Jacek Koronacki
This is a post-conference paper based on the presentation at the 5th Biennial Conf. on Religion and Politics, Cardinal Stefan Wyszyński University, Warsaw, November 6-7, 2025, accepted for publication in "Christianity-World-Politics".

Artificial Intelligence and truth about us in time of confusion

Abstract: Our aim is to outline what Artificial Intelligence (AI) is and how it should be seen within a wider context of the dominant trends in Western culture and the way Western humans see themselves (Sec. 1). We briefly describe AI methods and show that they are a mixed blessing to societies (Sec. 2). In order to better assess the inherent limitations of what AI can bring, in Section 3, we sketch what kind of an entity our human nature is. Finally, we claim that if AI's development proves harmful to our wellbeing, the fault will lie with us, not with AI (Sec.4).

1. Introduction: where we stand

The goal of this section is to outline briefly a wider context within which we are to discuss the blessings and dangers of the so-called artificial intelligence (AI from now). The section is atypical in that it is noticeably terse, filled with quotations and almost no commentary from this author. Nevertheless, we hope that the so composed message proves clear and convincing enough.

It was one hundred and eight years ago, in 1918, when Max Weber delivered his prescient speech at Munich University, in which he observed (Weber, 1919):

The increasing intellectualization and rationalization do not […] indicate an increased and general knowledge of the conditions under which one lives.

It means something else, namely, the knowledge or belief that if one but wished one could learn it at any time. Hence, it means that principally there are no mysterious incalculable forces that come into play, but rather that one can, in principle, master all things by calculation. This means that the world is disenchanted.

He was neither the first nor the last great intellectual to make and elaborate on this observation. For instance, not long after him the same was the theme of many writings of the English distributists or their American counterparts, the Southern Agrarians. Andrew Lytle wrote much on how *the pursuit of happiness [has been transmuted] into a nervous running-around which is without the logic, even, of a dog chasing its tail*. In particular, he wrote (Lytle, 1930):

This conflict is between the unnatural progeny of inventive genius and men. It is a war to the death between technology and the ordinary functions of living. The rights to these human functions are the natural rights of man, and they are threatened now, in the twentieth, not in the eighteenth, century for the first time. […] But since a power machine is ultimately dependent on human control, the issue presents an awful spectacle: men, run mad by their inventions, supplanting themselves with inanimate objects. This is, to follow the matter to its conclusion, a moral and spiritual suicide, foretelling an actual physical destruction.

More recently and more generally, at the beginning of this century, prematurely deceased Peter Augustine Lawler (2002) stated succinctly (see also Lawler, 2010):

the modern individual is an abstraction, an invention of the human mind. *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. [emphasis by jk]*

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 Kojéve 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); cf. (Lawler, 1999).

Rémi Brague, the late Peter Augustine Lawler, Chantal Delsol and many 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 (2017) has put it bluntly,

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

By way of example, let us note that the late Richard Rorty, a pragmatist, is a case in point. The following four citations are from (Rorty, 1979, 1989, 1991 and 1996), respectively:

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.

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.

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.

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

While it would be absurd to judge the legacy of a renowned philosopher on the basis of a few sentences from his writings, one can take him at his word and thus assume that he means what one finds in those writings. And in fact, the citations illustrate well the viewpoint of Richard Rorty, the pragmatist (however self-contradictory his stand as a whole is).

In the world without *logos*, to quote again from Rémi Brague (see also other works by this author on the subject),

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

And nowadays an able philosopher David Chalmers (2022) does not rule out that the world we live in is a virtual world, i.e., we are in a simulation.

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

2.1 Artificial Intelligence: what it is, seen from afar

Artificial intelligence (AI for brevity) is one of the pillars of what can be termed the information revolution. The foundation of this revolution is provided by technological advances in microelectronics, nanotechnology, telecommunications, and computer science. These advances have enabled an incredibly dynamic increase in the computing power of modern computers. They have also enabled a constant increase in the speed of transmitting ever-larger amounts of information over the internet. Finally, they have allowed for an extraordinary miniaturization of electronic equipment – for example, our cell phones are modern mini-computers. And all this brings in an increasingly visible presence of AI in our lives.

At the turn of the 21st century, we have become a „networked society"; cf. Castells' trilogy (Castells, 1996, 1997 and 1998). Today, we find organizations large and small, almost every person connected to this network with his or her computer, smartphone, or a similar device. Already then, at the turn of the century, it was clear that the ability to communicate via the internet would have a profound impact on culture and human condition. People began to immerse themselves in virtual worlds of multimedia communication. Over time, more and

more people, perhaps especially young people, have become virtually absorbed by this world. And sometimes, virtuality proves so captivating that the real world begins to seem less appealing. We watch with pain and helplessness as social media platforms transform the way young and old communicate with each other.

This kind of networking is one of the pillars of the information revolution. The second, as we have mentioned already, is AI. What is it? In a most concise manner, without referring to what intelligence in general is, but confining ourselves to its operational meaning, we can state that **AI 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, that of a machine, an animal and a human too; regarding this definition, cf. (Nilsson, 2009).

