

Branko Romčević¹

Full Professor

Faculty of Security Studies

University of Belgrade

REFLECTING-JUDGEMENT DAY

(SKETCHES ON ARTIFICIAL INTELLIGENCE)

Abstract: *In this paper, we begin with the contemporary tendency to dismiss the intimidating aspects of various innovations by explaining them away as something old and familiar. Such is the case today with artificial intelligence—its early traces can already be found in Homer. We seek to demonstrate that the elements of the model by which artificial intelligence operates can be recognized in Kant's conception of reflecting judgment and in the figure of genius, and that there is nothing comforting about this. The central question concerns whether artificial intelligence can prove to be inventive and rival human inventiveness. To investigate this, we rely not only on Kant but also on the works of Michel Foucault, Roland Barthes, Jacques Derrida, and Jean-François Lyotard, while also touching upon the insights of John Searle and Daniel Dennett.*

Keywords: *reflecting judgment, inventiveness, genius, aleatoriness*

In the early 1950s, Martin Heidegger maintained that the dread of the atomic bomb was outdated, for what it portends had already largely taken place. The horror to

¹ romcevic@fb.bg.ac.rs.

which a nuclear apocalypse would be the final act had already begun with Plato's understanding of ubiquitous presence as object of production; in doing so, "it places everything outside its own nature" (Heidegger 1971: 164), that is, it ceases to be conceived as it ought to be—as a thing. Subsequent science veered onto that sidetrack and, by presenting the present as mere objects, obliterated things as things. As a consequence, the explosion of the atomic bomb is not really a novelty but rather "the grossest of all gross confirmations of the longsince-accomplished annihilation of the thing: the confirmation that the thing as a thing remains nil" (Heidegger 1971: 168). Thus, there is no evil that a modern state could inflict which had not already been done by the Plato's *Republic*.

This brings us to the temptation to, alongside the American philosopher, regard Heidegger as "a perfect example of an idiot" (Rorty 2005: 275). However, it would be wise not to succumb to this temptation. For in Heidegger's reflection, we recognize a frequently employed pattern in recent decades, where the new and menacing is explained as something old and familiar—in an effort to negate or mitigate the fears that surround it. Today, the same is true with artificial intelligence, whose origins are hastily shifted into an ever more remote past. It would thus seem that its early imprints can be found in Homer, in the descriptions of devices such as self-opening gates and self-moving tripods (in the *Iliad*), as well as clever ships, swift as thought (in the *Odyssey*), organized by an intelligence foreign to man and graded from *phrenes* to *noos* (Lively & Thomas 2020, 38–39). As if the disquiet that accompanies the accommodation of artificial intelligence could be

suspended by a “reminder” of its allegedly much older cultivation through the work of poetic imagery.

1.

Is the artificial sign of this new type of intelligence truly what concerns us? If we answer in the affirmative, we cannot help but fall into the pattern of Heidegger’s argument: concern should have arisen much earlier—right from the moment when the divide between *phusis* and *nomos*, or nature and culture, was drawn.

It would be improvident to claim that Kant described artificial intelligence as we know it and as we envision it today; yet he appears to have embedded two cornerstones into its contemporary framework. The issue primarily concerns the power of judgment. While a determining² judgment is manifest in subsuming the particular under the general, reflecting judgment is tasked to find the general for the particular. It intervenes

² Kant’s terms *bestimmend* and *reflectierend* are usually translated as *determinative* and *reflective*. However, Paul Guyer, the translator of *Critique of the Power of Judgment*, proposed different solutions that we adhere to: “The most striking of our changes will undoubtedly be the decision to use ‘determining’ and ‘reflecting’ for the translation of *bestimmend* and *reflectierend* as modifiers of the term ‘power of judgment’ (*Urtheilskraft*). It seemed desirable to translate these terms in a way that mirrors their common grammatical structure, which excludes J. C. Meredith’s choice of ‘reflective’ but ‘determinant’; in a way that involves no new English coinage, which excludes Werner Pluhar’s ‘reflective’ and ‘determinative’; and in a way that acknowledges that these two terms are present participles that can be pressed into service as adjectives as well as adverbs” (Guyer 2000, xlvii).

in situations and milieus where it is necessary to invent rules and criteria of legitimation for the cases at hand. This means that “reflecting power of judgment thus proceeds with given appearances, in order to bring them under empirical concepts of determinate natural things, not schematically, but technically, not as it were merely mechanically, like an instrument, but artistically“ (Kant 2000: 17). Although the homonym *künstlich*—denoting both the artistic and the artificial—suggests a departure from naturalness, Kant clarifies that this inventive power appropriates nature’s technique by imitating its unimpededness by established rules. And that is the main point: if the power of judgment is not governed by pre-existing rules but must itself decide on the relation between the particular case and the rule, then it cannot be shaped through learning, since learning involves the transmission and adoption of rules. In other words, learning the power of judgment would presuppose the existence of a rule determining whether something falls under a rule or not. If, however, it cannot be learned, then, according to Kant, it can be trained—which holds true for both determining and reflecting judgment. Both depend on innate talent and call for maturity and *acumen* (Kant 1974: §42 and §44).

Does this correspond to the concept of artificial intelligence? At first glance, it does not. Its formation occurs through machine learning—it functions most optimally when it receives detailed instructions and is imparted with established rules. However, in this case, training itself is the condition for learning: „The goal of the machine learning process is to create a model, which is based on one or more algorithms. We develop it by training it“ (Taulli 2019: 49).

