Issue Brief

Chinese Critiques of Large Language Models

Finding the Path to General Artificial Intelligence

论大模型

Authors
Wm. C. Hannas
Huey-Meei Chang
Maximilian Riesenhuber
Daniel H. Chou

CSET CENTER *for* SECURITY *and* EMERGING TECHNOLOGY

January 2025

Executive Summary

Large language models have garnered interest worldwide owing to their remarkable ability to "generate" human-like responses to natural language queries—a threshold that at one time was considered "proof" of sentience—and perform other time-saving tasks. Indeed, LLMs are regarded by many as a, or *the*, pathway to general artificial intelligence (GAI)—that hypothesized state where computers reach (or even exceed) human skills at most or all tasks.

The lure of achieving AI's holy grail through LLMs has drawn investment in the billions of dollars by those focused on this goal. In the United States and Europe especially, big private sector companies have led the way and their focus on LLMs has overshadowed research on other approaches to GAI, despite LLM's known downsides such as cost, power consumption, unreliable or "hallucinatory" output, and deficits in reasoning abilities. If these companies' bets on LLMs fail to deliver on expectations of progress toward GAI, western AI developers may be poorly positioned to rapidly fall back on alternate approaches.

In contrast, China follows a state-driven, diverse AI development plan. Like the United States, China also invests in LLMs but simultaneously pursues alternate paths to GAI, including those more explicitly brain-inspired. This report draws on public statements by China's top scientists, their associated research, and on PRC government announcements to document China's multifaceted approach.

The Chinese government also sponsors research to infuse "values" into AI intended to guide autonomous learning, provide AI safety, and ensure that China's advanced AI reflects the needs of the people and the state. This report concludes by recommending U.S. government support for alternative general AI programs and for closer scrutiny of China's AI research.

Introduction: Generative AI and General AI

Achieving general artificial intelligence or GAI—defined as AI that replicates or exceeds most human cognitive skills across a wide range of tasks, such as image/video understanding, continual learning, planning, reasoning, skill transfer, and creativity[1]—is a key strategic goal of intense research efforts both in China and the United States.[2] There is vigorous debate in the international scientific community regarding which path will lead to GAI most quickly and which paths may be false starts. In the United States, LLMs have dominated the discussion, yet questions remain about their ability to achieve GAI. Since choosing the wrong path can position the United States at a strategic disadvantage, this raises the urgency of examining alternative approaches that other countries may be pursuing.

In the United States, many experts believe the transformative step to GAI will occur with the rollout of new versions of LLMs such as OpenAI's o1, Google's Gemini, Anthropic's Claude, and Meta's Llama.[3] Others argue, pointing to persistent problems such as LLM hallucinations, that no amount of compute, feedback, or multimodal data sources will allow LLMs to achieve GAI.[4] Still other AI scientists see roles for LLMs in GAI platforms but not as the only, or even main, component.[5]

Pondering the question of how GAI can be achieved is important because it touches on options available to developers pursuing AI's traditional holy grail—human-level intelligence. Is the path—or a path—to GAI a continuation of LLM development, possibly augmented by additional modules? Or are LLMs a dead end, necessitating other, fundamentally different approaches that are based on a closer emulation of human cognition and brain function?

Given the success of LLMs, the levels of investment,[6] endorsements by highly regarded AI scientists, optimism created by working examples, and the difficulty of reimagining new approaches in the face of models in which companies have great commitment, it is easy to overlook the risk of relying on a "monoculture" based on a single research paradigm.[7] If there are limitations to what LLMs can deliver, without a sufficiently diversified research portfolio, it is unclear how well western companies and governments will be able to pursue other solutions that can overcome LLMs problems as pathways to GAI.

A diversified research portfolio is precisely China's approach to its state-sponsored goal of achieving "general artificial intelligence" (通用人工智能).[8] This report will show that—in addition to China's known and prodigious effort to field ChatGPT-like LLMs,[9]—significant resources are directed in China at alternative pathways to GAI by

scientists who have well-founded concerns about the potential of “big data, small task” (大数据,小任务) approaches to reach human capabilities.[10]

Accordingly, this paper addresses two questions: What criticisms do Chinese scientists have of LLMs as paths to general AI? And how is China managing LLMs’ alleged shortcomings?

The paper begins (section 1) with critiques by prominent non-China AI scientists of large language models and their ability to support GAI. The section provides context for views of Chinese scientists toward LLMs (section 2) described in online sources. Section 3 then cites research that supports China’s public-facing claims about the non-viability of LLMs as a path to GAI. In section 4, we assess these claims as a basis for recommendations in section 5 on why China’s alternative projects must be taken seriously.