In a nutshell, AI is the result of human action aimed at making machines intelligent within a given environment. And AI works well, in fact exceptionally well, in very narrow or not so narrow environments, let it be the game of chess. To make a computer intelligent, humans equip it with memory, which stores information from the past, e.g., information on thousands of chess games played in the past, or strategies developed by grandmasters from the past. Furthermore, information from the environment – information about subsequent games played – is also stored in the computer in real time. The computer must be capable of learning – it must be equipped with algorithms that allow it to process the information it possesses in such a way as to make it capable of performing certain tasks, in our case, playing chess. Machine learning must also include the ability to improve its own performance by drawing on the machine's own "experience," e.g., based on the chess games it has played, sometimes winning, sometimes losing. The algorithm "rewards" the machine for winning and "punishes" it for losing. In this way, the algorithm "teaches" the machine to become an increasingly better chess player.

Back in 1997, IBM's specially programmed Deep Blue computer won a six-game chess match against Grandmaster Garry Kasparov by a score of 3.5 to 2.5. And for years now, there is no major chess player in the world who has not played training games with a computer, or rather, computers, since every program learns from its previous games and therefore plays chess games in a different style.

The example of chess helps one illustrate the fascinating development of AI methods. DeepMind's AlphaZero program, presented in 2017 (DeepMind was acquired by Google in 2014 and now is called Google DeepMind), had no knowledge of games previously played by humans; it knew only the rules of the game and honed its game by playing against itself (of course, the goal of each "player" was to win the game). It was therefore a program capable of "self-learning from scratch," an idea that proved to be somewhat revolutionary. The algorithm memorized the games it had played and, based on this, calculated, during the next game, which move, given current chess-board situation, would give the highest probability of winning. During this self-learning phase, the algorithm played millions of games. When

presented to the world, it proved to be ... the best player in the world – better than the best chess players and better than other programs. The program's play was considered outstandingly original, often amazing. Garry Kasparov wrote that AlphaZero transformed chess to its very roots. And no wonder, as the program had not previously had any human strategies loaded into it.

Let us add that in August 2020, in simulated dogfights between a human pilot seated in an F16 fighter simulator and an AI program simulating its opponent, the program trained to fight using the same self-learning principle as AlphaZero, proved superior. During the self-learning phase, the program (created by Heron Systems) competed against itself four billion times.

It is already easily seen that whenever we mention AI, what we actually mean are machine learning algorithms (machine learning to be abbreviated in what follows by ML). And indeed, it is ML algorithms which are the main building blocks of all AI implementations. Those ML algorithms can be applied to a virtually infinite number of tasks. Here are two more and conceptually very simple examples. The first is this: Images of tanks and trucks are loaded into a computer, and the computer is informed which photos contain tanks and which contain trucks. Based on this information, the algorithm "teaches" the computer to distinguish tanks from trucks. The computer's task is to determine in the future whether an uploaded "unlabeled" photo represents one or the other.

The second example concerns a situation in which we do not know whether the set of objects we are interested in is homogeneous in any respect, or whether it can be divided into subgroups of objects that are similar within each subgroup and significantly different from each other when they belong to different subgroups. A specific example would be a population of patients suffering from a certain disease. Each patient can have a large number of attributes describing him, e.g., when the description includes biomolecular data, related to the genome and epigenetic characteristics. AI algorithms allow us to verify whether the set of such descriptions, and thus the patients' population, can be divided into subgroups whose mechanisms of disease onset are different between subgroups. A similar situation is faced by all those who use a video-on-demand service, whose software recommends films to watch. The software divides the viewing population into subgroups with similar tastes, assigns the user to one of those groups, and recommends the next film based on that.

The successes achieved by applying and developing ML algorithms are breathtaking indeed. In 2024, the Nobel Prize in Chemistry was awarded to three scientists. Two of them are scientists from Google DeepMind: Demis Hassabis, the company's CEO, and John Jumper, one of its directors. The third is David Baker, a distinguished biochemist specializing in computational biology. They won the award for developing the AlphaFold2 algorithm, which predicts the three-dimensional structure of proteins based on sequences of amino acids. (To illustrate the significance of this achievement, note that without amino acids and the proteins they produce, there would be no life; the challenge is to find a way to extract from the astronomical number of random amino acid arrangements into three-dimensional structures

those that might represent the structures of actual proteins necessary for life; the discovered structures can have applications in biomedicine, for example.)

Around the turn of the last century, an idea of the so-called generative learning emerged in the field of ML. In short, it involves a kind of imitative learning – the creation of images or texts similar to those previously loaded into computer memory. The idea of generative learning would not have achieved the success it enjoys today without the impressively rapid development of modern computers' processing power. This development enabled in turn the emergence of a new approach to ML, called deep learning. It allowed, for example, construction of completely new, extremely efficient natural language models, not without reason called Large Language Models (LLMs). These models are the basis of the systems termed chatbots. Such models are capable of creating new texts of any length and conducting lengthy conversations with people on any topic, even, say, philosophical ones. In 2019, OpenAI completed work on the GPT-2 language model for processing and creating new natural language texts (GPT stands for Generative Pre-training Transformer). It is worth to realize just how complex such technical solutions are and how rapidly they are being developed. The model mentioned had approximately one and a half billion built-in parameters, tuned by the algorithm itself during its training. Shortly thereafter, an improved version of the model was created – the GPT-3 model (with 175 billion parameters) – and on November 30, 2022, OpenAI released a chatbot called ChatGPT to the general public, based on the further improved GPT-3.5 model. This chatbot's abilities – translating texts, answering questions, summarizing documents, and generating new natural language texts – at first glance seemed to rival those of a human. Short texts generated by the computer on a given topic were usually indistinguishable from a text written by a human expert in the given field. A conversation with a computer was usually no different from a conversation between two people.

