Could Neurocompositional Computing be the first actual catalyst in our road to human-level machine intelligence?

Christos Ioannidis (c.ioannidis@students.uu.nl)
MSc Artificial Intelligence Student, Utrecht University
BSc, MEng Electrical and Computer Engineer, Aristotle University of Thessaloniki

Philosophy of AI - Utrecht University
Supervisor: Dr. Emily Sullivan

1. Introduction

There have been some critical dilemmas surrounding AI research, and the debate about the method that will enable us to approach human-level machine intelligence, or strong AI [1], has probably been one of the most crucial and controversial ones. Two different schools of thought, the one of symbolic AI [2] – which focuses on using discrete symbols and rules to represent knowledge- and that of neural nets (sub-symbolic AI) [3]- which model data through interconnected layers and nodes-, have had alternating periods of blossom throughout the last decades, with the latter experiencing a surge in its popularity in recent times with the developments in Deep Learning. However, and despite the fact many prominent AI researchers are still being unwilling to concede that, it has been argued for a while that both of these methods might have some distinct limitations that diminish the prospects of reaching strong AI systems with either of them; or more concisely, either of them alone. As a consequence, the potential of hybrid models, that try to combine features and architecture from both, has been researched, with Neurocompositional Computing [4] being a particularly intriguing idea with encouraging early results.

Neurocompositional Computing, originally introduced as a concept from Cognitive Psychology, has led to a more holistic form of artificial intelligence that tries to mirror the duality of human cognition, described by the principles of Continuity and Compositionality. In fact, combining the fluidity of neural networks with the structured logic of symbolic systems, neurocompositional systems aim to achieve more robust learning. However, this is not the first time a new approach is introduced that has sparked a lot of hope and ambition in the AI research community. So is the attention and interest in this approach justified, or will it remain a concept that is useful in the narrow AI [1]– or, task-specific- domain? Motivated by its foundations, the problems it attempts to solve, as well as the first actual implementations – the target is to advance to the third generation of these systems in the time of writing – that have been very successful as will be discussed later, I firmly believe not only that there is more than enough merit to pursuing further development and progress in approaches in the Neurocompositional Computing domain for the purpose of reaching strong AI, but also that this hybrid approach is the first method with such a potential, as the traditional methods have hit what I consider to be an insurmountable roadblock.

To present the underlying reasons for this opinion, this paper will first describe in section 2 what we could expect from a strong AI system and the reasons why approaching it with the traditional methods seems all the more unlikely, before giving a brief description of neurocompositional computing and its novelty in section 3.1. Afterwards, I will provide arguments about why I believe that these systems have the potential to reach the aforementioned heights. Moreover, I make an attempt to argue about how we would evaluate these systems as well as how they would perform in the current machine intelligence tests.

2. Strong AI and why other approaches are not the answer

2.1 Defining Human-Level Intelligence

Human-level intelligence, or strong AI, is difficult to precisely define. However, it can be accepted that the target should be systems that can achieve higher level cognition and can understand not only concrete objects, but also abstract concepts as well. This has been supported strongly by the research community, with Aaron Sloman going as far as to claim that AI systems should be able to address deeper philosophical and scientific questions while comprehending the underlying principles of natural intelligence and cognition [5], and whereas that could be considered an exaggeration, as that is a trait that even humans sometimes do not possess, I acknowledge the merit of this statement.

Apart from that, it should also be anticipated by strong AI systems to be able to communicate effectively with a human. Sloman pointed out how replicating human-like intelligence may require machines that can integrate both discrete (symbolic) and continuous (sub-symbolic) forms of computation to perform that. I analyze this issue later on, connecting it with the Central Paradox of Cognition [6] that describes this particular duality of human minds, which are able to execute both types of computation. An important dilemma that should be answered, and has been a matter of conflict, is whether these two capabilities should be expected to be an inherent trait of AI systems, or something that can be learned with training, and I also refer to that in section 2.2.

Last but not least, one of the most important traits of human behavior since the first steps of our species' evolution is the ability to adapt to new situations. Considering how crucial it has been for our survival, it should be imperative that we should demand a similar attribute from any kind of intelligent machine.

2.2 Does any of the traditional approaches have any potential?

In order to advance toward the goal of human-level AI, the first question that should be addressed is why AI research has not only not managed to approach such a level of intelligence, but on top of that, especially with the traditional methods, why it has arguably not even come close to that. We can draw a few conclusions about certain existent speculations if we examine the development and the limitations of the two main approaches.

