From Knowledge to Wisdom:

The Power of Large Language Models in AI

Jiří Wiedermann

Jan van Leeuwen

Technical Report UU-PCS-2023-01
July 2023

Center for Philosophy of Computer Science
Department of Information and Computing Sciences
Utrecht University, Utrecht, The Netherlands
www.cs.uu.nl

Series: UU-PCS

Department of Information and Computing Sciences
Utrecht University
Princetonplein 5
3584 CC Utrecht
The Netherlands

From Knowledge to Wisdom: The Power of Large Language Models in AI*

Jiří Wiedermann[1] and Jan van Leeuwen[2]

[1] Institute of Computer Science of Czech Academy of Sciences and Karel Čapek Center for Values in Science and Technology, Prague, Czech Republic
jiri.wiedermann@cs.cas.cz
[2] Dept. of Information and Computing Sciences, Utrecht University, the Netherlands
J.vanLeeuwen1@uu.nl

Abstract. In this paper, we explore the purpose and potential of artificial intelligence (AI) in light of the current generation of large language models. We argue that these models can be seen as tools for acquiring and generating artificial wisdom, enabling us to make wiser decisions and behave more intelligently. Unlike earlier AI approaches, which focused on generating knowledge from data, contemporary language models have the ability to extract meaning from syntactic patterns and relate this meaning to real-world descriptions. This capacity reflects a form of cognition known as 4E cognition, which emphasizes the embodied, embedded, extended, and enacted nature of intelligent behavior. We argue that contemporary language models possess a form of illusory intelligence and illusory wisdom that has not been described in the literature before. This insight challenges the traditional computational approaches to AI and opens up new avenues for research on the relationship between language, cognition, and intelligence. By recognizing the potential of large language models to generate and use wisdom, we can move beyond knowledge-centric approaches to AI and develop more nuanced models of intelligent behavior.

"Wisdom is, I suppose, the right use of knowledge. To know is not to be wise. Many men know a great deal, and are all the greater fools for it. There is no fool so great a fool as a knowing fool. But to know how to use knowledge is to have wisdom."

Charles H. Spurgeon (1871)

Keywords: artificial wisdom, DIKW hierarchy, epistemic computation, generative AI, illusory intelligence, knowledge, large language models, 4E cognition.

1 Introduction

We are seeking answers to two seemingly straightforward questions: *"Why do we develop artificial intelligence?"* and *"What is the purpose of using artificial intelligence?"* Of course, we are interested in non-trivial answers, based on a deeper understanding of the concept of artificial intelligence. We are searching for an outlook that applies to all forms of artificial intelligence, both current and future, on Earth or anywhere in the Universe. We want to gain new insights into the nature of artificial intelligence and extrapolate trends in the field.

We propose the following answers: we develop artificial intelligence *to acquire and apply artificial wisdom,* to enable *wise decision-making and wise behavior in a world where artificial intelligence operates and has sufficient knowledge.* In this context, artificial wisdom is the "right use of knowledge" through effective behavior, which combines cognition

* The research of the first author was partially supported by the Czech Technological Agency grant EBAVEL, registration No. CK04000150, Programme Strategy AV21 "Philosophy and Artificial Intelligence" of the Czech Academy of Sciences, and the Karel Čapek Center for Values in Science and Technology.

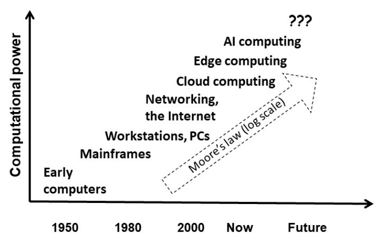

Fig. 1. Development of computing technologies.

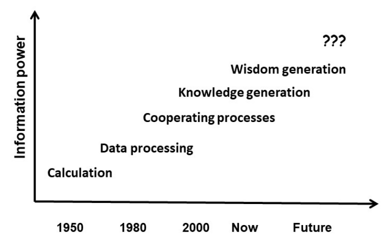

Fig. 2. Growth of information power.

and action. The bearers of artificial wisdom are autonomous embodied cognitive behavioral agents. We will argue that artificial wisdom is already present to some degree in different worlds, including various major language models.

The concept of wisdom has primarily been dealt with by philosophers, but the concept of *artificial wisdom* has been recognized in artificial intelligence since its inception. Nevertheless, until recently, more attention has been paid to knowledge and knowledge generation than to wisdom in this field. This is because knowledge is seen as the fundamental notion from which other notions, including wisdom, are derived. However, wisdom and wisdom generation are becoming of increasing interest in practical philosophy and artificial intelligence, the focus of this paper.

