AGI design

Tom Rochette <tom.rochette@coreteks.org>

July 24, 2025 — daae079c

0.1 Context

0.2 Learned in this study

0.3 Things to explore

1 Overview

- Opaque or transparent internals?
  - Neural networks are very opaque in term of understanding. It's possible to look at the resulting probability distribution and thus the actions that were chosen, but if we trace the output from the input, everything in between is more or less meaningless. We're talking about convolution of functions and weights. In other words, there's no "logic" per se, just the aggregation of probabilities learned on a training data set. In the end, this may reproduce very well how our brain works, but it makes it hard for us to do any type of inspection.
  - Transparent internals on the other hands most likely means to hand write components so that they may be understood: tasks, budgets, priorities, goals, objectives, etc. Even with transparent internals, looking at a decision may still result in an explanation that is probabilistic.
- Sequential vs parallel
  - Are the input/process/output tasks done in a sequential or parallel manner?
  - If done in a sequential manner, we need to manage how long the task may execute. Furthermore, we need to manage their execution quantum (how long they can execute within a iteration cycle).
  - In terms of how we understand neurons to work, we expect a task to propagate through the neural network in a feedforward manner. Some tasks may complete under a single pass (for things that appear "constant" to us such as vision, hearing and touch), while others may take many more passes (reasoning, math, problem solving, etc.). In this particular case,the process is done in a highly parallel fashion with the idea that the "signal" propagates through the neural network like it was a directed graph, thus an initial signal generating feedback loops that will maintain the process alive for a while.
- Internal vs external interruptions/Task eviction
  - Internal means that the process is taking care of how long it executes. Each process has to respect its deadline and it is expected that each and every process will act according to this policy.
  - External means that, like an operating system, the process will be evicted by the main process and that the process does not need to manage its execution duration. The process taking care of eviction is potentially able to save and restore the task state. If that is not the case, then when a task is evicted but did not complete, then it needs to start over, which means wasted computation.
- Interest/Attention/Focus management
  - How is attention managed? If something is being processed and said processing is taking a while, how can something more important stop it or take its place?
    - In any case we want to build a system that replicates the human brain, we need to be able to handle interruptions. If we fail to do so, we appear unresponsive, which may be socially unacceptable.

- If we interrupt some running process, how can we store the current state of the process so that it may be resumed later on?
  - To do so, we need to have access to the internals of the process. If we look at how operating systems work, this means being able to store current register values into a potentially slower memory system with more space.

2 See also

3 References