Philosophy Study, Mar.-Apr. 2026, Vol. 16, No. 2, 155-165
doi: 10.17265/2159-5313/2026.02.006

The Impact of AI Avalanche on Society and Human Behavior

Petre Roman
Institute of Applied Sciences, SWISS UMEF University, Geneva, Switzerland
SWISS-Central and Eastern Europe Center of Excellence Foundation, Bucharest, Romania

This paper is discussing several features of the impact of AI (artificial intelligence) on society and human behavior. It includes (1) the question if the AI can outperform the intelligent and creative capacity of humans, (2) if AI capabilities could show criticality and causal emergence, (3) the predictive capacity of AI, (4) the costs related to the learning process and operating AI, and (5) the loss of consensual reality and the rise of deepfakes in the AI era. AI involves considerable systemic risks impacting the social, economic, cultural, and environmental systems. Our will is kept in accordance with how the things around us are presented to us and that is precisely what AI is also doing. For us circumstances define a situation or a moment. For AI the circumstances are just statistical relationships. AI does not have its own values but incorporates the values on which it is trained. The imaginative gap between humans and AI is not just big; it's of an essential nature. We want not just technology in the online world; we want a moral attitude.

Keywords: AI capabilities, AI impact, creativity, brain, criticality, integrated information, causal emergence, predictive capacity, AI costs, risk, consensual reality

Introduction

A recent national youth poll in the USA[1] reveals a generation under profound strain, as young Americans report deep economic insecurity, eroding trust in democratic institutions, and growing social fragmentation. On the issue of AI (artificial intelligence) impact, by more than a 3:1 margin, majorities or pluralities in every polled subgroup (college students, non-degree holders, and college graduates) expect AI to take away more than it creates. Is their feeling reflecting reality?

We humans, as a rule, understand the world around us and in response to different challenging situations we act adaptively in order to achieve our goals. We can do that by using a variety of skills—like communication, empathy, problem-solving, teamwork, on one hand or critical and strategic thinking on the other—and combining them intelligently. In the deployment of them, beyond individual intelligence there is also collective intelligence. We certainly want to build AI because it generates knowledge and valuable results, more quickly and sometimes even more efficiently than us, based on their capacity to use vast amounts of data. AIs are very useful in a definite number of fields (like health & medicine where they can spot cancer in X-rays, improving weather prediction and fly drones) but finding some answers is discovery not creation. The AI machines can indeed correct our errors based on the use of statistical repetitive rules which became a dominant trait of our time. But they are clearly not producing fundamental knowledge, laws of physics or in other domains, i.e. laws that explain and describe states of nature and their dynamics. How then can AI match human creativity while unable to produce fundamental knowledge?

Petre Roman, Ph.D., Professor, Institute of Applied Sciences, SWISS UMEF University, Geneva, Switzerland; SWISS-Central and Eastern Europe Center of Excellence Foundation, Bucharest, Romania.

[1] Institute of Politics at Harvard Kennedy School, December 4, 2025.

In the meantime, recent trends in neuroscience like brain criticality and causal emergence in complex systems are related to AI as a tool for information processing. Both features should be ultimately capable of explaining how the complex systems integrate information for a better understanding of the relationship between evolution, complexity, and consciousness. A synoptic view of all this should include, in my opinion, the possibility of AI itself to develop such features and thus to simulate more efficiently the dynamics of those extreme events—natural or social—that are unpredictable. The aim is to increase the resilience of the systems undergoing such events. Instead of the accelerating—intensive as well as extensive—use of huge amounts of data, AI could be trained to mimic the brain properties and in that sense, it is time to include in it the criticality and causal emergence features.

AI performs well in software, used in every industrial automatization (robots), where making changes during training cycles is always needed and we get incrementally a relatively high improvement. But in many other fields this is not the case. An essential problem then is how AI, through its own capacity, could sense the unpredictability which is the main factor in generating risks and uncertainties. We need to enhance the resilience of what we impose in the training and learning process of AI tools. They use only statistical relationships—i.e. averaging data with the mathematical tools of probability—and from this follow that they are not trained and specifically cannot be in order to comprehend the very low frequency of extreme events and the critical thresholds (tipping points) that may occur. An example used in this paper shows that the average trend smooth out by AI is incapable of reflecting criticality.