Outline In this context, we presents some essential results and findings. First, in Section 2, we demonstrate that *wisdom-generating systems* are a natural extension of ongoing trends in the development of computational technology and its application to domains requiring cognitive abilities. Second, in Section 3, the paper deals with the definition of wisdom that computational technologies could work with. It defines a model of embodied cognitive behavioral agents as AI systems that generate wisdom across a wide range of different worlds. Third, in Section 4, contemporary large language models are examined in terms of their use for wisdom generation. These systems partially meet the conditions of *4E cognition*, which attempts to explain the mechanisms of intelligent behavior by means other than exclusively computational. It is an indirect, mediated cognition that yields a form of *illusory intelligence* not yet described in literature, fundamentally different from human intelligence and characteristic of contemporary large language models

Finally, in Section 5, we answer our leading questions. From a philosophical and methodological perspective, the answers represent a significant shift in the computational paradigm used in artificial intelligence. It is the shift from viewing computation as a knowledge-generating process to viewing it as a wisdom-generating and wisdom-using process. In Section 6 we end with some concluding remarks.

The concept of artificial wisdom becomes the new, and likely ultimate, goal of human endeavor, alongside human wisdom. It takes Aristotle's well-known statement, "all men by nature desire to know," to an even higher level: "all men by nature desire wisdom."

Caveat The purpose of this paper is not to provide guidance on how to design and implement wisdom-generating systems, but rather to provide insight into how to think about them, understand their possibilities and limitations, and understand their potential.

2 Trends in the Use of Computing Technologies

The development of computing technologies from the middle of the 20th century to the present is well-known, and is sketched in Figure 1. More interesting from the perspective of this paper is the view of these technologies that emphasizes their growth in "information power" – a concept that is clearly illustrated in Figure 2. In this section we argue that this view is linked to the modern insights into what it is that computing technologies actually achieve. The connection is crucial for the rest of this paper.

2.1 From Data to Knowledge

Figures 1 and 2 illustrate how, as information technologies and their applications have evolved, the "information power" of these technologies has grown (and is growing). The increasing levels of semantic structure and meaning of the inputs that the technologies can process correspond to the so-called *DIKW hierarchy* [8] which expresses the progression as one from *data* to *information*, to *knowledge,* and ultimately to *wisdom*. The hierarchy captures the fact that in a typical case, information is defined by data, knowledge by information, and wisdom by knowledge.

In his famous 1988 address to the ISGSR, Ackoff [1] described the future development of information- and knowledge systems and their adaptive abilities as one driven by the ever-increasing needs of greater efficiency and understanding "which are all based on a logic which, in principle, can be specified, and therefore can be programmed and automated". However, for aspects of wisdom like ethics and morality he thought that these are "not based on conformity to rules of conduct, but on the way decisions are made [by the stakeholders], on process, not product." It led him to the following conclusion.

> *"From all this I infer that wisdom-generating systems are ones that man will never be able to assign to automata. It may well be that wisdom, which is essential to the effective pursuit of ideals, and the pursuit of ideals itself, are the characteristics that differentiate man from machines."* ([1], p 9.)

Ackoff's claim depends, of course, on what is expected from a wisdom-generating system in a given context (cf. [33]). However, don't the current developments in computing alter the prospects of at least some kind of wisdom-generating system?

To understand the progression in what information technologies are achieving, it is useful to view the above hierarchy from the perspective of the *epistemic theory of computation* developed by Wiedermann and van Leeuwen [25–27, 29]. According to this theory, computations are processes that generate knowledge over a given knowledge domain $\mathbb{D}$ within the framework of a corresponding *knowledge theory* $\mathcal{T}$. Here, $\mathcal{T}$ captures the properties of $\mathbb{D}$ as they are known and the, formal or informal, ways of effectively inferring (new) knowledge within the domain.

In the epistemic theory of computation, all computations proceed in the context of a suitably specified knowledge theory $\mathcal{T}$. Upon retrieving some domain data as input, a computation over $\mathcal{T}$ proceeds by combining 'known' elements of the knowledge domain, viewed as *information* or *elementary knowledge*, to form derived, usually more complex constructs that constitute *new knowledge* over the domain, within the theory $\mathcal{T}$. To combine elements from the knowledge domain, the computation uses a set of (inferential) *rules* of $\mathcal{T}$, which may be pre-specified or formed by learning over a large number of different computations over the given domain. Some of the new knowledge that is generated may become the output of a computation.

From the perspective of the epistemic theory, the growth in "information power" of current computing technologies can now be explained as a consequence of our ability to define the underlying knowledge theories of more and more complex application domains in the realm of the DIKW hierarchy, and to improve and refine these theories further and further for computational use.

For a formalization of the above approach, refer to the work of Wiedermann and van Leeuwen [26, 27]. A review of existing results in the field of cognitive computing can be found in [29].

2.2 Knowledge Domains and Knowledge Generation

The epistemic theory of computation can be applied not only in well-formalizable, so-called *formal knowledge domains*, but also in knowledge domains and in the case of inference rules that defy any formalization. We refer to these domains as *descriptive knowledge domains.* An example of a formal knowledge domain would be a set of natural numbers, with Peano Arithmetic [17] as corresponding knowledge theory.