Symbolic AI had a period of blossom, especially during the period from the late 1950s to 1980s. During this period a lot of novelties and innovations were developed. A relatively successful category of them is the 'expert systems' in domains like medical diagnosis, with MYCIN a remarkable instance [7], and geological explorations. These systems attempted to mimic human expert decision-making using rule-based reasoning. A lot of

researchers, such as John McCarthy considered symbolic AI to be the future of machine intelligence, but nevertheless it became evident eventually that these systems have only limited potential. More specifically, researchers became aware that, as symbolic systems work on predefined rules, it is impossible to explain or navigate through but a fraction of the world, which has very complex composition [8]. I fully support this argument and believe, as mentioned before, that the lack of adaptability is detrimental to an AI system's potential for any kind of higher cognition. Furthermore, it is worth mentioning that we also live in a very fluid world with a lot of ambiguity in both concrete and abstract concepts and due to its rigidness, symbolic AI fails miserably in handling that.

Meanwhile, **Neural Networks**, or sub-symbolic approaches have been an unquestionable innovation ever since their conception due to the way they try to mimic the human neural system [3], and Deep Learning went a step further in helping this concept reach groundbreaking success. However, I am strongly convinced that it would be utterly improbable to see Deep Learning approaches perform well in anything but narrow and specific tasks.

Yann LeCun, Yoshua Bengio and Geoffrey Hinton wrote what has been often labeled as the 'manifesto' of Deep Learning [9], in a period where everyone was overwhelmed by the success its algorithms had. They cited the impressive results that Deep Learning Algorithms achieve in domains such as image recognition, and how impressed they were by their success in their ability to detect even very complex patterns in large datasets. It is worth noting especially that Hinton has been critical of symbolic manipulation in AI systems, and even in November 2020 declared in [10] that '*Deep Learning could do anything*', despite the fact that a lot of prominent researchers had already voiced their skepticism about this notion.

First of all, upon taking up a problem and despite being trained on it, Deep Learning systems cannot actually fully understand the concepts they handle. In fact, they struggle at completing tasks that require anything beyond pattern recognition and statistical analysis. They might provide an illusion of success in a lot of cases, which can be attributed to the fact that they are very capable at these mathematical functions. This can be proven by how rapidly their performance drops when any conditions or parameters of a problem they have been trained on are modified, as is stated by Gary Marcus [11], a vocal critic of Deep Learning. One way to interpret this issue is to advocate that Deep Learning systems do not understand the compositionality of information and how every structure or concept can be decomposed to simpler and more basic ones-and how could they, since they possess no mechanism that could enable it.

Additionally, Deep Learning systems lack the ability to grasp the complexities of language, as they are successful only when it comes to handling numerical data, which often contributes to an already opaque decision-making process with important lack of explainability.

Amazed by their rapid development, many researchers still support that these are just obstacles that will be overcome in the next years of Deep Learning research-but as aforementioned, I side strongly with the doubters and feel that these are insurmountable barriers that do not arise by limited development- on the contrary, I reach the conclusion that these problems have to do with the nature of the systems. Developers and researchers that focused on the development of the aforementioned 'hybrid approaches' shared this view.

The major question about these hybrid approaches concerned, as explained before, is whether we should focus systems where string manipulation is not an inherent trait, but one that can be learned. Jacob Browning and Yann Lecun in [12] acknowledge the necessity of enhancing Deep Learning approaches with skills such as common sense and reasoning, as well as the ability to handle linguistic data, a position that I consider to be the prominent one in the AI research community. However, they advocate that Deep Learning, with some enrichments, can actually learn to handle linguistic input through training. Others, such as Gary Marcus in [8] propose that symbol manipulated is a hard-coded function that is built into the system. We can extend this dilemma to whether, in human evolution, symbolic manipulation is something we can learn through exposure to it, or the mechanisms pre-exist in our perception. I propose a combination of the two; *that the mechanisms that allow us, humans, to be able to learn to handle language are inherent- and not the ability itself.* I base that notion on the fact that on the one hand, a human will not be able to communicate and process human language unless he is trained to do so, and on the other hand a human will vastly outperform a gorilla if exposed to symbol manipulation training at the same extent- the latter will only have very limited, if any, response, because the mechanisms that would allow them to learn to process language are not developed the same way as the human.