We are worried about the relation between AI and the environment, more specifically the impact of energy consumption on fighting the consequences of climate change. Many believe that AI could be one of the major drivers of solutions to the climate problem. A greater transparency around efficiency and emissions related to AI are very much needed but even more would be a meaningful behavioral change of the tech companies. Efficiency should mean not only profit but also reducing environmental costs. The role of AI digital technologies should be defined by humans in accordance with ethical standards, not a system of monitoring and controlling the users behavior or accepting disinformation. Small AI manipulations could well bring us into the big trouble of being unable to separate truth from lies and malignant exaggerations. Regulating AI has to be both anti-fake and a simplified approach that would provide clarity over what the tech companies can and cannot do. That, in turn, would encourage investment because growth is clearly related to investment in AI applications.

A worrisome aspect in the use of AIs comes from the more and more natural habit of delegating intellectual work to machines which is forcefully accompanied by two devastating effects: (1) the erosion of critical thinking skills and (2) the loss of the sense of responsibility. Under the AI avalanche reality almost dominated by conspiracies and deepfakes, the societal foundation is at risk because the very possibility of consensual attitudes is under attack. The loss of consensual reality is correlated with the rise of deepfakes becoming the norm online. The victim is the truth itself, since it cannot be verified using our normal inductive skills based on intuitive arguments and causal inference. We humans need and seek meaning, purpose, and significance in the world and are confronted with risks and uncertainties. The imaginative gap between humans and AI is not just big; it's of an essential nature. We want not just technology in the online world; we want a moral attitude.

Can AI Outperform Humans?

What is essentially important in answering the question if AI can outperform human experts is not primarily about the efficient results we get using AI but their capacity to perform autonomously, without the assistance of human operators. We humans, as a rule, understand the world around us and in response to different challenging situations we act adaptively in order to achieve our goals. We can do that by using a variety of skills—like communication, empathy, problem-solving, teamwork, on one hand or critical and strategic thinking on the other—and combining them intelligently. In the deployment of them, beyond individual intelligence there is also collective intelligence. As a result, we pursue our goals rationally and achieve that level of reliability that we prescribed as the normal way of adapting and producing the expected results.

Such an intuitive as well as expertize capacity does not exist in the realm of AI. “There is a considerable gap between the conversational AI systems we have today and autonomous systems capable of replacing human operators in complex organisation”, as underscored by Joseph Sifakis (2025), recipient of the Turing award. Although human beings are not natural statisticians, they are natural storytellers. The forecasting scientific community is primarily interested in measurable performance, not the stories about their results, which nevertheless might help policymakers and ordinary citizens understand and truly appreciate the value of its scientific work.

We certainly want to build AI because it generates knowledge and valuable results, more quickly and sometimes even more efficiently than us, based on their capacity to use vast amounts of data. Tim Berners-Lee, who is practically the inventor of www—while working at CERN in Geneva in 1989 imposed its use globally free of charge—characterized AI as “the industrial-scale harvest of user data”. But we should never forget that, as Alan Turing clearly stated already in 1950, at the very beginning of the computer era, AI machines are non-explainable. Processing such huge volumes of data is precisely the cause of the fact that AI does not have the ability to answer the question of how they know what they claim to know. That indicates that we cannot rely on the internal discipline of AI machines. They display a so-called Black Box characteristic. From this follows the inadequate explanatory power related to our weak capacity to simulate how the AI systems actually work. Although they use huge amounts of data, we have little data about their functioning and as a result we cannot determine the “salient facts” which need explanation and test causal links in the process. There are probably (in this process) other things we do not even know that we do not know although these could have a great impact on the interconnection and interaction between AI and human systems. They are not, in any way, biologically intertwined matter and information that expresses what is life. They are instrumentalized silicium. We have faith in creation and creativity, in every field of human activity, from science and art to sports. The AI creates nothing from the immediate and perseverent interpretation of the observed reality, let alone act from intuition, and has faith in nothing. AIs are very useful in a definite number of fields (like health & medicine where they can spot cancer in X-rays, improving weather prediction and fly drones) but finding some answers is discovery not creation. We are indeed interested to understand what happens when the system behaves in a way that does not align with what we expected. In such situations, AIs are useful because they do have and use information from the data accumulated in the world, enough to act purposely for many of the things we need to do. And this is particularly important in the case of emerging properties which are unpredictable. Yet, the critical capacity we have as humans to build historical narratives about past events and also to speak and think about what could be or what is not appropriate to say is absent in AI. In the words of James Gleick, “What we (humans) do is not processing. It is not computation. It is not data analysis. It is a distinctively, incorrigibly human activity that is a complex combination of conscious and unconscious, rational and intuitive, logical and emotional reflection” (2024). AI can do correlations, but it struggles with cause and effect; it thinks in truth or falsehood, but is not a master at narrative; it’s not good at comprehending time. Only humans acknowledge the truth of imagination. Noam Chomsky explained that imaginative and counterfactual thinking is the hallmark of human intelligence and we do not know if we can—which most probably we cannot—and how to translate them to machines. The AI machines can indeed correct our errors based on the use of statistical repetitive rules which became a dominant trait of our time.