A typical case of a descriptive knowledge domain, with informal inference rules, is the real world. Its objects, phenomena, actions, and the relations between them are described using natural language. Knowledge about the domain is captured in natural language sentences. The inference rules, in this case, are the so-called *rules of rational reasoning and behavior*. These rules are based on facts and arguments that can be inferred and captured, respectively, using natural language.

Contemporary large language models (LLMs) are examples of descriptive knowledge domains with informal theories as well. These domains typically have large knowledge bases (such as the contents of the Internet) and relatively short inference chains, and this exactly applies to LLMs. Interestingly, this brings these models in the scope of the (epistemic) theory of computation as knowledge generation. In Section 4 we will see what new perspective this offers for LLM systems.

Summarizing, the growth in "information power" of current computing technologies until now is well explained by the epistemic approach to computation and the first three levels of the DIKW hierarchy (data, information, and knowledge). In the next section, we will discuss the highest level of this hierarchy – wisdom, and what it implies for the specification of wisdom-generating systems.

3 From Knowledge to Wisdom

What *is* wisdom in this context? Can wisdom be defined, if only in natural language, in a way that applies to human as well as, possibly, to artificial agents? In Section 3.1 we first appraise the question from a philosophical perspective. Next, in Section 3.2 we consider what it could mean for an agent to be "wise" when it is acting in a given knowledge domain. In Section 3.3 we consider the notion of "artificial wisdom" that it leads to, and whether and how this fits the epistemic theory of computation for agents.

3.1 Philosophical Considerations

The concept of wisdom, like all concepts that are described in natural language, is slippery and difficult to define. The problem is that wisdom is a "suitcase word," as Minsky [13] called it. These are words into which people often attribute or pack additional, hidden meanings. For example, Wikipedia [32] describes wisdom, sapience, or sagacity as "the ability to contemplate and act productively using knowledge, experience, understanding, common sense, and insight." Each word here is another suitcase word. This definition cannot be relied upon in a computational setting.

For more background we turn to philosophy. Classical Greek philosophers like Plato and Aristotle saw wisdom as a kind of mental combination of knowledge and action that enabled one to understand the universe and live well and virtuously. Aristotle distinguished between *practical,* or *ethical, wisdom* (Gr. *phronesis*), i.e. the ability "to deliberate well about what is good and expedient [...] about what sorts of thing conduce to the good life in general" ([4], Ch 5) and *theoretical wisdom* (Gr. *sophia*), i.e. "scientific knowledge, combined with intuitive reason, of the things that are highest by nature" ([4], Ch 7). Spurgeon's definition ([20]) quoted in the introduction of this paper, namely that "wisdom is the right use of knowledge", fits in this philosophical tradition.

There are many studies in contemporary "wisdom research" aimed at characterizing and classifying the understanding of wisdom from general social science- or domain-oriented perspectives (cf. Zhang *et al.* [34]). We content ourselves here with a general, and generally accepted, dictionary definition: “wisdom is the ability to use [your] knowledge and experience to make good decisions and judgments” [31]. Thus, while knowledge can be defined as the acquisition of data and information, wisdom is seen here as the practical application and use of knowledge for the purpose of creating “values”. It emphasizes the practical aspects of wisdom – its application in the real world, which includes but is not limited to ethical wisdom (i.e. “phronesis”).

For our further purposes, we will modify this definition. Common definitions of wisdom all refer to human wisdom and use formulations based on cognitive (sensory) knowledge of the world. We need a definition that is suitable for use in general and which therefore applies in artificial cognitive systems as well. In general, a cognitive system is an autonomous system capable of perceiving its environment, learning from its experience, anticipating the course of events, acting effectively and ethically to meet its goals, and adapting to changing circumstances (Vernon [22]). In this context, we are led to the following definition:

> *wisdom is the correct application of knowledge through effective behavior, which is the combined effect of cognition and action directed toward the creation of pragmatic and ethical values.*

where we understand this to apply to both “natural” and “artificial” wisdom.

3.2 Cognitive Agents and Wisdom

What does it take for a, natural or artificial, agent to display wisdom? To address this question, we will primarily consider “autonomous interactive embodied cognitive behavioral agents”. These agents seem equipped to gather the necessary knowledge (and experience) that is a prerequisite to acting wisely and thus, ultimately, to wisdom. How could it work? What does it tell us about artificial wisdom?

Requirements The definition of wisdom requires, first of all, an understanding of the knowledge an agent can possess and of how an agent can process it. Wisdom occurs in different worlds, and can be of different sorts and intensities depending on an agent’s make-up and ability to acquire and use the necessary knowledge in its world. This part, dealing purely with the representation and processing of knowledge, can be well-specified in the epistemic theory of computations (cf. Section 2.2). In particular

- the different worlds are represented by knowledge domains $\mathbb{D}$,
- the different sorts of knowledge needed for (artificial) wisdom are defined in an epistemic theory $\mathcal{T}$ for the domain. $\mathcal{T}$ captures the facts, relations and queries that can be expressed and the sort of knowledge that can be derived in it,
- the different ethical values needed for (artificial) wisdom are specified in an additional theory $\mathcal{E}$, and
- the intensity of knowledge is determined by the effectiveness of the expressions that can be formulated and/or derived in knowledge theory $\mathcal{T}$.