Consequently, we can refer to my earlier point about Deep Learning being based on pattern recognition and mathematical functions and about it struggling to grasp general concepts. My assessment is that without the ability to decompose the input into encodings of linguistic data to then provide to a Deep Learning model, the ceiling is too low for any algorithm; too complex calculations will need to be made for way too simplistic language processing. Even discounting the factor of the energy consumption that would be required for this process, the inefficiency of this method glaring.

3. Neurocompositional Computing

3.1 Neurocompositional Computing and the Central Paradox of Cognition

Some of the shortcomings that were emphasized above can be highlighted by the Central Paradox of Cognition that describes the duality of the human mind, that is based on the Continuity and the Compositionality principle [4].

On the one hand, the Compositionality principle is tied to the idea that complex information can be represented and understood by composing simpler elements or structures together. An important power that is connected to this principle is the compositional generalization, which is shared by humans and embraced by symbolic systems. This skill enables us to understand novel situations by encoding them as novel composition of familiar parts. It is stated by Smolensky that human cognition gets tremendous power by understanding that the world is strongly compositional, and it is very fair to assume that human-level AI systems need the necessary compositional encodings to understand the world too.

On the other hand, the Continuity principle, which is respected by neural networks, is based on the notion that cognitive processes manifest through smooth, gradient-based transitions across a multitude of states or representations. By embodying this principle, humans, as well as neural networks, excel in tasks that require the discernment of intricate patterns, the processing of real-world sensory inputs, or ambiguity in either simpler or more intricate concepts. Smolensky referred to that as the Central Paradox of Cognition, which concerns the dual nature of the human mind that can act both as a continuous neural computer and a compositional-structure computer.

As far as I am concerned, a system that attempts to simulate human thinking should respect both principles. That could be enhanced as an argument by the fact that from its conception, Neurocompositional Computing has already had positive and encouraging signs, with its structure and the manner in which it is being developed revolves around both of them. The contrast between this and the foundations of Deep Learning on complex statistical analysis alone or Symbolic AI systems on predefined rules is hard to ignore. Neurocompositional systems' novelty is the combination of input decomposition, even linguistic one, through the use of a tool referred to as *tensor product representation* (TPR) encodings, and the ability to make continuous calculations, as they entail a neural network part. As a notion and model foundation I argue that this can be described as a more holistic approach that has the potential for robust learning and deeper understanding of the processed information- and it has been very hard to argue this about any AI system that has been built on the traditional approaches, even the most successful ones.

3.2 History comments, current status and why they have the potential to succeed in the future

The first generation neurocompositional systems became two of the most successful networks, namely the Convolutional Neural Networks (**CNNs**) [13,14]- which could be highlighted as the 'parents' of the neurocompositional approach - and the **Transformers** [15].

CNNs use many layers of feature detector filters, that firstly aim to capture lower-level features, and later combine those features to understand the more complex structure or information -in other words, the higher-level features- of the image. I firmly believe that, for a neural network on its foundation, to have the philosophy of deconstructing an image, and find the roots of what builds it up, incorporating the essence of the compositionality principle, should not go unnoticed.

The latter have been a revelation in the Natural Language Processing field, and work by decomposing the input into simpler units of linguistic information and creating graphs of these units and the links connecting them. It is worth mentioning that they have led to developments such as ChatGPT and BERT, who have arguably been the closest any member of the general population has been to interacting and communicating efficiently with a machine. Its impact, and even implications, have been a major topic of research in a lot of domains, such as Ethics, Philosophy, Education and others [16,17], but they can at least prove the gravity that these innovations can have. I contend that the fact that the first types of networks that followed this methodology have had this substantial an effect, should not be understated.

Nevertheless, higher cognition was not achieved, and further improvements had to be made to move to the second generation. Eventually, though, the systems, or better, case studies, that were developed have been, from a host of points of view, impressive.

Namely, **QANET** was a case study, developed by Palangi [18] and was tasked to answer questions about Wikipedia articles. This system had the additional hurdle that it had to communicate in a human comprehensible way – and its approach to manage that, as with other systems of this generation, was with the allocation of roles to the TPR encodings of the symbolic output. QANET in particular, based these roles in the linguistic structural properties of the encodings, with the called *wh*-restrictor-phrase (like *what* famous event in history), as is described by Smolensky in [4], an important instance of that. Other case studies, such as **CAPTIONNET** [19,20] that is tasked to generate image captions, experiment with roles that are tied to the sequence of parts of speech, such as noun, verb etc., or other methods. In my opinion, this mechanism shows an attempt for actual comprehension of the speech, syntax and meaning, not just a processing of information. CAPTIONNET's example is an indication of a system that can, with some success, understand the concept of an entity performing an act, and is not limited to performing a complex series of math functions on an

array resembling an image – like Deep Learning alone does, and that shows a move towards more robust and trustworthy learning.

