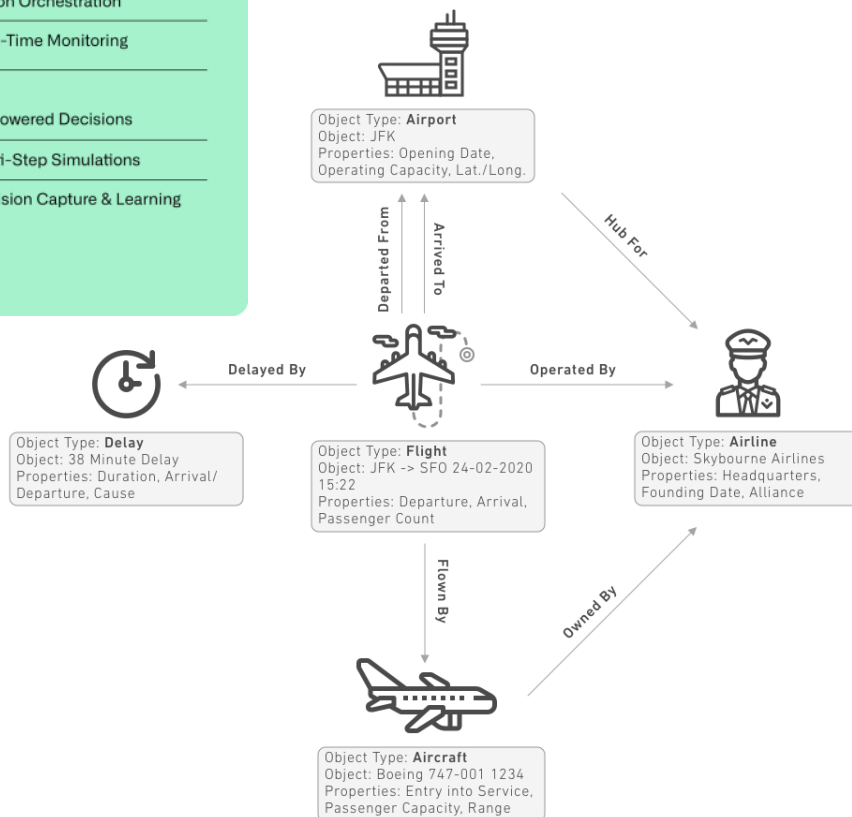
Sign up

Palantir and Anduril, two of the largest US defence technology companies, are in talks with about a dozen competitors to form a consortium that will jointly bid for US government work in an effort to disrupt the country's oligopoly of "prime" contractors.

The consortium is planning to announce as early as January that it has reached agreements with a number of tech groups. Companies in talks to join include Elon Musk's SpaceX, ChatGPT maker OpenAI, autonomous-ship builder Sarcos and artificial intelligence data group Scale AI, according to several people with knowledge of the matter.



AI Arms-Race



A simple ontology of 5 object types displays some of the properties and relationships within airline industry datasets.