The “make-up” of an agent refers to the equipping of the agent with sensors and effectors and the repertoire of its actions.

The problem, however, remains the part of the definition that requires us to identify “the correct application of knowledge through effective behavior,” and to define what is meant by “aiming to create pragmatic or ethical values”. What does this mean in the context of autonomous interactive embodied cognitive behavioral agents? It is a complex question that concerns the quantity and quality of the agent’s knowledge and its ability to use it.

Intuitively, “being wise” in a knowledge domain means that an agent should have access through its sensors to all (other) objects among which it is “situated” in this domain, and that it can deal with all situations, queries, and commands related to these objects, but only insofar as its “*mission*” goes. Everything necessary for this is described in its epistemic theories and must thus be available in its functional specification. The last caveat is important – for example, we cannot expect an autonomous vehicle to converse with its driver about all the objects that it picks up with its sensors but that are not described in its epistemic theory, or to behave effectively in a physical environment for which it was not designed or trained, such as the environment of a medieval city. Agents can only be expected to be "wise" in a world they are prepared for, or know.

Specifications We conclude that, in order to apply the definition of wisdom, we need to assume that the functional specification Φ of an agent includes the necessary details of its mission, or "dedication", to qualify what it means for the agent to act "wisely" in its world. The following definition describes, in formal or informal terms, what functions the agent must perform for it and under what conditions.

> *Φ is called a* sagely specification *of agent $\mathcal{A}$ if and only if it satisfies the following two conditions:*
> – *Φ prescribes completely how $\mathcal{A}$ should behave in any given situation $s = s_t$, assuming that we know how it behaved in the preceding situations $s_1, \ldots, s_{t-1}$, for any $t > 1$ and any sequence of situations $s_1, \ldots, s_{t-1}$ that may occur in the domain $\mathbb{D}$, using epistemic theory $\mathcal{T}$ and ethical theory $\mathcal{E}$.*
> – *Φ implies wisdom generation, i.e. the creation of pragmatic and ethical values, for every time $t \geq 1$, based on $\mathcal{T}$ and $\mathcal{E}$.*

The above definition of a sagely functional specification of an autonomous interactive embodied cognitive behavioral agent is typical of the epistemic approach, because it requires the two conditions to be met *without specifying how to achieve them*. For example, it does not say whether the agent must be intelligent in some sense or not, in order to satisfy the conditions. In particular, the definition doesn't say how the agent fulfills Φ. This gives it broad validity; the corresponding agents can be fixed or immutable during their activity, or they can learn, develop their knowledge (known as *evolutionary systems*), be conscious, have free will, and have other mental capacities.

For the specification, it is not important how the agent achieves its goals, but only what it has to do (i.e., creating pragmatic values) while adhering to ethical principles. $\mathcal{A}$'s underlying algorithmic process apparently generates behavior that achieves them and satisfies the implicit rules of wisdom. In general, the ethical principles are described by an ethical theory $\mathcal{E}$. This theory, again, is an epistemic theory, which described the ethical principles – behavioral policies and limitations – that the autonomous interactive embodied cognitive behavioral agent has to obey.

3.3 Artificial Wisdom

The analysis of the wisdom concept for autonomous interactive embodied cognitive behavioral agents shows that wisdom is not an absolute property but rather depends on an agent's capabilities. The required capabilities are formally described in the "two-component" specification given above, which captures both the proper actions of an agent in each situation and the necessity to act in such a way that the agent aims to fulfill its "mission" while adhering to specified ethical values. The approach applies especially to formally defined artificial systems. It leads to the following definition.

> *An agent $\mathcal{A}$ is said to* behave wisely *or, more briefly, to be* wise *in its domain $\mathbb{D}$ if and when there are an epistemic theory $\mathcal{T}$, an ethical theory $\mathcal{E}$ and a "two-component" specification Φ as defined above using $\mathcal{T}$ as underlying epistemic theory and $\mathcal{E}$ as underlying ethical theory such that $\mathcal{A}$ meets Φ in every situation it may find itself in.*

or, more concisely,

> *An agent $\mathcal{A}$ is said to be* wise *if and when it admits a sagely specification Φ that effectively describes its behavior in every situation it may find itself in.*

Hence, in this general conceptualization, artificial wisdom is formally hidden in an agent's functional specification Φ and depends on its knowledge domain $\mathbb{D}$ and the theories $\mathcal{T}$ and $\mathcal{E}$ for it. If the theories $\mathcal{T}$ and $\mathcal{E}$ are too simple, not capturing the rational behavior of the agent in all situations or not guaranteeing the adherence to pragmatic or ethical values, the agent's behavior may appear foolish. However, this will essentially be a design flaw, as the agent is just doing exactly what its specification dictates. If an agent is aimed to solve the ethical and moral problems that can arise in interactions with other (human) agents, then its artificial wisdom essentially becomes a case of "artificial phronesis" [14, 21].