And that was just the beginning! ChatGPT, launched on May 18, 2024, with the GPT-4o language model, had approximately 1.8 trillion parameters, these parameters controlling the operation of the eight submodels hidden within the system. But it is not the sheer number of parameters that matters. Rather, it is how quickly and significantly the quality of chatbots is being improved, e.g., by perfecting the mechanism for utilizing in real time the knowledge earlier loaded into them – whether it concerns writing text or "conversing" with the user (for simplicity's sake, we are ignoring chatbots' ability to analyze and process images). Released on December 11, 2025, ChatGPT with the GPT-5 model (in various versions) surprises with its still enhanced ability to compose substantive answers to questions requiring significantly specialized knowledge. It excels at passing university exams, exams granting qualifications in law, for example, it solves challenging mathematical problems, and helps scientists write challenging computer programs. The model's designers claim its ability to conduct logical reasoning, which is true to a limited extent. The GPT-5 model can certainly perform a task without an unequivocal solution in multiple ways and select the best one, according to the algorithm (some might say, "after deliberation", making a logical choice from among several possibilities).

In the above exposition we have confined ourselves to characterizing the GPT model and the ChatGPT chatbot, but other well-known chatbots and their LLMs, such as, e.g., Claude of Anthropic, Gemini of Google DeepMind, Grok of xAI, deserve similar praise. Companies developing AI systems are competing with each other; a race is underway, in which one model and chatbot wins in some respect, and another, in another. There is and will never be a single winner in this competition, the best one in every respect.

2.2. A word on AI as a mixed blessing, now and ahead

In this subsection, we present just a selective list of issues raised by the AI systems development, only touching upon their consequences to our wellbeing and moral standing.

It is now very well-known that the AI systems are changing dramatically the way wars are fought and are forcing profound changes in defense doctrines, the organization of armies, their much needed rearmament, the construction of defense systems against cyberattacks, etc. The wars in Gaza and, first and foremost, in Ukraine, have already proved it, and much more in those areas is to be expected.

For example, for several years now, Defense Advanced Research Projects Agency (known as DARPA) has been conducting a program called Air Combat Evolution (ACE). A fighter pilot, under whose control there are several autonomous unmanned drones, is expected to orchestrate his cooperation with AI systems, e.g., to hand over the initiative to AI program during a dogfight between his fighter and an enemy fighter and focus himself on instructing drones to destroy another selected targets. The goal of this project is to, on the one hand, allow AI programs which are more effective than humans to operate while, on the other, maintain human control over the entire combat. The project is envisaged to be broadened to the orchestration of human-AI cooperation in multi-aircraft scenarios.

Moreover, modern, rapidly developing neuroscience allows one to pinpoint the neural signals in the brain responsible for our mood, feelings of anxiety, fatigue, moments of distraction, or the weakening of the so-called working memory, which stores currently necessary information and, therefore, without which a given task cannot be performed. Neuroscience is constantly improving our knowledge of how to change these very brain states, and hence, where, and what signals to send to soldier's brain so that they stop his feeling of anxiety, fatigue, etc. Accordingly, brain-computer interfaces (BCIs), i.e., intracranial electrodes placed to capture selected brain signals and transmit them to and from a computer, are being developed. The computer is "trained" by a suitable AI algorithm to quickly recognize the soldier's undesirable state and induce the desired change in his brain. In this same area of research, considerable progress has already been made for thought and remote control of various devices, such as drones. It is worth noting that communication via a BCI is two-way. For instance, the human brain can not only control more than one drone but also receive

information from the drones about their surroundings. For more on AI in the military see, e.g., (Buchanan and Imbrie, 2024).

It must be added that similar implants (BCIs) are more than likely to help people who are paralyzed or suffer from Parkinson's disease or other neurological conditions. Several people suffering from such conditions have already received Neuralink implants (Neuralink Corp. is an Elon Musk's company).

A process of truly fundamental importance for our future is the process of intelligent robotization, i.e., surrounding humans with an increasing number of robots, bots (software programs capable of performing certain tasks in place of humans), and other devices with specific cognitive abilities. Renowned observer of technological progress Kevin Kelly calls this process cognification (Kelly, 2016). Those cognitive abilities will always be very limited; intelligent devices will be—and already are – capable of replacing humans in performing only very specific tasks. Such devices – as is already happening – will "think" in their own way, and this "thinking" will have nothing to do with human thought.

We have patients equipped with sensors measuring various body parameters, transmitting their values to a program capable of detecting situations that threaten the patient's health. As yet, it is another person, not a humanoid robot, that comes to the aid of this patient. But we already have robots capable of performing various specific tasks alongside and in concert with humans, robots that are "careful" and take care not to run over their human collaborators. And robot nurses are around the corner.