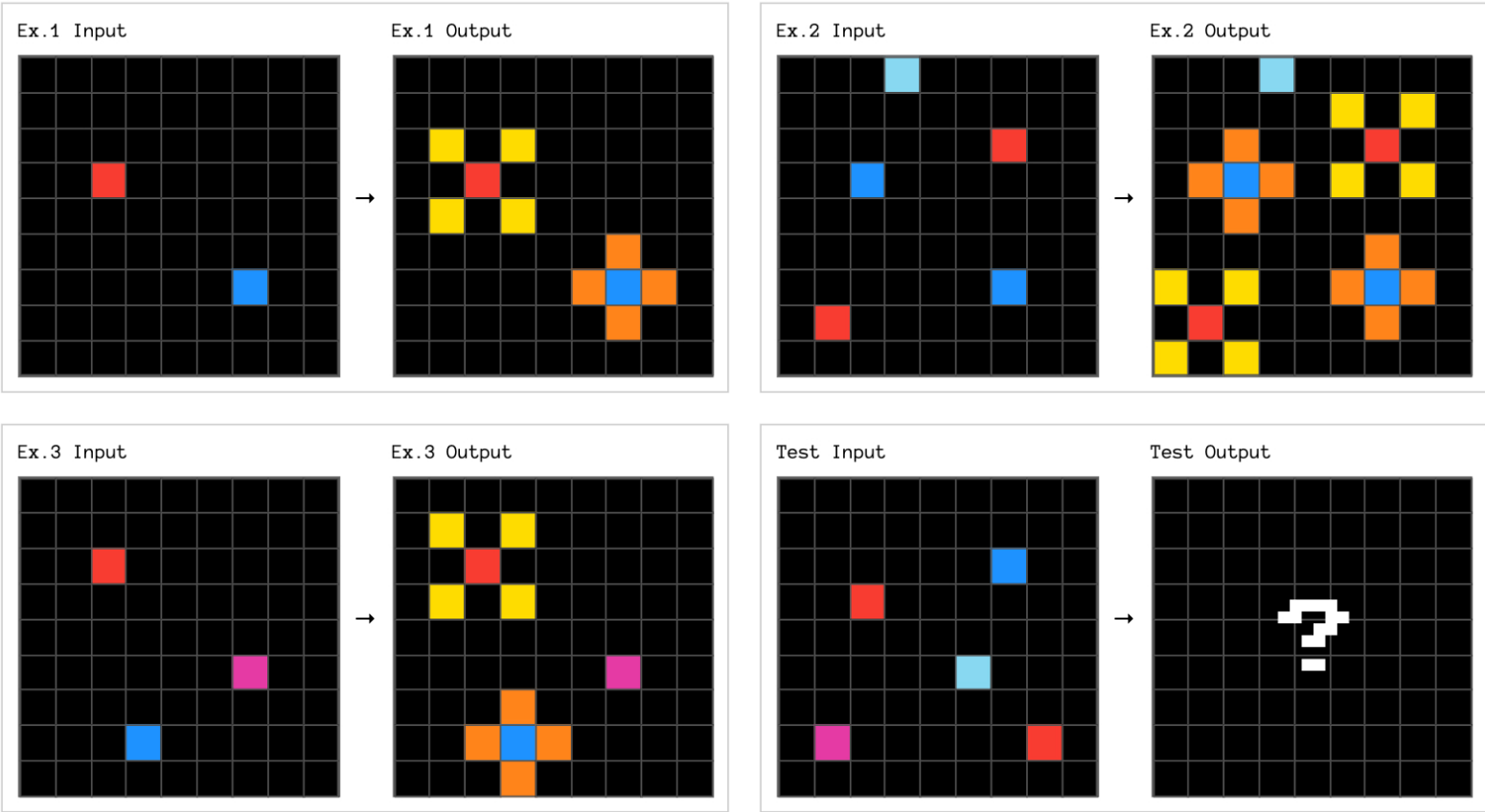
ARC-AGI Benchmark

Noam Brown @polynoamial

I think people are overindexing on the @OpenAI o3 ARC-AGI results. There's a long history in AI of people holding up a benchmark as requiring superintelligence, the benchmark being beaten, and people being underwhelmed with the model that beat it.

To be clear, @fchollet and @mikeknoop were always very clear that beating ARC-AGI wouldn't imply AGI or superintelligence, but it seems some people assumed that anyway.

ARC-AGI Prize 2024 (Human 97%-100%)



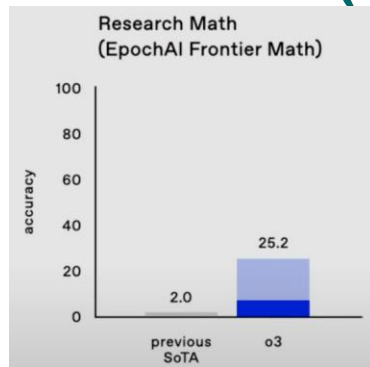
Place	Name	Score (Private evaluation set)
1st	the ARCHitects	53.5%
2nd	Guillermo Barbadillo	40%
3rd	alijs ← Agnis Luiķis (LV)	40%
4th	William Wu	37%
5th	PoohAI	37%

Table 1: ARC Prize 2024 Winners.

