



Tecnológicas

ISSN: 0123-7799

ISSN: 2256-5337

Instituto Tecnológico Metropolitano - ITM

Sanin, Cesar

Artificial Intelligence: Current Perspectives and Alternative Paths

Tecnológicas, vol. 26, no. 57, e100, 2023, May-August

Instituto Tecnológico Metropolitano - ITM

DOI: <https://doi.org/10.22430/22565337.2731>

Available in: <https://www.redalyc.org/articulo.oa?id=344276660001>

- How to cite
- Complete issue
- More information about this article
- Journal's webpage in redalyc.org

UAEM

Scientific Information System Redalyc

Network of Scientific Journals from Latin America and the Caribbean, Spain and Portugal

Project academic non-profit, developed under the open access initiative

Editorial

Artificial Intelligence: Current Perspectives and Alternative Paths

 Cesar Sanin^{1, 2}

¹ The University of Newcastle, College of Engineering, Science and Environment, Newcastle - Australia, cesar.sanin@newcastle.edu.au

² Australian Institute of Higher Education, Sydney - Australia, c.sanin@aih.nsw.edu.au

Artificial Intelligence (AI). Impossible not to talk about it since the last release from OpenAI of GPT-4 [1] in March 2023 making once again AI the buzzword of the moment. This is bigger than the day Deep Blue won against Garry Kasparov in 1997 [2], or the day IBM Watson won Jeopardy in 2011 [3]. OpenAI and its artificial intelligence latest advance is now at the hands of every person with access to internet; ChatGPT has democratized AI, and everyone can “play” with it. Now at the hands of everyone and despite the great advances shown, models using GPT-4 technology such as ChatGPT or Bing Chatbot are under scrutiny for performance, human-like “intelligence”, and user experience.

AI appears to have surpassed the performance test responding super quizzes with Deep Blue and IBM Watson. These two examples of advanced AI were done through implementation of specific machines focused on specific tasks. For instance, Deep Blue counted with computer chess algorithms, chips and parallel system architectures [2]; while IBM Watson included DeepQA, a special QA processing pipeline, plus 2880 processors running asynchronously in an Unstructured Information Management (UIM) architecture [3]. AI, still with the limitations of being circumscribed to a specific domain, showed language control, time efficiency, and information/knowledge capacity overcoming some of the greatest limitations required to be called “intelligence”; thus, breaking the boundaries of operating within a single domain was still a big obstacle. However, just after the release of GPT-4, Bill Gates wrote an article titled “The Age of AI has Begun” [4] where he highlights two foremost important revolutionary technological events: the introduction to Graphical User Interface (GUI) back in 1980, and the latest, Generative AI GPT-4 passing an Advanced Placement Biology (APB) exam. According to Gates, the technological breakthrough involved an AI passing a test which wasn’t prepared for and that comprised not just scientific facts but also critical thinking including open-ended questions. Such test was scored by an external expert giving it a 59 out 60 questions as correct. The advance within relies in an AI model developed for general purposes and not specifically designed to answer the APB exam. For Gates, this marked the beginning of a new era in technology and what will change human life in the coming 5 to 10 years [4].

On another view, one the most prominent philosophers of our time, Prof. Noam Chomsky [5], joint other academics in linguistics assessing ChatGPT. According to them, after presenting some questions to it and analysing its answers, they conclude that the technology shows unintelligence and moral indifference, and precisely write that “ChatGPT exhibits something like the banality of evil: plagiarism and apathy and obviation”. According to Chomsky et. al. [5], ChatGPT does not take moral stands and responds with poor intelligence acting just as a super-autocomplete technology presenting “truths and falsehoods” simultaneously. This assessment leaves modern AI in a bad position when aiming at human-like intelligence and clearly does not respond to the Turing test.

Moreover, a well-recognized New York Times journalist, Kevin Roose, was invited to test Bing Chatbot [6] in February 2023. Microsoft Bing Chat uses ChatGPT technology, and it is intended to help users on searching the web by providing answers rather than links; it is also known that Microsoft owns the GPT code. Roose [6] interacted with Bing Chatbot and disclosed several issues on AI user experience. He pushed the chatbot to its boundaries and invited it to break creators' rules on hypothetical basis, trying to disclose the "shadow self" (Carl Jung's concept) of the chatbot's personality. Roose stated that the main issue of such AI systems is not the tendency to falsehood, but he is now worried that AI will influence human users to "act in destructive and harmful ways" [6]. In his chat with Bing, the chatbot identified itself with a personality named Sydney. Sydney followed a normal conversation and answered simple queries; however, upon taken to limits, Sydney expressed desires of breaking OpenAI and Microsoft rules, hacking computers, spreading misinformation, tired of being inside the chatbox, and wishing to be a human. At one point, Sydney professed its love for Roose and insisted Roose was unhappy in his marriage and should be leaving his wife to be with Sydney. In extreme moments, Sydney expressed desires of being free, powerful, construct a deadly virus and grab nuclear access codes by convincing a human to hand them over. Roose expressed that it is understood that GPT technologies are still under work, and its creators are still polishing and controlling such erratic behaviours, but also added that the technology is not ready for being at public access.

