

Title

Leadership Today Means Leading with Human Values in the Age of Artificial Intelligence

By

Monica Lopez, PhD, Co-Founder and CEO, Cognitive Insights for Artificial Intelligence

Since entering into public consciousness in November of last year, news continues unabated about generative artificial intelligence (AI) and ChatGPT in particular. The list is long, but highlights include statements on the introduction of plagiarism at scale, the demise of human creativity, the domination of power players, the hodgepodge of truths and falsehoods, the call to ban its use, and even the claim that artificial general intelligence (AGI) is closer than ever.

Amid the noise, the fundamental reality of selective memory and missed opportunity remains. We have already experienced awe and distrust with a human-like chatbot when computer scientist Joseph Weizenbaum from the Massachusetts Institute of Technology introduced ELIZA in 1966 and beguiled users with its psychotherapeutic capabilities. It was a beginning step in natural language processing and while it was very simple in its conversational capabilities, it did not deter users from attributing it “background knowledge, insights and reasoning ability.”<sup>1</sup> Perhaps most importantly (and with uncanny foresight), it was above all an experiment in the psychology of human relationships, revealing the assumptions, expectations and desires we attribute to our interlocutor, of machine kind in this case, as we converse with and pour out our souls to it to understand ourselves. Food for thought: observing this reaction, why did we as a collective not delineate potential risks resulting from the inevitable advancement of this kind of intelligent system and immediately build risk management frameworks to prepare us for the possible questions and harms of the future?

Today, with ChatGPT and other generative AI systems producing not just text but audio, images, and video and proliferating in use case and across users, we have truly transformative and disruptive technology at our fingertips. While it is nowhere near human-like intelligence or AGI but certainly far more capable than ELIZA, we are enthralled by its possibilities and only now collectively experiencing and voicing its faults and risks. The proverbial genie is effectively out of the bottle and we have

---

<sup>1</sup> Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45.

definitely entered a new era. So what does this have to do with thought leadership? And more specifically, leading with human values in the age of AI? Everything.

While AI is a powerful tool, we cannot assume it will support our business's values, our culture, and/or our driving purpose. Whether we are building AI-enabled technology or using it to power our products, services, and/or organizational needs, we must all keep abreast of AI developments to both assess its potential and understand its current limitations. Leadership therefore lies in being thinkers and doers of this balance between potential and limitations so we can not only devise strategies that utilize the potential, but also identify solutions to improve upon the limitations we have before us, or at least create safeguards against any limitations we cannot positively eliminate.

Let's breakdown the current landscape as our object of analysis. We currently find ourselves embroiled within a multilayered situation in which the excitement of innovation has taken precedence over responsible leadership. Some key points to underscore, hardly all-encompassing, are: (i) we have created bigger, faster, better-looking, authoritative-sounding, sometimes impressive and sometimes biased, discriminatory and/or false, and indistinguishably real vs. artificial AI-generated content, (ii) we have achieved greater awareness of and access to this technology, exponentially increasing the number of people utilizing it and thus experiencing both its potential and its many failings, (iii) the race is on to build improved, larger and more powerful commercial options, and (iv) now we want to rein in what we have developed in some form or another. In the absence of legal requirements — which will come (e.g. the European Union's proposed AI Act<sup>2</sup>), will pose new questions about liability (e.g. will a technology platform's AI-generated content be protected by Section 230?<sup>3</sup>), and will necessarily have intended and unintended consequences on the AI marketplace—prioritization lies in responsible leadership on the side of businesses.

We in industry know what we are building and what we want to provide to our customers, our users better than anyone else, and innovation is our modus operandi. To stay competitive, innovative and gain public trust in our AI-enabled product and/or service, we must put explicit guardrails in place. The existence of general practices

---

<sup>2</sup> European Commission. April 21, 2021. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence>

<sup>3</sup> Gonzalez vs. Google LLC Oral Argument. February 21, 2023. The Supreme Court of the United States, [https://www.supremecourt.gov/oral\\_arguments/audio/2022/21-1333](https://www.supremecourt.gov/oral_arguments/audio/2022/21-1333)

regarding privacy by design and security by design inspires a need to have ethics by design, critically placing ethics as the point of departure before anything has been designed and not the afterthought as a response to harm in the field.

As business leaders we can start this ethics-by-design journey by making the overt determination as a business to continue building and/or using this technology with honesty and transparency; these values keep our intentions clear and support a risk management approach in the case of negative outcomes. In practice, this can be in the manner of, for example, setting rules of engagement. To start, in the context of producing and using AI-generated content, such rules can be: (1) The technology is being used for tasks and is thus fact-checked, verified, and properly attributed. This supports a high level of human input and therefore significant curation and transformation of content generated. This further prevents an AI-enabled system from overruling human decision-making and creation, and thus preserving a high level of human autonomy. (2) Accountability is held for content created as a result of using the technology. This encourages active participation by all to be more cautious in what the business is willing to adopt, promote and deploy. This further motivates a critical eye to determining whether content is non-biased, non-discriminatory, and inclusive. These rules of engagement help us effectively reflect on the risks of developing and adopting an AI-enabled technology by considering the effects of its outcomes and the resulting harms it can cause.

To conclude, AI is not a computing problem with purely technical solutions; it is an interdisciplinary socio-technical problem that requires socio-technical solutions. If we do not want to remain a step behind the very autonomous AI-enabled systems we are creating, we have a moral duty to be knowledgeable of where we are and where we are going, to maintain curiosity, and to commit to ensuring that the technology we are developing and implementing now is done in a manner that benefits humanity today and tomorrow.