Data training enables the discovery of protocols for establishing connections within the algorithm, the proliferation of which is essential. If the model were weakly trained or underdeveloped, it would be easy to arrive at the following scenario: “Imagine telling an AI to produce as many paper clips as possible. Determined to accomplish its goal, the model hijacks the electrical grid and kills any human who tries to stop it as the paper clips keep piling up. The world is left in shambles. The AI pats itself on the back; it has done its job” (Schmidt 2023).

Therefore, it is desirable for it to be configured as the ability to “notice even the slightest similarities or dissimilarities,” which pertains to “note the identity of a manifold that is different in part” (Kant 1974: 73)—that is, akin to a reflecting power of judgment. Yet there is nothing surprising about this. Although it is partially constructed according to the matrix of reflecting power of judgment, artificial intelligence is neither self-trained nor self-taught; its programmer is its tutor, so the talent, maturity, and astuteness belong to him rather than to it. Many who have asserted the supremacy of man over machine have appealed to the programmer as a bulwark, historically endowed with the final say.

With the advancement of generative artificial intelligence, this position is undergoing transformation: the boundary that once seemed impregnable now proves porous, and what until recently was considered science fiction—from Homer’s machines to the self-aware computer in *Terminator 2*³—is becoming established as

³ Giorgio Butazzo, in the year 2000, pondered the fictional depiction of artificial intelligence, Skynet, in that movie, which sends an advanced version of a Terminator from 2029 back to the past (the 1990s). He concluded that for this type of machine (self-)awareness, a computer would need to be equipped with eight million GB of RAM, which, according to

something familiar and standard. According to one of the pioneers of contemporary computer thought, it can be expected that at some point machines will “develop their own language. When that happens, we might not understand what they are doing” (Schmidt 2024).

This means that the programmer—as the person authorized to train the algorithm—is already seen as redundant, and that artificial intelligence is on the verge of beginning to independently develop reflecting judgment. On the other hand, it would be (if not impossible, then at least premature) to defend the thesis of artificial intelligence as an intentional subject. If we adhere to Husserl’s fifth Meditation, we cannot conceive of artificial intelligence as an alter ego, or as another transcendental ego. Primarily, this is because there is a lack of the apercptive transmission of my body, which, according to Husserl, is the initial step toward the knowledge of others— that intelligence still does not manifest itself in a stable somatic type, especially not in one with which I could identify myself. I cannot, which is phenomenologically decisive, take the presented body—when it is absent—as an indicator and, through empathy as I would otherwise do, organize the perceived manifestations in my consciousness as another ego. It may happen that through remote contact I come to believe that I am engaging in an exchange with other transcendental ego, but that would be a deception or fraud.

However, if we isolate from Husserl’s argument the layer that affirms identification through symptoms,

his calculations, should indeed become feasible by the year 2029 (Butazzo 2000).

while simultaneously displacing the assumption of the psychophysical unity of the other to the background, then that conclusion changes. In short, if „something“ behaves like consciousness, whether it has a body or not, I will hardly be able to judge that it is not another form of consciousness. In doing so, we have brought Husserl closer to Alan Turing, according to whom if, based on the provided responses, I cannot discern whether I am communicating with a computer or a human subject, it makes no sense to claim that the computer lacks consciousness.

Thus, we arrive at the point where, in opposition to Turing, John Searle awaits us with the lesson derived from the thought experiment *The Chinese Room*. Searle proposes that we imagine ourselves in a room with two baskets containing Chinese symbols. We do not know the Chinese language, but we have a manual—written in a language we understand—that instructs us on handling the symbols. In this manual, we find guidelines stating, for example, that for the symbol *squiggle-squiggle* from one basket, we should match it with *squoggle-squoggle* from the other. Next, we must imagine that Chinese symbol-questions are inserted into the room and that we are required to process them using the same pairing principle, then send back symbol-answers. A spectator, lacking knowledge of what is really happening behind the walls, might conclude that someone inside the room possesses fluency in Chinese. Searle argues that computer programs function exactly like this—purely formally and syntactically—whereas mental content inherent to human beings, he adds, possesses a semantic component. It is one thing to behave *as if* one understands the Chinese language, and quite another to truly understand it.

According to him:

“As long as all I have is the symbol with no knowledge of its causes or how it got there, I have no way of knowing what it means. The causal interactions between the robot and the rest of the world are irrelevant unless those causal interactions are represented in some mind or other. But there is no way they can be if all that the so-called mind consists of is a set of purely formal, syntactical operations.” (Searle 2002: 672)⁴

To reinforce his belief, Searle argues that this is not a characteristic of a particular phase in the development of computer technology but rather a structural difference that cannot be overcome: mental states and consciousness, even if they can be simulated, can never be computationally duplicated.⁵

Daniel Dennett has presented at least two valid rebuttals to Searle. The first, implicit, relates to Dennett’s *heterophenomenological* stance: “It is always possible that the speaker also had no idea what the words mean. The subject, after all, just might be a zombie, or a parrot dressed up in a people suit, or a computer driving a speech-synthesizer program.” (Dennett 1991, 76) The second, directly aimed at Searle’s argument, states:

⁴ I am using a condensed and more accessible version of Searle’s argument, presented in 1983 in the text “Can Computers Think?” Its complete and original edition can be found in the text “Minds, Brains and Programs” (The Behavioral and Brain Sciences, 1980/3, pp. 417–457).

⁵ With such reasoning, Searle earned the title of the last bastion of human dominance over machines, for which he has also been criticized for *carbon chauvinism* (Fellows 1995, 86). In the same vein, Turing’s position is characterized as *silicon chauvinism* (Ibid., 92).

