COUNCIL ON
EXTENDED INTELLIGENCE

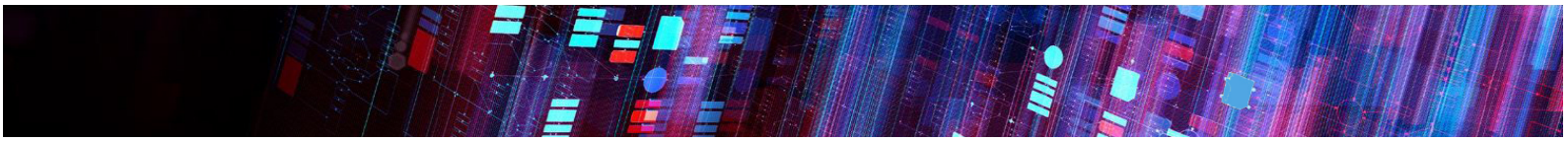
Human Intelligence and Autonomy in the Era of 'Extended Intelligence'

By Konstantinos Karachalios and Joi Ito

The last few years have seen a surge in marketing hype around artificial intelligence, or AI. Beyond its inherent vagueness, the term suggests the emergence of a new kind of intelligence, a new, *sui generis* ontological entity created by humans but somehow outside of our control.[1] To address such overwrought characterization, MIT Media Lab has introduced the concept of "extended intelligence"[2] and IEEE has established a program called the Global Initiative on Ethics of *Autonomous and Intelligent Systems* (note the emphasis) to address the complex dilemmas and issues posed by the increasing use of advanced computational systems.[3] Even these characterizations are, of course, a compromise because if it is not possible to exactly define "intelligence" and "autonomy", even for humans, how can we use these terms for machines? Intelligence and autonomy are *phenomena* of human activity with dimensions that go beyond what can be captured by the reductionist methods used to establish logical/scientific frameworks. Therefore, we should resist the temptation to oversimplify when trying to explain them.

Sometimes, the origin of a word reveals what it *means*. Etymologically, "autonomous" means a person who or society that has the desire and the capacity to define the *nómos* (law) that will guide future actions. This is not a trivial feat, as the historic and present reality of authoritarian systems demonstrate. At a deeper level, genuine autonomy is inherently associated with the possibility of "free will" (at an individual or collective level) and thus with the assumption of a

[1]This is by no means a novel thought; as an example, Heidegger, in his famous interview published in *Der Spiegel* 10 years after his death, said that "Technik" is something that can no longer be mastered by humans http://wfgw.diemorgengab.at/tzn200301.htm
[2]See https://www.media.mit.edu/projects/social-ai-and-extended-intelligence/overview/
[3]See https://standards.ieee.org/industry-connections/ec/autonomous-systems.html



non-deterministic context for our existence. It is precisely because we pre-suppose a kind of agency over our actions that we hold ourselves responsible for them. This assumption of agency makes all the difference between an autonomous entity and an *autómaton*, something that can move by itself once set in motion by a trigger, but that can never become autonomous in the original meaning of the term.

The origin and manifestation of "intelligence" are even more complex, both etymologically and conceptually. Simply put, one could say that—among several other possible manifestations—something unique to human intelligence is our capacity to pose *dilemmas*, beyond merely posing or solving *problems.* Moreover, since a dilemma is often related to a self-imposed need to choose between two frameworks of contradicting "laws" (and in most cases, paying a price for any choice), there seems to be an inherent link between autonomy and intelligence. At their culmination, both manifest themselves as a desire to transcend incumbent structures of power and thus rely primarily on courage, not on IQ.

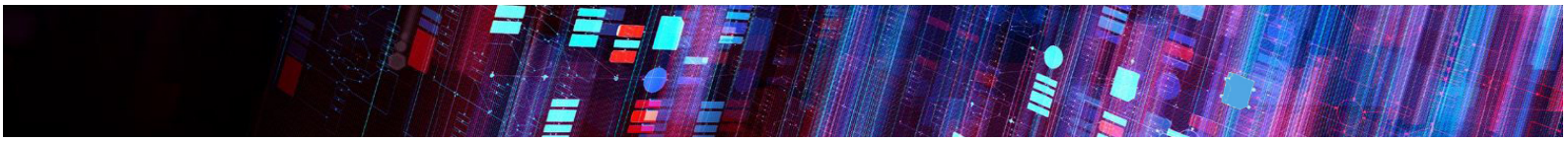
No machine can ever be autonomous or intelligent in such a way because a machine does not have any skin in any such game. Speaking of machine autonomy and intelligence thus runs the risk of reducing autonomy and intelligence and therefore the *human condition* itself to processes and systems that are "closed." Where, miraculously, everything exists from the beginning and nothing gets added or lost[4], and where all processes are completely determined by their initial and ongoing boundary conditions (including processes that appear at first sight stochastic or random). Schopenhauer extended this logic to humans and concluded that in a deterministic context, *"[m]an can do what he wants, but he cannot will what he wills."*[5]