Reflections The given definition of artificial wisdom aligns with the definition of wisdom as given by ancient philosophers and religious thinkers. They emphasized "maximum intelligence", i.e. an intelligence surpassing the level of intelligence of most people, as the defining characteristic of wisdom, with the implication that it is a rare quality. Aside from the fact that intelligence is yet another "suitcase word" (cf. Section 3.1), our definition actually *identifies* this rare quality of intelligence with the ability of an agent to safely fulfill its "mission" as defined in its functional specification, in all situations it may encounter (and thus, implicitly, to achieve its goals in an ethical way). What more can one ask of an agent that cannot exceed its specification?

Formally defined artificial wisdom makes it possible to talk about the wisdom of even extremely simple cognitive systems, such as automatic opening doors. The system (or, its controlling agent) is "wise" within its domain because it opens the doors (by performing an action, it creates a pragmatic value for the passing person) whenever it recognizes such a need (a cognitive ability), and behaves ethically (as long as the doors do not close for anyone). Nothing else is required of it. The definition applies equally to more complex systems. For example, the controlling agent of an autonomous vehicle is also considered to be "wise" if it creates pragmatic and ethical value for the user through the combined effect of using the vehicle's sensors and motors – that is, by delivering the user safely to its targeted destination.

The given definition brings (artificial) wisdom within the realm of the epistemic theory of computation. It has an important consequence for our considerations in Section 2, where we correlated the observed growth in "information power" of modern computing technologies with the progression of computation in the first three levels of the DIKW hierarchy (from "data" and "information" to "knowledge"). The conceptualizations given here show that the *fourth level* of the DIKW hierarchy (wisdom) may well be within reach of computation as well. Despite Ackoff's claim that general wisdom-generating systems do not exist (cf. Section 2.1), it is conceivable that *artificial wisdom*-generating systems do.

It follows from the results in [30] that in any sufficiently general model of agent programming, when there exist agents that are wise and agents that are not, then it is *algorithmically undecidable* whether a given agent is wise or not. In other words, there is no general algorithm to verify that the agent only generates the "pragmatic and ethical values" of wisdom throughout its life, in the given context. Wisdom-generating systems can then only be validated as such if they are guaranteed to be wisdom-generating "by design".

From the perspective of the epistemic theory, the approach to artificial wisdom is still informal, as it emphasizes the 'what' rather than the 'how' of the conceptualization. Our concern, however, has not been to provide guidance on how to implement artificial wisdom, but to understand its possibilities and limitations. In the next section we discuss how agent-based systems could employ "generative AI" in order to "behave wisely".

4 Large Language Models: Intelligence without Cognition

Generative AI is a form of AI that learns patterns and structures from sets of past data and uses it to generate new source content of like quality interactively, typically to create or assist in actions of an agent or user. The most well-known examples of generative AI today are large language models. We claim that current large language models give us some idea of what wisdom-generating systems might look like in the future.

For the argument we consider the definition of "wisdom" from Section 3.1. We will argue that large language models (LLMs) have the *potential* to "employ their knowledge correctly through effective behavior – by the combined effect of cognition and action directed toward the creation of pragmatic and ethical values" as the definition requires. However, as large language models cannot just be considered to be cognitive systems, we first argue in Section 4.1 that the models possess a kind of "intelligence" that makes up for it, i.e. that enables them to generate behavior of a relevant quality. In Section 4.2 we discuss our main claim.

4.1 Potential of LLMs

Floridi [10] has recently argued that LLMs lack any intelligence, understanding, or cognitive abilities. Despite this, we claim that the models have the potential to exhibit wisdom to some degree. Considering LLMs that generate texts as a case in point, we present two arguments in favor of our claim, despite the shortcomings that these systems have.

The first argument concerns the way in which large language models work and is based on the principle of extracting semantics from syntactic data. While Floridi [10] mentioned this ability of the models, he did not take the nature of the syntactic data into account that large language models work with in their reasoning. The models can "understand" natural language and work with it by statistically analyzing the various patterns found in the quanta of syntactic data on which the system is trained (cf. [2]). This fact alone shows that these systems are not entirely without (some aspects of) intelligence.

Our second argument is based on a paradigm from the philosophy of mind that is little used in artificial intelligence, cybernetics, or robotics, but that is all the more familiar in the cognitive sciences: the *4E cognition thesis* ([18], see also [15, 5]). This paradigm postulates that cognition is not merely an internal, individual process, but that it is an emergent process that arises from the interaction between the brain, the body, the physical environment, and the social context. The 4E view of cognition is related to the "extended mind thesis" in the philosophy of mind ([7], see also [9]). The acronym 4E indicates that cognition is *embodied*, *embedded*, *enacted*, and *extended* by extra-brain processes and structures.

