Hype, Media Frenzy, and Mass Societal Hysteria: Perspectives on Human-imitative Intelligence

Raeid Saqur[1,3] **Riccardo E. Sarpietro**[2,4] **Muhammad Shaheer**[5,7] **Marcel C. Buehler**[6,8]

[1]University of Toronto, Department of Computer Science [2]University of Catania, Department of Engineering
[3]Vector Institute for AI [4]PeRCeiVe Lab
[5]University of Luxembourg, Department of Automation [6]ETH Zurich, Department of Computer Science
[7]Automation & Robotics Group [8]AIT Lab

raeidsaqur@cs.toronto.edu, riccardo.sarpietro@phd.unict.it
muhammad.shaheer@uni.lu, buehlmar@ethz.ch

Topic

The chase for human-imitative artificial intelligence and the fear for it live a moment of hype, with the recent release of AI products such as semi-autonomous cars and chatbots. Are these the incarnation of human-imitative intelligence? Motivate your discussion with credited scientific evidence and peer-reviewed research work.

1 Introduction

> The original question, '*Can machines think?*' I believe to be too meaningless to deserve discussion.
>
> A. Turing, Mechanical Intelligence [1]

The quest for human-imitative artificial intelligence [2] has been an enduring topic in artificial intelligence (AI) research since its inception in the late 1950s, with the aspiration of realizing an entity possessing human-level intelligence. Since then, the history of AI (§section 2) has seen debates analogous to the question we explore here in different forms and variations, correlated with the promise and progress in the AI research frontier (Fig. 1). More recently, Jordan [2] uses the term "*human-imitative AI*" to refer to an AI entity that appears to be human, mentally if not physically. A few other interchangeable terms like *human-level* or, *human-like* AI, artificial general intelligence (AGI) have been used to convey the same idea.

The remarkable technical evolution of AI research over the past decade, coalesced with big data and computational hardware improvements, has seen the rise of powerful AI systems. The capabilities, risks, and opportunities of these AI systems have generated unprecedented mass interest and hysteria. Quintessential examples include the emerging capabilities [3] of the latest cohort of foundational [4] large language models (LLMs) [5–10] – especially Open AI's GPT series (ChatGPT [11], GPT-4 [12]) – that have reinvigorated the question whether these fascinating AI products are the incarnation of human-imitative intelligence beyond academia [13] to the cultural zeitgeist.

To explore the posed question, we stand on the shoulders of giants – influential thinkers, philosophers, AI researchers who have visited the question earlier that we aim to revisit in 2023. We arrive at our conclusion by scrutinizing arguments in favour and against both from historical and current perspective.

Submitted to the 17th International Computer Vision Summer School (ICVSS 2023) Essay Competition

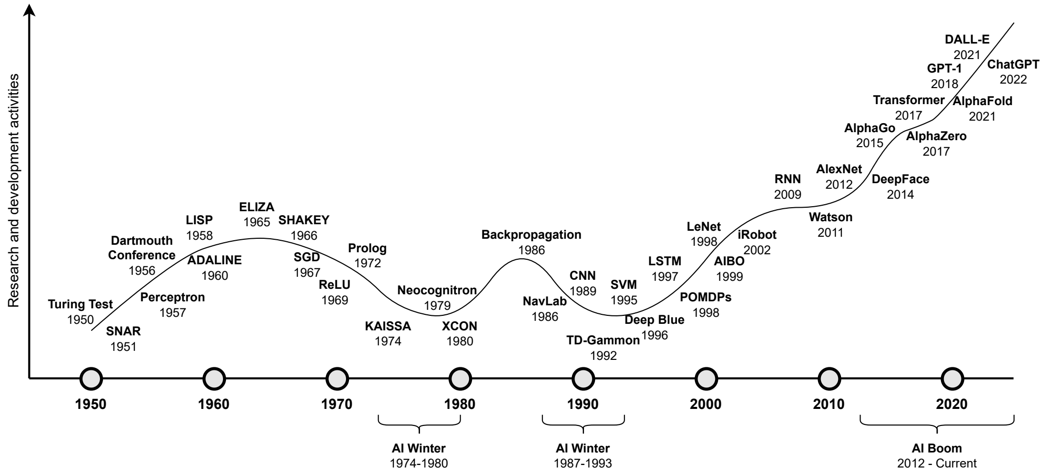

Figure 1: AI Research and development activities based on literature density since 1950 [14, 15].