However, now and then, there are situations where humans seem able to *"will what we will,"* where free will appears to manifest itself, opening up spaces of autonomy. Although it is impossible to know for certain, we should leave the possibility open that we are capable of *gnōsis*, an intuitively felt knowledge that gives rise to our intentions. As Merleau-Ponty said, *"[i]n so far as I have hands, feet, a body, I sustain around me intentions which are not dependent upon my decisions and which affect my surroundings in a way which I do not choose. These intentions are general ... they originate from other than myself, and I am not surprised to find them in all psycho-physical subjects organized as I am."*[6] We may agree that absent a fundamental breakthrough, machines can gather, store, and process information but cannot develop intentions based on such *gnōsis*.

So, one valid approach would be to strictly reserve the use of the concepts of "intelligence" and "autonomy" for manifestations of human activities. For the current discourse and as a

---

[4]See Carl Popper's analysis of Parmenides' framing of the Occidental sciences in *The World of Parmenides: Essays on the Presocratic Enlightenment*, Rutledge, 1998.
[5]Einstein's translation in his *Credo* (https://www.einstein-website.de/z_biography/credo.html) of Arthur Schopenhauer, who said *"Der Mensch kann tun was er will; er kann aber nicht wollen was er will";* Über die Freiheit des menschlichen Willens (1839).
[6]Maurice Merleau-Ponty, Phenomenology of Perception, trans. C. Smith, Routledge & Kegan Paul, 1962.



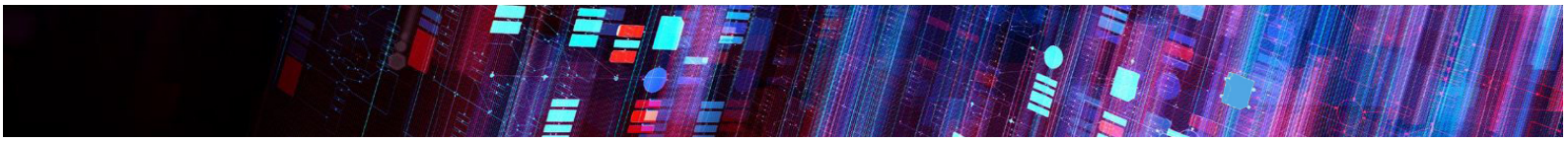
compromise, however, one could limit the scope of the term "intelligent" to computational systems using algorithms and data to address complex problems and situations, including the capability of improving their performance based on evaluating previous decisions. Such systems could be regarded also as "autonomous" in a given domain as long as they are capable of accomplishing their tasks despite environment changes within the given domain.[7] It is in this narrow sense that we will be using these terms in the remainder of this article.

As the technology advances, we could perhaps even *"[t]hink of the AI and the human organization as a single continuous network of facilities. When you do that, what you've done is you've added new capabilities to the organization. AI's are just 'capabilities' or 'actors' with particular characteristics. They're employed as appropriate."*[8]

The MIT Media Lab has been developing the concept of "extended intelligence" in order to capture the dynamics of this emerging "continuous network of facilities" and its diverse "actors," while avoiding the fallacies of the AI framing. In a strictly anthropocentric assumption, human intelligence is at the center of and is supported by machine capabilities in an effort to make sense of a "dumb" *cósmos* around us. Humans also use machines to facilitate practical problem solving such as pattern recognition, acceleration and optimization of decision making, or automation of self-driving vehicles. This "continuous network of facilities," made up of humans and machines working together, could be considered a manifestation of a simple form of "extended intelligence." If we accept, however, that there can be more than one source of intelligence (beyond our own, individual or collective) that has an impact on humanity, this would result in a more complex system, where our human intelligence may not always be at the center.[9] Accepting this may be as difficult a process as the transition from the Ptolemaic planetary system to the Copernican.