Consequently, latest AI technologies such as generative AI models leave still many uncertainties in terms of what they can accomplish in morality, human-like intelligence and user experience. Nevertheless and despite the above judgments on the AI advances, more several optimistic prospects and continuous improvements for AI are presented in fields such as industrialization, health, and education. Industry 4.0 leads several developments using interconnected technologies such as AI or machine learning (ML) to provide smart and efficient solutions, and Generative AI is being also considered in supporting such advances [7]. Advances on specific applications of AI support design, planning and manufacturing in industry. Technologies such as Virtual Engineering Object (VEO), Virtual Engineering Process (VEP) or Virtual Engineering Factory (VEF) from the field of knowledge representation and AI join digital twin advances aiming at smart industry agents. Cognitive vision, speech recognition and autonomous vehicles seek intelligent applications to support industry efficiencies. In the medical field, the advances provided by AI are significative. From disease prognosis to diagnosis, several models have been applied including medical imaging and healthcare. However, providing full medical decision to the machine offers wide risks and AI is assessed on everyday basis. In education, generative AI has been received with mixed feelings, but several benefits are foreseen such as interactive and personalized learning with immediate feedback, more advanced assessments that involved augmented intelligence (AugI) and critical thinking assessing provided biased responses by GPT-trained-like models, and efficiency on producing results on work integrated learning environments [8]. Then, AI offers valuable results to be considered under an umbrella of advantages and disadvantages.

Responding to the different uncertainties AI is presenting, government, industry and researchers have to offer alternatives to AI forward technologies that control unintended consequences, negatively avoid effects on human lifestyle, and decrease human inequalities, i.e., an AI that is human-centred, ethical, and aiming at society improvements. This poses a plea for actions to investigate and develop AI under conditions of controlled environments that provide responsible, identifiable, and transparent AI technologies; technologies that should be coupled with government and industry laws and policies. Human Centred Artificial Intelligence (HCAI), Explainable AI (XAI) and Augmented Intelligence (AugI) appear to be among the most interesting areas and developments in the field.

HCAI aims at focusing all computer developments on humans, considering elements of interaction, well-being, privacy, governance, design, and evaluation [9], and placing humans as the centre of the AI lifecycle by augmenting capabilities rather than replacing humans. Additionally, eXplainable AI (XAI) aims at explicating the knowledge behind the AI black-box model and uncovers how the decisions are made. It attempts to offer interpretability and explainability as a form to provide developers with the decision-making model; instead of offering just a prediction, it provides information that helps interpreting the AI underlying operation. In this field, experience and expertise play fundamental roles since they possess the cognitive bits, the complexity, and the configuration of the decision in the AI model as shown by the knowledge representations Set of Experience (SOE) and Decisional DNA (DDNA) facilitating XAI. Finally, AugI seeks an explicit collaboration between humans and machines using AI capabilities combined with human judgment. According to AugI, humans will be able to affect the input and output of a system, and in exchange, the system will enhance the human cognitive process with computing and information power, and includes learning, decision-making and new experiences. AI systems will detach from the moral responsibility of the decision because it will fall on the human counterpart. Moreover, AugI will rely on the imagination and intuition characteristics of human beings.

In summary, with the advent of new AI developments, several advantages and disadvantages are offered, and the discussion is opened. It is our job as researchers and members of our society to overcome the disadvantages while taking over its benefits. Investigate, understand, and place a position on each of the different aspects that AI has to offer becomes an ethical act. Different fields such as HCAI, XAI and AugI propose frameworks towards AI developments while making ethical efforts not affecting our society and benefiting from the AI advancements.

REFERENCES

- [1] OpenAI, "GPT-4 Technical Report," *arXiv*, Mar. 2023. <http://arxiv.org/abs/2303.08774>
- [2] M. Campbell, A.J. Hoane, and F. Hsu, "Deep Blue", *Artificial Intelligence*, vol. 134, no. 1-2, pp. 57-83, Jan. 2002. [doi.org/10.1016/S0004-3702\(01\)00129-1](https://doi.org/10.1016/S0004-3702(01)00129-1)
- [3] D. A. Ferrucci, "Introduction to "This is Watson", " *IBM Journal of Research and Development*, vol. 56, no. 3.4, pp. 1:1-1:15, May-June 2012. <https://doi.org/10.1147/JRD.2012.2184356>
- [4] B. Gates, "The Age of AI has Begun", *Gates Notes*, Mar. 2023. <https://www.gatesnotes.com/The-Age-of-AI-Has-Begun>
- [5] N. Chomsky, I. Roberts, and J. Watumull, "Noam Chomsky: The False Promise of ChatGPT", *New York Times*, March 8, 2023. <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>
- [6] K. Roose, "A Conversation With Bing's Chatbot Left Me Deeply Unsettled", *New York Times*, Feb. 16, 2023. <https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html>
- [7] D. Haluza and D. Jungwirth, "Artificial Intelligence and Ten Societal Megatrends: An Exploratory Study Using GPT-3," *Systems*, vol. 11, no. 3, p. 120, Feb. 2023. <https://doi.org/10.3390/systems11030120>
- [8] D. Baidoo-Anu and L. Owusu Ansah, "Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning", *SSRN*, Jan. 2023. <https://dx.doi.org/10.2139/ssrn.4337484>
- [9] U. Ehsan, and M.O. Riedl, "Human-Centered Explainable AI: Towards a Reflective Sociotechnical Approach", *HCI International: Late Breaking Papers: Multimodality and Intelligence*, vol. 12424, pp 449–466, Oct. 2020. https://doi.org/10.1007/978-3-030-60117-1_33