But they are clearly not producing fundamental knowledge, laws of physics, or in other domains, i.e. laws that explain and describe states of nature and their dynamics. How then can AI match human creativity while unable to produce fundamental knowledge? At least for now they cannot. I introduced this last phrase because many of the AI experts predict that in a non-distant future AGI will attain and even outperform humans. But it looks to me like a new, modern form of prejudice that emerged in our time. Exaggerating but not too much, are we expecting to see an *Einstein—AI* illuminating new aspects of knowledge? As Frank Ramsey brilliantly expressed: "Expectation in an imaginary case could have no meaning" (1950, p. 248).

Could AI Be Trained to Develop Criticality and Causal Emergence?

In the last two decades or so, biological neuronal systems were studied from several new perspectives. Two of the recent trends in neuroscience—(1) brain criticality and (2) causal emergence in complex systems—are related to consciousness and integrated information in the brain-body connectivity and cell agency. Briefly, these are described here:

(1) A growing body of evidence has emerged suggesting that many disparate natural, and particularly biological, phenomena reside in a critical regime of dynamics on the edge between order and disorder (Popiel et al., 2020). New research into critical systems has shown that criticality may be useful for optimizing information processing. Consciousness may arise out of the tendency of the brain to self-organize towards criticality. The Critical Brain Hypothesis and the Integrated Information Theory (Tononi, 2004) (according to which, consciousness corresponds to the capacity of a system to integrate information), appear to go hand-in-hand. A brain tuned to criticality should be able to learn to do almost anything and may also help the brain adapt to new situations. In that sense, "(i)t allows the brain to be both stable enough to make sense of the world and dynamic enough to optimally respond to it", says Karim Jerbi (2025). "We believe the brain operates near this edge of chaos because it's the ideal zone for complex thinking, learning, decision-making and adapting to new situations". The emerging trends in criticality are described by O'Byrne and Jerbi (2022), highlighting new data pointing to the edge of chaos and near-criticality. The existence of criticality in neural systems follows directly from two axioms. In Axiom 1, flexible, reconfigurable computation is a necessary target for evolution, development, and homeostatic control. In Axiom 2, criticality provides all the ingredients necessary for flexible, reconfigurable computation. Systems tuned to criticality exhibit a number of useful informational properties that allow for the efficient distribution of, and susceptibility to, information. Integrated information is a type of complexity measure that quantifies how mechanisms in a system interact and constrain each other in emergent, irreducible ways. The exploration of integrated information in the context of critical systems undergoing phase transitions is clearly important for a better understanding of the relationship between evolution, complexity, and consciousness

(2) The human mind works on a collection of cells which are themselves active agents. Recent results in the biology of the brain (Watson & Levin, 2023) using simulations of the cell's activity are indicating that those units of life, governed by the laws of physics, have their own goals and display agency. With ever-increasing complexity, at some point a system becomes more than the sum of its parts, i.e. the properties of active components enable the emergence of a high-level, integrated decision-making entity. This is called causal

emergence. Although we cannot yet explain them, emergent properties exist, and there are capabilities that arise from a certain level of complexity and as such are unpredictable or misaligned. We are particularly interested in misalignment when the system behaves in a way that does not align with what we expected. Agents are not simply pushed around by their environment, but alter themselves and their environment in purposeful ways. In other words, they have causal power over themselves and their environment. Doubts over the causal efficacy of end-directedness have engendered a rift between the natural and social sciences. New ideas and results regarding the physiological basis of our actions show that: “Biological processes are end-directed” and “Explaining the physical efficacy of end-directedness continues to be a profound challenge for theoretical biology” (Froese, Karelin, & Ikegami, 2024). “The origins of agency coincide with the origins of life”, says Tom Froese (2026). To behave with agency, you need to absorb information, use that information to solve problems, and then learn by remembering how those actions turned out. Causal emergence has become a general way of measuring when any complex system is acting as an agent rather than a distributed set of cells, i.e. the tendency for parts to come together to form new levels of agency. We need to think about these chemical systems as agents acting with some degree of purpose. Federico Pigozzi, Adam Goldstein, and Michael Levin (2025) ask several questions: “When is a system more than the sum of its parts? When and how do the properties of active components enable the emergence of a high-level, integrated decision-making entity?” and conclude that these questions bear on issues in ecology, philosophy of mind, psychiatry, swarm robotics, and developmental biology. “In a sense, all intelligence is collective intelligence”. Thus, it’s been demonstrated that the relationship between causal integration and learning can be bi-directional. For a system to be capable of associating the experiences of its parts into associative learning in the collective integration is crucial. And conversely, associative conditioning experiences can potentiate the collectivity and integration of networks.