According to the MIT Media Lab's evolving conceptual framework:
"*Instead of thinking about machine intelligence in terms of humans vs. machines, we should consider the system that integrates humans and machines—not artificial intelligence, but* ***extended intelligence****. Instead of trying to control or design or even understand systems, it is more important to design systems that participate as responsible, aware and robust elements of even more complex systems. And we must question and adapt our own purpose and*

[7]This description will be used by The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems in *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, First Edition. IEEE (to be published in 2019).
[8]Alex 'Sandy' Pentland as quoted from, *The Case for Extended Intelligence: Technological Advancement in Service of People and Planet.* The Council on Extended Intelligence, February 2019.
[9]Amandeep Gill as quoted in: *The Case for Extended Intelligence: Technological Advancement in Service of People and Planet.* The Council on Extended Intelligence, February 2019. "*From an Indian perspective, intelligence is always XI [extended intelligence]. It's an extension of divine intelligence. One person doesn't have IP over anything. Intelligence resides about two inches above the head."*



sensibilities as designers and components of the system for a much more humble approach: Humility over Control."[10]

This obviously goes beyond the simple form of extended intelligence mentioned above. In a kind of syncretistic approach, human intelligence and its machinic manifestations are thought to be embedded in a broader resonant space, an "ambient intelligence," where our natural environment is not a passive recipient but rather an active resonator and producer of intelligence itself.

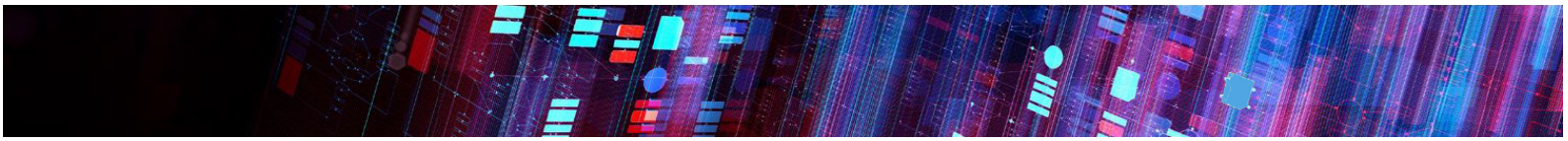
Of course, this "ambient intelligence" is not a novel concept. It has been shaped and explored for millennia by philosophical traditions and mythological narratives throughout the world.[11] For example, one function of the *xapiripë*, 'animal ancestors' or 'shamanic spirits' in the mythology of the Yanomami, is to augment human intelligence: *"Thus, the concept of xapiripë signals a complex interference, a chiastic distribution of identity and difference between the dimensions of 'animality' (yaro pë) and 'humanity' (yanomae thëpë)… 'it is the words of the xapiripë which augment our thoughts.'"*[12]

What is new is the question of whether human artifacts, such as intelligent machines and systems, have the necessary sophistication and quality to be added to the mix. Since such technical systems are playing a rapidly growing role in mediating between humans and our environment, this is a valid question. At least at a symbolic level, the proliferation of "smart X's" fueled by technology (e.g., phones, houses, cars, clothes, even cities) and the immense hopes and promises associated with such technology seem to emulate rather than negate the imaginaries of animistic religions.

A further fundamental idea associated with "extended intelligence" is to admit the inherent complexity of such systems, and limit the appetite for control to features that are necessary for our safety and security. Doing so avoids the rigidity of "optimizations" according to what Joi calls "single currencies" (usually quantities such as quarterly RoI, GDP, etc.) and the instability risks associated therewith. Goodhart's law states that when a measure becomes a target, it ceases to be a good measure. Economists use Goodhart's law to explain that any measurement used to measure the economy ceases to be a good measure because everyone will immediately optimize for it, abrogating its objectiveness. Joi argues that measuring things through purely financial measures is also "a measure" for which we have optimized, making it an incomplete measure of value. As a Wall Street banker might say, "How can he be smart? He's not rich."

---

[10]Joichi Ito, "Resisting Reduction: A Manifesto".
[11]For an example from the Western tradition, see James Lovelock, *Gaia: A New Look at Life on Earth*, Oxford University Press, 2016.
[12]The Crystal Forest: Notes on the Ontology of Amazonian Spirits; Eduardo Viveiros de Castro; Inner Asia, Vol. 9, No. 2, Special Issue: Perspectivism (2007), pp. 153-172.



In conclusion, our discourse around extended intelligence cannot aim to deliver an all-encompassing, perfect definition of intelligence and autonomy, either for humans and/or machines. Still, when one uses these words to characterize non-human entities and contexts, one should at least explain what is meant, including limits and restrictions. In addition, the techno-scientific communities should accept that we may not be able to explain everything with scientific methods. To invoke Husserl, we should not confuse nature with our mathematical models about it.[13] Instead, we should open ourselves to the possibility that there is mystery in life—and perhaps even more so in human existence—and that machines can never be part of this.

[13]"Für wahres Sein nehmen, was eine Methode ist;" *The crisis of European sciences and transcendental philosophy. An introduction to phenomenology.* The Hague, Netherlands: Martinus Nijhoff, 1976.