UDC 327:004.8:338.2
DOI: 10.5937/spm94-59849
Original research article

Српска политичка мисао
(Serbian Political Thought)
No 6/2025.
Vol. 94
pp. 25-42

Marko Pejković*

Institute for Political Studies, Belgrade

ARTIFICIAL INTELLIGENCE AND CENTRAL ECONOMIC PLANNING

Abstract

In this paper, the author will attempt to elaborate and complement the existing arguments about the impossibility of artificial intelligence (AI) ever enabling centralized, efficient economic planning at the state level in the form of a government plan. The rapid development of AI and other technologies in recent years has reawakened the debate about the upcoming technical capacities that could perhaps, in the near future, overcome all the weaknesses of centralized economic planning known from the previous century. However, the lack of conceptual prerequisites in the domain of AI that are necessary for successful centralized planning actually argues in favor of the fact that even any experiment on this topic is both scientifically and economically unjustified. The introductory part of the paper provides an overview of the main concepts and dilemmas within the debate, the next part discusses mathematical paradoxes as conceptual obstacles to economic planning with the help of AI (the problem of self-reference), while the final part of the paper will be a critique of the attempts of AI ideologists-architects to overcome these obstacles through the eventual fusion of humans and AI machines.

Keywords: AI, central economic planning, market, self-reference, Gödel's theorems, algorithms, superintelligence

* E-mail: marko.pejkovic@ips.ac.rs; ORCID: 0009-0004-5920-1381

INTRODUCTION

In the last few decades, it is an undeniable fact that people around the world have increasingly relied on digital technologies, whether for their business obligations or their private daily lives, but no current high technology has captured the world's attention as much as AI, precisely because of the controversies that AI has caused (Luknar 2025, 146–147). In the field of healthcare, more and more patients are using devices and sensors that register and analyze instantaneous changes in their bodies (calorie consumption, sleep rhythm, pulse) in order to provide physicians with recommendations for the optimal treatment method. In education, automated assessment, translation, and personalized learning (adapted to the cognitive abilities and work habits of students) are playing an increasingly important role. In the field of business correspondence and marketing, AI applications are increasingly used to automatically write personalized emails, tweets, or blog posts. "Smart homes" are being built with the help of AI systems (from room heating/cooling control to security systems). Management is experiencing an increasing trend of relying on AI software tools for decision-making (automated delegation of tasks to employees according to their work profile), while retail chains are happy to leave the processing of customer requests and complaints to the latest AI developments within their call centers. Some AI theorists list all these areas of human life and the role of AI in them when they talk about the "automated society" (Elliott 2022, 20–21).

On the other hand, the popularity and success of AI in solving "micro" problems in various spheres of society have fueled the hopes of some economists that it is already possible to talk about using AI – together with other information technologies (supercomputers) – to solve problems that are at a much higher level of complexity. So, supposedly, we could already have a central economic plan that would "optimize" the entire national economy with the help of AI. On the one hand, supercomputers would be responsible for calculating in the so-called input-output tables of the plan the value of each final product as the chain sum of each direct/indirect working time spent on its production (Cockshott and Cottrell 1993, 47–48,178–179). More precisely, supercomputers would be responsible for the supply side. On the other hand, AI tools would be the best technological solution for predicting demand patterns (Dapprich and Cockshott 2023, 420–421). The economic trend within which this idea of the relationship between digital technologies/AI and economic goals is generated is called cyber-communism (Nieto 2023, 254).

However, cyber-communism is not the subject of this paper for three reasons. First, there is already an extremely fruitful critique of its theoretical premises (Lambert and Fegley 2023; Phelan and Wenzel 2023; Moreno-Casas, Espinosa and Wang 2022; Wang, Espinosa and Peña-Ramos 2021). Second, all of these critiques focus only on the supply side, while none of them address the problematic use of AI in cyber-communism on the demand side. Third, we are interested in the potential role of AI not only in the domain of forecasting demand for the central economic plan, but also in its total role in the entire economic process – both on the demand side and on the supply side. And not at the current level of AI development (narrow AI), but at the one that currently exists only in the future (superintelligence). We start from the self-evident assumption that if there are no arguments for centralist economic planning by the projected and more advanced stage of AI, there can be none for the current one, which is inferior.

