

Beyond the Singularity

GPT-4 performance

2023 has seen exciting development in development of AI. These developments are markers in our quest to measure machine intelligence. The most exciting was introduction of GPT-4 [4] on March 15, 2023. Here is the introduction description: "GPT-4 is a Transformer-style model... pre-trained to predict the next token in a document... The model was then fine-tuned using Reinforcement Learning from Human Feedback (RLHF). Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar." While GPT-4 architecture is not disclosed, the GPT architecture is understood [5] and can be used for extrapolation.

Consider that GPT-4 has scored in the high 90's percentile of the LSAT [6]. This is a competitive test taken to qualify for law school. This is a test which opens or closes a future for the often young 20's test takers. The test matters to everyone taking it. Most participants prepared for the test. The score GPT-4 received was against prepared, bright, motivated competition.

Consider that GPT-4 scored 80-85% accurate in 15 languages: English, Italian, Afrikaans, Spanish, German, French, Indonesian, Russian, Polish, Ukrainian, Greek, Latvian, Mandarin, Arabic, and Turkish [7]. Random chance alone would yield only 25% accuracy score. There may be humans somewhere who have similar wide-ranging fluency but this is clearly beyond what most humans could achieve.

Following a wide ranging question and answer session with ChatGPT, 1974 ACM Turing Award winner, Don Knuth said "The most immediate impression is the quality of the wordsmithing. It's way better than 99% of copy that people actually write." [8]. Noting that Knuth had assessed GPT3.5, Optimistic AI rescored their understanding what Knuth said against their understanding of GPT-4. The result was that 10 poor answers of 20 with GPT-3.5 were now only 4 poor answers out of the same 20 [9]

It is the thesis of this paper that somewhere in the continuum in the development of Large Language Model (LLM) neural nets such as GPT-4, that a version of the Turing Test was passed. The traditional version of the Turing Test evaluates a machine's ability to exhibit intelligent behavior indistinguishable from that of a human, identified as AGI (Artificial General Intelligence). Of course, nobody is going to be fooled into believing a machine conversant in 15 languages is human. The "indistinguishable" criterion of the traditional test is wrong. We restate our version as we evaluate a machine's ability to exhibit general intelligent behavior no less than that of a human.

Humans have witnessed specialized programs progressively defeat world class human experts in checkers, chess, go, and many other competitive forums. Each specialized program is different. The checkers program plays only checkers and the chess program plays only chess. GPT-4 is different. It is trained on vast, unspecified amounts of human language material to respond conversationally to unstructured input. This allows GPT-4 to respond across a wide variety of prompts and topics.

Following, we compare information capacity of intelligent machine vs. human intelligence that we can use to project the future.

Information capacity

We accept a 2016 Scientific American published [11] estimate of 1000 TB [12] represents human brain capacity.

The number of parameters in GPT-4 and size is unknown to me. The 2020 GPT-3 model has 175 billion parameters and requires 0.8 TB to store [13].

GPT-3 size of 0.8 TB is well below the presumed 1000 TB of humans. Indeed, GPT-3 capacity would appear to be 0.08%, small indeed. We know model sizes are growing, as shown by the 13.5X yearly trendline in Fig. 2. We know GPT-3 dates to 2nd quarter of 2020.

We don't know how big GPT-4 is but we know it is more than a year later than the last entry Megatron-Turing (M-T). Allowing for the 13.5X growth 530B M-T parameters in late 2021 morphs into 7,155B = 7.155T GPT-4 parameters. So, we project 7.155T GPT-4 parameters if parameter growth of 13.5X per year holds through 2023.

Using 175 parameters and 0.8 TB storage, we deduce average storage per parameter is 4.57. Applying 4.57 bytes per parameter to 7.155T, we estimate GPT-4 occupies 32.7 TB, still only 3.27% of human capacity. We are mindful of M-T and GPT-3 deviations from the trend and plead only that the final result incorporates data from both prior models. Further, if there is an architectural sensitivity to storage per parameter, then using GPT-3 to characterize seems the best choice.

We sum up our comparison: consistent with 13.5 X/year trend continuation, early 2023 GPT-4 is expected to occupy 32.7 TB as compared to human brain 1000 TB.

Projecting Future Information capacity

If GPT-4 occupying 32.7 TB, is representative of continuing the trend to 2023, then 2 more years at 13.5X/year has the information capacity of 5.8 humans. This forecast is based on extrapolating a 13.5X trend line for which we are missing the last year and a half of data. The 13.5X is utterly unsustainable for the long term. Compare to the informal semiconductor growth rate of doubling every two years (sometimes taken as every 18 months) $\sqrt[2]{2} = 1.41X/year$. We expect long term for semiconductor growth to create an asymptotically approached limit. There are other limits such as electrical power limitations, value to society, emerging legal landscape, and other world conditions. But when?