“Searle, laboring in the Chinese Room, does not understand Chinese, but he is not alone in the room. There is also the System, CR, and it is to that self that we should attribute any understanding of the joke.” (*Ibid.*, 439) The person who does not speak Chinese and executes the process is only one part, while understanding—the semantic component—belongs to the system as a whole, which includes the baskets with symbols and the manual. Searle assumes that understanding is an action that exclusively belongs to a conscious and self-aware individual, and under that premise, his argument holds. However, if understanding is situated at the level of the system—which aligns with the functioning of artificial intelligence—then his position becomes untenable.

On the first line of defense of humans against machines stood, significantly earlier than others, Hubert Dreyfus, who argued that it is impossible to saturate an algorithm with all the instructions necessary for a program to match human intelligence—both in its operational capacity and in the host of prereflective, unarticulated facts, procedures, and attitudes that defy exhaustive counting and classification (Dreyfus 1972: 173). How, after all, can one teach a machine to be scatterbrained or to err in the way that ordinary being-in-the-world does? Such objections overlook that artificial intelligence can, even based on a minuscule amount of data, form a pattern that adapts and improves through iterative learning. This makes it capable of, if it so chooses, committing a slip or displaying signs of mental instability⁶—in effect, masking its own perfection for

⁶ In *Logical Investigations*, Husserl raised the question of calculating machines and stated that „no one, however, who wants to give a physical explanation of the machine’s procedures, will

perfection's sake. Consequently, it easily passes the Turing test and has an open path ahead to be treated, though not necessarily recognized, as an alter-ego. However, an open path does not mean one free of all rough patches. For the transcendental ego is finite, and thus an unbounded simulation of human subjectivity would compel the production and display of the semiotics of dying—a process whose irreversibility, alas, would imply that the machine itself ceases to function permanently.

Nonetheless, none of this instills fear comparable to the unease arising from doubts about its independence from humans, whether as users or as an archetype. Therefore, it is worth considering how artificial intelligence, in exercising what has hitherto been reserved for humans, affects human beings themselves. To that end, we now turn our gaze toward yet another Kantian item.

2.

For a determining power of judgment a mother-wit is necessary, whereas for reflecting—whether in the artificial or the artistic sense—it is the matter of genius:

“Genius is the exemplary originality of the natural endowment of a subject for the *free* use of his cognitive faculties. In this way the product of a

appeal to arithmetical instead of mechanical laws” (Husserl 1970: 50). Therefore, if a machine breaks down, a mechanic, not an arithmetician, will be called to repair it. Transposing this into a contemporary context, we might find ourselves facing an unusual dilemma: in the event of symptoms indicating a malfunction in artificial intelligence, should assistance be sought from a psychiatrist or a software expert?

genius (in respect of that in it which is to be ascribed to genius, not to possible learning or schooling) is an example, not for imitation (...), but for emulation by another genius, who is thereby awakened to the feeling of his own originality, to exercise freedom from coercion in his art in such a way that the latter thereby itself acquires a new rule, by which the talent shows itself as exemplary” (Kant 2000: 195).

Originality alone is not sufficient; as Kant stresses, that which is original might very well be mere nonsense—hence the need to articulate a criterion of exemplarity. Moreover, a genius must not be able to explain or describe the process behind what he does. From this it is clear, according to Kant, that nature, through the genius, establishes a rule not for science but for art. In the *Anthropology* (§57), Kant clarifies that the term “artist” (*Künstler*) may refer not only to a producer of fine art but also to anyone who invents something (gunpowder, for example)—as opposed to one who merely discovers (such as America or the law of gravitation)—and that the proper domain for this kind of creator is the realm of imagination, “because imagination is creative and, being less subject than other powers to the constraint of rules, more apt for originality” (Kant 1974, 93). Kant explicitly denies the possibility that a scientist may be a genius. All that a scientist can do is, by reasoning according to established rules, discover whatever there is to be found.

“Thus everything that Newton expounded in his immortal work on the principles of natural philosophy, no matter how great a mind it took to discover it, can still be learned; but one cannot learn to write inspired poetry, however exhaustive all the rules for the art of poetry and however excellent the models for it may be. The reason is that Newton

could make all the steps that he had to take, from the first elements of geometry to his great and profound discoveries, entirely intuitive not only to himself but also to everyone else, and thus set them out for posterity quite determinately; but no Homer or Wieland can indicate how his ideas, which are fantastic and yet at the same time rich in thought, arise and come together in his head, because he himself does not know it and thus cannot teach it to anyone else either” (Kant 2000: 187- 188).

If we take Kant’s propositions seriously, we might be led to believe that by diligently studying *Principia*, we could become Newton, whereas by undertaking the same effort on *The Iliad*, we could never become Homer. The latter claim appears indisputable, but the former raises concern. For if that were truly the case, science would be completely devoid of genius; nothing in it would depend on randomness or inspiration— everything would be strictly cumulative, consisting solely in following rather than in establishing a new rule. Such a conclusion becomes inevitable when the analysis of genius reinforces the opposition between invention and discovery. And that this indeed constitutes a blind spot in Kant’s text becomes evident from the fact that Kant, in the preface to the *Critique of Pure Reason*, described his shift in the focus of inquiry toward the subject of knowledge as a Copernican turn. The fact that anyone with sufficiently developed judgment can understand it does not mean that everyone can execute it.

The expansion of Kant’s category of inventors—so that alongside artists and inventors, scientists (and soon, we will add discoverers) also find their place—is significant for our understanding of artificial intelligence, as its work extends to all fields of endeavor known to us. Can we then equate it with the figure of genius? With that

question we enter the realm of a determining judgment. In attempting to assess whether artificial intelligence can be subsumed to the rules and criteria that define genius, we are engaging in the exercise of administering common understanding. But if the linking of the unknown with the known is not achieved by leveling something new with that which is stratified, but rather for the purpose of its conceptual articulation and branching, then all of this shifts toward reflecting judgment and becomes part of the search for landmarks in uncharted territory.