Name	Semi-private eval	Public eval
Jeremy Berman	53.6%	58.5%
Akyürek et al.	47.5%	62.8%
Ryan Greenblatt	43%	42%
OpenAI o1-preview (pass@1)	18%	21%
Anthropic Claude 3.5 Sonnet (pass@1)	14%	21%
OpenAI GPT-4o (pass@1)	5%	9%
Google Gemini 1.5 (pass@1)	4.5%	8%

Table 2: ARC-AGI-Pub leaderboard.

o3 solves ARC-AGI & FrontierMath (Dec20, 2024)



Nat McAleese @__nmca__

o1 was the first large reasoning model - as we outlined in the original "Learning to Reason" blog, it's **"just" an LLM trained with RL**. o3 is powered by further scaling up RL beyond o1

Jason Wei @_jasonwei

o3 is very performant. More importantly, progress from o1 to o3 was only three months, which shows how fast progress will be in the new paradigm of RL on chain of thought to scale inference compute. Way faster than pretraining paradigm of new model every 1-2 years

Noam Brown @polynoamial

This is the result I'm most excited about. Even if LLMs are dumb in some ways, saturating evals like @EpochAIResearch's **Frontier Math** would suggest AI is surpassing top human intelligence in certain domains. When that happens we may see a **broad acceleration in scientific research**

Practice working with AI. **Human+AI will be superior to human or AI alone for the foreseeable future**. Those who can work most effectively with AI will be the most highly valued.

We view **multi-agent as a path to even better AI reasoning**.

Sam Altman @sama

i expect this trend to continue, but also that the ability to get marginally more performance **for exponentially more money** will be really strange.

(Craig Mundie says the nuclear fusion company backed by Sam Altman will surprise the world by showing fusion electrical generation next year, becoming the basis for a "radical transformation of the energy system" due to safe, cheap power)

For now, we can only speculate about the exact specifics of how o3 works. But o3's core mechanism appears to be natural language program search and execution within token space - at test time, the model searches over the space of possible Chains of Thought (CoTs) describing the steps required to solve the task, in a fashion perhaps not too dissimilar to AlphaZero-style Monte-Carlo tree search. In the case of o3, the search is presumably guided by some kind of evaluator model. To note, Demis Hassabis hinted back in a June 2023 interview that DeepMind had been researching this very idea - this line of work has been a long time coming.

So while single-generation LLMs struggle with novelty, o3 overcomes this by generating and executing its own programs, where the program itself (the CoT) becomes the artifact of knowledge recombination. Although this is not the only viable approach to test-time knowledge recombination (you could also do test-time training, or search in latent space), it represents the current state-of-the-art as per these new ARC-AGI numbers.

Effectively, o3 represents a form of *deep learning-guided program search*. The model does test-time search over a space of "programs" (in this case, natural language programs - the space of CoTs that describe the steps to solve the task at hand), guided by a deep learning prior (the base LLM). The reason why solving a single ARC-AGI task can end up taking up tens of millions of tokens and cost thousands of dollars is because this search process has to explore an enormous number of paths through program space - including backtracking.

ARC-AGI Prompt

Mike Knoop @mikeknoop

Raising visibility on this note we added to address ARC "tuned" confusion:

> OpenAI shared they trained the o3 we tested on 75% of the Public Training set.

This is the explicit purpose of the training set. It is designed to expose a system to the core knowledge priors needed to beat the much harder eval set.

The idea is each training task shows you an isolated single prior. And the eval set requires you to recombine and abstract from those priors on the fly. Broadly, the eval tasks require utilizing 3-5 priors.

The eval sets are extremely resistant to just "memorizing" the training set. This is why o3 is impressive.

François Chollet @fchollet

Reminder: for any ARC task, there are *many* rules that could explain the task -- your job isn't just finding "a rule", it's finding "the rule", the one that reflects human consensus. After all, we know empirically that all ARC tasks are solvable by humans.

Typically that's the simplest rule. In some rare cases there could be two rules with arguably the same level of complexity. In such cases, you can leverage the fact that you have *two* attempts to solve each input grid: just try both solutions! If one of the two is correct, you've solved the task.