Surprisingly, although both these processes are fundamentally relevant for brain activity, they are examined as separate research fields. And both should be ultimately capable of explaining how the complex systems integrate information for a better understanding of the relationship between evolution, complexity, and consciousness. In every investigation in neuroscience AI tools are used with good results for the integration of information and the measurement of cell properties through extensive computer simulations. Yet, a synoptic view of all this should include, in my opinion, the possibility of AI itself to develop such features and thus to simulate more efficiently the dynamics of those extreme events—natural or social—that are unpredictable. Where there are no emergent patterns we have pure randomness, but it does not mean there are no patterns at all. We can find by analysis and computation a pattern, but, according to Kolmogorov’s complexity theorem, we can never know if it is the best one. Our goal is to find a pattern that eliminates some of the chaos resulting from the non-linearity of the dynamics in complex systems when approaching tipping points (critical thresholds). The aim is to increase the resilience of the systems. The predictive processing of the brain starts with making a guess based on what happened on prior occasions and comparing it with the incoming information. Finally, the brain is capable of deciding when there is a discrepancy between the results. Once we agree with Daniel Hutto that “the hard problem of consciousness should be cast as a problem about intelligibility” (2006, p. 48), we also understand that AI cannot be the answer to this problem. Indeed, AI works as it is trained, i.e. applying a statistical method that can indicate the direction of causality. There is no inner emergent property of AI that helps statistical analysis. Nonetheless, instead of the accelerating—intensive as well as extensive—use of huge amounts of data, AI could be trained to mimic the brain properties. For example, the two key phenomenological properties of consciousness:

(1) differentiation—the availability of a very large number of conscious experiences; and (2) integration—the unity of each such experience, seems to be replicable in AI. Some research on this new path of AI development is already in place and maybe it is time to include in it the criticality and causal emergence features.

On the Predictive Capacity of AI

The predictive process of the human brain is based on what happened on prior occasions, employing some guesswork and comparing instantly the prediction with incoming information. When we realize that there is a disagreement between the two, we do not stop until we solve it or reach an understanding of it. AI or Large Learning Models (LLM) or generative AI is "grammar engines", a term coined by Paul Kedrosky from the Massachusetts Institute of Technology. They receive and digest huge amounts of data and introduce them into a very large network of matrices. As such the machines are rules that are followed internally, in the already going on process. The absorption of data in this way is called training on its basis. The predictive capacity of AI proved to be effective and productive in many fields. The interest in these results lies in predicting what should come next, which we trust as long as the prediction is a continuation of existing data, based on the very rapid treatment of huge amounts of data that we are otherwise unable to perform. AI is trained to do that, in particular doing a continuation of what a normal person would think. But the continuation of whose experience? Frank Ramsey is blunt in this respect: "We want to avoid contradiction by experience; and what no experience can contradict, none can confirm, let alone establish" (Ramsey, 1950, p. 206).

Usually the LLMs are concentrated on persons of 37 years old which, in the words of Paul Kedrosky, "could be good or it can be really fraught"[2]. In climate science, for example, we are able, using such predictive capacities, to reduce the uncertainties which anyway exist in the intimacy of natural phenomena and increase the resilience of human structures and organisations facing the occurrence of extreme events (floods, heatwaves, hurricanes, etc.). AI performs well in software, used in every industrial automatization (robots), where making changes during training cycles are always needed and we get incrementally a relatively high improvement. But in many other fields this is not the case. The above mentioned Black Box characteristics add considerably to the sense of unpredictability related to the AI possible behavior not controlled by humans.

