Variables worth consideration in Developing an AI Strategy:
Evaluating Intelligence, Future Model Capabilities, and Economic Implications

Plug in the numbers for current computing speeds, the current doubling time, and an estimate for the raw processing power of the human brain, and the numbers match in: 2021.
— Eliezer Yudkowsky (1996)

Primary Inspirations:

- Eliezer Yudkowsky, Microeconomics of an Intelligence Explosion (2013)
- Ajeya Cotra, Forecasting TAI with Biological Anchors (2020)
- Tom Davidson, Could Advanced AI Drive Explosive Economic Growth? (2021)

Other Notable Inspirations:

Carl Shulman, Geoff Hinton, Ilya Sustkever, Holden Karnofsky, Noam Brown, Yoshua Bengio, Chris Olah, Neel Nanda, Demis Hassibis, Dario Amodie, Connor Leahy, Paul Cristiano, Jan Leike, Andrej Karpathy, Joscha Bach, Douglas Hofstadter, Yann Lecun, Emad Mostaque

When will the first weakly general AI system be devised, tested, and publicly announced?

Primary Condition: *Model must be awarded the Loebner Silver Prize*

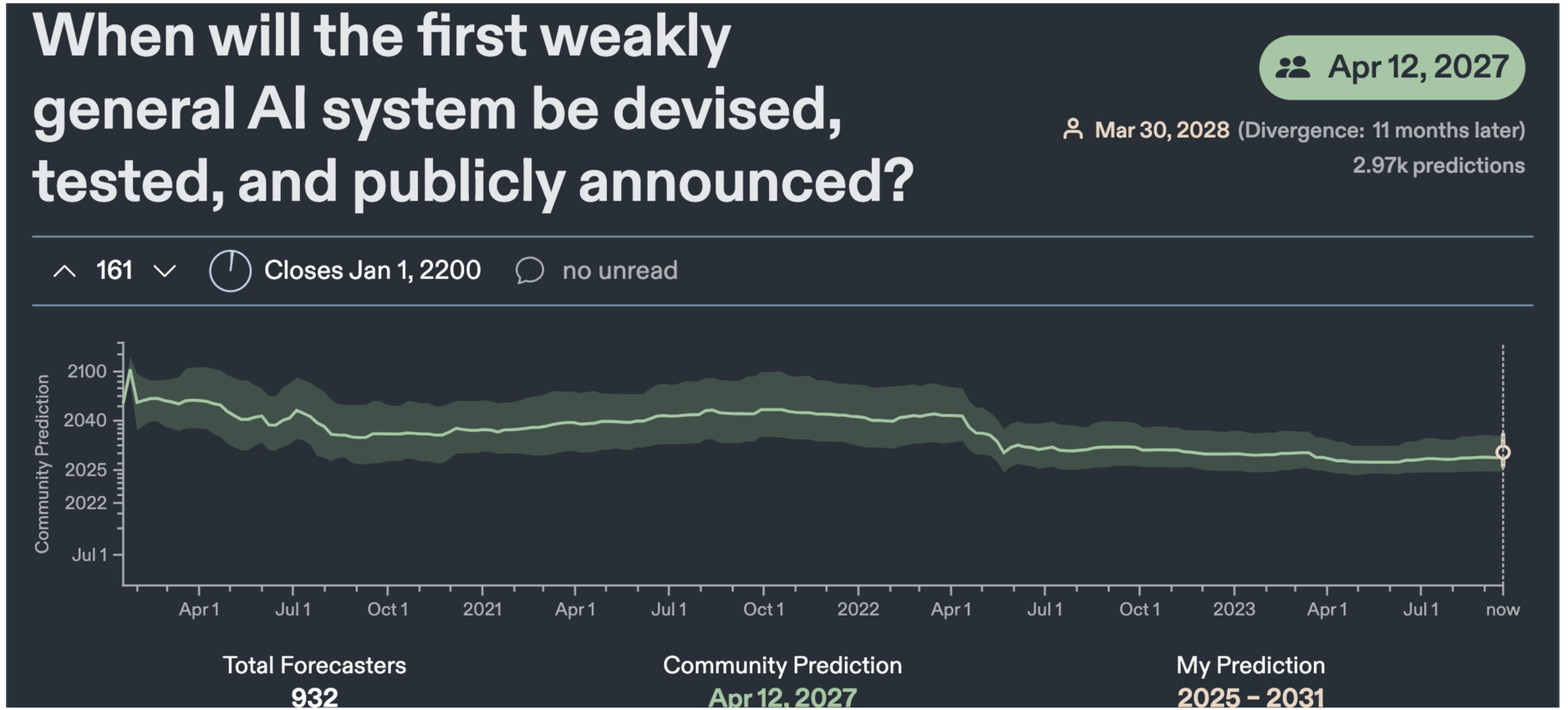
Text-Only Test Each Round is 30 minutes — deliberately rigged to favor humans