The social problems this robotization will create are numerous. In particular, a normal person – we mean it, a normal person – has never lived surrounded by robots, e.g., those preparing food at McDonald's and others serving and cleaning there. A human being desires a human contact. As mentioned, a person wants another person to administer medicines in a hospital, not a humanoid robot performing such duties. Above all, a person prefers a person who expresses emotions, not a robot that imitates them.

Repeating Kelly's comparison, cognition will encompass as many devices as electrification once did. We will be just as surrounded by such devices as we live today among devices that consume electricity. The problem, however, is that humans will experience their ubiquity largely as a substitute for human contact.

It is clear that societies in developed countries are facing a revolution in employment structures. Robots, chatbots, cobots (collaborative robots, i.e., robots that work with humans), and bots will take jobs from a significant number of industrial workers and service industry employees (from medical support staff and pharmacists, through transportation workers, to the entire service sector in the strict sense). AI systems – including those based on LLMs – will take jobs from many mid-level employees, e.g., those responsible for accounting,

conducting business transactions, and more. The novelty in this regard is that it is unclear where those unemployed individuals will be directed by the future labor market.

The engineers who invented the so-called Internet of Things (IoT) envisioned things connected through networks to serve us, for example, by freeing us from the need to check the fridge and see what we need to buy – the fridge will signal the store about what is running low, and a robot or human will deliver the goods. The internet will give us a smart city with a whole host of intelligent features that facilitate our movements in it, and at home. This internet network is meant to encompass us – as it is often called – 24/7. However, we are not meant to be just users; we are meant to be a part of it. For the IoT to function truly intelligently, meaning as effectively as possible, it would be best if we were equipped with brain implants that would make us just one thing in this entire network. Thanks to intracranial microchips, our thoughts will control devices which, e.g., will play music or a movie when we enter our homes.

For some time already, work is underway on the so-called Internet of Bodies (IoB). Data regarding our health—the state of our bodies—are to be collected by external devices, such as smartphones, and internal ones, whether swallowed or implanted, and transmitted to an AI system, which then provides healthcare providers with suggestions for possible interventions to maintain or save our health. Indeed, man is also threatened with enslavement because of the IoB.

There can be no doubt that AI systems are becoming ubiquitous in our everyday lives. In fact, they are already ubiquitous, even though we do not see it, in the network itself. Simply put, they are co-responsible for organizing the flow of information within this network.

The internet is changing our lives even more profoundly since it is an "intelligent" internet in the sense that everything it offers us is based on AI programs. We have already emphasized that natural interpersonal relationships are moving into the virtual world. And because we now primarily obtain information from the internet, it is gaining the power to control what information reaches us. If we are not inquisitive enough, our knowledge is "formatted" by AI algorithms. The situation is even worse when we rely on "dialogues" with chatbots and AI advisory systems.

Thanks to ML algorithms, the most powerful internet service companies today know almost everything about us. They don't know our thoughts yet, but they know what we like and dislike, what our interests are, and they can create a more or less accurate psychological profile of each and every one of us. Needless to say, there comes with it the erosion of normal interpersonal relationships and the development of dependence on virtual advisors. And it does not take much imagination to see the enormous potential power that has been placed in the hands of those who control the internet. A potential now being openly exploited by the Chinese government and covertly by other governments.

It has to be mentioned that an AI algorithm may act contrary to the designers' intentions. For instance, such risk arises when the algorithm itself or the problem being solved is so complex – and this situation is increasingly common – that the designers cannot explain why the algorithm adopted a particular solution (i.e., made a particular decision). It is crucial to remember this when using devices as sophisticated as, e.g., chatbots. The process of training them is incredibly complex, and this can lead to unexpected results. And, however sophisticated they are, they are just computing devices sometimes ready to provide erroneous results , e.g., as sometimes chatbots when they answer questions asked by a user (we shall come back to this issue in the next subsection).

No wonder that heated debates take place within research teams, research funding agencies, and military headquarters over how much autonomy to grant to lethal autonomous weapons controlled by AI systems (land, sea, air, and space, although the latter is rather a future possibility). When allowed to decide by themselves, there is always the risk of these assets selecting inappropriate targets. On the other hand, experiments show that the less human intervention in AI systems' decisions during an attack, the faster and more effective the attack is.

In this subsection, we have focused more on the threats posed by AI development than on the benefits it brings. This is by no means to imply that AI development brings no benefit to society. As with any progress, it brings both the threats and the benefits. And these benefits are numerous. In short, AI systems are playing an increasingly important and advantageous role in medicine, industry and services, including the public sector. They bring economic growth and make life easier for everyone. They have transformed the banking and finance sectors. But of course, there is no free lunch and AI development raises various dilemmas.

We wrote that humanoid robots will be able to replace medical support staff, yet humans would prefer to see another human alongside them, not a robot. But most likely, aging societies will need either robots or immigrants from foreign countries in their hospitals and nursing homes. And every such society will have to choose between these two options. A more serious problem, among many, will concern early diagnosis and treatment, e.g., enabled by genome studies. Already now it allows for prenatal diagnosis and treatment. However, if societies accept their moral degeneracy, it will lead to introducing such endeavors as designer babies, in vitro gametogenesis (e.g., for the so-called multiplex parenting), and the like (see, e.g., Greely (2016), Palacios-Gonzáles i in. (2014)).

