

The Branding Dilemma of AI: Steering Towards Efficient Regulation

Zeynep Engin, Data for Policy CIC and UCL Computer Science

Many experts in the AI field now realise that the term 'Artificial Intelligence,' originally chosen to define this rapidly evolving domain, unwittingly laid the groundwork for some of its most significant challenges. This early [nomenclature](#) misguided the conversation from the outset, setting questionable objectives for the AI community and shaping a [public perception](#) that strayed from reality. This foundational misstep has led to widespread misconceptions about AI's capabilities, characteristics, potential impact, and the regulatory expectations surrounding the technology.

After participating in two high-level panels on AI regulation in December, held at [TechUK's Digital Ethics Summit](#) in London and the United Nations [UNCTAD eWeek](#) in Geneva, I was inspired to further develop and clarify my thoughts on the subject through writing. Both sessions encompassed a range of governmental and multilateral initiatives, including the [G7 Leaders' Statement on the Hiroshima AI Process](#), the [Biden Administration's Executive Order on AI](#), the UK AI Safety Summit's [Bletchley Declaration](#), and the [EU's AI Act](#). This piece aims to crystallise and elaborate upon the perspectives I shared during these discussions, offering a deeper exploration of the broader challenges in AI regulation and inviting feedback on the insights I provided.

The AI conversation started on the wrong foot...

To recap, when John McCarthy coined "Artificial Intelligence" in the 1956 [Dartmouth Conference proposal](#), the field's goal was ambitious:

"The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer."

Fast forward 66 years, and ChatGPT came along. For the first time, significant milestones appear to have been achieved towards fulfilling the ambitions of the original Dartmouth summer project. Machines finally appear to be using language flexibly, producing creative material, and improving themselves. Similar developments are also happening in multimodal data processing and content creation - thanks to what we now call 'foundation models' or 'frontier AI'.

AI development¹ initially revolved around rule-based systems, where human programmers endeavoured to encode human reasoning into machines. This approach gradually evolved as the focus shifted towards enabling machines to 'learn' and 'understand' in a manner akin to humans. In the absence of any substantial understanding of the actual human cognitive capabilities, and constrained to only binary electric signals to produce anything useful from these systems, the best shot for the field was to try established statistical models to facilitate any 'machine learning' from data. More complex statistical optimisation schemes came next in the form of 'biologically inspired' neural network models. In popular media, AI was frequently depicted as humanoid robots, and conversational AI systems often used human pronouns in their interactions with human users. This progression was in sync with the field's original branding, which ambitiously aimed to achieve human-like intelligence in machines. Consequently, the public began to perceive these systems as increasingly human-like. They were viewed as autonomous 'intelligent' agents, capable of simulating human emotions and possibly attaining consciousness, even surpassing human intelligence and potentially dominating in future scenarios. The Turing Test² epitomised this phase, positing the ultimate AI achievement as the ability to make its responses indistinguishable from those of a human.

Undoubtedly, the term 'Artificial Intelligence' has captured the public imagination, proving to be an excellent choice from a marketing standpoint (particularly serving the marketing goals of big AI tech companies). However, this has not been without its drawbacks. The field has experienced several 'AI winters' when lofty promises failed to translate into real-world outcomes. More critically, this term has anthropomorphized what are, at their core, high-dimensional statistical optimization processes. Such representation has obscured their true nature and the extent of their potential. Moreover, as computing capacities have expanded exponentially, the ability of these systems to process large datasets quickly and precisely, identifying patterns autonomously, has often been misinterpreted as evidence of human-like or even superhuman intelligence. Consequently, AI systems have been elevated to almost mystical status, perceived as incomprehensible to humans and, thus, uncontrollable by humans.

Rebranding AI and setting up the 'right' objectives...

A profound shift in the discourse surrounding AI is urgently necessary. The quest to replicate or surpass human intelligence, while technologically fascinating, does not fully encapsulate the field's true essence and progress. Indeed, AI has seen significant advances, uncovering a vast array of functionalities. However, its core strength still lies in computational speed and precision – a mechanical prowess. The 'magic' of AI truly unfolds when this computational capacity intersects with the wealth of real-world data generated by human activities and the environment, transforming human directives into computational actions. Essentially, we are

¹ The 2023 Royal Institute Christmas Lectures by Michael Wooldridge, titled ['The Truth about AI'](#), provide further insights and positive examples related to AI's development for readers interested.