Large Language Models and Their Critics

The term "large language model" captures two facts: they are *large* networks typically with billions to trillions of parameters, and they are trained on natural *language*, terabytes of text ingested from the internet and other sources. LLMs, and neural networks (NN) generally, are typologically distinct from "good old-fashioned" (GOFAI) symbolic AI that depends on rule-based coding. In addition, today's large models can manage, to different degrees, multimodal inputs and outputs, including images, video, and audio.[11]

LLMs debuted in 2017, when Google engineers proposed a NN architecture—called a *transformer*—optimized to find patterns in sequences of text by learning to "pay attention" to the co-occurrence relationships between "tokens" (words or parts of words) in the training corpus.[12] Unlike human knowledge, knowledge captured in LLMs is not obtained through interactions with the natural environment but depends on statistical probabilities derived from the positional relationships between the tokens in sequences. Massive exposure to corpora *during training* allows the LLM to identify regularities that, in the aggregate, can be used to *generate* responses to human prompts after the training. Hence, the OpenAI product name "GPT" (generative pre-trained transformer).

The ability of LLMs to "blend" different sources of information (which plays to traditional strengths of neural networks in pattern matching and uncovering similarities in complex spaces) has given rise to applications in areas as diverse as text summarization, translation, code writing, and theorem proving.

Yet, it has been hotly debated whether this ability to find and exploit regularities is sufficient to achieve GAI. Initial enthusiastic reports regarding the "sentience" of LLMs are increasingly supplemented by reports showing serious deficits in LLMs' ability to understand language and to reason in a human-like way.[13]

Some persistent deficits in LLMs, as in basic math,[14] appear correctable by plugins,[15] i.e., external programs specialized for areas of LLM weaknesses. In fact, such an approach—of a network of systems specialized in different aspects of cognition—would be more like the brain, which has dedicated modules, e.g., for episodic memory, math, reasoning, etc., rather than a single process as in LLMs.[16]

Some scientists hope that increases in complexity alone might help overcome LLMs' defects. For instance, Geoffrey Hinton, crediting an intuition of Ilya Sutskever (OpenAI's former chief scientist, who studied with Hinton), believes scale will solve some of these problems. In this view, LLMs are already "reasoning" by virtue of their ability "to

predict the next symbol [and] prediction is a pretty plausible theory of how the brain is learning."[17] Indeed, increases in complexity (from GPT-2 through GPT-4) have led to increased performance on various benchmark tasks, such as "theory of mind" [18] (reasoning about mental states), where deficits were noted for GPT-3.5.[19]

Other such deficits are harder to address and persist despite increases in model complexity. Specifically, "hallucinations," i.e., LLMs making incorrect claims (a problem inherent to neural networks that are designed to interpolate and, unlike the brain, do not separate the storage of facts from interpolations) and errors in reasoning have been difficult to overcome,[20] with recent studies showing that the likelihood of incorrect/hallucinatory answers increases with greater model complexity.[21]

In addition, the strategy of increasing model complexity in the hope of achieving novel, qualitatively different "emergent" behaviors that appear once a computational threshold has been crossed likewise has been called into question by research showing that previously noted "emergent" behaviors in larger models were artefacts of the metrics used and not indicative of any qualitative changes in model performance.[22] Correspondingly, claims of "emergence" in LLMs have declined in the recent literature, even as model complexities have increased.[23]

Indeed, there is the justified concern that the high performance of LLMs on standardized tests could be ascribed more to the well-known pattern matching prowess of neural networks than the discovery of new strategies.[24]

Still other criticisms of LLMs center on fundamental cognitive and philosophical issues such as the ability to generalize, form deep abstractions, create, self-direct, model time and space, show common sense, reflect on their own output,[25] manage ambiguous expressions, unlearn based on new information, evaluate pro and con arguments (make decisions), and grasp nuance.[26]

While these deficits are discussed in the western research literature, along with others such as LLMs' inability to easily add knowledge beyond the context window without retraining the base model, or the high computational and energy demands of LLM training, most current investment of commercial players in the AI space (e.g., OpenAI, Anthropic) is continuing down this same road. The problem is not only that "we are investing in an ideal future that may not materialize"[27] but rather that LLMs, in Google AI researcher François Chollet's words, "sucked the oxygen out of the room. Everyone is just doing LLMs."[28]

Chinese Views of LLMs as a Path to General AI (or Not)

A review of statements by ranking scientists at China's top AI research institutes reveals a high degree of skepticism about LLMs' ability to lead, by themselves, to GAI. These criticisms resemble those of international experts, because both groups face the same problems and because China's AI experts interact with their global peers as a matter of course.[29]

Here follow several Chinese scientists' views on LLMs as a path to general AI.