To fully understand narrow AI and superintelligence, it is necessary to first define AI itself. The diversity and multiplicity of definitions of AI are a consequence of the ubiquity of its applications (Girasa 2020, 8). We believe, however, that there is one definition that best describes what AI researchers have been doing for the past 50 years and what they will continue to do in the future (Ertel 2017, 2) – AI is the study of how to make computers solve problems (those that humans currently solve better) in a more efficient manner (Rich, Knight and Nair 2010, 3). Indeed, AI has sought since its inception to create a comprehensive replica of human cognitive abilities to solve various problems and then to multiply them quantitatively and qualitatively. And this is whether we are talking about a multitude of applications, each of which, as a replica of exactly one of these human abilities, solves only one problem (which would be narrow AI) or about a single application that, as a replica of all these abilities, solves, in a better way than humans themselves, all human problems (superintelligence).[1]

[1] In addition to these two types of AI, there is also general AI that is structurally identical to superintelligence (a single application that solves all problems), but whose results (at least in a qualitative sense) are at the same level as those of humans (Bostrom 2014, 14, 16, 22, 52, 62, 70; Yampolskiy 2016, 41–50). However, since the projections of AI engineers are that general AI will be a relatively short-lived and transitional phase towards superintelligence, we will not mention it in the

Although superintelligence does not yet exist as a developed technology (and there are no indications that it will in the near future, in the coming years, or even decades), we can already test its conceptual foundations despite this fact. More precisely, we are interested in whether it is algorithmically absolutely consistent, i.e., are its algorithms able to always provide a logically meaningful answer to any question that would concern the subject of its work (which in our case is the national economy)?[2] If it is not algorithmically omnipotent *a priori*, then there is no point in further investigating its role in the domain of centralized economic planning in that direction. This dilemma is extremely relevant to social science because redirecting financial resources from studying this issue would achieve a lot in terms of savings, rationalization, and higher-quality investment in scientific projects. And since in the rest of the paper we will only examine superintelligence, for the sake of brevity, when we mention AI, we will refer exclusively to superintelligence (unless we emphasize otherwise that it is a narrow AI).

At the end of this introductory part, it is necessary to explain what we mean in this paper by centralist economic planning. If it has long been customary in science to define this phenomenon as "the replacement of the free market by a central monopoly that makes decisions about how and by whom what will be produced, as well as how the results of production (rewards) will be distributed" (van den Haag 1964, 27–28), there is no reason not to adopt this definition for the purposes of this paper. However, we specify this central monopoly precisely as AI.

rest of the paper, i.e., all the conclusions we draw for superintelligence also apply to general AI. Also, it is of utmost importance for the continuation of the work to keep in mind that superintelligence is one (and not many) and as such centralist in itself.

2 An algorithm is defined as "a set of well-defined rules for solving certain computational problems" (Roughgarden 2017, 1). An algorithm can also be understood as a "Turing machine" that, like a function, receives inputs and produces outputs according to a finite number of instructions. If a "Turing machine" continuously works without interruption (stopping), the problem is considered unsolvable (non-computational), and if it stops working by providing any answer – the problem is solvable (computational). Moreover, a "Turing machine" is designed so that it does not necessarily have to perform computational tasks in the sense of arithmetic, but can provide "yes-no" answers to mathematical questions or demonstrate proofs to complex questions regarding mathematical theorems (Nguyen 2024, 25).

AI AND THE PROBLEM OF SELF–REFERENCE

There is a belief among neoclassical economists that the mathematical-logical foundations of their theories are quite reliable (Winrich 1984, 987). And since neoclassical economics is the one that has been characterized as the "mainstream" of economic science (Boettke 2012, xvii, 104, 157, 268, 271–272, 284, 290, 324, 378), then such a belief applies to almost the entire discipline of economics. However, this belief is on extremely shaky grounds, due to the mathematical-logical insights into the problem of self-reference, which, admittedly, have not yet been incorporated into the domain of the social sciences (Nguyen 2024, 24; Winrich 1984, 987).

Self-reference occurs when we think not only about the world around us and its objects, but also when we think about our own thoughts, when we are self-aware – when we do not simply use symbols to denote external objects, but when we use symbols to denote the symbols themselves, i.e., our own thought processes (Winrich 1984, 988). This can be illustrated by the famous linguistic paradox: "this statement is false" (Winrich 1984, 988). As Winrich points out, this paradox leaves us in an eternal dilemma about the truth of this statement, because if the statement is true, then what it communicates must be true. And since it communicates that the statement itself is false, then the statement is at the same time false. Conversely, if the statement is false, then what is true is the opposite of what is communicated in the statement, and this would actually be the truth of the statement itself. Therefore, self-reference can also be understood as a statement about a statement – as soon as a statement refers to itself ("this statement is…") in terms of its truth or falsity (Winrich 1984, 989).