3.3 How we would evaluate these systems

A critical question that would define the whole matter is how we could evaluate these systems. Of course, defining the exact tests that should be created to determine strong AI is beyond the scope of this paper. However, we can base off of what has been developed in the past and draw our own conclusions.

First of all, I feel that we have reached the point where we can reach the consensus that tests like the Turing Test (TT) can be considered obsolete. I could still argue many neurocompositional-based systems could have huge success in it, especially considering how a system like Eugene Goostman- with significantly less functionality than, for example, ChatGPT, appears to have passed the TT, as did the latter [21]. After all, as Jamie Cullen supports in [22], TT belongs to a category of challenges that are based on deceiving the judges that the system approaches human-level intelligence. Other traditional tests like the Winograd Schema Challenge can also be completed by models that are specifically trained to handle it, as is detailed in [23], and therefore do not pose a holistic form of evaluation.

If new tests are to be created, it should, as far as I am concerned, be more of a collection of tests that aim to capture some key elements of human cognition. Firstly, speech generation and comprehension would be pivotal for it, to prove the capability of human-interaction, and it has been a quality that has been sought by previous evaluation methods as well. Furthermore, as mentioned, the comprehension of both abstractions and concrete objects should be taken into account for sure- after all, it is a main part of human cognition. Moreover, seeing how systems approach novel situations and how advanced their general problem-solving skills are would be important as well. The important part is that they try to evaluate human-level thinking and not how well a system can solve a single specific task, as with earlier challenges, where we had “successful” attempts from AI systems that were trained specifically to solve a particular problem

I am of the opinion that Neurocompositional Systems have to potential to excel in all the aforementioned domains. As far as the human communication is concerned, I would like to mention the experiment that is carried out and explained in [18], with the QANET excelling against other models in discerning the difference between the popular British TV character Dr. Who from the question “who”. I believe that this is the essence of the features that a lot of tests- such as the Winograd Challenge, presented in [24]- aim to capture: linguistic decomposition and thorough understanding of its meaning. Another case study, STORYNET [25], was designed to answer questions about short synthetic narratives and showed remarkable success, when compared towards other implementations, in handling

abstract storytelling elements alongside more concrete narrative structures, as well as in the compositional generalization attribute. Each particular study is bringing AI research closer to higher cognition. I am convinced that if we see the bigger picture, we should acknowledge these experiments as important steps towards our end goal of human-level machine intelligence.

4. Conclusion

All things considered, it should be unanimous that neurocomputational computing is a pivotal element in the advancement towards human-level AI, but a host of arguments can point us to the conclusion that it might be more than just that. To this day, higher cognition has been a distant dream for every developed system, whether symbolic, sub-symbolic or hybrid. However, I have come to the belief that in this particular type of models, the limitations are not blockers, and the mechanisms exist to surmount many of the existent obstacles and to cover a lot of ground in our journey towards human-level intelligence. Whether it will actually manage remains to be seen, but I reckon that it is the first time we can be this optimistic about it.

If we see improvements in the complexity of the probable TPR encodings to handle sequential input data more efficiently, or research attempts for even more advanced role assignments between the nodes of linguistic information, we could be very close to genuine human-computer communication. That among other points for further improvement can propel this architecture success in the road to strong AI. Whether it will actually manage to achieve that remains to be seen, but I reckon that it is the first time we can be this optimistic about it.

References

[1] K. Arkoudas and S. Bringsjord, “Philosophical foundations,” in *The Cambridge Handbook of Artificial Intelligence*, K. Frankish and W. M. Ramsey, Eds. Cambridge: Cambridge University Press, 2014, pp. 34–63

[2] J. McCarthy, "Programs with Common Sense," in *Proc. of the Teddington Conference on the Mechanization of Thought Processes*, vol. 1, pp. 75-91, 1959.

[3] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115-133, Dec. 1943. [Online]. Available: https://doi.org/10.1007/BF02478259

[4] P. Smolensky, R. McCoy, R. Fernandez, M. Goldrick, and J. Gao, "Neurocompositional computing: From the Central Paradox of Cognition to a new generation of AI systems," *AI Magazine*, vol. 43, no. 3, pp. 308-322, 2022.