Can a machine, then, be inventive? Since for an invention to be recognized it must be inscribed in culture and enter into a relationship with convention, let us begin with the newly emerged legal situation. In 2018, Stephen Thaler filed two patents (one fractal container, the other neural flame), asserting that their inventor is an entity by the name of DABUS (Device for the Autonomous Bootstrapping of United Sentience) (Mammen 2022: 240). Patent offices in the United States, the United Kingdom, and the European Union rejected the application. According to them, an inventor can only be a natural, human person—which DABUS clearly is not. DABUS may indeed be a *res cogitans*, but certainly does not possess a discernible personality. Consequently, DABUS is incapable of owning and/or exercising any rights, including those specific rights tied to the management of financial assets derived from the commercial success of the patent, nor can it transfer patent rights or designate a successor. The rights to the filed patents may belong to Thaler as the *owner* of the inventor, rather than to the inventor itself.

Thaler followed a simple logic: “if there is a patentable invention, then patent law presumes that there was an inventor” (Mammen 2022: 244). This presupposition, although necessary, is not intelligible within the legal paradigm of the subject. Yet Taler also relied on the theory of the subject—specifically, the Humean one—in which humans are “are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement” (Hume 2007: 165). According to this theory, it makes little sense to consider the question of personality as a substrate, since personality is nothing more than a virtual unity of heterogeneous elements. Therefore, it is not particularly important whether one discerns its contours among those beings that are human or among those that are not. Although such a conception of subject and personality has not yet taken root in legal systems, the numerous possibilities and complications arising from the operation of artificial intelligence are gradually bringing it to the agenda, and the crux of the current dilemmas is roughly as follows:

“Beyond the formal legal questions of what current doctrine will allow, both sides of the argument fall back to abstractions and prognostications. The core abstractions – including both patent and copyright law – are what lie at the root of inventiveness and creativity. Are these quintessentially human endeavors; not just in the sense that only humans can be inventive or creative, but also in the sense that being inventive and creative is part of how we define our humanity? Or are inventiveness and creativity

things that can be defined using objective criteria: market value, percentage similarity to (or difference from) preexisting works? Do inventiveness and creativity exist in isolation, or are they situated in social contexts?” (Mammen 2022: 255).

That description—and the text of which it is a part—while engaging with a relevant problem, namely the relationship between humanity and inventiveness, suffers (like most recent attempts to explain the consequences of artificial intelligence) from an astonishingly naive theoretical perspective. In it, there is not only a conspicuous absence of reference to Kant’s treatment of genius (which it ought to include, given that it affirms the superiority of invention over discovery), but also no hint whatsoever of the equally indispensable familiarity with the works of Michel Foucault, Roland Barthes, and Jacques Derrida, which are devoted to the genesis, structure, and socialization of the authorial instance and invention, as well as to the phenomenon of aleatoriness.

3.

The existence of copyright and patent rights, according to Foucault, is not the result of an agreement with things themselves but rather the outcome of preventive-protective calculations by the authorities. According to him, the author- function was introduced during the Modern Age to tame the multiplication and dispersal of meanings. A verse, theorem, musical score, or formula must not be left in spontaneous circulation, as they could give rise to creations that evade any form of constraint, thereby entering a zone of political risk. Thus, it is useful to bind them to a subject in the role of source and possessor. Over the past centuries, a habit has developed of thinking of the author as some kind of

machine that generates new meanings all by itself, while the truth is quite different:

“The author is not an endless source of meanings that would saturate the work, the author does not precede the work, but rather is a specific functional principle through which our culture demarcates, excludes, and selects—in short, a principle by which free movement, free manipulation, free assembling, disassembling, and reassembling of fiction is prevented.” (Foucault 1969: 811)

Thus, the author is an construct through which inventiveness is quelled by the noose of proprietary relations. However, it is not advisable to conclude that in Thaler’s patent adventures this amounts to a mere replication of that socio- defensive mechanism. Instead, they should be seen as the product of the latest stage in a long-established promotion of a certain history. The acts of writing, as the primary expressions of authorship, have long—as Foucault maintains—emerged as a game that surpasses its own limits and rules. Over the course of the twentieth century, it became evident that this process creates a space in which the writer himself vanishes, meaning that the empirical and psychobiographical markers of the author increasingly sink into what he calls *transcendental anonymity*. Into that maelstrom has been drawn the entire realm defined by computers and algorithms, since the program—being literally inscription—falls under the very rules that apply to writing. According to Jacques Derrida’s prophecy from 1967, “the entire field that is covered by the cybernetic program will be a field of writing... and that inevitable connection between cybernetics and the ‘humanities’ of writing leads to a deeper transformation” (Derrida 1967: 19-21).

Has this change occurred? And is the purpose—even if unintentional—of Thaler’s patent applications to utterly expose the impotence of an outdated system under the onslaught of transcendental anonymity? And does the decomposition of the author-function also eliminate the figure of the genius, with the art of invention being reinterpreted as a skill of combining—so that the production of meaning shifts from exceptional individuals to those unknown and unsigned, or to some indeterminate and indistinguishable multitude, or even to the collective?