Starting from such linguistic paradoxes, mathematics itself has also been preoccupied with the problem of self-reference, as attempts have been made to examine whether even within mathematics itself we have paradoxes like the above-mentioned linguistic ones (Winrich 1984, 989). Winrich reminds us that one of the mathematical paradoxes was detected by Bertrand Russell, when he noticed that there are two types of sets in mathematics – some that do not contain themselves as members of the set (normal sets) and others that do contain themselves as members of the set (abnormal sets). Thus, the set of all cats would be normal, because it is not a cat in itself and does not contain itself. On the other hand, the set of all normal sets is paradoxical. For, if such a set is normal, it must not contain itself as a member of a set, but this is contrary to what its determinant says – all normal sets are part of it as a set (therefore, as a normal set it is also part of itself because it is normal, regardless of the fact that it is a set of sets). And if it is abnormal, it must contain itself as a member of its own set, but since all members of its set are normal sets, this would be contrary to the thesis that it is an abnormal set. And so on, there is an eternal circularity of indeterminacy (inconsistency) of the answer to the question of whether the set of normal sets is really normal or not (Winrich 1984, 989). Russell tried to remove these paradoxes from mathematics, but without success.[3]

Although some mathematicians besides Russell believed that it was possible to cleanse all mathematical expressions of the paradox of self-reference (completeness and inconsistency), Kurt Gödel proved that completeness and consistency are mutually exclusive (if mathematics is complete, it is not consistent, and *vice versa*). He went a step further and merged statements about mathematical expressions (metamathematics) with the arithmetic branch of mathematics, by encoding an entire system of mathematical notation through which it is possible to obtain any mathematical expression and showed that not only statements about mathematics (metamathematics) are subject to problems of self-reference, but even indirectly mathematics itself as a system in itself (Winrich 1984, 990–991). Thus, the statement "this expression is not provable" has its counterpart in the world of numbers. If the statement is true, it is not provable. In other words, if one insists on the consistency (accuracy) of a mathematical system, one must sacrifice its completeness (the belief that everything is provable) and vice versa.[4] Following Lowenheim-Skolem,

3 Russell attempted to eliminate these paradoxes of self-reference from mathematics. Elementary mathematical objects would be type 0, sets of such objects type 1, sets of sets type 2 (and so on). A higher set could not be contained in a set of a lower type, in order to avoid self-reference. But Russell realized that the price of avoiding self-reference would be the complete unusability of mathematics. For example, maximization, which is of crucial importance to the "mainstream" of economics, would be impossible - on the one hand, the maximum must always be from the set of real numbers (type 1), but if one insisted on eliminating self-reference, then the concept of maximum (which is by definition type 2) could not imply any maximum that was a real number (which is impossible to imagine in mathematics as a science). Therefore, he introduced the principle of reducibility, which softened the sharp division between "types", thereby effectively returning Russell fruitlessly to the beginning from which he had embarked on his endeavor (Winrich 1984, 990).

4 Theorem I states that there are arithmetic expressions for which no algorithm can decide whether they are true, and theorem II states that no mathematical system can prove its own consistency within itself (Nguyen 2024, 26).

Winrich concludes, quite correctly, that the very fact that some statements about mathematical expressions cannot be proven opens up the possibility for a non-dualistic interpretation of the world – axiomatic systems always allow for interpretations different from those initially intended (Winrich 1984, 992). That is, no matter how hard we try to categorize phenomena, there is always a chance that our categorizations will prove inadequate, because a phenomenon may meet all the criteria to fall under a given category (type), although it evidently does not belong there.

If the "total interconnectedness of nature" is something that cannot be completely and consistently captured by any axiomatic system (even mathematics), then this is all the more true of economic life, regardless of the efforts of neoclassical economists to make economics an exact science by applying axiomatic methods to economic variables (Winrich 1984, 992). In accordance with Gödel's theorems, it follows, according to Winrich, that it is not possible to make a strict logical distinction between economic concepts such as "utility" and "disutility", "efficiency" and "inefficiency", or "abundance" and "scarcity".[5] This is even better seen in the example of neoclassical choice theory, which assumes that individuals rationally choose the options available to them in order to maximize their preferences, but does not engage in any consideration of the preferences themselves (why individuals have a preference at all, whether they reconsider, change, or abandon it), precisely because it does not want to deal with self-reference, because self-reflection of one's own preferences (desires) is exactly that – self-reference.[6] According to Winrich, people are not programmed machines with fixed preferences, but are in a tense dialectical relationship with their preferences, so much so that they can even give them up altogether – depending on the cultural context in which they are immersed and which cannot be ignored (Winrich 1984, 996, 1000).