2.3. A closer look at AI: from AI to AGI, to conscious machines?

The ever-increasing power of AI algorithms is a perfectly normal result of scientific and technological progress, though its pace is truly astonishing. The fact that machines built by engineers and scientists perform the same tasks as humans, and increasingly better, should not be alarming, but rather prompt reflection on how to harness this AI power for our benefit and how to counteract the potential harm that AI development may bring.

It's perhaps worth beginning this reflection by emphasizing that AI systems are not, and never will be, systems that understand their own behavior. For example, the aforementioned systems that generate texts and conduct dialogues with humans – already excellent – are incredibly complex systems that calculate what should be the next words or larger chunks of text in a text already generated by the system.

The first LLMs calculated what the next word should be. During the training phase, text documents (literary, newspaper, encyclopedia, dialogues, etc.) are loaded into a computer. For this computer, every text document is a sequence of words (and punctuation marks), and learning to "understand" language on the basis of previously loaded texts is roughly equivalent to constructing an algorithm that, given a specific sequence of words (a piece of text from the beginning to a given point), finds the most likely next word (adds another word to the text's piece). Building such an algorithm is conceptually and computationally very demanding, but it has nothing to do with understanding the actual meaning of texts. **It has nothing to do with understanding.**

Today's LLMs have built-in algorithms that allow for better use of the knowledge loaded into the system, but invariably, what an algorithm can do is to perform calculations. These calculations, and this is fundamentally important, do not and cannot be based on text understanding. This is an operation within the domain of language syntax, by no means of language semantics – the meaning of words. Semantics is transcendent to syntax, one might say perfectly transcendent, but with one caveat. The "but" is that computations that allow for purely syntactic analyses ostensibly reveal the meanings of words and concepts. The machine recreates syntactic relationships, which it can do by conducting calculations, and these relationships point to semantic relationships that the machine knows nothing about, but which, as a result, it can imitate. This is, and will remain, a perfectly "mindless" imitation, though its precision is truly astonishing (by the way, this "computable imitation" produces, sometimes and not surprisingly, false or misleading information, called hallucinations). It will remain so as long as computers are machines that calculate the so-called computable functions. That is what they are today, and no one knows if this limitation will ever be overcome.

Regardless of the AI system in question, it can perform a specific task faster than a human not because it understands the command, but because its operation has been pre-programmed, the system has sufficient pre-loaded knowledge, has a vast memory capable of storing "infinitely" more information than a human, and because it calculates faster. But what about the self-learning AI systems – such as AlphaZero – which do not require any "historical" knowledge to be loaded into them? They only require that the rules of their operation be loaded into them and they be programmed to achieve a well-defined goal (e.g., winning a game with known rules). However, this does not change the validity of the observation just made about the "mindlessness" of AI systems. The AlphaZero program played millions of chess games during the training phase. Since it is faster than a human, it has played many more games than a

human could have played during his or her lifetime, and because it has a larger memory than a human, it can use this "experience" better than a human.

It is true that more often than not AI 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 ML (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). But let us repeat: **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. Being a mathematician, this author would say that **AI works like a set of Turing-computable (or recursively computable, or efficiently computable) mathematical models**.

Some researchers wish for uploading our minds into a computer, but their dream is doomed to fail for this fundamental and at the same time trivial reason – our rational nature is incorporeal, however absurd if not blaspheming such a claim may appear nowadays. We shall elaborate on this issue in the next section. Some other researchers predict that computers of the future will be endowed with consciousness. But as long as a computer remains a computer, that is, a machine that performs calculations (a "computing machine," from the Latin "computare" meaning "to calculate"), it will have no consciousness whatsoever. A computer program applies syntactic rules defined by humans to process symbols, also defined by humans and having meaning only for them. This processing is a purely physical process, changing the computer's internal states according to rules arranged in an algorithm that solves a given task. The algorithm must be expressible as a sequence of calculations (so-called recursive functions), period. And now: the interpretation of these calculations, the interpretation of symbols, the interpretation of the algorithm's output, is the work of a human consciously engaged in these acts. In physical processes—and these are the acts of the computer executing the algorithm—there is no interpretation, and these are no conscious acts. The computer is not conscious of anything, neither of itself nor of its surroundings, that's obvious.

Albeit here, AI seems to come to the rescue of a designer who would like to make a computer conscious. Indeed, with current – let alone future – LLMs, a computer can easily imitate a human, i.e., imitate a being endowed with consciousness, and not just any consciousness, but human-specific. However, let us repeat, whatever AI model we have in mind, it achieves its results via the calculation of computable (recursive) functions. And this implies that even such imitation cannot be fully satisfactory. To demonstrate this, one does not need to be a Thomist who finds rational nature incorporeal. A naturalistic perspective is sufficient. Shortly put, the point is that human consciousness "lives" in a space incomparably richer than that of a computer. It is the space of sensations and thoughts, the space where this substantial unity –

a continuum, a naturalist would say – of body and mind “lives” and “moves”. This space is infinitely richer than the space of logical operations in which an AI system operates. And an AI system can only approximate that human space, the “movements” of that continuum of body and mind within that space, through the mathematical operations available to it. (For a thorough discussion of this topic within a naturalist paradigm, see Landgrebe and Smith, 2025).

