

Jacek Koronacki

Artificial Intelligence and truth about us in time of confusion

A note for the 5th Biennial Conf. on Religion and Politics 2025

Where we stand

Twenty three years ago, the late Peter Augustine Lawler observed that

„the modern individual is an abstraction, an invention of the human mind. [my emphasis] That individual is made more free from social and political constraints, and less directed toward duty and goodness by God and nature, than a real human being ever could be. The modern individual is distinguished from the political animals — the citizens, statesmen, and philosophers — described by the Greek and Roman philosophers, and from the social, familial creatures described by Christian theologians. The modern individual is liberated from the philosopher’s duty to know the truth about nature, from the citizen’s selfless devotion to his country, from the creature’s love and fear of God, and even from the loving responsibilities that are inseparable from family life.”

From his essay in *The Intercollegiate Review* (Fall 2002)

Already then and nowadays as well, **the dominant trains of thought have been those of the philosophies of finitude.**

With only mild exaggeration, we are being convinced that we live after the end of history as envisaged by Alexandre Kojève and his student Allan Bloom (needless to say, their understanding of the end of history had nothing in common with the looney and abandoned idea of Francis Fukuyama).

Rémi Brague, the late Peter Augustine Lawler, Chantal Delsol and others are right – **all those lines of thought are unable to legitimate human existence, let alone assign any dignity to it.** At best, they can make man his own project. However, as Brague has put it succinctly,

“Self-creation turns into the destruction of the self by the self.”

The Legitimacy of the Human (2017)

A pragmatist, the late Richard Rorty is a case in point:

“The wholehearted behaviorism, naturalism and physicalism I have been commending [...] help us avoid the self-deception of thinking that we possess a deep, hidden, metaphysically significant nature which makes us ‘irreducibly’ different from inkwells or atoms.”

Philosophy and the Mirror of Nature (1979)

“I do not think there are any plain moral facts out there in the world, nor any truths independent of language, nor any neutral ground on which to stand and argue that either torture or kindness are preferable to the other.”

Contingency, Irony, and Solidarity (1989)

What seems to me most worth preserving in Dewey's work is his sense of the gradual change in human beings' self-image which has taken place in recorded history – the change from a sense of their dependence upon something antecedently present to a sense of the utopian possibilities of the future, the growth of their ability to mitigate their finitude by a talent for self-creation.

Objectivity, Relativism, and Truth (1991)

“No organism, human or nonhuman, is ever more or less in touch with reality than any other organism.”

Relativism – Finding and Making (1997)

In the world without *logos*, to quote again from Brague,

“From the point of view of technology, man appears as outdated, or at least as superfluous.”

And nowadays an able philosopher David Chalmers does not rule out that the world we live in is a virtual world, we are in a simulation (see his *Reality+: Virtual Worlds and the Problems of Philosophy*, 2022).

Artificial Intelligence – what it is, how its (and our) future may look like

Artificial Intelligence is that quality that enables machines to function appropriately and with foresight in its environment. Write “an entity” in lieu of the word “machines”, and you get an operational definition of intelligence as such.

Cf. Nils J. Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (2010)

AI works well in very narrow or not so narrow environments, let it be the game of chess (AlphaZero of Deep Mind) or the game of go (AlphaGo), AI-powered fighter pilots, drones and much more for the military (Heron Systems in 2020, Shield AI, and other companies), including the use of Brain-Computer Interfaces (BCI being used for curing diseases too, e.g., by Neuralink Corp.), LLMs like ChatGPT, Claude, Gemini and others, discovering new medicines via molecular studies (e.g., AlphaFold of Deep Mind – Nobel Prize in Chemistry in 2024), medical imaging, genome studies (e.g., for prenatal diagnosis and in IVF labs), etc.

Possible blessings it offers and dangers it poses are obvious. Regarding the latter, in a nutshell:

- Becoming enslaved by the state due to its ways to surveil us
- Succumbing to AI systems' dictates (AI "advisors" included) and becoming a thing in the Internet of Things
- Succumbing to AI systems' dictates in the workplace
- Succumbing to AI system's dictates in combat
- Accepting our moral degeneracy by introducing such endeavors as designer babies, in vitro gametogenesis (e.g., for multiparenting), and the like

In any case, **AI is a tool**, capable of assisting us in our endeavors.

More often than not it surpasses our human abilities in more and more domains. But, so what, unless we do not apply it for immoral or otherwise wrong goals.

AI is a tool which has its roots in Machine Learning (founded on mathematical logic and mathematical statistics) and (Artificial) Neural Networks. Already now it is brilliantly successful in surprisingly many applications thanks to such breakthrough innovations as Deep Learning and Generative AI (with a kind of self-learning from scratch included).