² Also popularly known as the 'Imitation Game', the Turing Test was proposed by Alan Turing in 1950 as a measure of a machine's ability to exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human. In this test, a human evaluator would engage in a natural language conversation with one human and one machine, both of which are hidden from view. If the evaluator cannot reliably tell the machine from the human, the machine is said to have passed the test. This test has been a fundamental benchmark for AI development, although it has also been subject to criticism for its focus on deception and imitation rather than genuine intelligence.

now outsourcing complex processing tasks to machines, moving beyond crafting bespoke solutions for each problem in favour of leveraging vast computational resources we³ have. This transition does not yield an 'artificial intelligence', but poses a new challenge to human intelligence in the knowledge creation cycle: the responsibility to formulate the 'right' questions⁴ and vigilantly monitor the outcomes of such intricate processing, ensuring the mitigation of any potential adverse impacts.

By moving away from the narrative of an 'uncontrollable mystical artificial intelligence,' we enter a more pragmatic realm. In this space, we can realistically assess the capabilities of these advanced machines at our disposal and explore effective ways to govern their development and application. Once we establish the fact that what we are dealing with is computational processing and advanced statistical optimisation, the technology itself becomes pretty controllable⁵. The true scope of regulation then becomes the interplay of human-data-machine interactions, particularly in socio-economic contexts.

We stand at a crossroads with this truly remarkable technological capacity, facing both historic opportunities and unprecedented risks. The path we choose now will shape whether these capabilities become transformative tools to innovatively tackle our most persistent problems or, conversely, exacerbate them. While these technologies hold the potential to contribute to a fairer and more sustainable world, realising this potential is far from automatic. It requires a dedicated commitment to careful oversight and responsible use. In the absence of these efforts, there is a significant risk that the trajectory of these technologies will lead us towards a more troubling reality. The path to positive outcomes is not the path of least resistance; rather, it demands continuous and conscious steering to ensure technology serves the greater good. A precise understanding and description of the technology, coupled with an accurate diagnosis of the challenges we face, are essential in setting a beneficial course.

The rebranding of AI as a computational and statistical tool can revolutionise our approach to global regulation. This perspective makes AI more understandable and manageable, shifting the regulatory focus to the effective translation of social objectives into technological applications. The true challenge lies not in comprehending AI's computational intricacies, but

³ While this article discusses the collective computational capacity at humanity's disposal, it's important to acknowledge that access to these resources is not uniformly distributed. The 'we' referred to is, in reality, often limited to entities with significant resources, such as large corporations or well-funded research institutions. This uneven access raises important questions about inclusivity and equity in the development and application of AI technologies, which deserves its standalone discussion.

⁴ The term 'right questions' in this context carries multiple interpretations. It encompasses the need for setting the appropriate scope in AI research, the thoughtful formulation and design of AI inputs and outputs to align with human and environmental welfare, and the imperative for users to ask informed questions while comprehensively understanding AI responses. Stefaan Verhulst, in their article ['Debate: ChatGPT reminds us why good questions matter'](#), elaborates further on these aspects.

⁵ When AI is recognized as a form of computational processing and statistical optimization, it becomes evident that its operations and outputs can be strategically directed, monitored, and adjusted according to specific parameters and objectives. This understanding emphasises the role of human decision-making in determining the scope and application of the technology. It underscores that humans bear the ultimate responsibility for guiding these computational tools and addressing any potential misuses or unintended consequences in socio-economic contexts.

in effectively governing its interactions with humans and the environment. This is where the real impact of AI is felt, and where regulatory efforts must be concentrated.

The future of AI should be rooted in complementing and augmenting human efforts, not competing with them. This is not just an ideal but a realistic trajectory for AI. It seems implausible that our collective goal would be to engineer something akin to Frankenstein's creature. Similarly, focusing our efforts on creating machines that rival human intelligence is as impractical as it is unnecessary. We have to steer the narrative towards more constructive ends. This redirection is imperative, particularly as current misconceptions pose significant barriers and obscure the true potential and purpose of AI - a technology that, when rightly applied, can serve as a powerful ally in our ongoing quest for knowledge and innovation.

Looking ahead to [Data for Policy 2024](#), we are setting a proactive agenda for 'trustworthiness' in AI-empowered human actions. Our focus is on the public sector and governance implications of data-driven and algorithmic decision-making. Join us at Imperial College London in July for a pivotal conversation on reshaping governance 'with' AI.

Acknowledgements: *I extend my sincere thanks to David Hand, Jon Crowcroft, and Stefaan Verhulst for their valuable feedback and insightful reviews of this piece prior to publication.*

About the author: *Dr. Zeynep Engin is the Chair and Director of Data for Policy CIC and an Editor-in-Chief of the Data & Policy journal. She is also affiliated with UCL Computer Science as a Senior Researcher.*

DOI: 10.5281/zenodo.10462686

This article is to be published in the Data & Policy Blog, and made available on Zenodo open-access repository – January 2024.