[5] A. Sloman, "A Philosopher-Scientist’s View of AI," *J. Artif. Gen. Intell.*, vol. 11, pp. 91-96, 2020.

[6] P. Smolensky, "On the Proper Treatment of Connectionism," *Behavioral and Brain Sciences*, vol. 11, no. 1, pp. 1-23, 1988.

[7] R. Davis, B. Buchanan, and E. Shortliffe, "Production rules as a representation for a knowledge-based consultation program," *Artificial Intelligence*, vol. 8, pp. 15-45, 1977.

[8] G. Marcus, "Deep learning alone isn't getting us to human-like AI," *Noema*, 2022.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436-444, 2015. doi: 10.1038/nature14539.

[10] K. Hao, "AI godfather Geoffrey Hinton: Deep learning will do everything," MIT Technology Review, Nov. 3, 2020. [Online]. Available: https://www.technologyreview.com/2020/11/03/1011616/ai-godfather-geoffrey-hinton-deep-learning-will-do-everything.

[11] G. Marcus, "Deep learning: A critical appraisal," arXiv preprint arXiv:1801.00631, 2018. [Online]. Available: http://arxiv.org/abs/1801.00631

[12] J. Browning and Y. LeCun, "What AI Can Tell Us About Intelligence," *Noema Magazine*, June 16, 2022. [Online]. Available: https://www.noemamag.com/what-ai-can-tell-us-about-intelligence/

[13] K. Fukushima and S. Miyake, "Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Visual Pattern Recognition," in *Competition and Cooperation in Neural Nets*, S. Amari and M. A. Arbib, Eds. New York: Springer, 1982, pp. 267-285.

[14] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," *arXiv preprint arXiv:1706.03762*, June 2017, revised August 2023. [Online]. Available: https://doi.org/10.48550/arXiv.1706.03762

[16] B. C. Stahl and D. Eke, "The ethics of ChatGPT – Exploring the ethical issues of an emerging technology," *International Journal of Information Management*, vol. 74, 2024, Art. no. 102700. ISSN 0268-4012. [Online]. Available: https://doi.org/10.1016/j.ijinfomgt.2023.102700

[17] M. Montenegro-Rueda, J. Fernández-Cerero, J. M. Fernández-Batanero, and E. López-Meneses, "Impact of the Implementation of ChatGPT in Education: A Systematic Review," *Computers*, vol. 12, no. 8, Art. no. 153, 2023. [Online]. Available: https://doi.org/10.3390/computers12080153

[18] H. Palangi, P. Smolensky, X. He, and L. Deng, "Question-Answering with Grammatically-Interpretable Representations," arXiv preprint arXiv:1705.08432, 2017. [Online]. Available: http://arxiv.org/abs/1705.08432

[19] Q. Huang, P. Smolensky, X. He, L. Deng, and D. Wu, "Tensor Product Generation Networks for Deep NLP Modeling," in *Proc. of the North American Association for Computational Linguistics*, 2018.

[20] Q. Huang, L. Deng, D. Wu, C. Liu, and X. He, "Attentive Tensor Product Learning," in *Proc. of the American Association for Artificial Intelligence*, vol. 33, pp. 1344-1351, 2019.

[21] C. Biever, "ChatGPT broke the Turing test — the race is on for new ways to assess AI," *Nature*, vol. 619, pp. 686-689, July 25, 2023. [Online]. Available: https://www.nature.com/articles/d41586-023-02361-7

[22] J. Cullen, "Imitation Versus Communication: Testing for Human-Like Intelligence," *Minds and Machines*, vol. 19, pp. 237-254, 2009. [Online]. Available: https://doi.org/10.1007/s11023-009-9149-3

[23] V. Kocijan, E. Davis, T. Lukasiewicz, G. Marcus, and L. Morgenstern, "The defeat of the Winograd schema challenge," Artificial Intelligence, 2023, Art. no. 103971.

[24] H. J. Levesque, "The Winograd Schema Challenge," in *Proc. AAAI Spring Symposium: Logical Formalizations of Commonsense Reasoning*, 2011.

[25] I. Schlag and J. Schmidhuber, "Learning to Reason with Third Order Tensor Products," in Proc. of the Advances in Neural Information Processing Systems, pp. 9981-9993, 2018.