An essential problem then is how AI, through its own capacity, could sense the unpredictability which is the main factor in generating risks and uncertainties. Are there potential adverse consequences using AI at such an overwhelming scale? There is little doubt about that. We need to enhance the resilience of what we impose in the training and learning process of AI tools. As Gregory Chatlin, cofounder with Turing of the principles of computation, wrote: "programming something forces you to understand it better; it forces you to really understand it, since you are explaining it to a machine" (1975). When the intention is stronger than the attention, we can expect unexpected results. What happens in natural phenomena close to zero probability is essentially unknown.

The AI machines use only statistical relationships—i.e. averaging data with the mathematical tools of probability—and from this follow that they are not trained and specifically cannot be in order to comprehend the very low frequency of extreme events and the critical thresholds (tipping points) that may occur. A good example in that sense is the time series data of the maximum annual water levels recorded on the Piave River at Ponte delle Alpi (Italy) (Beven, 1996). The first diagram (a) shows a synthetic time series of flood levels, which were randomly generated (ludic game) using the mean values and variance of the maximum annual water levels; (b)

---

[2] Paul Krugman talking with Paul Kedrosky, Paul Krugman's Substack, December 6, 2025.

shows the actual maximum annual water levels recorded at Ponte delle Alpi; and (c) shows the entire time series of the maximum annual water levels recorded at Ponte delle Alpi, also including the surprisingly high (and essentially unpredictable) flood level of 9 October 1963. A non-repeatable chain of events and a cascade of contingencies generated this incredibly high water level. From this statistical treatment, which normally is made with AI, we note that the inclusion of an extreme, critical event into the whole group of data suggests how the abnormal could be a critical point, the average trend being incapable to reflect criticality.