2 Historical Perspective

Half a century ago, Alan Turing dissected AI's human-like abilities most famously by asking: *Can machines think?* using the 'Imitation Game' [1, 16], and subsequent formulation of the seminal Turing Test. Turing rebuttals a slew of emblematic arguments for the case against AI ever possessing human-like abilities (e.g., the agency of thinking). Of these contrary views, the one most relevant, even after 50+ years, is the '*Arguments from consciousness*' (AC). It has seen reincarnations in various forms and arguments in AI research and philosophy by luminaries following Turing, like Dreyfus, Searle, Harnad, Haugeland [17–20] to name a few.

The gist of AC is that machines can never attain human-like intentionality [20] or, *a fortiori* 'human phenomenology' (emotions, ego, imagination, moods, consciousness etc.), and (thus) cannot ever attain human-imitative intelligence. However, these AC become largely moot if we use the imitation game (Turing Test) setup, where outward physiological human attributes are excluded by design. To exemplify, a robotic lab assistant's (textual) response: "*I'm terribly ashamed and sorry for burning down the lab*" is human-imitative irrespective of whether any 'feelings' behind are. Turing alludes to this by rhetorically implying that putting on 'artificial flesh' is not necessary on a hypothetical 'thinking' machine. Thus, if human-imitative intelligence entails generating human-like language responses, then the capabilities of current SOTA LLMs (like GPT4 [12]) enables us to answer 'yes' to our question.

'*Arguments from Various Disabilities*' (AVD) is another archetypal category that takes the form: "*AI systems can do all these tasks, however, they still cannot do X*" – where X can be a task or, a collection of tasks. Examples of the AVD are still prevalent – to exemplify, take the various faux pas of voice assistants like Siri, or, the generated hallucinations in LLM outputs. An early prominent example – the *Lady Lovelace's objection* [21] – that machines are unable to do anything novel outside what is programmed, is weak if not debunked in the current perspective. Adding to Turing's formidable counter to this argument, the AI research sub-domain of 'library-learning' [22] is a counter-example: Machines learn to form new (unknown *ex-ante*) routines from programmed primitives.

Yet another timeless and popular culture (e.g., 2015 Hollywood movie: *Ex-Machina*) contrarian view is the issue of '*human-like*' vs. '*<u>simulation</u> of human-like*' in negating such human-imitative intelligence evidence. Historically, John Searle's '*Chinese room argument*' [18] (CRA) is illustrative of such takes. The gist of CRA is a human, with no knowledge of Chinese, endeavouring to simulate an operational understanding of Chinese in a locked room with a look-up table (e.g., dictionaries, instructions) to mechanically output correct responses to Chinese inputs. Behaviourally, the human with a look-up table will pass the Turing Test for understanding Chinese! There are many classical responses to this argument, including the systems and solipsistic point of views: although the man in the room does not learn Chinese, (functionally) the system as a whole does; or you need to be the machine to know (and make claims of) whether it feels etc. Regardless, the preceding ‘functionality’ argument suffices against such takes.

Perhaps a much stronger argument negating machine’s ability to ever reach human-like intelligence hinges on the Chomskyan universal grammar (UG) [23] proposition and the poverty of stimulus (POS) [24] argument. UG implies innate intelligence (like language faculty) in humans independent of sensory experience. POS stipulates that children can acquire language understanding largely without negative samples – a feat currently impossible for machines – that supports UG. If we agree to these, then exhibiting human-imitative intelligence would require innate human components and human-like amalgamation and learning from experience. Further, the introduction of innate genetic components raises the ‘evolution’ problem, like how we imbue the hereditary material in machines, and human-imitative (requiring innate mechanisms) machine learning. Thus, even if machines can exhibit close to human-like language faculty trained and fine-tuned using essentially the entire web (which subsumes the data contamination argument that says LLMs memorize answers in training data [25, 26]), it can never achieve human-imitative ‘generalized’ intelligence.