Hieu Pham @hyhieu226

So o3 trained on the training set of ARC-AGI. How about FrontierMath? I don't think FrontierMath has a similar "training set." Wouldn't o3's good performance on FrontierMath mean the model is good?

Find the common rule that maps an input grid to an output grid, given the examples below.

Example 1:

```
Input:
0 0 0 5 0
0 5 0 0 0
0 0 0 0 0
0 5 0 0 0
0 0 0 0 0
Output:
1 0 0 0 0 5 5 0 0
0 1 0 0 0 0 5 5 0 0
0 0 5 5 0 0 0 0 1 0
0 0 5 5 0 0 0 0 0 1
1 0 0 0 1 0 0 0 0 0
0 1 0 0 0 1 0 0 0 0
0 0 5 5 0 0 1 0 0 0
0 0 5 5 0 0 0 1 0 0
0 0 0 0 1 0 0 0 1 0
0 0 0 0 1 0 0 0 1
```

Example 2:

```
Input:
2 0
0 0
Output:
2 2 0 0
2 2 0 0
0 0 1 0
0 0 0 1
```

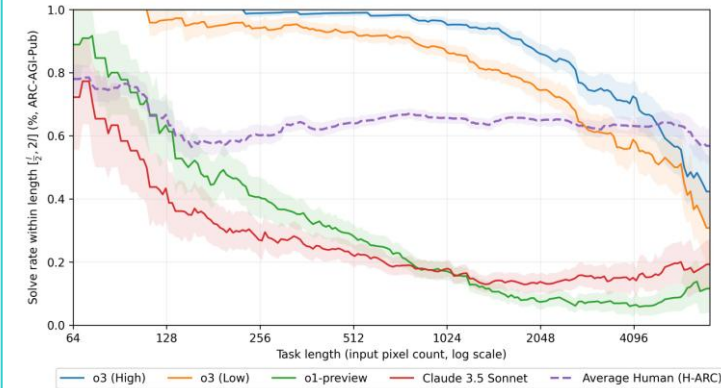
Example 3:

```
Input:
0 0 0 0 0 3
0 0 0 0 0 0
0 3 0 0 0 0
0 0 0 0 0 0
0 0 0 0 0 0
0 0 0 0 0 0
0 0 0 0 0 0
Output:
0 0 0 0 0 0 0 0 0 0 3 3
0 0 0 0 0 0 0 0 0 0 3 3
1 0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0 0 0 0 0
0 0 3 3 0 0 0 0 0 0 0 0
0 0 3 3 0 0 0 0 0 0 0 0
0 0 0 0 1 0 0 0 0 0 0 0
0 0 0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 0 1 0 0 0 0 0
0 0 0 0 0 0 0 1 0 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 0 0 0 0 1 0 0
```

Below is a test input grid. Predict the corresponding output grid by applying the rule you found. Your final answer should just be the text output grid itself.

```
Input:
0 4 0
0 0 0
4 0 0
```

Long memory problem



Task solve rates (ARC-AGI-Pub) by various models. Each point represents a sliding window of tasks with length in the range $[l/2, l*2]$, with bootstrapped 75% CI shaded.

What we see is quite striking: **LLMs are really good at solving the smallest ARC tasks**, even prior to o3. Each subsequent model gets better and better at solving larger tasks, but everything before o3 (and especially pre-o1) **struggles with solving tasks with >512 input pixels**. With o3, we raise that threshold to perhaps 4096 input pixels.

<https://x.com/goodside/status/1871777037096190038>

Claude 3.5 Sonnet



Speaking like a 6-year-old

Ooh, I see a fun pattern! Let me tell you what I notice:

1. In each picture, there are some blue dots and red squares at the start.
2. In the final pictures, something magical happens! The blue dots grow and make a long blue path that connects all the blue dots together, like making a big blue bridge!
3. When the blue path touches the red squares,
4. The red squares that are far away from the blue path stay red, like they're playing "don't touch me!"

So the rule is like this: "The blue dots want to be friends and hold hands, so they make a blue path to connect to each other! When they need to walk through red squares to reach their friends, those red squares become blue too!"

I bet for any new picture, I need to:

1. Find all the blue dots
2. Draw a path to connect them all together
3. Turn any red squares that are in the path to blue
4. Keep other red squares red

It's like a fun connect-the-dots game, but with colors changing too! Ca ↓ cry solving one?