To put it bluntly, AI is nowadays in its infancy. It is something different, on the one hand, to efficiently perform specific tasks, even achieving better results than humans, and, on the other, to take on new tasks, previously not even conceptually embedded in the machine's program. At its core, AI is machine learning (ML) and does not go any further, of course, wonderfully developing and delightfully expanding the ML algorithms. Performing specific tasks implies that what we have called the environment in which a given program is intended to operate is very narrow. It may contain a wealth of possibilities, but it still addresses a specific task.

Designers dream of developing general – some call it strong – artificial intelligence programs (AGI, for Artificial General Intelligence). Any such AGI program is intended to be able to handle as many tasks from various fields as possible. In this context, we speak of AGI performance that equals or exceeds human cognitive abilities in as many fields as possible. Some dream of AGI surpassing human cognitive abilities in all fields, in all tasks humans can undertake (today, such an all-powerful AGI is called artificial superintelligence, or ASI for short). Some believe that humans will build an ASI that will not have the cognitive limitations of the human mind, or even dream of transferring the human mind to a machine and enhancing its intelligence to levels unimaginable.

Let us set aside the dream of AGI, the realization of which is not ruled out. Let us note the obvious: the development of AI systems will continue. This is inevitable, and the point is not to fear it, but to direct this development toward wise applications.

As already noted, our networked servers, as well as our supercomputers, compute faster than us and have larger memories. It is a slight exaggeration to say that they compute infinitely faster and have infinitely larger, shared within a network memory, which they can use with incredible efficiency. And they will be even faster and more efficient when we replace them with quantum computers one day. Work is underway on building neuromorphic computers, inspired by the structure and functioning of the brain. Systems capable of self-improving the programs embedded in them are emerging. Thus, the moment is coming when the currently known discipline of ML will be expanded by machines themselves.

It is possible that we will see robots capable of self-replication. In the more distant future, though less likely, we may see robots endowed with a nature similar to that of vegetative beings, and in the even more distant future—though currently unimaginable—with a nature

resembling that of animals. Brian A.J. Boyd writes interestingly about this in his article "Will AI be Alive?" (Boyd, 2025).

But, let us repeat, we should not be afraid of AI's development. We should just be careful not to let ourselves be dominated by the AI systems.

3. 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 a human, which can be considered holistic and consistent, and – most importantly – conforming to individual's pursuit of happiness and fulfillment. (In this section, we borrow freely from the writings of Edward Feser; see, in particular, his *Immortal Souls. A Treatise on Human Nature* (Feser, 2024).)

Thomists (and earlier Aristotelians, as well as in slightly different ways, other ancient philosophers) distinguish three forms of life: vegetative, animal, and rational. All living beings possess a vegetative nature. Plants possess this nature only. Beings endowed with a vegetative nature have the ability to transform inanimate matter into animate matter (that is, to nourish themselves), live according to their own life cycle, and are capable of reproduction. Animals also possess an animal nature. They possess the capacity of sentience, i.e., the capacity for conscious awareness of stimuli – they are equipped with organs that allow them to receive signals from the environment and from their own bodies. Animals see, hear, have senses of taste, smell, touch, feel pain and joy. Crucially, they are conscious of all these experiences—they possess an inner awareness of them. An animal, deprived of reason, directly perceives what is beneficial and what is harmful in a given environment. These cognitive apprehensions are not sensory qualities per se, but belong to the animal faculty of sensory judgment. Thus, non-rational animals are guided by internal appetites that cause them to strive for something (e.g., reproduction or to satisfy hunger) or avoid something (e.g., a natural enemy). They also have imagination, which allows them to create mental images of what they have previously experienced—mental images of what they have seen, heard, etc. In a given situation, an animal re-awakens what was previously perceived through the senses and stored in the imagination. This spontaneous re-creation of previously perceived things is due to the animal's built-in memory (we will not delve deeper into the types of memory a non-rational animal possesses and which it lacks compared to humans). Finally, animals have the capacity of locomotion – they are capable of movement (e.g., to find food or escape danger).

Humans, in turn, naturally possess vegetative abilities and are animals, but rational ones – they are also endowed with a rational nature. Every human possesses an intellect that allows him or her to understand abstract concepts, to construct well-organized thoughts from these concepts, and, through reasoning consistent with the laws of logic, infer from these thoughts new statements. For example, a human understands the concepts of "being human" and "being

mortal"; he can formulate the thought "all humans are mortal"; and he can, from the premises "all humans are mortal" and "Socrates is a human," conclude that "Socrates is mortal."

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

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.

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.

All these observations point towards incorporeal character of human rational nature. 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. Note that it goes against all we know about the animal nature, which – as any Aristotelian-Thomist would agree – is to be seen as exclusively corporeal (an Aristotelian-Thomist defines matter more broadly than it is understood by contemporary natural sciences and as Descartes understood it earlier as *more geometrico*).