3 Current Perspective

Current SOTA LLMs are able to pass the original Turing Test and show sparks of AGI [5, 3, 13]. However, this seminal and once apt test has long passed the expiration date as a viable metric for assessing human-imitative intelligence [27, 21], with various extensions suggested by researchers over the years[28–30].

Further, none of the current AI systems are able to embody human-like wholesome intelligence or knowledge. Their remarkable feats are limited to fine-tuned verticals on domain-specific data (e.g., LLMs for NLP tasks; separate, targeted computer vision models for self-driving cars, medical diagnosis). Even within their specialized domains, these systems suffer from various (relevant) disabilities. To exemplify, self-driving vehicles are unable to comprehend or attune to environmental changes like weather and illumination conditions [31–34]. While these are rapidly shrinking limitations (or, disabilities), the current state is far from human-like capabilities to be deemed as incarnation of human-imitative intelligence. In the NLP domain, the SOTA LLMs are still incapable of doing distinct human-like tasks, like *creative problem solving* [35, 36]. Brute-force extrapolation of scaling factors [37] is an unlikely answer, and indefinite scaling has various impeding limits [38]. Alternative technologies like quantum computing or neuromorphic algorithms [39] still remain in niche research stages.

We have been historically poor at estimating technological advances, including AI capabilities, and extrapolating subsequent societal impacts. We need not regress to the 19th century Luddites [40] to find supporting examples. In the 1940s, ‘experts’ had estimated the arrival of self-driving cars by 1960 thanks to the then rapidly developing automotive sector. In the 2000s and 2010s, Kurzweil [41] predicted the creation of a super-intelligent algorithm or AGI, and the occurrence of the so-called *singularity* event as very near. Cursory literature search reveals large swath of works that make highly inaccurate predictions about singularity or the incarnation of human-imitative intelligence[42, 43]. Methodical frameworks like [44] objectively argue that the advent of such transformative AGI is highly unlikely even year 2043! More recently, circa 2020, there was a massive investment surge in robotics companies, hyped by the prospect of humanoid robots with human-like skills or capabilities. Perversely, current SOTA humanoid robots (e.g., from Boston Dynamics) and their capabilities – demonstrated by captivating, viral videos – predominantly rely on classical control theories rather than deep learning [45]. We have also seen subsequent divesting and mass transfer of ownership in the robotics industry after the premature hypes were stymied [46].

4 Conclusion

The hype, media frenzy, and mass societal hysteria around AI’s ability to mimic humans is understandable due to the translation of such abilities into objective risks and impacts on society. Examples and dire scenarios include job displacements, generative fake news, propaganda, privacy issues, effects of inherent biases in training data, further marginalization of selective groups, tools for Orwellian surveillance. The noise around AI today is further exacerbated by quick mass consumption channels (e.g., social media, free-form news outlets). The flashy, captivating achievements of

ChatGPT [11], AlphaGo [47], AlphaZero [48], AlphaStar [49] and scientific AI like AlphaFold [50], AlphaTensor [51] and AlphaDev [52] have (rightfully) thrust the topic in the global human psyche.

Upon considering both the historical and current perspective, we argue that there is a collective tendency to overestimate AI's human-like abilities. While recent developments in AI are remarkable, we conclude that these developments should not be labelled as the incarnation of human-imitative intelligence. Perhaps, our discussion of the very question will be looked back (in a few decades) as a repetition of the question famously asked by Turing: *Can machines think*? – upon the advent of a technology that was less powerful than a modern-day pocket calculator!

References

[1] A. Turing, "Computing machinery and intelligence-am turing," *Mind*, vol. 59, no. 236, p. 433, 1950.

[2] M. I. Jordan, "Artificial intelligence—the revolution hasn't happened yet," *Harvard Data Science Review*, vol. 1, no. 1, pp. 1–9, 2019.

[3] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler *et al.*, "Emergent abilities of large language models," *ArXiv preprint*, vol. abs/2206.07682, 2022.

[4] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill *et al.*, "On the opportunities and risks of foundation models," *ArXiv preprint*, vol. abs/2108.07258, 2021.

[5] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong *et al.*, "A survey of large language models," *ArXiv preprint*, vol. abs/2303.18223, 2023.

