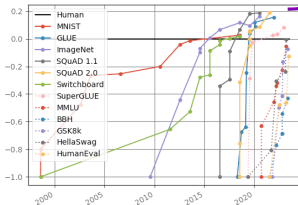


Recursive Self-Improvement

Tero Keski-Valkama

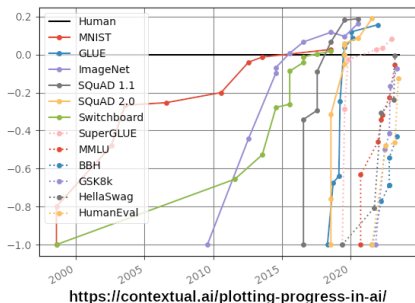
February 2, 2024



Introduction: Tero Keski-Valkama

- ▶ Tero Keski-Valkama is an AI generalist with over 25 years of experience spanning four countries, currently living in Spain.
- ▶ Worked with machine vision, complex control, SLAM, sensor fusion, semantic web, perceptrons, genetic algorithms, simulated annealing, SVMs, gradient boosting, deep learning, deep reinforcement learning, GANs, CNNs, Transformers, **LLMs**, **AGI**, embodiment, meta-learning, ...
- ▶ Robotics, pre-LLM chatbots, **LLM chatbots**, facial emotion recognition, automatic mapping, logistics, supply chain, **healthcare**...
- ▶ He has authored over 20 patents in the topic among countless other publications.
- ▶ Authored the first correct open source Google WaveNet implementation, the first communist AI, cofounded the second largest recurring AI event in Finland (AI Morning), ...

Capping at the Human-Level: Imitative Benchmarks



ACTIVITYNET A woman is outside with a bucket and a dog. The dog is running around trying to avoid a bath. She...

A. rinses the bucket off with soap and blow dry the dog's head.
B. uses a hose to keep it from getting soapy.
C. gets the dog wet, then it runs away again.
D. gets into a bath tub with the dog.

wikiHow How to determine who has right of way.

A. Stop for no more than two seconds, or until the light turns yellow. A red light in front of you indicates that you should stop.
B. After you come to a complete stop, turn off your turn signal. Allow vehicles to move in different directions before moving onto the sidewalk.
C. Stay out of the oncoming traffic. People coming in from behind may elect to stay left or right.
D. If the intersection has a white stripe in your lane, stop before this line. Wait until all traffic has cleared before crossing the intersection.

HellaSwag: Can a Machine Really Finish Your Sentence?

- ▶ All the **common AI benchmarks** we have tend to saturate at human-level, why?
- ▶ Because they are inherently imitative: They pick some tasks which are typically trivial for humans, but in which AIs still struggle. These tasks have correct answers produced by humans.
- ▶ Yes, even **BIG¹-Bench Hard** is largely imitative.

¹Beyond the Imitation Game

Imitative vs Non-Imitative Training

- ▶ Examples of **imitative** training: Pre-training of LLMs from web corpuses, instruct-tuning, anything with human generated labels.
- ▶ Examples of **non-imitative** training: **RLHF** (but specific in scope and limited by human bandwidth), **Code Llama** (but just one small task).
- ▶ We need a suite of open-ended, non-imitative tasks involving generalist skills with preferential machine judges.
- ▶ The tasks can be judged procedurally (like chess), or by LLMs (like social agent tasks).
- ▶ If a task is judged by an LLM, the act of judgement must also be judged \Rightarrow **Recursive self-improvement**.
- ▶ The training itself is just fine-tuning for example with **DPO**, but it requires access to weights.

Recursive Self-Improvement Suite

- ▶ I predicted **“unambiguous AGI”** to be reached before the end of 2023, because the step to do it is so easy. Sadly the large labs didn't start properly working on it.
- ▶ So I started my own project, [Recursive Self-Improvement Suite](#)², to show how it's done.
- ▶ It's a work in progress with lots of collected references. It will be a suite of tasks and related preference judgement processes, which creates a large corpus of synthetic data which can be used to fine-tune any LLM with e.g. DPO.
- ▶ Implemented tasks so far: “Create a programming task”, “Rank programming tasks”, “Rank programming task rankings”, “Create an evaluation code for a programming task”, “Rank evaluation codes”, “Rank evaluation code rankings”, “Solve a programming task”, “Rank programming task solutions”, “Rank programming task solution rankings”,

...

²Team: Tero Keski-Valkama, Asli Yaman

Try This at Home!

- ▶ The step to **recursive self-improvement** and **unambiguous AGI** is so small that anyone can now do it at home!
- ▶ You'll just need a large corpus of synthetic task performances involving generalist skills, preference evaluations on those, and apply DPO on some LLM with this data.