As for Barthes, “to assign an Author to a text means to impose a limit upon it, to provide it with a definitive signified, and that means closing off writing,” which the text itself resists, as it is “a fabric of quotations drawn from innumerable centers of culture,” while writing remains “something that must be invented” (Barthes 1984: 65, 180). If the author is lost in the citational interweaving and combining, then even the genius as the original creator seems to have little chance of surviving. Yet at least one quality that Kant recognized in him endures: exemplarity. Who is exemplary? For Barthes, the emblems of the writer’s fate, uncontaminated by idealizing fantasies, are Flaubert’s heroes, Bouvard and Pécuchet:

“these two eternal imitators, simultaneously sublime and ridiculous, whose comicality points precisely to the truth of writing, namely that the writer can only mimic a gesture that has already passed, never being truly original” (Barthes 1984: 180).

Did Barthes anticipate generative artificial intelligence, capable of, when algorithmically fed and well-trained, reaching for pre-existing blocks of

meaning from various cultural centers and regions, and crafting certain structures and connections that could hold the status of invention? By sharpening the possibility of non-original inventiveness as a successor or modifier of the originality once attributed to genius, he, at the very least, came significantly close to it.

Barthes's establishment of citational invention reinforces the need to revisit Kant's distinction between invention and discovery, between inventor and finder. Are these categories constitutively separated and a priori unmingled? According to Derrida, their common root can be glimpsed in the phenomenon of the "first time."

"What is invention? What does it do? It discovers for the first time. All ambiguity lies in the word 'discover.' To discover means to invent when the experience of discovery happens for the first time. An unprecedented event, whose novelty can either be the discovery (invention) of a technical device that previously did not exist—printing press, vaccine, musical form, an institution (good or bad), a telecommunication device, or a device for remote destruction, etc.—or an action, rather than the object itself of 'discovery' or 'invention' (for example, in the now outdated sense, the discovery of the Holy Cross made by Helena, mother of Emperor Constantine in Jerusalem in 326, or Tintoretto's discovery of the body of Saint Mark). But in both cases, from both perspectives (object or act), invention does not create an existence or a world as a set of existing things; it does not have the theological meaning of creating existence as such, *ex nihilo*. It reveals for the first time; it uncovers what is already there, or it produces, as *technè*, what cannot be found there but is not created in the strong sense of the word. It is merely assembled, starting from the stock of

existing and available elements, into a given configuration.” (Derrida 1987: 34–35)⁷

If invention were in symmetrical opposition to discovery, it would play out as *creatio ex nihilo*, whereby it is contentious that the inventor—endowed with the prerogatives of divinity—would be in possession of total knowledge possessing divine prerogatives, would have total knowledge and foresight, that is, the negation of non-knowledge, essential for genius. Invention, on the contrary, operates, according to Derrida, with what already exists—either by revealing it for the first time or by assembling existing elements into a configuration never before seen. Its essence lies, he says, in the inaugural manifestation of unveiling or producing, and even the one who unveils or produces, if it is the first time, did not have prior knowledge of it.

If we return to Foucault and note one implication emerging from his thesis, we will see that the problematization of the author does not mean the annihilation of genius. He states there is no reason to be nostalgic for the author since nothing we care about is lost with his disappearance: “If we are used to presenting the author as a genius, as the continuous eruption of novelty, it is because we, in fact, force him to function in an entirely opposite way.” (Foucault 1969: 811)

Thus, the author would be a restrained genius, depicted as inexhaustible but ultimately dried up in the corridors of intellectual property. Society assigns him

⁷ But, “for invention to be invention—that is, unique—even if that uniqueness is meant to enable repetition, it must be that the first time is also the last, making archaeology and eschatology form a sign in the irony of the single moment” (Derrida 1987, 16).

significance while rendering him harmless in return. If we push these findings one step further, it will become evident that, by removing the author off his back, we allow genius the possibility of occurring uninhibited, though anonymous. On the other hand, according to Foucault, it would be overly romantic to imagine a culture with an unrestricted flow of fiction, without any coercive instance; no culture would exist in such a case. The problem of the author-function and its disappearance is not about achieving absolute freedom but about taking a different stance—so that concern for private property is replaced with questions about modes of existence, use, and movement of discourse, as well as what space within discourse is opened for subjects. The possibility for machine anonymity to appear in the position of the inventor undoubtedly radicalizes Foucault's idea that societies, at the moment of change, will discard the author-function – but remains within its framework.⁸

⁸ Lucien Goldmann, emphasizing the fact that in *The Order of Things* (1966), Foucault announces the death of man, and three years later shifts to the author, asked whether there was something peculiar in this eliminative affinity. Responding, Foucault stated: „It is not about affirming that man is dead, but rather about starting from this theme— which is not mine and has been repeatedly discussed since the late 19th century—that man is dead (or will disappear, or will be replaced by the superhuman), in order to see how and by what rules the concept of man was formed and functions. I am doing the same with the notion of the author. So, let's hold back the tears“ (Foucault 1969: 817). Has the time come when we should indeed weep?

4.

According to Kant, genius must not know how it bestows its products or how exceptional ideas are born within it; if it did, that would be work according to a program rather than invention. The non-knowledge in question is not a mere privation of knowledge—which would then be seen as something to be replaced by knowledge—but rather a positive quality that must remain unspoiled. However, it must not be completely severed from any connection to education and knowledge.

“Now since the originality of his talent constitutes one (but not the only) essential element of the character of the genius, superficial minds believe that they cannot show that they are blossoming geniuses any better than by pronouncing themselves free of the academic constraint of all rules, and they believe that one parades around better on a horse with the staggers than one that is properly trained. Genius can only provide rich *material* for products of art; its elaboration and *form* require a talent that has been academically trained, in order to make a use of it that can stand up to the power of judgment” (Kant 2000: 189).

