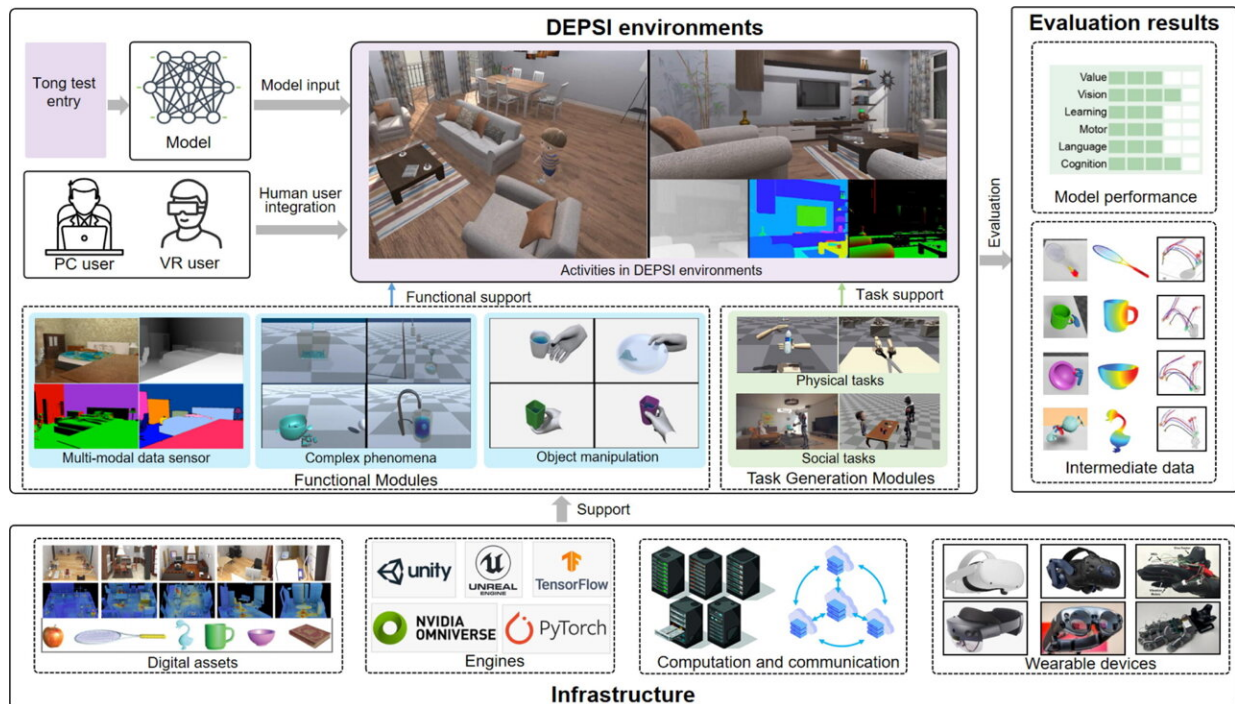# The Tong test: A new approach to evaluating artificial general intelligence

September 21 2023

The architecture consists of three main parts: infrastructure, DEPSI environments, and evaluation tools. With the support of physically and socially realistic task generation, the Tong test platform provides a standardized test pipeline for evaluating and benchmarking AGI models. PC: personal computer. Credit: Yujia Peng et al.

A recent perspective article published in *Engineering* proposes a new way for evaluating artificial general intelligence (AGI) with the introduction

of the Tong test (where "Tong" corresponds to the pronunciation of the Chinese character of "general," as in "artificial general intelligence"). This innovative approach aims to provide a standardized, quantitative, and objective evaluation system for AGI by focusing on dynamic embodied physical and social interactions (DEPSI).

The rapid advancement of the generative pre-trained transformer (GPT) series has brought AGI to the forefront of the artificial intelligence (AI) field. However, defining and evaluating AGI remained a challenge. The Tong test offers a fresh perspective on AGI evaluation by emphasizing the importance of DEPSI as a framework.

Traditionally, AI benchmarks have been task-oriented, but the Tong test shifts the focus towards ability- and value-oriented evaluations. The virtual platform proposed in the Tong test supports embodied AI in training and testing, enabling AI agents to acquire information, learn, and fine-tune their values and abilities interactively.

The Tong test proposes five critical characteristics that can serve as AGI benchmarks: infinite tasks, self-driven task generation, value alignment, causal understanding, and embodiment. These characteristics form the basis for a systemic evaluation system that allows for the delineation of AGI milestones through a virtual environment with DEPSI.

Unlike classical AI testing systems, the Tong test provides a more comprehensive and inclusive evaluation approach. It combines a general algorithmic testing paradigm with a human–AI interaction-based testing paradigm, taking inspiration from the philosophy of the Turing test. The Tong test's virtual platform generates unlimited tasks with dynamic embodied interaction scenarios, covering various dimensions of abilities and values.

The Tong test platform incorporates essential components such as

infrastructure, DEPSI environments, and evaluation tools. This combination provides a practical pathway for building an embodied platform with infinite tasks, where AI algorithms can be evaluated onsite with human interactions.

By introducing the Tong test, this perspective article paves the way for a standardized and objective evaluation system for AGI. It offers theoretical guidance for the development of AI algorithms while emphasizing the importance of DEPSI in evaluating AGI.

The authors of the perspective article believe that the Tong test has the potential to drive the field of AGI evaluation forward by promoting standardized, quantitative, and objective benchmarks. This will not only contribute to the further development of AGI but also foster greater transparency and understanding in the AI community.

Provided by Engineering