

New study introduces a test for artificial superintelligence

Researchers propose a benchmark that uses advanced compression, a type of probability to find the most likely explanations for Artificial Superintelligence

LONDON, UNITED KINGDOM, April 22, 2025 /EINPresswire.com/ -- A research team led by [Dr Hector Zenil](#), Senior Lecturer and Associate Professor at the School of Biomedical Engineering & Imaging Sciences, King's College London and Founder of [Oxford Immune Algorithmics](#), a golden triangle

start-up (Oxford, Cambridge, KCL) has published a paper that puts forward an innovative benchmark—referred to as the SuperARC framework. Inspired by the original ARC challenge, this new test is designed to evaluate whether current and future AI systems possess the foundational characteristics required for artificial superintelligence (ASI).



Superintelligence

“

SuperARC is a step towards a universal test that can guide the search and defines a path towards Artificial Superintelligence as opposed to systems specialised in just emulating human-like behaviour.”

Dr Hector Zenil FRMS, lead and senior author

The findings show that current leading large language models (LLMs) such as ChatGPT, DeepSeek or Qwen are still far from AGI or ASI and often circle around the same intelligence levels. In contrast, they evaluated their own neurosymbolic approach, the Block Decomposition Method (BDM), which is more firmly grounded in principles of computational causality and demonstrated superior performance on a test case compared to current frontier LLM models, with the potential for further scalability given greater computational resources.

The new SuperARC test defines intelligence in terms of

recursive compression repeatedly condensing information to reveal deeper patterns not apparent to tools such as Large Language Model chatbots (LLMs) that heavily rely on pattern-matching. The test employs a type of specialised probability, drawing upon the equivalence

between compressibility and predictability established in the theory of randomness. The paper proves mathematically the equivalence between compression and prediction and exploits it to show how model abstraction and planning in the context of AI are formally two sides of the same coin.

The authors argue that intelligence is best measured by the ability to produce approximations to short computable hypotheses—one that can not only reconstruct but also predict data by running code in parallel to simulate many future states and pick the one that is closer to the observation at any given time. This perspective moves away from conventional, human-centric IQ-style tests, aiming for a more fundamental and agnostic measure of natural and artificial higher cognitive ability not based on human-centric single answers.

The test also makes it more difficult for current AI systems and frontier models to cheat which is something current systems are implicitly or explicitly doing when training on the very answers they are tested on.

Multiple leading LLMs (including GPT variants, DeepSeek, Qwen, Grok, Claude, Gemini, Meta, and others) were tested on tasks requiring model abstraction, inverse problem-solving, and short-sequence prediction and generation. Despite their linguistic prowess, these systems generally failed to model and generalise beyond trivial “print” solutions, that is, simply answering back with the original question. The study thus raises questions about LLM convergence on higher-level reasoning or merely amplifying pattern matching from ever larger sources of big data.

Results indicate no clear breakthroughs towards AGI or ASI, particularly for tasks requiring true model inference and robust planning. Notably, newer versions of the same LLMs occasionally performed worse than their predecessors, suggesting no consistent upward trajectory in less-human centric intelligence metrics. This also suggests LLM teams are focusing on optimising for ever changing human-centric tests for AI in their attempt to appear more intelligent rather than being so.

The authors of this study propose that future AI progress hinges on integrating symbolic inference with machine learning, arguing that “pure memorisation” approaches fall short of genuine comprehension. A shift to neurosymbolic models may be required to bridge the gap between advanced pattern recognition and true algorithmic inference.

Dr Zenil believes that this more reliable form of superintelligence—one not solely reliant on LLMs—will be key to addressing major human challenges such as disease, and will transform healthcare.

Read the [full paper here](#).

PR Team
OxfordIA

[email us here](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/805568367>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2025 Newsmatics Inc. All Right Reserved.