---

5 Although every language and self-reference is subject to imprecision and dialectics, Winrich believed that science should not transfer these "imperfections" from a "perfect" system to the domain of religion or dogma, much less adapt them to itself through sterile simplification and reduction, but rather adapt to them by reformulating its basic premises (Winrich 1984, 993).

6 The best example Winrich gives is the statement of a classic cigarette smoker: "I don't like the fact that I have the desire to smoke." This is absolutely impossible to fit into the choice theory, which assumes that individuals consciously and rationally make choices to maximize the realization of their desires that are positive for them. However, this statement clearly demonstrates that some desires can be experienced negatively, as something that is rationally undesirable, but which cannot be rationally resisted (Winrich 1984, 999).

Moreover, the problem is not only the self-reference of individual economic entities that would tomorrow be an insurmountable obstacle to the computational activity of AI, but also the self-reference of the entire economic system that would try to "explain" itself in the form of a centralist plan. This situation can be illustrated by Hayek's argument that any apparatus that performs classification (explains) must have a structure of a higher level of complexity than the lower-order objects that are subject to classification (Hayek 1952, 185). Here, complexity is defined as the number of possible classifications that come into consideration (Hayek 1952, 186). Although Hayek postulated this to prove that our brain will never be able to provide a complete explanation of all the individual ways in which it classifies neural stimuli (i.e., to explain itself), we believe that there is no reason why we should not apply this to any other entity that seeks to explain anything of the same level of complexity (itself), including therefore the central economic plan as the personification of the entire economic system. And although it is reminiscent of Russell's attempt to avoid self-reference with sets of different types, Hayek based his argument on Cantor's theorem that any classification system understood as a power set must always have a greater number of classes than the number of basic elements of the power set (subject to classification), and therefore, a power set cannot have itself as a member of its own set, because this would lead to the self-referential paradox that it is at the same time larger than itself (Hayek 1963, 340; Koppl 2010, 865–866). So a national economy that would "think" about itself in the form of a central plan would actually be an insoluble enigma for any algorithm.

But, regardless of all this, Gödel seemed to anticipate that someone in the future might argue the following – the computational limits mentioned in Gödel's theorems could be overcome when a new technology of such capacity and quality (such as AI) appears. Gödel did not deny this possibility, but he clearly pointed out that if it were possible, it would have nothing to do with humans. Namely, Gödel believed that the human mind is capable of formulating only a part of the mathematical intuition that it otherwise constantly uses while considering mathematical topics (Wang 2001, 184–185). One of the implications of Gödel's theorems is that every computer that operates with mathematical expressions sometimes generates certain truths (accuracies) that only we humans interpret without rigorous proof as such (not the computer itself), i.e., that the human brain somehow surpasses the computer itself in the sense that it can decide on some theoretical questions that are unsolvable for a computer (Wang 2001, 184–185).[7] However, based on what has been proven so far, the possibility of the existence of a computer (machine) that would be equal to mathematical intuition as a whole cannot be ruled out. But even if such a computer existed, there is no way to demonstrate evidence through which a person would gain scientific confidence that it is indeed such a computer. More precisely, Gödel says that if there were a computer that operated with axioms that did not obey his (Gödel's) theorems, we humans would not be able to reliably recognize that computer in that capacity, i.e.. we would not know with mathematical certainty whether all of its outputs (solutions) were correct or not (Wang 2001, 185). Even if humanity could build such a computer, it would either not work properly or, even if it did work properly, we would not be able to understand its solutions (Nguyen 2024, 27).

These considerations of Gödel are of extraordinary importance for our topic, because if there is an unbridgeable epistemological gap between the capacity of human knowledge and the capacity of a hypothetical supermachine, then the question arises of any sense in investing in and researching AI that would supposedly have the answer to all of humanity's problems (including economic ones). Admittedly, someone like Yuval Harari could argue that today perhaps 1% of people understand how the financial system works, but that does not prevent 99% of the others from voluntarily obeying the "rules of the game" of that same system, regardless of the fact that they understand very little or nothing about them (Harari 2020). However, there is an essential difference between these two epistemological gaps. The first, which relates to our futuristic AI, is absolutely imperative and unbridgeable. There is no possibility that any individual (no matter how talented or diligent) will understand anything about how such an AI works. Moreover, others do not have this opportunity. If we were to accept the decisions of such a machine as logically authoritative, we would do so not on the basis of rational insight into its workings, but solely on the basis of faith or superstition

[7] A problem that is insoluble for a computer algorithm but is solvable for the human mind concerns, for example, the aforementioned ordinal preferences, which are not quantified in intensity but are simply our judgment of what we prefer in relation to something else. Ordinal preferences cannot be transposed into cardinal units in any way, because the evaluation of two states or objects is an absolutely physical and personal act that is not subject to projection into the outside world, an act that is perceived and understood only by the individual and that he cannot communicate as rationally understandable – much less instill – in any other human being (Nguyen 2024, 32; Mises 1998, 97).