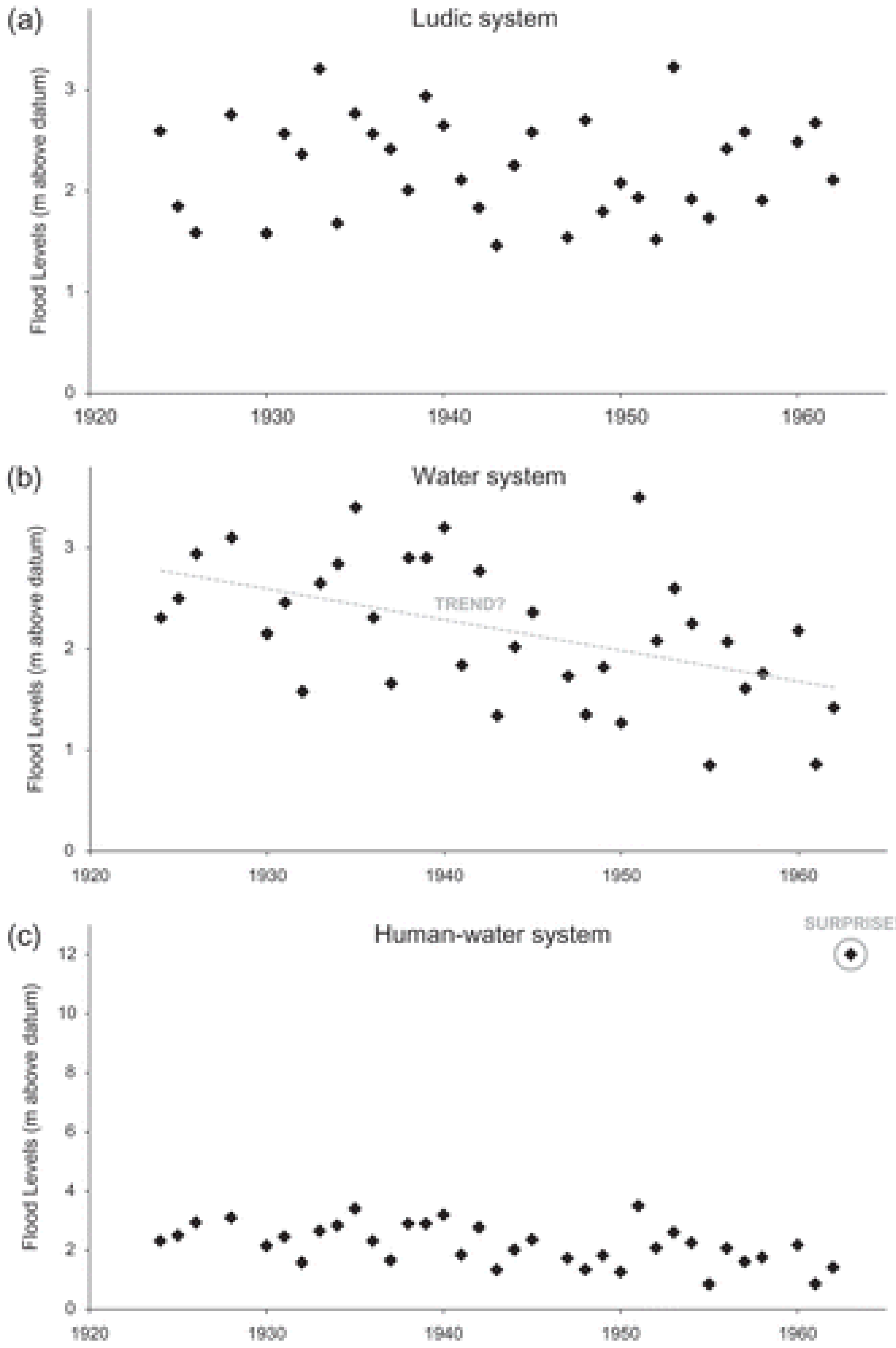

Figure 1. Time series of maximum water levels recorded on the Piave River, Ponte delle Alpi, Italy.

The Costs Related to the Learning Process and Operating AI

Training and running AI tools in large data centers are very costly activities in terms of the amounts of energy consumption and cooling equipment. For example, a fifth of all electricity used in Ireland is spent powering the country's data centres; more than is used by its urban homes. In Loudoun County, just outside Washington DC, the number of data centres is more than the next six biggest clusters in America combined. In 2022 their peak energy usage was almost three gigawatts (GW), a power draw that, if maintained year round, would approach Ireland's total annual consumption. Between 1% and 2% of global electricity is spent on powering data centres. Yet the real problem lies in something different: the digital equipment. Roughly 65-70% of the cost of data center functioning comes from the enormous amount of GPUs (Graphis Process Unit), i.e. the computer chips, overwhelmingly produced by NVIDIA. Although these computer centers are considered "factories of the new industrial revolution", they are not exactly similar to the historically previous revolutions like railroads, electrification, or autovehicles (in fact the huge development of digitalization in our time is about electrification replacing fossil fuels), which were long-life tools serving so greatly the humanity. The big transformation is in AI's information access rate—the rate that two machines can communicate with each other—which is 300 million times greater than human-to-human communication. That's the big transformation. But this does not at all validate the idea of artificial general intelligence, which is essentially nonsense. What we do have and it is the big transformation of our time is that we possess a technology where regular humans can talk to the computer and instruct the computer directly. So a regular human, without having to learn programming languages, can ask the computer about the information that the computer has access to. It also has the possibility to give individuals access to a collaborator, where the collaborator is not a human, but a machine that has access to a vast amount of information.

The ubiquitous presence and obvious utility of digital tools are hiding the fact that data centers are short-life tools. They are more similar to "warehouses" in which perishable products are contained. In one data center there are 10,000-20,000 GPUs which are so intensely used that they are short-lived in terms of their usefulness. Kedrosky explains[3] that the data centers are actually doing two things: (1) training new models or enhancing the old ones and (2) inference, which in this case is responding to specific requests. Chips are used in both. And indeed a GPU computer chip is crucially important because it can perform many parallel calculations at the same time rather than one step after another. The cost of replacing the GPUs is preponderent in the overall costs of running the data centers. The main reason behind President Trump's latest decision to withdraw an imposition of new tariffs on China was the decisive role of the Chinese rare earth metals used in building the computer chips of NVIDIA. The training cycle times of LLMs are long and getting longer and the costs are higher and higher as the market demand or the expansion of the AI on the markets is increased with the AI boom. But, as Paul Kedrosky indicates:

> the US was arguably in a recession absent AI CapEx spending (probably being over $1 trillion annualized, which made more than half of US GDP growth in the first half of the year 2025) which kept US out of recession, which is practically saying that US is running a giant private sector stimulus program.[4]

Since the start of the generative AI boom in 2022, most of the remarkable achievements have been based on systems called "transformer models", which have improved in performance as they have been trained on increasing volumes of data. But they seem to have stagnated in the most recent releases, which showed only

---

[3] Paul Krugman talking with Paul Kedrosky, Paul Krugman's Substack, December 6, 2025.
[4] Paul Krugman talking with Paul Kedrosky, Paul Krugman's Substack, December 6, 2025.