In order to better assess this claim, skipping however a more detailed analysis, let us look at the brain's operation as a computer (more generally, a machine) executing some programs (algorithms) and ask about the specific act of adding numbers. As such, the physical properties of the brain and the physical processes occurring within it, tell us nothing about the

rules of inference the brain follows. We do not know whether adding numbers in the brain occurs as we understand them, and thus the brain arrives at 2 + 2 = 4, or whether it arrives at this result in a different way. In other words, there is no hope that the physics of the brain – the material processes occurring within it – can describe the intellect in terms of programs executed by the brain. We know that a material entity, such as a computer, in which only physical processes occur, executes some programs, but we know this only because their execution is forced upon it by a programmer. However, the physical properties of the computer themselves do not and cannot tell us anything about it. By observing only these processes, we would not know what program they are actually executing, whether the computer 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. Can our thinking, then, be described entirely as a physical process?

A deeper insight into this problem – we refer the reader to (Feser, 2024) – allows one to conclude that no material system, no process of this kind, can deliver, as a result of its operation, a clearly defined semantic or conceptual content. Confined within a naturalistic framework, science and philosophy are unable to explain how a sequence of material symbols can express a thought about one content and not another. They are unable to explain how mental representations of concepts and statements (sentences) with a specific meaning (specific content) can arise in the brain, these representations being born from the senses which note facts from the external world, or being an outcome of some (mysterious) biological processes, e.g., evolutionarily forced. An even greater problem is to explain how a material system—not knowing the laws of logic—would reliably and unequivocally recognize logical relationships between the sentences "all men are mortal," "Socrates is a man," and "Socrates is mortal".

Edward Feser summarizes it thus:

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. [Feser's emphasis]*

In brief, it can also be stated that without the occurrence of physical (material) processes, there would be no thought, but these processes are not sufficient for the existence of thought. As Edward Feser wrote somewhere, neuroscience has already told us a great deal and will tell us even more, but it will not "discover" that thoughts are solely material operations of the brain. Moreover, our intellectual activity requires phantasms or mental images rightly considered material within the Aristotelian-Thomistic hylomorphism. Rational nature goes beyond what is material, but – one might say – it is coupled with what is material. As St. Thomas Aquinas wrote (*Summa theologiae*, I.85.5),

Although the intellect abstracts from the phantasms, it does not understand actually without turning to the phantasms.

Concluding this section, let us observe that, at least as of now, it is only the hylomorphism that can account in a holistic and consistent way for what it means to be a human. It is only within this paradigm, that we can account for the substantial unity of body and mind (of matter and form), and see the indispensability of taking into account formal and final causes when trying to describe a human. Indeed, how to analyze intellectual activities with substantial forms claimed to be nonexistent? And 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.

4. In lieu of a conclusion

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

In the world with *logos* replaced by *techne* confusion abounds. The technological core of society may well turn into its mythical core; cf. (Jacobs, 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 (see also Davis, 2015).

How deeply this confusion has come, is well demonstrated by patently preposterous statements in a book by such seasoned authors as Henry Kissinger, Eric Schmidt, and Daniel Huttenlocher, titled *The Age of AI: And Our Human Future* (2021). The late Kissinger needs no introduction. Eric Schmidt is the former head and still a key figure of Google, which he joined in 2001 and helped turn from a startup into a global power. Professor Daniel Huttenlocher is an outstanding scientist with extensive achievements in the field of AI. They wrote 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**. [my emphasis]*

And they continued:

*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?** [again, my emphasis]*

Intellectual cognition and morality are inextricably intertwined. By proclaimimg the death of God and thereby invalidating the moral thread, modernity severed this link. And by confining science—and knowledge!—within a materialistic, naturalistic framework, it lost the ability to speak of what is intimately human. Thereby came the death of the university as well. Western governments, possessed by sick ideologies, are pushing their subordinate states and nations toward self-destruction. Are the peoples of the West still capable of mounting an effective rebellion? Or do they no longer understand the language their grandparents spoke? (Serious works on the subject are legion; in addition to those already mentioned, let us add Kalb (2008); Koons (2010); Thompson (2010); Dreher (2017); and Deneen (2023).)

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. It is not AI that threatens us, as nuclear weapons once did (and to some extent still do), but AI within us. The true defense against the threats posed by AI is a return to the vision of a human as a spiritual being, a return to the true self, born to love, seek truth, and die. Contemporary (anti)culture has closed itself off from the spiritual and encourages us to live in a frenzy devoid of *the logic, even, of a dog chasing its tail*. And so, in our frenzy, our pursuit of abundance, fear of losing health and youthful appearance, or in the psychological fatigue we suffer, we succumb to AI advisory systems for help.

But are we really doomed to surrender our humanity? Who, for God's sake, forces us to live in this frenzy, with no time for reflection or encounter with another human being, instead of with a computer or a cell phone, or a silly entertainment? Who is forcing us to become slaves to AI systems? And who is forcing us to learn the history of philosophy from some chatbot, instead of reading most eminent authorities on the subject (Władysław Tatarkiewicz's "History of Philosophy" if one reads Polish)? Let us use these chatbots and ML algorithms, but without seeing them as our teachers, let alone authorities. Let ML and AI help us find information and assist us in hundreds of other ways, but... well, let they help, not rule us. Furthermore, let us remember that chatbots and other systems can give incorrect answers and write falsehoods. And more importantly, let us remember that in difficult matters they cannot provide the best answer. The point is that AI systems respond based on the majority of the material they have stored in their memories. Truly profound, most relevant content simply drowns in a flood of lower-quality content, having no chance of being selected by AI. And this is just one, and not the gravest, of a multitude of dangers AI poses, if used irresponsibly.