Most importantly: **AI rests on calculations and engineering genius, and nothing else, by no means it understands anything.**

Some would say that AI has been trained to work like the left hemisphere of our brains. As a mathematician, I would say that **AI works like a set of Turing-computable mathematical models.**

Jobst Landgrebe and Barry Smith summarize it well, by the way, remaining within a naturalist paradigm:

“[...] any Artificial Intelligence, no matter to what problems it is applied, would have to reach its solutions by executing a set of mathematical functions that are each computable in the Church-Turing sense. [...] *Let logic systems be all those that can be modelled comprehensively and adequately on a Turing machine.* [...] When such a model is executed on a computer, it becomes **a process of the logic system which is the computer itself**, which is realised in the way the binary logic of the microprocessors and the other components of the computer operate. **But this system is not a subsystem of the complex system it approximates. [This complex system] is the human mind-body continuum.** Our mental experience (including introspection) and our overt physical behavior are all emanations [of this continuum].”

Why Machines Will Never Rule the World (2025)

Yet, in the world without *logos*, confusion abounds and, e.g., Henry Kissinger, Eric Schmidt and Daniel Huttenlocher write in their otherwise valuable book:

“Even if advances in AI do not produce artificial general intelligence (AGI) — that is, software capable of human-level performance of any intellectual task and capable of relating tasks and concepts to others across disciplines — **the advent of AI will alter humanity’s concept of reality and therefore of itself.**

And they continue:

We are progressing toward great achievements, but those achievements should prompt philosophical reflection. Four centuries after Descartes promulgated his maxim, a question looms: **If AI “thinks,” or approximates thinking, who are we?”**

The Age of AI: And Our Human Future (2020)

Let us repeat: **AI, AGI included, „approximates thinking” by calculation and by no other means. It does not understand anything.**

However, due to pervasiveness of Information Technologies and AI within them, **our societies become prone to succumb to technopoly** (Neil Postman’s term), **cease thinking and acting in a human and humane way, and give it over to machines.**

And, in our time of confusion, the technological core of society may well turn into its mythical core (cf. Allan Jacobs, *The New Atlantis*, 58, Spring 2019). Indeed, one day, it was to be the Cult of Reason, later Religion of Humanity, now, e.g., Yuval Harari promotes dataism – a kind of religion with information flow of data as its supreme and mythical value (cf. also *Techgnosis: Myth, Magic & Mysticism in the Age of Information* by Erik Davis).

Apparently, Western societies are being pushed right up to the edge:

[...] indulging in trains of reflection opposed to your nature, you will prove yourself foolish; and if you persevere in such a course, you will fall into utter madness

Saint Irenaeus, *Against Heresies*, Book 2, Ch. XXV

All this is not against AI, whose progress is both awesome and fascinating. **What has been stated so far, is against using AI for wrong goals and, even more importantly, cherishing reflection opposed to human nature and pushing the West into utter madness.**

The progress of AI will be continued, from the AI’s today’s status to computer programs and perhaps robots capable of self-replication, what is less probable and in a more distant future, to robots capable of thriving in some sense (as, e.g., a cactus is), and – what is much less probable and in a much more distant future – with an animal self-consciousness of sorts.

Why not assume that we shall be able to produce artificial beings with vegetative and, **to some extent** and however unlikely this assumption’s fulfillment seems today, **with animal nature as well** (cf. Brian A.J. Boyd, *Will AI be Alive?*, *The New Atlantis*, 80, Spring 2025).

But, as I shall argue, **with nothing akin to even rudimentary elements of the rational nature of humans.**

What to say about human nature, as opposed to the nature of other species?

Whether one likes it or not, it is only the Aristotelian-Thomistic account of what it means to be human, which can be considered holistic and consistent, and – most importantly – conforming to individual's pursuit of happiness and fulfillment. (In this section, I follow the exposition by Edward Feser; see, in particular, his *Immortal Souls. A Treatise on Human Nature* (2025).)

In general, we distinguish between three basic forms of life: vegetative, animal, and rational.

Intellect is that faculty by which we grasp abstract concepts, put them together into judgments, and reason logically from one judgment to another.

It is to be distinguished from **imagination**, the faculty by which we form mental images, and from **sensation**, the faculty by which we perceive the goings on in the external material world and the internal world of the body. **Thought** is irreducible to sensation and imagination.

The concepts – which are the constituents of intellectual activity – are **universal** while mental images and sensations are always essentially **particular**. Moreover, mental images are always to some extent **vague** or **indeterminate**, while concepts are most often **precise** and **determinate**.

We have many concepts – like man or triangle – that have some connection with mental imagery. While we cannot visualize humanness or triangularity, we can visualize a particular human and a particular triangle.

We also have concepts – such as law, logical consistency, square root, infinity – that can strictly be associated with no mental image at all.

Humans are rational animals endowed with language that performs four functions (distinguishment by Karl Popper): **expressive, signaling, descriptive and argumentative.**

Some non-human animals are capable of the first two functions (i.e., they have language of sorts), but only human beings are capable of the latter two functions, and hence possess language which expresses concepts, thoughts, and arguments. In fact, language is required to have these three – concepts, thoughts and arguments.

Seemingly, **it is only humans who possess such language.** Animals feel pain, but have no concept of it. Had it been otherwise, they would have told us about it in some way.