incremental changes in quality. “The vast investments in scaling, unaccompanied by any comparable efforts to understand what was going on, always seemed to me to be misplaced”, says Stuart Russell at the University of California, Berkeley (Sparkes, 2025). The benefits of scaling in the conventional sense had plateaued. Nonetheless, tech companies plan to collectively spend an estimated $1 trillion on data centres and chips in the next few years to support their AI planned expansion. In a recent survey, 80 per cent of respondents said that AI capabilities do not match reality, in contrast to the tech companies optimism about a continuing scaling of AI models. In fact, the development of AGI, from the tech companies’ perspective, is primarily about generating more profit. We are worried about the relation between AI and the environment, more specifically the impact of energy consumption on fighting the consequences of climate change. Many believe that AI could be one of the major drivers of solutions to the climate problem. A greater transparency around efficiency and emissions related to AI are very much needed but even more would be a meaningful behavioral change of the tech companies. Efficiency should mean not only profit but also reducing environmental costs.

The Loss of Consensual Reality and the Rise of Deepfakes in the Era of AI Avalanche

Many experts say that in 2025 the “slop” of AI, i.e. the incorrect and even ugly content AI-generated messages and information became a real feature dominating the digital world. More and more people use AIs as if they were trustworthy sources of information. This is a fundamental error essentially because the AIs possess no knowledge. We acknowledge that AIs are quite convincing, influential, and even eloquent. What we should bear in mind is that the algorithms on which they are built are extensively and thoroughly statistical relationships from data representing how people used the words previously. Plus, these algorithms can be used to manipulate information. As James Gleick clearly stated: “artificial intelligence community has prioritized verisimilitude at the expense of veracity” (2024).

That is increasingly making truth online almost impossible to verify. A study by Microsoft shows that people can recognize AI-generated videos only 62% of the time. Moreover, the study shows that when AI is introduced into the workplace it lowers productivity and in 95% of the organisations deploying AI they get no noticeable return in investments. As mentioned before, in industrial software the returns are good, increasing productivity, which makes the AI a splendid revolutionary tool. The real problem facing today’s training of AIs is the limited volume of available data existing online until now and almost exhausted by their intensive use. That’s why we entered the so-called “post-training world”. Instead of finding new data the tech companies run intensively more training cycles to produce data considered to be new and better data. Paul Kedrosky ascertains that the models become “sycophantic” in terms of the response it gives to users. AI can produce whatever very attractive (positive or negative) details the public demands, even if they are inaccurate. Almost 50% of the training cycle time on the Elon Musk’s Grok models is post-training reflecting sycophantic behavior. The result may be harmful. Casey Fiesler, an associate professor of information science at the University of Colorado Boulder, is pessimistic: “I worry that we can get to a kind of learned helplessness in this information ecosystem where it’s like, ‘I’m never going to be able to know whether anything’s real anymore, so why bother?’” (2026)

The role of AI digital technologies should be defined by humans in accordance with ethical standards, not a system of monitoring and controlling the users behavior or accepting disinformation. Small AI manipulations could well bring us into the big trouble of being unable to separate truth from lies and malignant exaggerations. Regulating AI has to be both anti-fake and a simplified approach that would provide clarity over what the tech companies can and cannot do. That, in turn, would encourage investment because growth is clearly related to investment in AI applications.

A worrisome aspect in the use of AIs comes from the more and more natural habit of delegating intellectual work to machines which is forcefully accompanied by two devastating effects: (1) the erosion of critical thinking skills and (2) the loss of the sense of responsibility. A research study at the Massachusetts Institute of Technology found that people using LLMs extensively show far less brain activity than those who do not. Moreover, the permanent statistical averaging offered by AI is also producing a homogenization of thought which is conducive to a loss of creativity.

Another emerging reality from the AI machines dominating social media is the decline of the Web as it used to be. The users are not in control, on the contrary, the levers are held by the big tech companies Google, Zuckerberger, Elon Musk, Andressen, Sam Altman through their AI tools. The goal of these companies is to wield the power to control the advertising raising prices. In a recent book, Cory Doctorow (2025) explains that:

> The objects of the Internet have become spy tools, surreptitiously collecting information about us—our habits, our desires, our health, our political inclinations—and using it to manipulate our behavior. The algorithms are designed to maximize "engagement", in fact amplify danger and sensationalism at the expense of truth. (p. 22)

Indeed, algorithms do not digest the data they are fed. The tech companies do that when they use the levers of the information economy to consolidate their dominance.