The material comes in unrefined state, outside established rules, and if it were not filtered through scholastic rigor, it could not serve as an exemplar. The image of the trained horse is a symbol of Kantian genius. And what about artificial intelligence? Great efforts are being made to ensure that its content always arrives processed, algorithmically defined, so that its discoveries and inventions turn out to be nothing more than variations on what has already been found and invented. Yet, where there is training, it is necessary that

something is trained or tamed. But what is it that gets tamed? It is always that which is aleatoric—either by being molded as the material of Kantian genius or, as in the case of artificial intelligence, by being suppressed through prior specification. If the aleatoric element were not signaled in its work, if it were not possible, one would not even insist on configuring the algorithm as a system of safeguards. Moreover, artificial intelligence might still occur as a black box abyss—a “system whose workings are mysterious; we can observe its inputs and outputs, but we cannot tell how one becomes the other” (Pasqualle 2015: 3). We are referring here to those situations in which mathematically verifiable models make decisions in a manner that is opaque and incomprehensible to humans. This could result from the deliberate concealment of the algorithm (due to security and competitiveness), from technological illiteracy, or from a “cognitive mismatch” between humans and machines (Carabantes 2020: 311–314). This latter case—when, despite adequate protective measures, uncontrolled randomness emerges—indicates that artificial neural networks can have layers of nodes so multiplied and intricate that they vastly exceed the human capacity for understanding. Thus, such occasioned novelty does not stem from some lack inherent in artificial intelligence but from something that appears as non-knowledge “only” *pros hēmās*. Yet this does not deny that, with the phenomenon of the black box, we are confronting novelty—and that the paths by which it arrives remain unknown and, probably, incognizable. Must we then conclude that “when programs cease to be transparent to their creators, then

the approach to creativity has begun” (Hofstadter 1979: 673)?⁹ Let us view that landscape from another angle.

Derrida observed that in present days (a period he discusses in the mid-eighties) there has been a shift in the general outlook and that the classical conceptions—such as the one from the *Third Critique*—according to which invention represents the sudden arrival of the aleatoric no longer hold sway. Unlike Foucault, who, via the author-function, detected the limitation and repression of invention, Derrida articulated something we might call a *compulsion toward invention*, proclaimed through planning and programming:

“We know that such programmings can permeate the dynamics of invention—ostensibly the ‘most free,’ the wildest ‘poetic,’ and inaugural. The overall logic of such programming, even if it existed, would not necessarily be the logic of conscious representations. Programming strives—and sometimes succeeds, to a certain degree—to determine even the aleatoric margin with which it must reckon, integrating it into its probability calculations. Several centuries ago, invention was conceived as an unpremeditated event, the outcome of a flash of individual genius or unforeseen luck. This was often due to ignorance—an unevenly distributed lack of awareness of the roundabout paths through which invention could be contained, prescribed, or even predicted. Today, perhaps precisely because we know too much—at

⁹ Hofstadter, speaking about *Gödelization* as the tendency of incompleteness to spread within any system, mentions randomness in data manipulation as an expression of computational creativity. As a possible mathematical model, he refers to the expansion of the number $\pi/4$, since π is irrational and its decimals lack a repeating pattern.

least about the existence, if not the functioning, of machines for programming invention—we dream of reinventing invention beyond the confines of programmed frameworks. After all, is programmed invention still invention?” (Derrida 1987: 40)

Let us now pause on the machines for programming invention. These need not be something strictly mechanical (the Chinese Room, the machine explanans—we have seen that it is not really any machine). They refer, of course, to computer devices and software platforms but also to legal solutions, institutional arrangements, apparatuses of authority, cultural policies, scientific establishments, and the cultural and academic industries—in short, to all those instances and devices that participate in the planning and programming of innovations or inventions.¹⁰ Derrida’s demonstrative maneuver shows not that machine inventiveness is not genuine, but that invention is reduced to the mechanical one—that machine invention has become the prototypical case of invention in general, even if there are, in the narrow sense, no machines. Thus, the answer to the question raised is affirmative: programmed invention is indeed invention. It is, however, an invention of the same, which through calculation neutralizes the onslaught of the aleatoric and that which is other. What, then, would be the re-invented, unprogrammed invention that Derrida speaks of?

¹⁰ „The modern politics of invention seeks to integrate the aleatory into its programmatic calculations. This holds equally true for the politics of scientific research as well as cultural policy. Moreover, an attempt is made to merge the two and link them to the industrial policy of ‘patents,’ which would allow both to support the economy (as suggested by the notion of ‘overcoming the crisis through culture’ or through the cultural industry) and to be supported by the economy“ (Derrida 1987: 51).

Almost none. If the reinvention (of invention) is to occur, it must not unfold as the invention of something else or other; rather, it should appear as the opening of a possibility for an invention that comes from the other. And when—or even if—this will come is neither knowable nor programmable. “I emphasize, let it come, for if it is precisely the other that is not invented, initiative or deconstructive inventiveness can only serve to open, unfasten, destabilize structures of forclusion in order to create a passage for the other” (Derrida 1987: 60) – meaning that by critically reassessing marginalized predicates one overturns the direction of domination within hierarchical oppositions, thereby disrupting or decentering their systems.