Material things are never determinate or exact in the way universals or concepts are. Any material triangle, for example, is always only an approximation of perfect triangularity.

In general, universals are determinate and exact in a way material objects and processes cannot be.

All this points toward stating incorporeal character of human rational nature.

Indeed, the physical properties of any material representation are indeterminate or ambiguous with respect to its content. Whatever conceptual content it turns out to have must be determined by something outside of these properties.

To put a long story short, e.g., when we are adding numbers, the physical processes in the brain cannot tell us if it is carrying out addition or some other process which, in this way or another and at least so far, has given us a proper outcome.

Moreover, naturalistic theories of thought are unable to explain in a convincing and coherent way how material processes inside our brains can produce thoughts of this and not another meaning, this meaning being unambiguous and unequivocal. The more so with logical inferences.

On top of that (I borrow here freely from Feser's *Immortal Souls*), within naturalism, one must take thinking to consist of nothing more than the transition from one material symbol to another in accordance with causal laws.

The neurophysiological properties of material symbols alone are what determines which further symbols are generated. However, a symbol could serve as a rational justification of another symbol only by virtue of the meaning or conceptual content associated with the symbols.

In sum, Edward Feser states:

“Causal relationships of the kind a naturalistic theory posits in order to account for our thought processes could at most explain why, as a matter of *psychological* fact, we consider an inference to be reasonable. But they cannot underwrite the supposition that, as a matter of *logical* fact, it really is reasonable.”

The problem with naturalistic theories is that they reject Aristotelian – Thomistic hylomorphism, formal and final causes. How to analyze intellectual activities with substantial forms claimed to be nonexistent? How to analyze them, as well as our volitions, without formal and final causation?

In fact, **without formal and final causation, efficient causes do not suffice to infer from the observed (from the past) about the yet unseen.**

To put it mildly, as of now, it is only the hylomorphism that can account for **the substantial unity of an immaterial principle – form, soul, intellect – and body.** As St. Thomas Aquinas writes:

“Although the intellect abstracts from the phantasms, it does not understand actually without turning to the phantasms.” (*Summa theologiae*, I.85.5).

A word on the foundations of mathematics and the alleged nonmechanical character of human mind

The 1930s brought a revolution in our understanding of the limits of mathematics and our mathematical abilities.

Kurt Gödel, among others, proved several most fundamental facts in the area, in particular and roughly:

The First Incompleteness Theorem: Any consistent formal system F within which a certain amount of elementary arithmetic can be carried out is incomplete, i.e., **there are statements of the language of F which can neither be proved nor disproved in F .**

The Second Incompleteness Theorem: For any consistent system F within which a certain amount of elementary arithmetic can be carried out, **the consistency of F cannot be proved in F itself.**

Another eminent mathematician, Alfred Tarski, proved in particular the following, again informally stated:

Tarski's Undefinability Theorem: Let F be a consistent formalized system that contains a sufficient amount of arithmetic. Then there is no formula in the language of F such that for every sentence A of the language of F this formula assumes the true value if and only if A is true (put otherwise, **the concept of truth cannot be defined by a formula in the system**).

While the concept of truth is undefinable within system F , it can be defined for F in a stronger (or wider) system, which includes F as its part.

In fact, independently of Tarski, Gödel proved his incompleteness theorems using the fact that truth must be undefinable in the system.

We have already hinted to the fact that **all the machines (computers) can do today** is to perform calculations – they obtain solutions by executing sets of mathematical functions that are each computable in the Church-Turing sense (Alonzo Church and Alan Turing being another eminent mathematicians of the era). These **are purely syntactic operations**.

It is only us who ask about truth, which is a semantic category. Indeed, it is only us who perform semantic operations. **Semantics (e.g., the concept of truth) transcends syntactics**. It is another matter what our limitations are, as beautifully depicted by the aforementioned work of great mathematicians.

(Cf., this is a recommendation for the Polish reader only, Jan Woleński, “O filozoficznym sensie metamatematycznych twierdzeń limitacyjnych” *Wiadomości Matematyczne* 45 (2), 2009, and Roman Murawski, “Twierdzenia limitacyjne”, Konferencja ChFPN Nauka – Etyka – Wiara 2013.)

Remark: Mathematics as a whole is inexhaustible in that it cannot be expressed within one consistent system of axioms. We need wider and wider systems, but each of them will stay incomplete. Perhaps this is what contributed to Kurt Gödel's decision to prove a theorem on the necessity of the existence of God.

In lieu of a conclusion

“Is it not reasonable to suspect that if existence were pointless and the universe devoid of meaning, we would never have achieved not only the ability to imagine otherwise, but even the ability to entertain this very thought – to wit, that existence is pointless and the universe devoid of meaning?”

Leszek Kolakowski, *Horror Metaphysicus* (1988)

To paraphrase Peter Augustine Lawler from the essay quoted at the beginning:

Love is not an illusion, and we have been fitted by our nature to know the truth. But then it follows that both love of each other and love of the truth spring up from our spiritual souls.