in its superiority. Or, for fear of possible punishment if the machine were to operate in the environment of some technocratic dictatorship (Nguyen 2024, 37). On the other hand, the second epistemological gap is bridgeable for at least two reasons. First, it is partly a matter of personal choice how much one will become familiar with the functioning of the financial system. Many people choose to devote themselves to something else (instead of finance), although a relatively the same level of cognitive abilities is required for both. Second, even those people who do not want to or do not have the intellectual abilities to deal with finance can still gain a certain level of understanding of the financial system indirectly, precisely through people who possess such field of expertise. If they do not want or cannot make the effort to understand the basic outlines of the topic that financial experts briefly present to them, they can rely on their reputation (which they need as experts or politicians in order to obtain business success/political mandates).

At the end of this part of the paper, it is extremely important to respond to the objection of the advocates of central planning that even if we accept the argument that central planning with the help of information technologies (in our case, AI) is defective (because perfect equilibrium is impossible to establish), then this applies at least to the same extent to markets that are not regulated by a central plan (Cottrell, Cockshott and Michaelson 2009, 224). However, this would be correct only on the condition that we accept a reductionist understanding of the market that views economic relations exclusively through the prism of algorithmically solvable tasks. And as we have already implied above by mentioning the cultural context, economic relations largely escape algorithmic logic (Nguyen 2024, 38–39). Furthermore, as Koppl notes, a market that is not regulated by the center is not the same as an economic entity with a clearly defined goal in advance. It is only a framework within which many different goals exist that interact. In other words, while central planners must know in advance what exactly they are planning, a market free from a central plan does not plan anything in advance (Koppl 2010, 862). Therefore, the equilibrium of a market understood in this way is something completely different from the neoclassical one. It is established gradually and evolutionarily, in the very process of its emergence. The reality of life is revealed at least in part during the very implementation of a large number of plans (decentralization), and the plans are corrected on the fly, by adding missing elements that have been discovered in the meantime (Devereaux, Koppl and Kauffman 2024, 503; Koppl 2010, 862). Moreover, the correction or optimization of plans is not carried out only with the help of algorithmic methods, but with local knowledge that can be both tacit knowledge and exclusively intuitive, capable of being encoded, decoded, and transmitted only in the context of a specific culture (Devereaux, Koppl and Kauffman 2024, 492, 502). However, bearing in mind that even when this tacit knowledge is exhausted, individuals are left with residual knowledge in "their worlds" that are absolutely unique and beyond the reach of even local cultures (Devereaux, Koppl and Kauffman 2024, 502).

Therefore, no algorithmic technology can legitimately plead for central planning of the national economy, because it cannot solve the problems of self-reference that are inherent to man as an economic subject. Problems of self-reference can only be solved by man, but with other, intuitive tools that cannot be subsumed under the algorithm. If, however, planning with the help of AI is carried out by ignoring the problems of self-reference, this is possible only with the artificial reduction of economic life and man in abstract models that, by the very fact that they simplify reality, can only have extremely negative repercussions on that same reality. And even if at some point in the future AI (or any other technology) could adequately plan the national economy, there is no conceptual assumption that could bridge the epistemological gap between the solutions brought by AI and human cognitive abilities – at least while in the physical sense the machine representing AI and the human himself are two separate entities. It remains to be seen in the next part of the paper what the prospects are for the eventual merging of man and AI machine into some new entity, i.e., what implications this could have for the argument about the alleged justification of futuristic AI technologies in economic planning.

THE MERGING OF HUMANS AND ARTIFICIAL INTELLIGENCE – IS CENTRALIST ECONOMIC PLANNING JUSTIFIED AFTER ALL?

In order to understand the attempt of apology for centralized economic planning from the perspective of the projected integration of man and machine into a new entity, we must examine the most basic features of the predicted principles of operation of AI understood as superintelligence. Beyond its definition that we have already given, superintelligence can best be understood figuratively as a circle with an outer shell and an inner core. The outer shell would consist of all precautionary measures, anticipated controls and regulations, or the most general frameworks (physical capacities and principles) that programmers would assign to superintelligence as a system of permanent or initial coordinates, within which all its activity generated by the aforementioned core would then flow. This core is precisely the internal logic of superintelligence.