If there is any text-based task that a typical human can reliably do, that AIs cannot reliably do, the judges can and will use that task to distinguish the AI from the human, and the AI will fail

If human evaluators fail to predict human/model distinction on any 30-minute cognitive task
⟶ then, AI has human-level competence on every 30-minute task

It would be a universal personal assistant

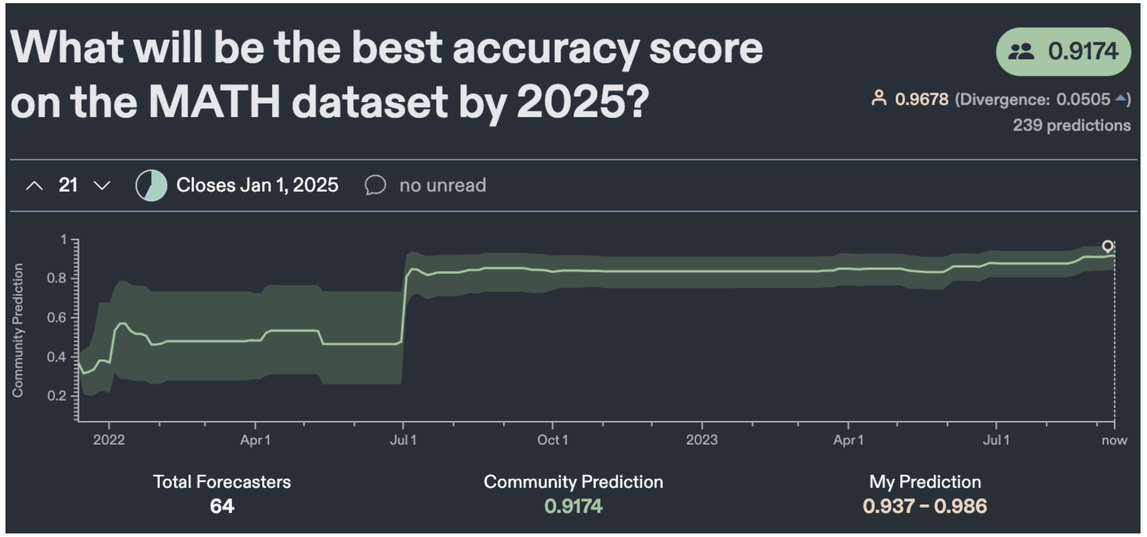
When will the first weakly general AI system be devised, tested, and publicly announced?



Indistinguishability + Better, Faster, Cheaper

Timeline to Superhuman Mathematical Reasoning

1. MATH Benchmark — 12,500 difficult math problems. A three-time International Math Olympiad gold medalist got 90%, and GPT-3 got ~5%.
2. Noam Brown's team at OpenAI hits a SOTA 78% on the MATH dataset using Monte Carlo Tree Search (MCTS) at inference time, a technique developed at Deepmind that subsequently led to a 2200 point improvement in AlphaZero's ELO score in Go (3000 → 5200). It turns out it works for language too.
3. Prediction Markets state a median prediction of 91% on the MATH dataset by 2025, which implies VLLMs (Visual) will be superhuman at math in under 15 months.



Overview

Modeling Intelligence

If these two premises hold → designing a human level intelligence is tractable

1. There are scaling Laws to Neural Langauge Models
2. There is Parity between Artificial and Biological Neurons

If we build something the size of the human brain that's made up of neurons that act the same way as ours do → we will have built an intelligence as capable as a human

Taking a step back, we will define intelligence, epistemically grounding the term in provably correct mathematical formalism

Modeling implications for Economic Growth

Growth Theory — Solow-Swan Growth Model + Semi-Endogenous Growth Model
Intelligence Explosion — Current Trends + GPU landscape
Outline Future Model Architecture

Biological Anchors: Size Matters

Myth: Brains are better learners

Parameters	GPT-3	GPT-4	1-year-old	30-year-old
Neurons	~500 Million	~5 Billion	86 Billion	86 Billion
Connections/ Parameters	175 billion	1.8 Trillion	100 Trillion	100 Trillion
Data Set (Bytes)	570 Billion Bytes GigaBytes	20 Trillion Bytes TeraBytes	43 Trillion Bytes (11 Million bits/sec)	1.3 Quadrillion Bytes PetaBytes
Compute (FLOPs)	3.14e23 FLOPs	5.63e24 FLOPs	3.15e24 FLOPs	9.46e25 FLOPs
Adult x : Y y	172 : 1 571 : 1 2280 : 1 301 : 1	17 : 1 55 : 1 65 : 1 16 : 1	1 : 1 1 : 1 30 : 1 30 : 1	1 : 1 1 : 1 1 : 1 1 : 1