Tang Jie (唐杰) is professor of computer science at Tsinghua University, the founder of Zhipu (智谱),[30] a leading figure in the Beijing Academy of Artificial Intelligence (BAAI),[31] and the designer of several indigenous LLMs.[32] Despite his success with statistical models, Tang argues that human-level AI requires the models to be "embodied in the world."[33] Although he believes the scaling law (规模法则)[34] "still has a long way to go," that alone does not guarantee GAI will be achieved.[35] A more fruitful path would take cues from biology. In his words:

"GAI or machine intelligence based on large models does not necessarily have to be the same as the mechanism of human brain cognition, but analyzing the working mechanism of the human brain may better inspire the realization of GAI."[36]

Zhang Yaqin (张亚勤, AKA Ya-Qin Zhang) co-founded Microsoft Research Asia, is the former president of Baidu, founding dean of Tsinghua's Institute for AI Industry Research (智能产业研究院) and a BAAI advisor. Zhang cites three problems with LLMs, namely, their low computational efficiency, inability to "truly understand the physical world," and so-called "boundary issues" (边界问题), i.e., tokenization.[37] Zhang believes (with Goertzel) that "we need to explore how to combine large generative probabilistic models with existing 'first principles' [of the physical world] or real models and knowledge graphs."[38]

Huang Tiejun (黄铁军) is founder and former director of BAAI and vice dean of Peking University's (PKU) Institute for Artificial Intelligence (人工智能研究院). Huang names three paths to GAI: "information models" based on big data and big compute, "embodied models" trained through reinforcement learning, and brain emulation—in which BAAI has a major stake.[39] Huang agrees that LLM scaling laws will continue to operate but adds "it is not only necessary to collect static data, but also to obtain and process multiple sensory information in real time."[40] In his view, GAI depends on integrating statistical models with brain-inspired AI and embodiment, that is:

LLMs represent "static emergence based on big data" (是基于大数据的静态涌现). Brain-inspired intelligence, by contrast, is based on complex dynamics. Embodied intelligence also differs in that it generates new abilities by interacting with the environment.[41]

Xu Bo (徐波), dean of the School of Artificial Intelligence at University of Chinese Academy of Sciences (UCAS) (中国科学院大学人工智能学院) and director of the Chinese Academy of Sciences (CAS) Institute of Automation (CASIA, 中国科学院自动化研究所),[42] and **Pu Muming** (蒲慕明, AKA Muming Poo), director of CAS's Center for Excellence in Brain Science and Intelligence Technology (CEBSIT, 脑科学与智能技术卓越创新中心)[43] believe embodiment and environmental interaction will facilitate LLMs' growth toward GAI. Although the artificial neural networks on which LLMs depend were inspired by biology, they scale by adding "more neurons, layers and connections" and do not begin to emulate the brain's complexity of neuron types, selective connectivity, and modular structure. In particular,

"Computationally costly backpropagation algorithms … could be improved or even replaced by biologically plausible learning algorithms." These candidates include spike time synaptic plasticity, "neuromodulator-dependent metaplasticity" and "short-term vs. long-term memory storage rules that set the stability of synaptic weight changes."[44]

Zhu Songchun (朱松纯, AKA Song-Chun Zhu) dean of PKU's Institute of Artificial Intelligence and director of the Beijing Institute for General Artificial Intelligence (北京通用人工智能研究院) founded BIGAI on the premise that big data-based LLMs are a dead-end in terms of their ability to emulate human-level cognition.[45] Zhu pulls no punches:

"Achieving general artificial intelligence is the original intention and ultimate goal of artificial intelligence research, but continuing to expand the parameter scale based on existing large models cannot achieve general artificial intelligence."

Zhu compares China's LLM's achievements to "climbing Mt. Everest" when the real goal is to reach the moon. In his view, LLMs are "inherently uninterpretable, have risks of data leakage, do not have a cognitive architecture, and lack causal and mathematical reasoning capabilities, and other limitations, so they cannot lead to 'general artificial intelligence'." [46]

Zeng Yi (曾毅), director of CASIA's Brain-inspired Cognitive Intelligence Lab (类脑认知智能实验室)and founding director of its International Research Center for AI Ethics and

Governance,[47] is building a GAI platform based on time-dependent spiking neural networks. In his words:

"Our brain-like cognitive intelligence team firmly believes that only by mirroring the structure of the human brain and its intelligent mechanism, as well as the laws and mechanisms of natural evolution, can we achieve artificial intelligence that is truly meaningful and beneficial to humans."[48]

Criticisms of LLMs by other Chinese AI scientists are legion.