The outer shell, therefore, contains a series of methods that can be divided into two groups. One consists of measures that programmers continuously take to "restrain" the internal logic of the superintelligence, i.e., its freedom of maneuver so that it does not go in this or that wrong direction for the welfare or even the survival of humanity (permanent coordinates), while the other consists of methods that would only give the initial impulse to the internal logic of the superintelligence (initial coordinates) from which it would then, without further intervention by programmers, somehow "arbitrarily" develop the trajectory of its actions on its own with full freedom of maneuver, but in such a way that the vital interests of humanity are not put in question (Bostrom 2014, 127–143). If the first group may play a greater role in the initial testing phases of the development of superintelligence, only the second group of methods allows the full potential of superintelligence to be exploited (Bostrom 2014, 185).

However, the initial coordinates can never cancel the fact that the internal logic of superintelligence will always be a kind of "black box", i.e., no one will be able to know how the superintelligence came to a certain solution, or what that solution will ultimately be – at least until the human-AI merger is complete (Bostrom 2014, 216). The "black box" is actually the aforementioned epistemological gap between AI and the human mind. Because superintelligence theorists assume and prefer to summarize all the initial coordinates into one central principle that could be reduced to the following: let the superintelligence do exactly what we as humans would do if we had its superhuman cognitive capacities (Bostrom 2014, 209–227; Russell 2019, 172–179, 314). In other words, there is no guarantee that a superintelligence in the economy would necessarily apply some superior method of central planning. It might even prescribe the inappropriateness of the plan and the continuation of the current market framework.

But even if we were to assume that AI programmers and architects could "direct" AI in the direction of central economic planning, by precisely specifying permanent or initial coordinates in the outer shell, so that the final outcome would be some kind of centralist economic planning – the problem of the "black box", i.e., the epistemological gap between the AI computer and the human mind, would still not be solved. A person would still not be able to understand the solutions that AI has come up with. Furthermore, it is not clear how AI, physically separated from humans, could overcome the problems of self-reference that only humans are capable of, regardless of the amount of information and the processing speed of their machines.

It would seem that the only way out for the proponents of planning would be in some kind of direct integration (or at least indirect connection) of humans with the machines that represent AI. Although Elon Musk advocated this integration for the sake of better control over the work of AI, this reason is again purely epistemological in nature, as in our case of economic planning. Namely, in March 2023, Musk was one of the signatories of an open letter from numerous leaders from the AI industry that called for a moratorium on the development of AI, because the upcoming AI systems are a great risk, given the possibility that they will outnumber, outsmart, make us redundant and replace us – thereby risking losing control over the civilization in which we live (Samuel 2024). Supposedly, in order to follow the trajectory of AI development, we as humans must have improved cognitive abilities regarding the speed of data processing in our brains, at least indirectly, through a high-throughput brain interface that would read data from our brains at a speed identical to that already connected to AI systems. Musk has already begun experimenting with a brain implant (*Neuralink*) that makes direct contact with neurons via electrodes. It is still unclear to what extent it is possible to reduce the discrepancy between the algorithmic capacities of AI and humans through these attempts by Musk, but it is also unclear whether there is an alternative to the union of man and machine AI – if the epistemological gap is to be bridged.

However, let us assume that the epistemological gap has been resolved, that a seamless integration of humans with AI has been achieved, i.e., that humans can easily understand all of AI's decisions (and how it arrived at them). Moreover, let us assume that AI is now not only able to be understood by humans, but also that, through access to neurons (and therefore all of the sensations from the human's immediate environment), it can gain insight into everything that has so far eluded algorithms – into humans' complete intuition, tacit knowledge of the local cultural context, and the ability to interpret all of the problems of self-reference. Would that then mean that all of the obstacles to effective centralist planning that we have outlined in this paper have been removed? In all likelihood, no. If AI were to acquire the ability to solve problems of self-reference, all humans would lose that ability, and AI would only be able to "solve" these problems maybe once. We have already seen above that self-reference is not a one-time, but a continuous process, full of dynamics and tension at the border between the subject of self-reference and its environment (culture) that surrounds it. Self-reference is not just a matter of a one-time interpretation, but above all a continuous, evolutionary reinterpretation by many subjects who, immersed in the cultural-local context, constantly change, complement, or abolish their decentralized economic plans of a "small range", without any total and rigid foresight in advance (before the process of implementing the plan itself begins in practice). Therefore, the fact that the AI would acquire the ability to interpret self-reference once would have practically no significance for the AI itself, if it is known that this was done at the cost of the disappearance of the entire context that is the *conditio sine qua non* for the reinterpretation of self-referential problems, and without which there is no effective economic planning. Context would disappear, because we have seen that its existence requires a large number of different self-referential subjects that interact, and here we would have only one single actor (AI). In other words, the problem of self-reference seems to be insoluble not only for AI machines that have the ability to work only algorithmically, but also for a futuristic AI that could hypothetically think non-algorithmically (in the form of some artificial amalgamation of all previous human self-referential interpretations into a single central type) – and this is only because centralizing self-reference would automatically mean paralysis of its reinterpretative capacities. Self-reference is not possible as a centralized plan, at least as a multiple process that can only make sense, but only as a highly decentralized system of numerous plans of self-conscious human subjects.

