Systems Neuroscience and AGI

Tom Rochette <tom.rochette@coreteks.org>

August 30, 2025 — 861fb9d0

0.1 Context

This article contains most of the content of the slides of the presentation by Demis Hassabis available at https://www.youtube.com/watch?v=IjG_Fx3D0o0.

0.2 Learned in this study

0.3 Things to explore

1 Overview

- How can we know/measure we're making progress toward AGI
- Non-biological approach vs biological approach
- Issues with the non-biological approach
  - Brittle
  - Time-consuming to train
  - Poor at general learning
  - Difficult to acquire/generate new symbols
  - How do you refer to things outside of the agent? (symbol grounding problem)
- Biological approach
  - The brain as a blueprint
  - Covers a large class of approaches
- Different search spaces of possible AGI solutions
  - Regime 1: Small and dense search space
    - Not worth too much relying on the human brain design
  - Regime 2: Large and sparse search space
    - Worth a lot to rely on the human brain design
- Evidence points to regime 2:
  - Evolution has only produced human level intelligence once
  - Large non-biological projects failed to make progress

1.1 Approaches to AGI (from abstract to biological)

- Cognitive science architectures: SOAR (Laird/Newell), ACT-R (Andersen), OpenCog (Goertzel)
  - Unsatisfactory because they're besed on introspection and when changes in knowledge occurs, they have to modify their model to fit in this new understanding
- System neuroscience: the brain algorithms
- Brain emulation: Blue Brain (Markram), SyNAPSE (Modha)
  - Not telling us about the internal processes/functions going inside the brain
  - Relying on very intricate imaging techniques (at what level do we need to stop? Calcium ion channels? Atoms?)

1.2 Marr's three levels of analysis

- Computational: What - the goals of the system
- Algorithmic: How - the representations and algorithms
- Implementation: Medium - the physical realisation of the system

1.3 Rapid advances in neuroscience

- Revolution in cognitive neuroscience
- New experimental techniques
- Sophisticated analysis tools
- Exponential growth in understanding
- Actively conduct neuroscience research useful for building AGI

1.4 Role of neuroscience

- Likely that neuroscience will have a big role in building AGI
- As an orthogonal source of information to Machine Learning
- Provides direction: inspiration for new algorithms/architectures
- Validation testing: does an algorithm consistute a viable component of an AGI system?
- How can it not be a net benefit in the quest for AGI systems to add neuroscience knowledge into the mix?

1.5 The hybrid approach

- Combine the best of machine learning and neuroscience
- Where we know how to build a component
  - Use the latest state-of-the-art algorithms
- Where we don't know how to build a component
  - Continue to push pure machine learning approaches hard
  - In parallel, also look to systems neuroscience for solutions

1.6 Systems neuroscience procedure

- Extract the principles behind an algorithm the brain uses
- Creatively re-implment that in a computational model
- Result: a state-of-the-art technique and AGI component

1.7 Intermediate goals

- Full embodied physical robots: throws up complex engineering problems whilst distracting from the main problem of intelligence
- Toddler AGI: AI-controlled robot that display qualitatively similar cognitive behaviours to a young human child (~3yo)
- Massive breadth of capabilities required = extremely hard

1.8 Core AGI

- Core capabilities:
  - Conceptual knowledge acquisition/representation
  - Planning and prediction abilities

1.9 Concepts are key

- Knowledge in the brain

  - Symbols
  - Conceptual
  - Perceptual
- Equivalent machine learning algorithms
  - Logic networks
  - ???
  - DBN, HMAX, HTM
- So how does the brain acquire conceptual knowledge?

1.10 Hippocampal-neocortical consolidation

- Hippocampus sits at the apex of the sensory cortex
- High-level neocortex: association and prefrontal cortex
- Stores the memories of recent experiences or episodes
- Replays those memories during sleep at speeded rates
- Gives high-level neocortex samples to learn from
- Memories selected stochastically for replay
- Rewarded: emotional or salient memories replayed more
- Circumvents the statistics of the external environment
- (Hypothesis) Leads to abstraction and semantic knowledge

1.11 Interim milestones

- Build knowledge on top of existing knowledge
- Abstract classification: classification of empty/full containers
- Discovery of higher-order structures (eg. 123456789101112131...) What is the next number? Statistics is not enough
- Algorithms that can build sophisticated models of the environment (eg. play any card game just by observing a raw perceptual stream)
- Transfer learning: learning a response in one perceptual context, abstracting a rule, and applying it correctly in a new context
- Some impressive things have already happened:
  - MoGo - first program to beat a professional human go player
  - IBM's Watson - taking on human champions at Jeopardy quiz show

1.12 How to measure progress?

- One approach: measure success across a suite of tasks
- Ideally we'd like a more integrated measure of progress
- Algorithmic Intelligence Quotient (AIQ)

1.13 Predictions

- Systems neuroscience understanding will help inspire to several key components of the overall AGI puzzle
- System with transfer learning and conceptual knowledge acquisition capabilities will appear in the next 5 years
- Measurement tools charting progress are improving all the time
- Once interim milestones have been achieved, we will have a better understanding of of intelligence and the safety issues involved
- Probably ~20+ years for full human-level AGI but lots of interesting technologies will be built on the way

2 See also

- Marr's tree levels of analysis: https://en.wikipedia.org/wiki/David_Marr_(neuroscientist)
- Algorithmic Intelligence Quotient: http://arxiv.org/pdf/1109.5951.pdf

3 References

- https://www.youtube.com/watch?v=IjG_Fx3D0o0