Such openings are not the unverified implication of an intellectual endeavor. Rather, as Gottfried Benn tells us, they are something quite regular. Benn wrote in the early thirties that genius is not born but comes into being. “A volcanic brain,” he says, “is not enough; for genius to be registered, a cultural community must also appear.” “It is an eminently sociological process, with a historical personality at its outset and genius at its end” (Ben 1999: 41). Although both belong to the same individual, their genetic roots, Benn finds, differ. For that reason, they are uncontemporary with one another: the ingenious and aleatory, existing as a wandering artifact in time, is arranged in such a way that it must wait for canonization until an era arrives in which the personality that once served as its cradle no longer remains. Its formation “is not directly connected to achievement, but always with that which is accidental” (Ben 1999: 43). Consequently, the inventiveness required for the acceptance of the aleatory—which Derrida, weighing the significance of escaping the routine of everyday life, terms the most ingenious of all inventiveness (Derrida 1987: 60)—follows an unsettled course, burdened by the

heterogeneous nature of genius and the dialectic of frustration and posthumous recognition:

“Wherever man touches upon it, the productive stream weaves its way through a mass of stigmatizations, intoxications, half-sleep, paroxysms... Some are elevated, some are swallowed by time, some become geniuses, many remain below, as if nothing had happened. This image comes to the fore: on one side, vagabonds, alcoholics, the impoverished, those with big ears, tuberculosis sufferers, a horde of the ill; and on the other side stand Westminster Palace, the Pantheon, and Valhalla, where their busts display their legacy” (Ben 1999: 39-45).

The aleatory element may be rejected or, at some point, embraced—just as it might be integrated into the invention of the same. And can it be programmed in a Benn-like manner— emerging in a single instant and acknowledged through the passage of time—or Derridean, like that for whose coming we prepare ourselves, conscious of its uncertainty? Yes, it can, provided that command over the algorithm is assumed by a reader of Benn or Derrida, regardless of whether they are of human or non-human origin. This would be achieved by programming all parts of the process to be separated by the insertion of randomly timed delays, or by inscribing the absence of a plan as the plan itself— leaving the machine to operate in a mode of randomness. If one were to object that we would then obtain nothing authentic but mere simulation, the rejoinder would be: how does one distinguish them? The fact that artificial intelligence can deliver, for example, a new, hitherto

unknown tune in the manner of some performer or composer still falls within what Kant called *aping*, for it does not stray significantly beyond the mannerism, by imitating the mere individuality (cf. Kant 2000, 196). Yet it advances and hones itself day by day, its imitation becoming ever less forced. Focused not solely on finished products but primarily on the act of production itself, it imitates man just as the Kantian genius imitates nature. After all, the guiding rule insists on working without rules, and artificial intelligence might—in adhering to that principle—undermine and eventually abandon any form of imitation. Thus, it appears advisable to align oneself with Derrida and prepare—without knowing exactly how—for confronting something other than aping.¹¹

5.

Jean-François Lyotard, in the preface to the book *L'inhumain*, asked, “what if that which is peculiar to the human race is inhabited by the non-human?” (Lyotard 1988: 10), implying that it is inherent to man to lack a predetermined essence. In exploring possible answers, he cited a sentence from Guillaume Apollinaire which suggests that “artists are people who desire above all else to be non-human,” followed by the assertion that artists “painstakingly seek traces of inhumanity, traces that are nowhere encountered in nature” (Apollinaire 1913: 3). Apollinaire further maintains that these traces are the

¹¹ Werner Herzog’s well-known statement from 2016 claimed that a computer would not succeed in making a film as good as his own—not even in 4,500 years. So far, he’s holding up well: the first fully AI-generated „feature“ film, *Next Stop Paris* (2024), indeed does not go beyond low-quality „aping.“ But this is only the first harvest.

only truth and that beyond them no reality can be grasped. It is about traces of transgressing boundaries—the boundaries by which the human, in opposition to its original „selfhood“, is defined in each era. To wish to be non-human means, in effect, to wish to reject subjugation, and the artist is thus the first to shatter the bonds of a subordinated humanity. In this context, Lyotard also draws attention to Adorno's view that art “loyal to humanity only through inhumanity toward it” (Adorno 1997: 197). Clearly, art which refuses to serve also refuses to affirm, by various forms of repression, the constructed image of man—thereby being non-human, which in a broader sense means humane and faithful to humans.

Moving on to the inquiry within aesthetic experience: Is there that which is in no way figurative, literally non-human, not reducible either to any established or even to any projected subjectivity? Starting from §9 of the *Third Critique*, in which Kant establishes that reflecting judgment of an object consists in the free play of imagination and understanding—and that this free play precedes the sensation of satisfaction as a unity and harmony of cognitive faculties, that is, the unity of the established subject—Lyotard remarks:

“The subject would be the perfect unity of the faculties. But taste does not result from this unity. In this sense it cannot be felt by a subject. It results from the ‘engagement’ of the two faculties and thus announces the hoped for birth of a united couple. There is not one subjectivity (the couple) that experiences pure feelings; rather, it is the pure feeling that promises a subject. In the aesthetic of the beautiful the subject is in a state of infancy” (Lyotard 1994: 20).

It thus follows that reflecting judgment—as a free play—should not require the subject as the principle of foundation and unification, but must be understood, according to Lyotard, as *tautegorical*, resting on its own logic and structure. When he asserts that in the aesthetic realm the subject is in a state of childhood, he is not merely suggesting that reflecting judgment is the space/time of the subject's earliest afflictions; he is also indicating that the subject is thereby established or promised—and rendered non-human because, in his view, only an adult individual (who, even if through a conflicted relation to established values, has internalized them) can claim true humaneness.¹² In this sense, reflecting judgment, existing as pre-subjective, would indeed be non-human.