[6] M. Shoeybi, M. Patwary, R. Puri, P. LeGresley, J. Casper, and B. Catanzaro, "Megatron-lm: Training multi-billion parameter language models using model parallelism," *ArXiv preprint*, vol. abs/1909.08053, 2019.

[7] J. Wei, M. Bosma, V. Y. Zhao, K. Guu, A. W. Yu, B. Lester, N. Du, A. M. Dai, and Q. V. Le, "Finetuned language models are zero-shot learners," in *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022.

[8] S. Zhang, S. Roller, N. Goyal, M. Artetxe, M. Chen, S. Chen, C. Dewan, M. Diab, X. Li, X. V. Lin *et al.*, "OPT: Open pre-trained transformer language models," *ArXiv preprint*, vol. abs/2205.01068, 2022.

[9] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, P. Barham, H. W. Chung, C. Sutton, S. Gehrmann *et al.*, "PaLM: Scaling language modeling with pathways. 2022," *ArXiv preprint*, vol. abs/2204.02311, 2022.

[10] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar *et al.*, "LLaMA: Open and efficient foundation language models," *ArXiv preprint*, vol. abs/2302.13971, 2023.

[11] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray *et al.*, "Training language models to follow instructions with human feedback," *Advances in Neural Information Processing Systems*, vol. 35, pp. 27 730–27 744, 2022.

[12] OpenAI, "GPT-4 technical report," *ArXiv preprint*, vol. abs/2303.08774, 2023.

[13] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg *et al.*, "Sparks of artificial general intelligence: Early experiments with GPT-4," *ArXiv preprint*, vol. abs/2303.12712, 2023.

[14] T. e. a. Noguchi, "A practical use of expert system "ai-q" focused on creating training data," in *2018 5th International Conference on Business and Industrial Research (ICBIR)*. Bangkok: IEEE, May 2018, p. 73–76.

[15] A. T. et al., “A brief history of AI: How to prevent another winter (a critical review),” *PET Clinics*, vol. 16, no. 4, pp. 449–469, oct 2021.

[16] B. J. Copeland, *The essential turing*. Clarendon Press, 2004.

[17] H. L. Dreyfus, *What computers still can’t do: A critique of artificial reason*. MIT press, 1992.

[18] J. R. Searle, “The chinese room revisited,” *Behavioral and brain sciences*, vol. 5, no. 2, pp. 345–348, 1982.

[19] S. Harnad, “The symbol grounding problem,” *Physica D: Nonlinear Phenomena*, vol. 42, no. 1-3, pp. 335–346, 1990.

[20] J. Haugeland, *Artificial intelligence: The very idea*. MIT press, 1989.

[21] S. Bringsjord, P. Bello, and D. Ferrucci, “Creativity, the turing test, and the (better) lovelace test,” *The Turing test: the elusive standard of artificial intelligence*, pp. 215–239, 2003.

[22] K. Ellis, C. Wong, M. Nye, M. Sable-Meyer, L. Cary, L. Morales, L. Hewitt, A. Solar-Lezama, and J. B. Tenenbaum, “Dreamcoder: Growing generalizable, interpretable knowledge with wake-sleep bayesian program learning,” *ArXiv preprint*, vol. abs/2006.08381, 2020. [Online]. Available: https://arxiv.org/abs/2006.08381

[23] R. Montague *et al.*, “Universal grammar,” *1974*, pp. 222–46, 1970.

[24] S. Laurence and E. Margolis, “The poverty of the stimulus argument.” *British Journal for the Philosophy of Science*, vol. 52, no. 2, 2001.

[25] I. Magar and R. Schwartz, “Data contamination: From memorization to exploitation,” in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Dublin, Ireland: Association for Computational Linguistics, 2022, pp. 157–165.

[26] A. Jacovi, A. Caciularu, O. Goldman, and Y. Goldberg, “Stop uploading test data in plain text: Practical strategies for mitigating data contamination by evaluation benchmarks,” *ArXiv preprint*, vol. abs/2305.10160, 2023.

[27] J. Moor, *The Turing test: the elusive standard of artificial intelligence*. Springer Science & Business Media, 2003, vol. 30.

[28] P. Hayes and K. Ford, “Turing test considered harmful,” in *IJCAI (1)*, 1995, pp. 972–977.

