Aaron Saporito | Russell Kosovsky | Matthew Lee | Jay Nash |Derin Gezgin
COM316: Artificial Intelligence | Answers to Standard Questions #11
Meeting Date & Time: 12/2/24, 5:00-6:00 PM

What other types of problems can we solve using this method? In other words, the problem probably deals with a particular situation. Can we categorize what general category of problems this method can solve?

With the sum of our approaches, we can deal with a large variety of problems, however, all of our approaches (with the possible exception of GPT) require a deterministic environment to be fully effective. We do not need the problem to have a fixed "size" or fixed complexity, but it does need to be deterministic, not stochastic. Some of our approaches (such as MCTS) can deal with stochastic environments but are rendered less effective by it.

Assuming that this solution does not give us a fully autonomous artificial mind, what is holding us back?

In our approaches that attempt to simulate logical reasoning (semantic networks etc.), we lack an effective way of easily training on an unknown problem. On the other hand, approaches that can learn easily (neural networks, GPT), lack a logical structure to their answer and instead rely on probability to determine the correct action or solution. A neural network cannot show why it arrived at a solution (beyond a bunch of tensor math), unlike a human that could lay out the steps taken to arrive at that solution. This is why the technology to bridge knowledge-based AI is important. A 'black box' solution will always have limited credibility which is what knowledge-based AI can solve.

One of the peaks of current AI technology is generative AI. However, these models rely on repeated connections within their input data in order to build a probabilistic model. The result of this is that it can be difficult to prune the training dataset to create a mostly accurate model, but even at its peak, there is a small chance that the randomness of the model will produce incorrect solutions. I suppose this is somewhat similar to the human mind.

Can we restate this problem and/or add more tools to gain more ground in our search for the artificial mind? What small change will

force us to develop a solution that is one step closer to a fully autonomous artificial mind?

It could be helpful to not only consider the success of an artificial agent in an environment but also rate the success of an agent based on how well it "explains" its solutions. For example, a neural network could output both its solution and some other output that explains why that solution was chosen. The network could be evaluated both on the accuracy of the solution if the explanation made sense, and if the explanation followed with other previous explanations.

If we were to combine natural language processing sentiment analysis with an LLM of sorts, we could attempt to make a generative model with emotional factors in its decisions. This could look like a significant number of additional neurons as a dense layer. This could allow our artificial mind to make decisions based on emotional factors, which generally doesn't seem useful but could be beneficial in the realms of healthcare.

While these approaches can be helpful in a general scale of problems it is certain that there is still room for improvement and maybe a huge breakthrough in this area. A general intelligence that can learn by itself can be a possible breakthrough; at that point, we will not need to develop anything further and let the AI learn by itself. This can be dangerous as AI can run out of control.

Another approach of validating the correctness of LLM could be rather than creating output from a single pass, having a self-critic knowledge-based model that detects facts from LLM outputs.