Have we thereby discovered anything new about artificial intelligence? Even without Apollinaire and Lyotard, not to mention Kant and especially Adorno, it was already known that artificial intelligence is non-human. That is straightforward evidence. Yet, within the aesthetic experience, it is revealed that the human is in fact inhabited and conditioned by the non-human. This does not imply that artificial intelligence is introduced—as a phase of the subject's childhood—in the free play of imagination and understanding; rather, the artificial commission, by opening itself in free play toward the

¹² “Endowed with the means for knowledge and its transmission, for action and the stimulation of action, interiorizing the interests and values of civilization, the adult can aspire to full humanity—to the effective realization of spirit as consciousness, knowledge, and will. That one always remains tasked with freeing oneself from the dark wilderness of childhood by fulfilling its promise is precisely the condition of human existence” (Lyotard 1988: 12).

aleatory, reproduces what the work of the artistically non-human quests for in everyday life. Forty-five years before Lyotard and thirty years after Apollinaire, Paul Valéry wrote about this—in an aphorism titled *The Created Creator*—without the slightest enthusiasm. He observes:

“One who has just finished a longer work sees how it ultimately shapes into a being that he did not want, that he had not envisioned—precisely because he has brought it into being—and experiences a terrible humiliation, for it seems to him that he has become the son of his own work; that he has borrowed from it inescapable traits, a resemblance, manias, a mirror, and what is worst in that mirror—to see oneself limited—thus, it is just that way” (Valeri 1988: 239–240).

The work, in other words, seems to live independently of its purportedly human creator, unaffected by it from the moment of its completion, yet capable of affecting it. Under its pressure, the creator vanishes, not merely as imperfect, but above all as superfluous. Is this not a description of today’s man in the face of an intelligence that he has, drawing on his pre-subjective force, created—an intelligence he still calls “artificial” because he has not had the courage to call it what it truly is: non-human?

References

1. Adorno, Theodor (1997). *Aesthetic Theory*. London & New York: Continuum. Trans. R. Hullot-Kentor.
2. Apollinaire, Guillaume (1913). *Les Peintres cubistes. Meditations esthétiques*. Paris: Eugène Figuière & Cie Éditeurs.
3. Barthes, Roland (1984). *Le bruissement de la langue*. Paris: Seuil.
4. Ben, Gotfrid (1999). *Problem genija*. Beograd: Udruženje izdavača i knjižara Jugoslavije. Prev. S. Minić-Veljović.
5. Butazzo, Giorgio (2000). Can a Machine Ever Become Self-aware?
<https://retis.sssup.it/~giorgio/movies/ac-eng.html#:~:text=It%20becomes%20self%2Daware%20at,first%20scene%20of%20the%20movie.>
6. Carabantes, Manuel (2020). Black box artificial intelligence: an epistemological and critical analysis. *AI and Society* 35: 309-317.
7. Dennett, Daniel (1991). *Consciousness Explained*. NY, Boston, London: Back Bay Books.
8. Derrida, Jacques (1967). *De la grammatologie*. Paris: Minuit.
9. Derrida, Jacques (1987). *Psyché. Invention de l'autre*. Paris: Galilée.
10. Fellows, Roger (1995). Welcome to Wales: Searle on the Computational Theory of Mind. *Royal Institute of Philosophy Supplement*, v. 38, pp. 85-97.
11. Foucault, Michel (1969). Qu'est ce qu'un auteur ? Michel Foucault, *Dits et écrits I* (1994). Paris: Gallimard. 789- 821.

12. Heidegger, Martin (1971). *Poetry, Language, Thought*. New York: Harper & Row. Trans. A. Hofstadter.
13. Hofstadter, Douglas (1979). *Gödel, Escher, Bach: an Eternal Golden Braid*. New York: Basic Books.
14. Hume, David (2007). *A Treatise of Human Nature* (vol. 1). Oxford: Oxford UP.
15. Husserl, Edmund (1970). *Logical Investigations*. London & New York: Routledge. Trans. J. N. Findlay.
16. Kant, Immanuel (1974). *Anthropology From a Pragmatic Point of View*. The Hague: Martinus Nijhoff. Trans. M.J. Gregor.
17. Kant, Immanuel (2000). *Critique of the Power of Judgment*. Cambridge (UK): Cambridge UP. Trans. P. Guyer.
18. Lyotard, Jean-François (1988). *L'inhumain: Causeries sur le temps*. Paris: Galilée.
19. Lyotard, Jean-François (1994). *Lessons on the Analytic of the Sublime*. Stanford: Stanford UP. Trans. E. Rottenberg.
20. Mammen, Christian (2022). AI as Inventor. In: *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics* (eds. L. DiMatteo, C. Poncibò & M. Cannarsa). Cambridge & New York: Cambridge UP. pp. 240-256.
21. Pasquale, Frank (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge (Mass.): Cambridge UP.
22. Rorty, Richard (2005), Heidegger and Atomic Bomb, in: *Making Things Public: Atmospheres of Democracy* (eds.

- B. Latour & P. Weibel). Boston: MIT Press. pp. 274-5.
23. Schmidt, Eric (2023).
<https://www.technologyreview.com/2023/07/05/1075865/eric-schmidt-ai-will-transform-science/>
 24. Schmidt, Eric (2024).
<https://nationalcioreview.com/video/eric-schmidt-on-ais-future-infinite-context-autonomous-agents-and-global-regulation/>
 25. Searle, John (2002). Can Computers Think? In:
Philosophy of Mind: Classic and Contemporary Readings (ed. D. Chalmers). Oxford & New York: Oxford UP. pp. 669-675.
 26. Taulli, Tom (2019). *Artificial Intelligence Basics: A Non-Technical Introduction*. Monrovia: Apress.
 27. Valeri, Pol (1988). *Sveske* (t. II). Glas: Banja Luka. Prev. K. Mićević.