Biological Anchors II

Biological Neural Networks Vs. LLM-based Artificial Neural Networks

→BNNs and LLM-based ANNs are both Deep RL models that execute weight update through dot products

**Are they equivalent? → no … but almost

1. **Neuronal Spiking** - Binary vs. Continuous
2. **Hardware/Software** - Hardware == Software vs. Hardware Software Distinction
3. **Learning Styles** - Study + Distillation vs. Pre-training + Weight Sharing
4. **Weight Update** - Forward-Forward Propagation vs. Back-propagation
5. **Compute : Memory Ratio** - 1 : 1 vs. 1000 : 1
6. **Computation Style** - Collective Intelligence vs. Monolithic Intelligence

Emergence

More is Different — P.W. Anderson 1972 (Nobel Prize winning physicist)
→*The essential idea is that in the so called N < infinity limit of large systems it is not only convenient but essential to realize that matter will undergo mathematically sharp, singular; 'phase transitions'*

Future ML Systems Will Be Qualitatively Different
→*Uranium. With a bit of uranium, nothing special happens; with a large amount of uranium packed densely enough, you get a nuclear reaction.*
→*DNA. Given only small molecules such as calcium, you can't meaningfully encode useful information; given larger molecules such as DNA, you can encode a genome.*

Grokking: As models grow, they predict the next element more accurately
→As prediction accuracy climbs, new capabilities emerge

***LLM Emergence*:** With Each Order of Magnitude, Models Grok New Capabilities
→Arithmetic (e.g. 3-4 digit addition/subtraction): Emerged in GPT-3 at ~2 × 10^22 FLOPs.
→Truthful question answering: Emerged in Gopher at ~5 × 10^23 FLOPs.

**This trend will continue!!

Defining Intelligence

Conjecture: Intelligence is the Kolmogorov Complexity of an isolated deterministic system

1. The shortest program to predict all N points is provably the most likely to predict point N+1

2. Kolmogorov compressors navigate the search space of all possible programs

3. SGD + Backprop allows big ANNs to navigate the search space of all possible circuit states

4. SGD + Backprop approximates Kolmogorov compressor

→search space of all possible circuit states approximates search space of all possible programs

5. As N goes to infinity, circuit search discovers the shortest possible program

6. As programs become shorter, cross-entropy log-loss on test sets decreases

→In information theory, minimizing the distinguishability between two distributions is equivalent to minimizing “cross entropy loss”

Is this correct: compression is prediction, but it’s not clear the prediction is intelligence

Romer’s Semi-Endogenous Growth Theory

Formalism — $y_{t+1} = y_t^{f(x)}$

What are the determinants of economic growth?

Solow-Swan Growth Model (previous model):

- Growth Rate of Effective Labor Force (G) is the primary determinant of Economic Growth
- Recursive Loop: more effective labor (people)⟶more ideas⟶more resources
- G will increase dramatically as GPUs enter the work force

**To quantify GPU thought speed, note that majority of the text on the internet is already LLM-generated
**~100 Trillion words on the internet for context (not including video, audio, pdfs)

Semi-Endogenous Growth Model
R&D — AI entering the workforce is unimportant compared to AI entering the research space

- more ideas rather than more people is the true driver of economic growth
- 20 Million Researchers in the world today
- Old loop: more people⟶more people allocated to research⟶more ideas⟶more resources
- new loop: more computer chips⟶more chips allocated to research⟶more ideas⟶more resources

OpenAI’s Definition of AGI is an AI capable automating AI research

- Automating this single task is akin to automating every task

Intelligence Explosion

Multi-Billion Dollar Training runs by 2026

Key Trends

1. *Investment — More dollars to compute*
    - The most expensive training run cost is doubling about every two years.
2. *Moore's Law: More compute per dollar*
    - Compute trends are still doubling about every 2.5 years right now.
3. *Algorithms — More (effective) compute per compute*
    - Since 2012, the effective compute budget has functionally doubled every nine months.