CONCLUSION

In this paper, we have shown, citing other authors, that AI cannot legitimately claim to be a centralized economic planning system, because it cannot solve the problems of self-reference as long as it operates according to algorithmic principles. These problems can only be solved by numerous market participants who alone possess the intuition to reinterpret self-reference in the complex local cultural environment during the implementation of decentralized economic plans, and within an evolutionary process that is subject to constant changes, additions, and cancellations (replacements) – decentralized plans can never be completely prescribed in advance, as is the case with centralized ones. On the other hand, in the hypothetical example of Musk's projected integration of man and machine AI, we would see that even such an AI that might somehow acquire the ability to interpret self-reference once by exploiting and amalgamating all human thoughts and sensations, would actually be impotent to repeatedly reinterpret self-reference (and reinterpretation is the only true meaning of dealing with self-reference). Namely, by renouncing their right to self-reference in favor of some kind of centralist resultant embodied by AI, people would actually abolish the decentralized cultural framework of local knowledge. And self-reference cannot be repeatedly processed without such a decentralized framework, in some kind of vacuum of the exclusive monopoly position of AI that would produce the central plan.

In addition, it turned out that experimenting with the role of AI in a planned economy is not only unjustified from the point of view of pointless investment and waste of money. There are also ethical issues at stake. If the state decides to wholeheartedly embark on the path of experimenting and putting AI systems into operation for the purposes of a central economic plan, what will happen to individuals who refuse to be gradually or abruptly drawn into the new economic order? Will they be left with the opportunity to do business in the old way (and to what extent)? Or will they be punished, perhaps even eliminated as "obstacles" to a bright future led by "omnipotent" AI? All these and many other questions regarding AI are waiting to be answered by the scientific community and the general public.

REFERENCES

Boettke, Peter J. 2012. *Living Economics*. Oakland: The Independent Institute.

Bostrom, Nick. 2014. *Superintelligence – Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Cockshott, W. Paul, and Allin Cottrell. 1993. *Towards a New Socialism*. Nottingham: Spokesman Books.

Cottrell, A., Cockshott, W.P., and Greg Michaelson. 2009. "Is economic planning hypercomputational? The argument from Cantor diagonalisation." *International Journal of Unconventional Computing* 5 (3–4): 223–236.

Dapprich, Jan Philipp, and William Paul Cockshott. 2023. "Input-output planning and information." *Journal of Economic Behavior and Organization* 205 (2023): 412–422. DOI: 10.1016/j.jebo.2022.10.043

Devereaux, Abigail, Roger Koppl, and Stuart Kauffman. 2024. "Creative evolution in economics." *Journal of Evolutionary Economics* 34 (2024): 489–514.

Elliott, Anthony. 2022. *Making Sense of AI*. Cambridge: Polity Press.

Ertel, Wolfgang. 2017. *Introduction to Artificial Intelligence*. Cham: Springer.

Girasa, Rosario. 2020. *Artificial Intelligence as a Disruptive Technology*. Cham: Palgrave Macmillan.

Harari, Yuval. 2020. "Read Yuval Harari's blistering warning to Davos in full." *World Economic Forum*. Last accessed on June 21, 2025. https://www.weforum.org/stories/2020/01/yuval-hararis-warning-davos-speech-future-predications/

Hayek, Friedrich. 1952. *The Sensory Order*. Chicago: The University of Chicago Press.

Hayek, Friedrich. 1963. "Rules, Perception and Intelligibillity." *Proceedings of the British Academy* 48 (1962): 321–344.

Koppl, Roger. 2010. "Some epistemological implications of economic complexity." *Journal of Economic Behavior & Organization* 76 (2010): 859–872.

Lambert, Karras J., and Tate Fegley. 2023. "Economic Calculation in Light of Advances in Big Data and Artificial Intelligence." *Journal of Economic Behavior and Organization* 206 (2023): 243–250. DOI: 10.1016/j.jebo.2022.12.009

Luknar, Ivana. 2025. „Društvena dinamika u eri veštačke inteligencije." *Srpska politička misao* 90 (2): 145–162. DOI: 10.5937/spm90-56968

Mises, Ludwig. 1998. *Human Action*. Auburn: Ludwig von Mises Institute.