- Shen Xiangyang (沈向洋, Harry Shum AKA Heung-Yeung Shum), former Microsoft executive VP and director of the Academic Committee of PKU's Institute of Artificial Intelligence, laments that AI research has no "clear understanding of the nature of intelligence." Shen supports a view he attributes to New York University professor emeritus and LLM critic Gary Marcus that "no matter how ChatGPT develops, the current technical route will not be able to bring us real intelligence." [49]
- Zheng Qinghua (郑庆华), president of Tongji University and a Chinese Academy of Engineering academician, stated that LLMs have major flaws: they consume too much data and computing resources, are susceptible to catastrophic forgetting, have weak logical reasoning capabilities, and do not know when they are wrong or why they are wrong.[50]
- Li Wu (李武), director of the State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University, stated his belief that "current neural networks are relatively specialized and do not conform to the way the human brain works. If you desperately hype the large model itself and only focus on the expansion of parameters from billions or tens of billions to hundreds of billions, you will not be able to achieve true intelligence."[51]

Recognition of the need to supplement LLM research with alternative paths to GAI is evidenced in statements by China's national and municipal governments.

On May 30, 2023, Beijing's city government—within whose jurisdiction much of China's GAI-oriented LLM research is taking place—issued a statement calling for development of "large models and other general artificial intelligence technology systems" (系统构建大模型等通用人工智能技术体系).[52] Section three has five items (7-11), the first four of which pertain to LLMs (algorithms, training data, evaluation, and a basic software and hardware system). Item 11 reads "exploring new paths (新路径) for general artificial intelligence" and calls for:

Developing a basic theoretical system (基础理论体系) for GAI, autonomous collaboration and decision-making, embodied intelligence, and brain-inspired (类脑) intelligence, supported by a unified theoretical framework, rating and testing standards, and programming languages. Embodied systems (robots) will [train in] open environments, generalized scenarios, and continuous tasks.

The plan also mandates the following:

"Support the exploration of brain-like intelligence, study the connection patterns, coding mechanisms, information processing and other core technologies of brain neurons, and inspire new artificial neural network modeling and training methods."

Alternatives to LLMs were cited at the national level in March 2024, when CAS vice president Wu Zhaohui (吴朝晖, formerly vice minister of China's science ministry and president of Zhejiang University),[53] stated that AI is moving toward "synergy between large and small models" (大小模型协同), adding that China must "explore the development of GAI in multiple ways" (多路径地探索通用人工智能发展). The latter include "embodied intelligence, distributed group intelligence, human-machine hybrid intelligence, enhanced intelligence, and autonomous decision making."[54]

The following month Beijing's Haidian District government, with jurisdiction over 1,300 AI companies, more than 90 of which are developing big models,[55] issued a three-year plan to facilitate research in embodied (具身) AI. The plan defines "embodiment" as "the ability of an intelligent system or machine to interact with the environment in real time through perception and interaction" and is meant to serve as a platform for nationwide development. Its details include plans for humanoid robots facilitated by replicating brain functionality.[56]

Our analysis of public statements by government institutions and ranking Chinese AI scientists indicates that an influential part of China's AI community shares the concerns and misgivings held by western critics of LLMs and seeks alternative pathways to general artificial intelligence.

What Does the Academic Record Show?

Public statements by scientists are one measure of China's approach to GAI. Another is their record of scholarship. Prior reviews of Chinese technical literature determined that China is pursuing GAI by multiple means, including generative large language models,[57] brain-inspired models,[58] and by enhancing cognition through brain-computer interfaces.[59] Our present task is to examine the literature for evidence that Chinese scholars—beyond what positive features brain-based models have—are also driven to seek alternative paths by LLM's shortcomings.