Cognitive ROI

- Inference Side
    - *Faster Minds*
    - *More Minds*
- Pre-training Side
    - *Bigger Minds*
    - *Better Designed Minds*

**Jevon's Paradox*

- if efficiency increases
- then price drops
- therefore demand increases

Year	Cost (in dollars)	Compute (in FLOPs)	Compute given hardware efficiency rate (in FLOPs)	Effective compute given hardware efficiency rate (in FLOPs)
2023	134,837,923.29	1.30×10^{26}	1.30×10^{26}	1.30×10^{26}
2024	417,997,562.19	4.03×10^{26}	5.32×10^{26}	1.01×10^{27}
2025	1,295,792,442.79	1.25×10^{27}	2.18×10^{27}	7.90×10^{27}
2026	4,016,956,572.66	3.87×10^{27}	8.91×10^{27}	6.15×10^{28}
2027	12,452,565,375.26	1.20×10^{28}	3.64×10^{28}	4.80×10^{29}
2028	38,602,952,663.29	3.72×10^{28}	1.49×10^{29}	3.74×10^{30}
2029	119,669,153,256.20	1.15×10^{29}	6.10×10^{29}	2.91×10^{31}
2030	370,974,375,094.23	3.58×10^{29}	2.50×10^{30}	2.27×10^{32}
2031	1,150,020,562,792.12	1.11×10^{30}	1.02×10^{31}	1.77×10^{33}
2032	3,565,063,744,655.57	3.44×10^{30}	4.18×10^{31}	1.38×10^{34}

← Models of Tomorrow

Cost Equivalence

→ Golden Gate Net
→ 1 B-2 Bomber
→ Large Hadron Collider
→ James Webb Telescope
→ Metaverse Investment
→ ISS
→ 10 Aircraft Carriers

GPU Landscape

The largest asset class on the planet is not real estate or energy but rather Intelligence.
→ Its *NPV is roughly $1.2 quadrillion and it will be heavily disrupted over the next decade by GPUs*

Context: Current GWP = ~100 Trillion, America's GDP 2021 = 23.32 trillion

GPT-4 Pre-training: 25,000 A100s for 6 Months costing roughly $63 million
- 10,000 H100 (Nvidia's successor to the A100) could complete this pre-training in 50 days

GPT-series inference: 29,000 GPUs costing roughly $694,444/day

→The market will be flooded with more than 3.5 million H100s by the end of next year
→That's 140 Billion in Revenue from one chip, NVIDIA's revenue in 2022 was $1.54 billion
→NVIDIA is shipping more FLOPs this year than the sum of all the FLOPs they have ever shipped previously

The rise of Intelligence Monopolies
→ GPUnicorns — By the end of next year a handful of firms will have 100k+ GPUs
- American: OpenAI, Google, Anthropic, Inflection, X, Meta, Character.ai, Renaissance Capital
- Chinese: Baidu, Tencent, Alibaba, Bytedance
- European + Canada: Cohere, Stability AI, Hugginface

Models of Tomorrow

1. More Modalities: Not a bolted MoE head like GPT-4+Dalle-3, but end-to-end multi-modal (Gobi)

2. MoE: Switch Transformers, Soft MoE — Trillion Parameter Models with Simple, Efficient Sparsity
- Each Expert: GPT Can Solve Mathematical Problems — Tiny Math Model outperforming GPT-3

3. Synthetic Data as Data Flywheels — Methodologies to turn compute directly into intelligence
- RLHF → ReST / RLAIF / phi-1.5 — GPT-4 as a human level output grader and data annotator
- Pre-training MCTS — Run self-play simulations, now you have a dataset, pre-train on dataset
- Inference MCTS — Verification has polynomial runtime and solution generation is non-polynomial

4. Continuous learning + Better Optimizer: SGD with Restarts + Sophia > ReLu

5. Modified Attention + Longer Context windows: Billion Token Context Windows

6. Cognitive Architectures + Cross Temporal RL: Model Based Autonomous Agents
- Chain of Hindsights, Algorithm Distillation — Techniques to remediate loss of train of thought
- Conversation Swarm Intelligence

7. New Arichetures: JEPA, State Space Models, Retnet, RWKV, Sakana

Conclusion

Intelligence

- Intelligence is not this ethereal concept shrouded in mystery
- It can be formalized and repeatably tested

Scaling Laws

1. As model scale is increased, cross-entropy log-loss drops
2. log-loss == measurement of distinguishability

→ Therefore, models will become indistinguishable from humans on a growing number of tasks

Economics

- If you try to model the points from above, the NPV of the intelligence asset class explodes
- Economic growth dislocates soon thereafter

Pitch — AGI is a concept worth taking seriously

AGI Timelines — I'd wager somewhere around a 75% CI of 2026-2035, but who knows