Moreno-Casas, Vicente, Victor I. Espinosa, and William Hongsong Wang. 2022. "The political economy of complexity: The case of cyber-communism." *Journal of Economic Behavior and Organization* 204 (2022): 566–580. DOI: 10.1016/j.jebo.2022.10.042

Nguyen, Hai-Trieu. 2024. "The Incompleteness of Central Planning." *Quarterly Journal of Austrian Economics* 27 (4): 22–43. DOI: 10.35297/001c.126016

Nieto, Maxi. 2023. "Economic Calculation, Complexity, and Cyber-Communism: Bad News for the Austrian School." *World Review of Political Economy* 14 (2): 234–262. DOI: 10.13169/worlrevipoliecon.14.2.0234

Phelan, Steven E., and Nikolai G. Wenzel. 2023. "Big Data, Quantum Computing, and the Economic Calculation Debate: Will Roasted Cyberpigeons Fly into the Mouths of Comrades?" *Journal of Economic Behavior and Organization* 206 (2023): 172–181. DOI:10.1016/j.jebo.2022.10.018

Rich, Elaine, Kevin Knight, and Shivashankar B. Nair. 2010. *Artificial Intelligence*. New Delhi: Tata McGraw Hill Education Private Limited.

Roughgarden, Tim. 2017. *Algorithms Illuminated (Part 1: The Basics)*. San Francisco: Soundlikeyourself Publishing.

Russell, Stuart. 2019. *Human Compatible – Artificial Intelligence and the Problem of Control*. New York City: Viking.

Samuel, Sigal. 2020. "Elon Musk wants to merge humans with AI. How many brains will be damaged along the way?" *Vox*. Last accessed on June 21, 2025. https://www.vox.com/future-perfect/23899981/elon-musk-ai-neuralink-brain-computer-interface

Van den Haag, Ernest. 1964. "The Planners and the Planned." In *Central Planning and Neomercantilism*, ed. Helmut Schoeck and James W. Wiggins, 19–39. New Jersey: D. van Nostrand Company.

Wang, Hao. 2001. *A Logical Journey: From Gödel to Philosophy*. Hong Kong: Asco Trade Typesetting.

Wang, William Hongsong, Victor I. Espinosa, and José Antonio Peña-Ramos. 2021. "Private Property Rights, Dynamic Efficiency and Economic Development: An Austrian Reply to Neo-Marxist Scholars Nieto and Mateo on Cyber-Communism and Market Process." *Economies* 9 (165): 1–17. DOI: 10.3390/economies9040165

Winrich, J. Steven. 1984. "Self-Reference and the Incomplete Structure of Neoclassical Economics." *Journal of Economic Issues* 18 (4): 987–1005.

Yampolskiy, Roman V. 2016. *Artificial Superintelligence – A Futuristic Approach*. Boca Raton: CRC Press-Taylor&Francis Group.

Марко Пејковић[*]

Институт за политичке студије, Београд

ВЕШТАЧКА ИНТЕЛИГЕНЦИЈА И ЦЕНТРАЛИСТИЧКО ЕКОНОМСКО ПЛАНИРАЊЕ

Сажетак

У овом раду, аутор ће настојати да разради и допуни већ постојеће аргументе о немогућности да вештачка интелигенција (ВИ) икада омогући (централизовано) ефикасно економско планирање на нивоу националне економије у виду државног плана. Нагли развој ВИ и осталих технологија, као да је последњих година поново пробудио дебату о надолазећим техничким капацитетима који би можда могли у скоријој будућности да превазиђу све слабости централистичког економског планирања познатих из претходног века. Међутим, недостатак концептуалних предуслова у домену ВИ који су неопходни за успешно централистичко планирање, заправо говори у прилог да је чак и било какав експеримент на ову тему и научно и економски неоправдан. Уводни део рада даје преглед главних појмова и дилема у оквиру дебате, наредни део разматра математичке парадоксе као концептуалне препреке за економско планирање уз помоћ ВИ (проблем самореференције), док ће завршни део рада бити критика покушаја идеолога-архитеката ВИ да те препреке савладају кроз евентуални спој човека и машина ВИ.

Кључне речи: вештачка интелигенција, централистичко економско планирање, тржиште, самореференција, Геделове теореме, алгоритми, суперинтелигенција

[*] Имејл: marko.pejkovic@ips.ac.rs; ORCID: 0009-0004-5920-1381

[*] This manuscript was submitted on June 30, 2025, and accepted by the Editorial Board for publishing on November 10, 2025