Reasoning AI (RAI), Large Language Models (LLMs) and Cognition

Victor Senkevich

Abstract

Do Large Language Models have cognitive abilities? Do large language models have understanding? Is the correct recognition of verbal contexts or visual objects, based on pre-learning on a large training dataset, a manifestation of the ability to solve cognitive tasks? Or is any LLM just a statistical approximator that compiles averaged texts from its huge dataset close to the specified prompts?

The answers to these questions require rigorous formal definitions of the cognitive concepts of ”knowledge”, ”understanding” and related terms.

Legend: “∆ ” – definition; “►” – statement / declaration; “●” – proposition; “○” – remark / explanation / clarification.

1. Basic Definitions

∆ **Perception** is a distinguishable part / subjective projection of any kind of entities / relationships of the real world or virtual environments.

∆ A **datum** is a representation of any kind (for example, an unstructured text / signs / machine view / visualization) of a single element / unit of perception.

∆ Any selection of datums forms data (dataset).

Data are elements of perception stored in any form, including verbal, machine and neurophysiological.

∆ **Meaning** is a representation of any kind (for example, awareness or verbal description, including formula, algorithm, program code) of a single act of relationship.

○ Elementary meaning is a representation of some relationship between objects of the real world or virtual entities / words / concepts / signs / symbols.

∆ **Knowledge** is a collection of meanings.

∆ **Understanding** is a gaining of the meaning.

∆ **Reasoning** is a mechanism for understanding. Reasoning is a discrete / transactional sequential / recursive process of forming knowledge from data, using feedback to remove uncertainty.

∆ **Intelligence** is an operator of meanings (Senkevich, 2022).

I.e. intelligence is a subject operating with meanings, forming, creating meanings, i.e. determining the existence of relations between real objects or virtual entities. Intelligence uses understanding to manipulate meanings in order to create knowledge / personal ontology.

∆ **AGI (Artificial General Intelligence) / HLAI (Human Level Artificial Intelligence)** is an entity capable of understanding.

Intelligence forms knowledge by solving cognitive tasks / discovering relationships between real or virtual entities.

∆ A **cognitive task** is any task initially containing uncertainty.

○ **Uncertainty** is the main characteristic of cognitive tasks.

Data is a source of knowledge. Intelligence is a processor generating knowledge from data.

2. Bayesian and Boolean Understanding

LLMs may have a “partial understanding”, which I would call a “Bayesian understanding”. This understanding is “partial” precisely because of its probabilistic nature. Such a “Bayesian understanding” arises as a result of Bayesian inference forming statistical relationships between entities / words in datasets.

”Real” AI / AGI / HLAI may have two types of understanding: logical / Boolean (or 3VL, 3-valued logic, which better represents uncertainty), responsible for rigorous reasoning, and statistical / ”Bayesian” / ”fuzzy”, similar to intuition / reflexes / instincts / patterns of behavior resulting from accumulated experience / learning / evolution.

► 3-valued logic (3VL) allows to build iterative and recursive procedures for forming propositions in a more natural way. The values “yes” or “no” complete the process of cognition. The value “unknown / undetermined” initiates a new stage of the cognitive process.

Patterns of behavior, understood in a broad sense, also become patterns of thinking. Instincts, unconditioned reflexes, conditioned reflexes, reactions, habits, mechanical actions, skills, intuition and even “common sense” are all just different types of behaviors that are ”slowly” formed by multiple repetitions and then “packaged” into a ”fast” system of thinking for rapid reproduction in real time when competitiveness and survival depend on the reaction speed. Such patterns of behavior are described by the terms “heuristics” / “mental shortcuts” in the Kahneman system 1 “fast” thinking (Kahneman, 2011). The concept of ”behavioral patterns” is more general, since it also covers unintelligent living beings whose behaviors / reflexes / instincts are formed by evolution

• “Bayesian understanding” may well reflect “common sense” / averaged “public opinion” or even intuition, but in general it does not correlate with the truth deduced logically.

• Bayesian (”statistical”/ ”fast”) and Boolean (”logical”/ ”slow”) understandings are both important, completely integrated and complement each other.

Learning by example (as well as pre-training on a large dataset) as the basis of ”intuition” and ”common sense” in statistically significant cases help a lot to make the right choice without much thought. But in infrequent, but important cases, it’s wrong.

► Using statistics instead of logic, we will always get averaged answers instead of correct ones. In simple cases, they will coincide almost always, in complex cases - almost never.

► Bayesian LLM’s ”partial” understanding is an analogue of Kahneman’s System 1 “fast” thinking.

The Bayesian understanding, which is responsible for “common sense” / intuition, contains statistically accumulated results of logical reasoning in order to quickly reproduce them as ready-made solutions. Text prompts entered by LLM users are triggers for reproducing the most contextually similar ”behavioral patterns” / ready-made text patterns from a large training dataset.

► Boolean AGI’s understanding is an analogue of Kahneman’s System 2 “slow” thinking.

