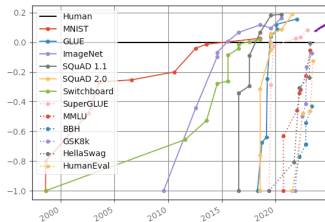# Recursive Self-Improvement

Tero Keski-Valkama
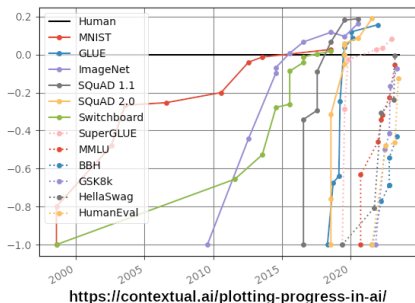
January 27, 2024

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

https://contextual.ai/plotting-progress-in-ai/

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 BIG[1]-Bench Hard is largely imitative.

_____

[1]Beyond 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 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 ⇒ **Recursive self-improvement**.

- ▶ The training itself is just fine-tuning for example with **DPO**, but it requires access to weights.

Recursive Self-Improvement Suite