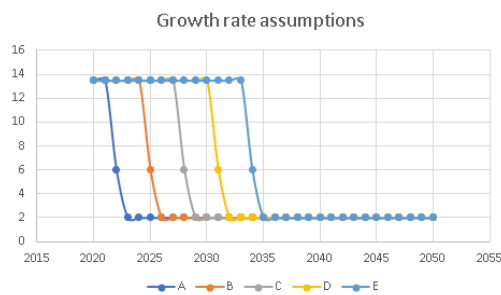


Fig. 3 Growth rates dropping from 13.5X to 2X.

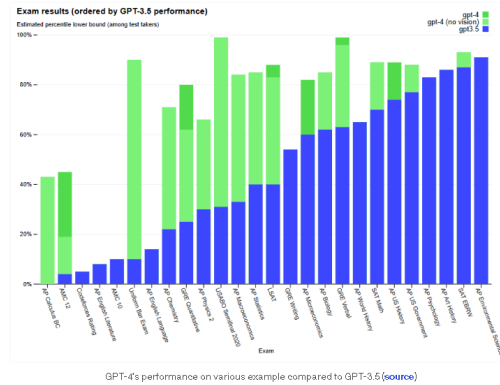


Fig. 1 Scoring GPT4 performance, shown against GPT3.5 results. ChatGPT free version is powered by GPT3.5. A \$20/month subscription substitutes GPT-4. There is substantial improvement achieved over a year's time. Interestingly, in some of the poorest scoring GPT3.5 challenges, no visible improvement is seen by GPT4. Progress there is stalled for now. Sources: as discovered [1], discussed [2], and source [3]

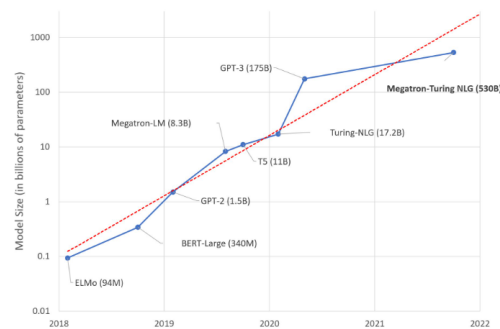


Fig. 2 Exponential growth of 13.5X yearly in Large Language Models from 2018 to 2021. [10]

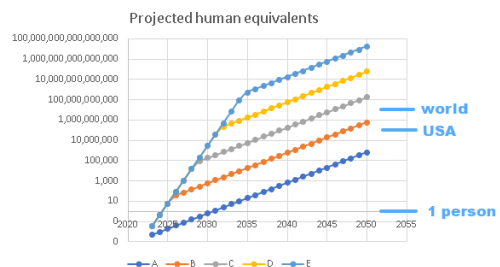


Fig. 4 Projecting Model size where 1000 TB = 1 human. source: [author spreadsheet](#)

We now extend our previous examination into the future, considering the infeasibility of projecting a 13.5X growth forward. We account for this infeasibility by exploring five scenarios labeled A through E and shown graphically as Fig. 3. Each scenario begins with 13.5X growth and transitions to 2X growth within two years, with the transition set at 6X. Scenario A's transition began in 2022, while E's is anticipated in 2034, and B through D are spaced at three-year intervals between these two.

For the purposes of our analysis, the choice between scenarios B through E is inconsequential, as they all predict 13.5X growth and a crossover in information capacity equivalence in 2025.

Running these scenarios reveals in Fig. 4 that the entire information capacity of the USA population, estimated at 300 billion TB [14], will be reached between 2049 (B) and 2032 (E), assuming we remain in the 13.5X growth regime. The crossover point of model size and human capacity, projected as occurring around 2024-2025 based on 13.5X growth, is taken as a marker for the onset of the singularity and what lays beyond.

References and notes

- 1. <https://www.assemblyai.com/blog/introduction-large-language-models-generative-ai>
- 2. <https://openai.com/research/gpt-4>
- 3. <https://arxiv.org/abs/2303.08774>
- 4. <https://arxiv.org/abs/2303.08774>
- 5. <https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/>
- 6. <https://openai.com/research/gpt-4>
- 7. <https://openai.com/research/gpt-4>
- 8. <https://cs.stanford.edu/~knuth/chatGPT20.txt>
- 9. <https://optimystic.ai/knuth>
- 10. <https://huggingface.co/blog/large-language-models>
- 11. <https://www.scientificamerican.com/article/new-estimate-boosts-the-human-brain-s-memory-capacity-10-fold/>
- 12. quadrillion bytes from article = 10^{15} bytes = 1000×10^{12} bytes = 1000TB
- 13. <https://arxiv.org/abs/2005.14165>
- 14. 300×10^{24} bytes equal to population 3×10^8 times human brain 1000 TB = 1000×10^{12} bytes