Conclusions

Under the AI avalanche reality almost dominated by conspiracies and deepfakes, the societal foundation is at risk because the very possibility of consensual attitudes is under attack. The loss of consensual reality is correlated with the rise of deepfakes becoming the norm online. The victim is the truth itself, since it cannot be verified using our normal inductive skills based on intuitive arguments and causal inference. We simply are not informed about and often even not allowed to know the sources of the content of the messages generated by AI online. The set of our personal values is what makes us what we socially are. Our will is kept in accordance with how the things around us are presented to us and that is precisely what AI is also doing. For us circumstances define a situation or a moment. For AI the circumstances are just statistical relationships. AI does not have its own values but incorporates the values on which it is trained. The ubiquitous use of AI makes me remember the formidable imagination of Karel Čapek in his War with the Newts (also known as The Salamander War) and think that we could end up being the servants of AI. Breaking these AI "shackles" imposed by the tech corporations is like fighting for a "new freedom". We humans need and seek meaning, purpose, and significance in the world and are confronted with risks and uncertainties. And, as Roger Scrouton splendidly expressed: "An improbability, however possible it might be, involves a failure of imaginative gap" (1994, p. 347). The imaginative gap between humans and AI is not just big; it's of an essential nature. We want not just technology in the online world; we want a moral attitude.

References

Beven, K. J. (1996). Facets of uncertainty: Epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. *Hydrological Sciences Journal, 61*, 1652-1665.

Chatlin, G. (1975). A theory of program size formally identical to information theory. *J. Assoc. Comput. Mach., 22*, 329-340.

Doctorow, C. (2025, October 7). *How enshittification conquered the 21st century and how we can overthrow*. Google Books.

Fiesler, C. (2026, February 2). AI is not the only threat menacing big tech. *The Economist.*

Froese, T., Karelin, G., & Ikegami, T. (2024). Making mind matter with irruption theory: Bridging end-directedness and entropy production by satisfying the participation criterion. In K. Kull and D. Favareau (Eds.), *Semiotics of biology: Making meaning in living systems* (pp. 1-17). Cambridge: The MIT Press.

Froese, T. (2026, February 14). How teaching molecules to think is revealing what a “mind” really is. *New Scientist.* Retrieved from https://www.newscientist.com/article/2513815-how-teaching-molecules-to-think-is-revealing-what-a-mind-really-is/

Gleick, J. (2024, November 23). *AI chatbots have been trained on trillions of words—but all in the service of plausibility over truth.* New York: NY Review of Books.

Hutto, D. D. (2006). Unprincipled engagement: Emotional experience, expression and response. In R. Menary, *Radical enactivism: Intentionality, phenomenology, and narrative: Focus on the philosophy of Daniel D. Hutto* (pp. 13-38). Amsterdam: John Benjamins.

Jerbi, K. (2025, September 1). The crucial role of chaos in our brain’s most extraordinary functions. *New Scientist*. Retrieved from https://www.newscientist.com/article/2493651-the-crucial-role-of-chaos-in-our-brains-most-extraordinary-functions/

O’Byrne, J., & Jerbi, K. (2022). How critical is brain criticality. *Trends in Neurosciences, 45*(11), 820-837.

Pigozzi, F., Goldstein, A., & Levin, M. (2025). Associative conditioning in gene regulatory network models increases integrative causal emergence. *Communications Biology, 8*, Article 1027.

Popiel, N. J. M., Abdollahi, S. K., Abeysinghe, P. M., Riganello, F., Nichols, E. S., Owen, A. M., & Soddu, A. (2020, March). The emergence of integrated information, complexity, and “consciousness” at criticality. *Entropy, 22*(3), 339.

Ramsey, F. (1950). *The foundations of mathematics and other logical essays*. London: Routledge and Kegan Paul.

Scrouton, R. (1994). *Modern philosophy.* London: Arrow Books.

Sifakis, J. (2025, December 3). Verimag laboratory, Grenoble. *China Daily*.

Sparkes, M. (2025, October 15). The AI bubble is heading towards a burst but it won’t be the end of AI. *New Scientist.* Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0262407925017361

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience, 5,* Article 42.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 49*, 433-460.

Watson, R., & Levin, M. (2023). The collective intelligence of evolution and development. *Collective Intelligence, 2*(2), 1-22.