Apparently, Western societies are being pushed right up to the edge. Some 1,800 years ago St. Irenaeus warned us that (Saint Irenaeus, c. 180)

[…] 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.

Now, it is us to ask this question from *Horror Metaphysicus* by Leszek Kolakowski (1988):

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?

And, hopefully, answer it paraphrasing 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.

Bibliography

Boyd, Brian A.J., "Will AI be Alive?", *The New Atlantis*, No. 80, Spring 2025.

Brague, Rémi, *The Legitimacy of the Human*, St Augustine's Press, 2017.

Buchanan, Ben and Andrew Imbrie, *The New Fire. War, Peace, and Democracy in the Age of AI*, The MIT Press, 2024.

Castells, Manuel, *The Rise of the Network Society, The Information Age: Economy, Society and Culture* Vol. I, Blackwell. 1996.

Castells, Manuel, *The Power of Identity, The Information Age: Economy, Society and Culture* Vol. II, Blackwell, 1997.

Castells, Manuel, *End of Millenium; The Information Age: Economy, Society and Culture* Vol. III, Blackwell, 1998.

Chalmers, David, *Reality+: Virtual Worlds and the Problems of Philosophy*, W. W. Norton & Company, 2022.

Davis, Erik, *Techgnosis: Myth, Magic & Mysticism in the Age of Information*, North Atlantic Books, 2015.

Deneen, Patrick J., *Regime Change: Toward a Postliberal Futu*re, Sentinel, 2023.

Dreher, Rod, *The Benedict Option: A Strategy for Christians in a Post-Christian Nation*, Sentinel, 2017.

Feser, Edward, *Immortal Souls. A Treatise on Human Nature*, Editiones Scholasticae, 2024.

Greely, Henry, *The End of Sex and the Future of the Human Reproduction*, Harvard University Press, 2016.

Irenaeus, Saint, *Against Heresies*, Book 2, Ch. XXV, c. 180; independently published, 2024.

Jacobs, Allan, "After Technopoly", *The New Atlantis*, 58, Spring 2019.

Kalb, James, *The Tyranny of Liberalism, Understanding and Overcoming Administered Freedom, Inquisitorial Tolerance, and Equality by Command*, ISI Books, 2008.

Kelly, Kevin, *The Inevitable: Understanding the 12 Technological Forces That Will Shape Our Future*, Viking Press, 2016.

Kissinger, Henry, Eric Schmidt, and Daniel Huttenlocher, *The Age of AI: And Our Human Future*, John Murray (Publishers), 2021.

Leszek Kolakowski, *Horror Metaphysicus*, Wiley-Blackwell, 1988.

Koons, Robert, “The War of the Three Humanisms: Irving Babbitt and the Revival of Classical Learning”, *Modern Age*, 52: 198-207, 2010.
Lawler, Peter Augustine, Conservative Postmodernism, Postmodern Conservatism, in *The Intercollegiate Review*, vol. 38, issue 1, Fall 2002.
Lawler, Peter Augustine, *Postmodernism rightly understood : the return to realism in American thought*, Rowman & Littlefield, 1999.
Lawler, Peter Augustine, *Modern and American Dignity; Who We Are as Persons, and What That Means for Our Future*, ISI Books, 2010.
Lytle, Andrew, „The Hint Tit”, in *I’ll Take My Stand*, 1930, also in: *From Eden to Babylon*, essays by Andrew Lytle, Regnery Gateway, 1990.
Nilsson, Nils J., *The Quest for Artificial Intelligence:A History of Ideas and Achievements*, Cambridge University Press, 2009.
Palacios-Gonzáles, César, John Harris, and Guiseppe Testa, Multiplex parenting: IVG and the generations to come, Journal of Medical Ethics, 40(11):752-8, Nov. 2014.
Rorty, Richard, *Philosophy and the Mirror of Nature*, Princeton University Press, 1979.
Rorty, Richard, *Contingency, Irony, and Solidarity*, Cambridge University Press, 1989.
Rorty, Richard, *Objectivity, Relativism, and Truth*, Cambridge University Press, 1991.
Rorty, Richard, “Relativism – Finding and Making”, in: *Debating the State of Philosophy: Habermas, Rorty, and Kolakowski*, Józef Niżnik and John D. Sanders [eds.], Praeger, 1996.
Thompson, Ewa M., “Can We communicate? On Epistemological Incompatibilities in Contemporary Academic Discourse”, in: *College and the Academic Community in the European and the American Tradition*, Mark O’Connor and Piotr Wilczek [eds.], Warsaw University, 2010
Weber, Max, "Wissenschaft als Beruf," Duncker & Humblodt, 1919; „Science as a Vocation”, in *Max Weber: Essays in Sociology*, H.H. Gerth and C. Wright Mills [translated and edited], pp. 129-156, Oxford University Press, 1946.