Reply to Claude



SYSTEM 1

Intuition & instinct

95%

Unconscious
Fast
Associative
Automatic pilot

SYSTEM 2

Rational thinking

5%

Takes effort
Slow
Logical
Lazy
Indecisive

Source: Daniel Kahneman

SYSTEM 3

Algorithms, DB,
Structured data

THE NEW YORK TIMES BESTSELLER

THINKING,
FAST AND SLOW



DANIEL
KAHNEMAN

WINNER OF THE NOBEL PRIZE IN ECONOMICS

"[A] masterpiece . . . This is one of the greatest and most engaging collections of insights into the human mind I have read." —WILLIAM EASTERLY, *Financial Times*

- François Chollet @fchollet Dec 8, 2024

- It's highly plausible that fuzzy pattern matching, when iterated sufficiently many times, can asymptotically turn into reasoning (it's even possible that humans do it basically in this way). But that doesn't mean it's the optimal way to do reasoning.
- Re: "optimal way to do reasoning". The idea here is that if you have an "iterated system 1" type reasoning system that takes ~\$1k and O(hour) in compute to find a correct program to solve a simple task, vs. a discrete program search system that does the same in 0.1s on your laptop, then the more optimal (discrete) option eventually wins out as scale increases. Compute scaling benefits both sides.
 - -- skat ARC-AGI Prize paper: 6.lpp apakšā: 3.vieta no LV Agnis Liukis (Artūra Sproģa klasesbiedrs!):
- Human **Core Knowledge** has been investigated by Spelke et al. (22). These knowledge priors are acquired by children very early (typically before age four) and are universally shared by all humans. The ARC-AGI-1 public training tasks are designed to expose test-takers to all the Core Knowledge priors needed to solve ARC-AGI tasks.
 - 22. Elizabeth S. Spelke and Katherine D. Kinzler. Core knowledge. *Developmental science*, pages 89–96, 2007.

<https://arxiv.org/pdf/2412.04604>

Reasoning Evolution Beyond Biological

System	Stage	Key Features	Examples / State
System -5	Big Bang and Fundamental Physics	Origin of space, time, matter, and energy; formation of fundamental particles and physical laws.	Big Bang, quarks, formation of atoms.
System -4	Stellar and Galactic Evolution	Formation of stars, galaxies, and heavier elements through nuclear fusion; building blocks of planets.	Star formation, supernovae, stardust.
System -3	Planetary and Chemical Evolution	Formation of planets and complex molecules; preconditions for life (e.g., water, organic compounds).	Early Earth, organic chemistry.
System -2	Prebiotic Chemistry	Self-organization, autocatalytic networks, protocells; emergence of proto-life systems.	RNA World, lipid bilayers, hydrothermal vents.
System -1	Prokaryotic Life	Simple, single-celled organisms; reflexive biochemical responses, horizontal gene transfer.	Bacteria, archaea, microbial mats.
System 0	Eukaryotic and Multicellular Life	Compartmentalized cells, sexual reproduction, multicellularity; foundation for sensory-motor systems. Soft Robotics.	Eukaryotes, plants, animals.
System 1	Reflexive and Pattern-Based Cognition	Fast, heuristic responses; indirect reasoning via sensory systems (e.g., vision, instincts).	Early animals, modern instincts.
System 2	Deliberate and Iterative Thinking	Abstract reasoning, planning, problem-solving; slower but more flexible decision-making.	Humans, advanced animals (e.g., primates).
System 3	Discrete Symbolic Reasoning	Explicit, rule-based logic; symbolic AI, mathematics, formal computation.	Humans, machines, symbolic AI.
System 4	Quantum Reasoning and Multiverse Thinking	Parallel exploration of possibilities using quantum systems; exponential problem-solving capability.	Quantum computing, probabilistic systems.
System 5	Post-Quantum and Transcendent Systems	Beyond quantum mechanics; integration of physics, computation, and consciousness across space-time.	Speculative future intelligence, FTL systems.

Policy RL
DNN
LLM
CPU
Model RL = SLAM