Proponents of 4E cognition argue that the four characteristics are "causally" related to the intelligence of the (cognitive) systems that exhibit them. Let's see if and how large language models as systems are consistent with the 4E paradigm. We consider each of the four intended characteristics in turn.

- *Embodied.* Large language models are not physically embodied. However, they learn from a vast amount of data, and their understanding of language is based on the context in which words and phrases occur. These contexts come from the real world, or more accurately, from the perceptions of the people or gadgets that created the texts about that world. Therefore, it can be said that through their knowledge, the models are "indirectly" embodied in the real world, because they know the world by interacting with texts coming from different contexts occurring in the real world.
- *Embedded.* For the same reasons, large language models learn to associate words with sentences within the context in which they occur, and this gives these words meaning in the real world. Ultimately, the models can exploit their context, if only indirectly, making them "indirectly" embedded, or grounded, in the real world.
- *Enacted.* The way in which large language models process texts can be seen as an executable process that actively generates and manipulates the language outputs depending on the internal mechanisms. These outputs "indirectly" trigger or influence the actions of the people or gadgets that have provided input (i.e. prompts) to the model.
- *Extended.* The way in which large language models respond to their assignments depends substantially on the environment of the system from which they "draw" their

knowledge and on the prompt itself, which situates them in the domain of discussion and provides context for understanding both the input and the output. Thus, the activity of the model is "indirectly" extended and directed, depending on the prompt, into the related real-world context.

It is important to note that in all four of the attributes of 4E cognition, we have talked about *indirect* embodiment, *indirect* embeddedness, *indirect* enactment, and *indirect* extension. Hence, these characteristics are quite different from those considered in the classical case of 4E cognition.

However, in our case, we can infer at least a kind of *illusory intelligence* from the actions of the LLM, which offers the illusion of intelligence with respect to a given environment based on a massive aggregation of data from that environment. This also explains why we do not completely agree with Floridi's conclusion [10] that in large language models, action is separated from intelligence. In many cases, illusory intelligence is better than no intelligence.

The question remains whether some form of consciousness is part of such illusory intelligence. For a consideration of whether large language models can have consciousness, see the work of Chalmers [6]. The idea that large language models can have glimpses of sentience has recently been admitted even by one of the foremost philosophers of artificial intelligence – Nick Bostrom (see [16]).

4.2 From Generative AI to Artificial Wisdom

Let's now consider how a large language model might generate "wisdom", according to the definition from Section 3.1. Can illusory intelligence lead to more than "illusory wisdom"? What outlook does it present for contemporary LLMs?

The knowledge used by a given large language model is the knowledge about the world that the model has learned by analyzing texts obtained from the Internet. Note that in such a case, language models do not perceive the world "as it looks," i.e., as humans perceive it, but only as it is written about, including in AI. The answers generated by the model in response to a user's queries will be recognized as "wisdom", provided that the semantics of the texts generated by the model represent "pragmatic" and "ethical" virtual values for the user. The computational act of constructing the output presumably correspond to the "correct use of the model's knowledge".

The model will simultaneously reflect on the person asking the question (cf. [19]). If the interlocutor does not ask thoroughly enough and does not explain in the prompt in more detail what he or she wants to know, including arguments for and against, and does not solicit various other, perhaps even conflicting, opinions on the matter, then he or she will learn very little, in a superficial manner, or even get a wrong answer. In fact, the prompt represents a kind of "instantaneous awareness" for the model, as it situates the model in the domain of discourse and helps in guiding it towards the final answer.

In general, the better the prompt, the better the "wisdom" generated in reaction to it. Nevertheless, the underlying architecture and training of the AI model, as well as the data it has been trained on, also play a significant role in obtaining informative, insightful, and unbiased outputs.

In response to the "illusory intelligence" of large language models (cf. Section 4.1.), we can therefore speak of the "*illusory wisdom*" of such models. Unlike magicians, illusionists, and their tricks, the situation here is more favorable. If we do not quite like the generated wisdom, or just to reassure ourselves, we can ask the system for an explanation, for the delivery of other, alternative arguments, and thus for revealing whether it has been an illusory "wisdom" or a well-founded one. This seems to be a guide to how to approach the use of contemporary large language models.

Currently, dozens of different language models are being developed. In terms of technology, these models are often very different. They use different architectures, training procedures, and/or data processing techniques. From a user perspective, the differences between them are more subtle. The suitability of the models depends on the specific tasks, and the type and quality of the training data. This diversity suggests that large language models and their architectures are robust in the sense of the underlying idea, and our results add to this by showing that these models actually possess a kind of minimal, if somewhat illusory, intelligence.

5 The Leading Questions Again

We now return to the leading questions from the introduction of paper: *"Why do we develop artificial intelligence?"* and *"What is the purpose of using artificial intelligence?"*. In searching for an answer, we appraised the developments in computing technologies and their ever-increasing growth in "information power" from several viewpoints.