[29] D. L. Dowe and A. R. Hajek, “A computational extension to the turing test,” in *Proceedings of the 4th conference of the Australasian cognitive science society, University of Newcastle, NSW, Australia*, vol. 1, no. 3. Citeseer, 1997.

[30] J. Hernandez-Orallo, “Beyond the turing test,” *Journal of Logic, Language and Information*, vol. 9, pp. 447–466, 2000.

[31] S. G. Narasimhan and S. K. Nayar, “Vision and the atmosphere,” *International Journal of Computer Vision*, vol. 48, 2002.

[32] K. Garg and S. K. Nayar, “Vision and the rain,” *International Journal of Computer Vision*, vol. 75, 2007.

[33] T. T. Robby, “Visibility in bad weather from a single image,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[34] D. Dengxin, T. . Robby T, P. Vishal, M. Jiri, S. Bernt, and G. Luc Van, “Special issue on “computer vision for all seasons: Adverse weather and lighting conditions”,” *International Journal of Computer Vision*, vol. 129, 2021.

[35] F. Shi, X. Chen, K. Misra, N. Scales, D. Dohan, E. Chi, N. Schärli, and D. Zhou, “Large language models can be easily distracted by irrelevant context,” *arXiv preprint arXiv:2302.00093*, 2023.

[36] S. Naeini, R. Saqur, S. Mozhgan, J. Giorgi, and B. Taati, "Large language models are fixated by red herrings: Exploring creative problem solving and einstellung effect using the only connect wall dataset," *arXiv preprint arXiv:2306.11167*, NeuRIPS 2023-under review.

[37] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, "Scaling laws for neural language models," *arXiv preprint arXiv:2001.08361*, 2020.

[38] K. Flamm, *Measuring Moore's Law: Evidence from Price, Cost, and Quality Indexes*, Cambridge, MA, Apr 2018, no. w24553. [Online]. Available: http://www.nber.org/papers/w24553.pdf

[39] C. D. e. a. Schuman, "Opportunities for neuromorphic computing algorithms and applications," *Nature Computational Science*, vol. 2, no. 1, p. 10–19, Jan 2022.

[40] D. Silver, "Rebels against the future: The luddites and their war on the industrial revolution: Lessons for the computer age/silicon snake oil: Second thoughts on the information..." *Journal of Popular Culture*, vol. 31, no. 4, p. 180, 1998.

[41] R. Kurzweil, *The singularity is near: when humans transcend biology*. Viking, 2005.

[42] J. Cordeiro, "The singularity is nigh," *Engineering & Technology*, vol. 5, no. 1, pp. 27–29, 2010.

[43] S. Mortuza, "Nearing singularity," *Crossings: A Journal of English Studies*, vol. 4, pp. 224–226, 2014.

[44] A. Allyn-Feuer and T. Sanders, "Transformative agi by 2043 is <1% likely," 2023.

[45] S. e. a. Kuindersma, "Optimization-based locomotion planning, estimation, and control design for the atlas humanoid robot," *Autonomous Robots*, vol. 40, no. 3, p. 429–455, Mar 2016.

[46] J. Majko, "Softbank's vision fund or how to lose 5 billion dollars in a car's seat," in *OVERCOMING CRISIS: Case Studies of Asian Multinational Corporations*. World Scientific, 2023, pp. 95–109.

[47] D. e. a. Silver, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, p. 484–489, Jan 2016.

[48] D. S. et al., "Mastering chess and shogi by self-play with a general reinforcement learning algorithm," 2017.

[49] O. e. a. Vinyals, "Grandmaster level in starcraft ii using multi-agent reinforcement learning," *Nature*, vol. 575, no. 7782, p. 350–354, Nov 2019.

[50] J. e. a. Jumper, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, no. 7873, p. 583–589, Aug 2021.

[51] A. e. a. Fawzi, "Discovering faster matrix multiplication algorithms with reinforcement learning," *Nature*, vol. 610, no. 7930, p. 47–53, Oct 2022.

[52] D. J. e. a. Mankowitz, "Faster sorting algorithms discovered using deep reinforcement learning," *Nature*, vol. 618, no. 7964, p. 257–263, Jun 2023.