Thus the inherent LLM Bayesian understanding can provide the part of “real” AGI responsible for averaged “common sense”, but only the “Boolean” understanding guarantees logical reasoning and the absence of contradictions / ”hallucinations” in AI-generated texts.

LLMs are not able to solve cognitive tasks, because it requires the ability to fully functional logical / Boolean reasoning. Learning by example is not enough to solve cognitive tasks, it requires an understanding that LLM does not have and that AGI should have.

• Learning is important and ”useful”, but it cannot compensate for the lack of understanding. This is the whole essence of LLMs.

LLM is just a statistical approximator that compiles averaged texts close to the specified prompts from its huge dataset. Meaningless approximation of contexts from different semantic domains leads to ”hallucinations” / mixing contexts of different meanings.

• Thus, the reason for LLM ”hallucinations” is simple, contexts from different semantic domains are mixed. This is a fundamental weakness of LLMs that cannot be solved by scaling.

A real understanding of a text is impossible without a semantic / ontological model that is at least partially compatible with the semantic model on the basis of which this text was created. Simply put, the ontological models / semantics of the writer and the reader should overlap. Real intelligence is mainly engaged in creating personal subjective ontologies, including those based on perceived texts. Such an activity requires the ability to reason, statistical learning by example is not enough.

Statistical Bayesian inference, which only LLMs are capable of, is an asymmetric unidirectional logical implication A→ B. This is not enough for a full-fledged reasoning, since this is only a subset of possible logical relations. Therefore, the understanding of LLM is partial, ”statistical”.

• Statistical ”understanding” / approximation: – is able to interpolate / average quite well; – is able to extrapolate / predict very mediocre; – is completely unable to generalize / understand.

• LLM is trained to recognize the known. HLAI / AGI should be able to understand the unknown.

• Statistical inference finds the most likely solution. The logical reasoning finds the right solution

The issue of seamless integration between LLMs and AGI is extremely important and fundamental. This is an integration between approximation and generalization. Between probability and reasoning. LLMs will provide associative big data, AGI will provide its logical processing.

3. Language and the Reasoning AI (RAI)

“The world is unpredictable, but understandable”.

► The trivial truth is that real AGI is a reasoning AI (RAI).

∆ Language is a universal tool for describing, modeling and knowledge representation.

Communication is based on language, broadly understood as any generally recognized symbolism. Language reflects collective knowledge. Individual knowledge is largely non-verbal and imaginative.

Any knowledge other than primary sensory data needs language as a symbolic system for representing, storing and sharing it. Any knowledge is based on ontology / hierarchical dictionary / semantic model / collection of relationships between real objects or virtual entities. Any ontology is a symbolic system that uses a language for referential representation of membership relations / hierarchical knowledge.

Definitely, we can think without language. Any symbolism and language appeared as a result of the ability to think, this is an elementary truth. But we cannot create and share knowledge (other than behavioral skills and habits) without language. Any personal ontology / object recognition / representation of the world requires the designation (symbolic, mental) of hierarchical / recognizable concepts, i.e. elementary language in any form.

• The process of understanding the natural language text consists in determining the relationships represented by this text.

A relationship is not a concept. But a concept is always a relationship. A relationship becomes a concept when it gets a name / symbol / designation. A symbol / word / sign always denotes some concept representing some relationship. Relationships are primary, symbols are secondary. Symbols without relations / meanings they represent are meaningless by definition.

• A text is meaningful if it represents a sufficiently complete order / hierarchy of relationships.

∆ Phrases are order relations that form meanings as a superposition of ontological / semantic representations of words.

• Verbal reasoning is based on the “patterns of thinking” / “patterns of meaning” and the representation of symbols / words / concepts as relationships preserved in a personal subjective ontology / semantic model.

Predictive ”world models” / ”language models” are fundamentally limited due to both the unpredictability of the real world and natural language texts that fully reflect this unpredictability. An adequate world model should not be ”predictive”, but ”explanatory”. It is useless to ”predict” the sequence of decimal places of the number Π. It is necessary to understand and explain the formula for calculating it. There is no need to “predict” the next word in a phrase you understand, as opposed to the one you are trying to guess.

• Prediction is based on probability. Understanding is based on certainty.

► Explanation is the best prediction.

4. Cognition of the Unknown and Cognition of the Unexplored

“Induction never can originate any idea whatever. No more can deduction. All the ideas of science come to it by the way of Abduction. . . Deduction proves that something must be. Induction shows that something actually is operative. Abduction merely suggests that something may be” (Peirce et al., 1934) .

Is the discovery of a new mathematical formula an act of cognition? Apparently, yes.

Is the creation of a chemical formula for a new organic macromolecule an act of cognition? Probably, yes.

Is finding the best move in a chess game an act of cognition? Probably not.

Is searching for information in the knowledge base an act of cognition? Apparently not.