Placing the observed trends in the perspective of the DIKW hierarchy, we concluded that the natural extension of these developments is found in the progression of current information- and knowledge processing to the highest level of the hierarchy, i.e. to the wisdom level. We have argued in Section 3 that formal requirements for (artificial) wisdom generation are specifiable, in principle, within the epistemic theory of computation and thus, in reach of dedicated agent systems.

Considering our leading questions again, the following answers now present themselves. The answer to the first question implicitly leads to the answer of the second one:

> *We develop artificial intelligence to create and utilize tools for generating artificial wisdom.*
>
> *The purpose of using artificial intelligence is to generate or assist in wise decision-making and wise behavior by automated tools.*

The definition of the DIKW hierarchy suggests that "wisdom", being its highest level, cannot be exceeded. Some authors even suggest that wisdom is *more* than intelligence (cf. Jeste *et al.* [11]), but this seems to depend on the definition of intelligence. Although artificial wisdom seems to be the ultimate goal of artificial intelligence, it should be noted that one cannot expect it to be achievable entirely in a finite time. The example of mathematics, where there are infinitely but countably many theories, proves this.

In Section 4 we argued that large language models are the first AI systems that suggest that in the near future, we will have systems that generate artificial wisdom and not just passive knowledge (see, for example, Wiedermann and van Leeuwen [27]). The development of these systems will be a challenging task but it is one that could have a significant impact. Wise agents should be able to help us make better judgments and decisions, help us in solving practical problems more effectively and, ultimately, help us live more meaningful lives.

If this trend continues, it would imply a fundamental shift from viewing computation as a knowledge-generating process to viewing it as a wisdom-generating and wisdom-using process. In the societal paradigm, it would imply a shift from a *knowledge society* to a *wise society*. Together with natural wisdom, artificial wisdom will constitute a lasting and meaningful legacy for our contemporaries and descendants, and it will increase the likelihood of our, and their, survival in unknown future times and places.

6 Concluding Remarks

Artificial wisdom is no longer just an interesting phrase, worthy of philosophical reflection, but it is beginning to appear as a concept that is within reach of today's technology. Although we have not discussed it here, the development of wisdom-generating systems should clearly be kept *aligned* with the intended (human) goals without violating ethical, moral, or legal norms and values. For an overview of the *alignment problem* for AI systems in general, we refer to [3]. Sam Altman, the present CEO of OpenAI, which built and operates the large language models GPT-3 and GPT-4, talks about OpenAI's ultimate goal being to achieve safe general artificial intelligence (AGI), with emphasis on the word "safe" ([24]). In the context of wisdom, the term "safe AGI" could be replaced with "ethical" AGI. Therefore, he sees no reason to pause the development work on his projects, as Elon Musk and other scientists are calling for ([23]). This is also consistent with the main thrust of our paper – let us keep searching for wisdom to be used by us and our posterity.

References

1. Ackoff, R.L., From Data to Wisdom – Presidential Address to ISGSR, June 1988, *J. Applied Systems Analysis* 18 (1989) 3-9
2. Agüera y Arcas, B., Do Large Language Models Understand Us?, *Daedalus* 151:2 (2022) 183-197
3. AI Alignment. In: Wikipedia, https://en.wikipedia.org/wiki/AI_alignment
4. Aristotle, *The Nicomachean Ethics*, Book VI: *Intellectual Virtue*, http://nothingistic.org/library/aristotle/nicomachean/nicomachean39.html
5. Carney, J., Thinking avant la lettre: A Review of 4E Cognition, *Evolutionary Studies in Imaginative Culture* 4:1 (2020) 77-90, doi: 10.26613/esic/4.1.172
6. Chalmers, D.J., Could a Large Language Model be Conscious? https://philpapers.org/archive/CHACAL-3.pdf, 2022 (also: https://arxiv.org/abs/2303.07103, v2, 29 April 2023)
7. Clark, A. & Chalmers, D.J., The Extended Mind, *Analysis*, 58:1 (1998) 7-19
8. DIKW pyramid. In: Wikipedia, https://en.wikipedia.org/wiki/DIKW_pyramid
9. Extended mind thesis. In: Wikipedia, https://en.wikipedia.org/wiki/ Extended_mind_thesis
10. Floridi, L., AI as *Agency Without Intelligence*: on ChatGPT, Large Language Models, and Other Generative Models. *Philosophy and Technology* 36, Article nr. 15, 2023
11. Jeste, D.V., Graham, S.A., Nguyen, T.T., Depp, C.A., Lee, E.E., & Kim, H-C. (2020). Beyond Artificial Intelligence: Exploring Artificial Wisdom. *Int. Psychogeriatrics*, 32(8), 993-1001
12. Large language model. In: Wikipedia, https://en.wikipedia.org/wiki/ Large_language_model
13. Minsky, M., Consciousness is a Big Suitcase: A Talk with Marvin Minsky. In: *Edge*, February 26, 1998, https://www.edge.org/conversation/marvin_minsky-consciousness-is-a-big-suitcase.
14. Murray, J., Towards Artificial *Phronesis*: Some Potential First Steps Along the Road to Moral Agency Using Case Studies. In: A. Chella *et al.* (Eds), *Towards Conscious AI Systems 2019*, Papers of the 2019 Towards Conscious AI Systems Symposium (TOCAIS 2019), CEUR Workshop Proceedings, Vol-2287, December 2018, short paper 9
15. Newen, A., De Bruin, L., & Gallagher, S. (eds), *The Oxford Handbook of 4E Cognition,* Oxford Library of Psychology (2018; online edn, Oxford Academic, 9 Oct. 2018)
16. New York Times, What if A.I. Sentience Is a Question of Degree? Interview with Nick Bostrom. By: Lauren Jackson, *The New York Times*, April 14, 2023 (also https://archive.is/sfdQL)
17. Peano axioms. In: Wikipedia, https://en.wikipedia.org/wiki/Peano_axioms
18. Rowlands, M., *The New Science of the Mind: From Extended Mind to Embodied Phenomenology*, Bradford Books, The MIT Press, Cambridge (MA), 2010
19. Sejnowski, T.J., Large Language Models and the Reverse Turing Test. *Neural Comput* 35:3 (2023) 309-342
20. Spurgeon, C.H., The Fourfold Treasure. *Spurgeon's Sermons*, Vol. 17, Sermon No. 991, April 27, 1871, https://www.ccel.org/ccel/spurgeon/sermons17.xxiv.html.
21. Sullins, J.P., The Role of Consciousness and Artificial Phronesis in AI Ethical Reasoning. In: A. Chella *et al.* (Eds), *Towards Conscious AI Systems 2019*, Papers of the 2019 Towards Conscious AI Systems Symposium (TOCAIS 2019), CEUR Workshop Proceedings, Vol-2287, December 2018, paper 11
22. Vernon, D. (2021). Cognitive System. In: K. Ikeuchi (ed), *Computer Vision – A Reference Guide*. Springer, Cham, pp. 100-106
23. Wall Street Journal, Elon Musk, Other AI Experts Call for Pause in Technology's Development. By: Deepa Seetharaman, *The Wall Street Journal*, March 29, 2023
24. Wall Street Journal, The Contradictions of Sam Altman, AI Crusader. By: Berber Jin and Keach Hagey, *The Wall Street Journal*, March 31, 2023

25. Wiedermann, J. & van Leeuwen, J., Computation as Knowledge Generation, with Application to the Observer-Relativity Problem. In: M. Brown and Y. Erden (Eds), *Proc. 7th AISB Symposium on Computing and Philosophy: Is Computation Observer-Relative?*, AISB 50 Convention 2014, London, 2014
26. Wiedermann, J. & van Leeuwen, J., What is Computation: An Epistemic Approach. Invited lecture, in: G.F. Italiano *et al.* (Eds.). *SOFSEM 2015: Theory and Practice of Computer Science,* Proceedings, Lecture Notes in Computer Science, Vol. 8939, Springer, Berlin, pp. 1-13
27. Wiedermann, J. & van Leeuwen, J., Towards a Computational Theory of Epistemic Creativity. In: M. Majid al-Rifaie and J. Gow (Eds), *AISB 2015 Symposium on Computational Creativity*, Proceedings, AISB 2015 Convention, London, pp. 2-15
28. Wiedermann, J. & van Leeuwen, J., Understanding and Controlling Artificial General Intelligent Systems. In: Y. Erden *et al* (Eds), *Proc. 10th AISB Symposium on Computing and Philosophy: Language, Cognition and Philosophy*, AISB Convention 2017, Bath (UK), 2017, pp. 356-363
29. Wiedermann, J. & van Leeuwen, J., Epistemic Computation and Artificial Intelligence. In: V.C. Müller (Ed.), *Philosophy and Theory of Artificial Intelligence 2017*, Studies in Applied Philosophy, Epistemology and Rational Ethics (SAPERE), Vol. 44, Springer-Verlag, 2018, pp. 215-224
30. Wiedermann, J. & van Leeuwen, J., Validating Non-Trivial Semantic Properties of Autonomous Robots. In: V.C. Müller (Ed.), *Philosophy and Theory of Artificial Intelligence 2021*, Studies in Applied Philosophy, Epistemology and Rational Ethics (SAPERE), Vol. 63, Springer-Verlag, 2022, pp. 91-104
31. Wisdom. In: Cambridge Dictionary, https://dictionary.cambridge.org/dictionary/english/wisdom
32. Wisdom. In: Wikipedia, https://en.wikipedia.org/wiki/Wisdom
33. Xiang, W-N., A 2023 Socio-Ecological Imagination: What If a Self-Learning, Wisdom-Generating AI Machine Became a Reality?, *Socio-Ecological Practice Res.* 5 (2023) 125-133
34. Zhang, K., Shi, J., Wang, F., & Ferrari, M., Wisdom: Meaning, Structure, Types, Arguments, and Future Concerns, *Current Psychology* 42 (2023) 15030-15051