The correct answers to these questions depend on our interpretation of the concept of ”unknown”. If finding the unknown boils down to finding the best alternative on a set of known alternatives, then such a procedure can not be considered cognition. If, in order to find at least one acceptable alternative, it is necessary to redefine the space of possible alternatives itself, then such a procedure is clearly an act of cognition.

Neural networks solve combinatorial problems well, albeit approximately. Where there is a large but well-defined space of alternatives and the evaluation / utility function is known. Creating a chemical formula for a new organic macromolecule and finding the best move in a chess game require going through a large number of alternatives. Such a procedure is not a full-fledged act of cognition, but since the search algorithms used by the neural network are adaptive, the impressive results obtained are very similar to cognitive ones.

► Thus, LLMs can quite successfully solve problems that can be considered cognitive due to the uncertainty associated with the need to sort through a large number of alternatives. However, in order to solve cognitive tasks that require redefining the space of alternatives itself, the ability to logical reasoning is necessary, which LLM lacks.

Redefining the space of alternatives is usually a generalization procedure, an essential part of the process of cognition. Abduction is the process of forming such generalizing hypotheses, creating a representation of a set / class based on known data about its elements. The very fact of having a common property is not a proof that an element belongs to a set / generalized class, nor is it a proof of the existence of the generalized class itself. But such facts make it possible to formulate a hypothesis of generalization, i.e. to implement abduction as part of a sequential process of cognition, because any confirmed hypothesis is knowledge.. Which, in fact, was claimed by the author of the term abduction, who believed that abduction is a mechanism for the emergence of scientific hypotheses (Peirce et al., 1934).

5. Cognition and Optimization

As mentioned above, LLMs can quite successfully solve combinatorial tasks associated with the need to iterate over a large number of alternatives defined on a large but limited domain. If all such alternatives correspond to the value of some optimization criterion / utility function, then such tasks are optimization tasks. In general, such tasks are not cognitive, although the results of solving these tasks may well be interpreted as new knowledge, for example, it may be a new chemical formula.

• Cognition is not optimization, although optimization can be part of the cognition process.

All optimization tasks are formulated in a rigorously defined space of alternatives with a given optimization criterion. However, the very formulation of such a definition is a cognitive task that is not an optimization one.

► The essence of the Moravec's paradox, which is the observation that reasoning requires very little computation than processing perception and recognition, is very simple – real intelligence is not computational.

Reasoning and cognition are generally not computations. The Kolmogorov's complexity of algorithms is uncomputable. Any available formula for the number Π is computable, or potentially computable, bearing in mind that the decimal notation of the number Π is infinite. But computing the shortest formula, the length of the description of which is actually the Kolmogorov's complexity, is fundamentally impossible. This does not mean that reasoning and cognition cannot be implemented in software. This means that such an implementation should not be computational / optimization, but a logical / sequential / recursive process capable of changing the conditions of the task at each step of the recursion.

The paradigm of "modeling" a reasoning / cognitive AI is futile. Models are computable. The architecture of the cognitive process is not a "model", but a "meta-model" capable of building models of the world / environment / domain itself, and therefore it is not a model.

Thus:

• Statistical training on huge datasets with billions of parameters will not help to reason logically. Computational recognition is not sequential cognition.

• LLMs will not be able to scale up to the AGI level by simple quantitative expansion. This is a cognitive problem that requires a qualitatively different approach.

Therefore:

► All approaches to AGI based only on optimization are wrong a priori.

6. Cognition, Reasoning and Feedback

As stated in the above definition, reasoning is a sequential / recursive process. It uses real-time feedback to eliminate uncertainty and shape the next step of cognition. Philosophers of the past called such a procedure of cognition a hermeneutical circle. Norbert Wiener considered feedback to be the basis of cybernetics. Feedback is the driving force behind a sequential cognition process. And, although the purpose of the cognition process is to get rid of uncertainty, uncertainty can both decrease and increase at each new step of the cognition process. Moreover, it is the creation of uncertainty that triggers the process of cognition, since any question creates uncertainty, and the answer removes it.

Norbert Wiener believed that all intelligent behavior is the result of feedback mechanisms. "I repeat, feedback is a method of controlling a system by reinserting into it the results of its past performance. If these results are merely used as numerical data for the criticism of the system and its regulation, we have the simple feedback of the control engineers. If, however, the information which proceeds backward from the performance is able to change the general method and pattern of performance, we have a process which may well be called learning." (Wiener, 1950)

► Thus, only sequential logical reasoning is able to perform the full-fledged process of cognition.

References

Kahneman, D. Thinking, Fast and Slow. Macmillan, 2011.

Peirce, C., Hartshorne, C., and Weiss, P. Collected papers of Charles Sanders Peirce. volume 5, Cambridge, MA, 1934. Harvard University Press.

Senkevich, V. Existence and Perception as the basis of AGI (Artificial General Intelligence). arXiv:2202.03155 [cs.AI], 2022.

Wiener, N. The Human Use of Human Beings: Cybernetics and Society. The Riverside Press, Cambridge, Massachusetts, 1950.