Toward that end, we ran keyword searches in Chinese and English for "AGI/GAI + LLM" and their common variants in CSET's Merged Corpus[60] for papers published in 2021 or later with primary Chinese authorship. Some 35 documents were obtained. A separate query using web-based searches recovered 43 more papers.[61] 15 of the 78 papers were rejected by the study's lead analyst as off topic. The remaining 63 papers were reviewed by the study's subject matter expert, who highlighted the following 24 as examples of Chinese research addressing LLM problems that stand in the way of large models achieving the generality associated with GAI.[62]

1. CAO Boxi (曹博西), HAN Xianpei (韩先培), SUN Le (孙乐), "Can Prompt Probe Pretrained Language Models? Understanding the Invisible Risks from a Causal View," arXiv preprint arXiv:2203.12258v1 (2022).
2. CHENG Bing (程兵), "Artificial Intelligence Generative Content (AIGC) including ChatGPT Opens a New Big Paradigm Space of Economics and Social Science Research" (以 ChatGPT 为代表的大语言模型打开了经济学和其他社会科学研究范式的巨大新空间), *China Journal of Econometrics* (计量经济学报) 3, no.3 (July 2023).
3. CHENG Daixuan (程岱宣), HUANG Shaohan (黄少涵), WEI Furu (韦福如), "Adapting Large Language Models to Domains via Reading Comprehension," arXiv preprint arXiv:2309.09530v4 (2024).
4. DING Ning (丁宁), ZHENG Hai-Tao (郑海涛), SUN Maosong (孙茂松), "Parameter-efficient Fine-tuning of Large-scale Pre-trained Language Models," *Nature Machine Intelligence*, March 2023.
5. DONG Qingxiu (董青秀), SUI Zhifang (穗志方), *LI Lei* (李磊), "A Survey on In-context Learning," arXiv preprint arXiv:2301.00234v4 (2024).
6. HUANG Jiangyong (黄江勇), *YONG Silong* (雍子隆),[63] HUANG Siyuan (黄思远), "An Embodied Generalist Agent in 3D World," Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria, PMLR 235. 2024.

7. JIN Feihu (金飞虎), ZHANG Jiajun (张家俊), “Unified Prompt Learning Makes Pre-trained Language Models Better Few-shot Learners,” IEEE International Conference on Acoustics, Speech and Signal Processing, June 2023.
8. LI Hengli (李珩立), ZHU Songchun (朱松纯), ZHENG Zilong (郑子隆), “DiPlomat: A Dialogue Dataset for Situated Pragmatic Reasoning,” 37th Conference on Neural Information Processing Systems (NeurIPS 2023).
9. LI Jiaqi (李佳琪), ZHENG Zilong (郑子隆), ZHANG Muhan (张牧涵), “LooGLE: Can Long-Context Language Models Understand Long Context?” arXiv preprint arXiv:2311.04939v1 (2023).
10. LI Yuanchun (李元春), ZHANG Yaqin (张亚勤), LIU Yunxin (刘云新), “Personal LLM Agents: Insights and Survey about the Capability, Efficiency and Security,” arXiv preprint arXiv:2401.05459v2 (2024).
11. MA Yuxi (马煜曦), ZHU Songchun (朱松纯), “Brain in a Vat: on Missing Pieces towards Artificial General Intelligence in Large Language Models,” arXiv preprint arXiv:2307.03762v1 (2023).
12. NI Bolin (尼博琳), PENG Houwen (彭厚文), CHEN Minghao, ZHANG Songyang (张宋扬), *LING Haibin* (凌海滨), “Expanding Language-image Pretrained Models for General Video Recognition,” arXiv preprint arXiv:2208.02816v1 (2022).
13. PENG Yujia (彭玉佳), ZHU Songchun (朱松纯), “The Tong Test: Evaluating Artificial General Intelligence through Dynamic Embodied Physical and Social Interactions,” *Engineering* 34, (2024).
14. SHEN Guobin (申国斌), ZENG Yi (曾毅), “Brain-inspired Neural Circuit Evolution for Spiking Neural Networks,” *PNAS* 39 (2023).
15. TANG Xiaojuan (唐晓娟), ZHU Songchun (朱松纯), LIANG Yitao (梁一韬), ZHANG Muhan (张牧涵), “Large Language Models Are In-context Semantic Reasoners Rather than Symbolic Reasoners,” arXiv preprint arXiv:2305.14825v2 (2023).
16. WANG Junqi (王俊淇), PENG Yujia (彭玉佳), ZHU Yixin (朱毅鑫), FAN Lifeng (范丽凤), “Evaluating and Modeling Social Intelligence: a Comparative Study of Human and AI Capabilities,” arXiv preprint arXiv:2405.11841v1 (2024).
17. XU Fangzhi (徐方植), LIU Jun (刘军), Erik Cambria, “Are Large Language Models Really Good Logical Reasoners?” arXiv preprint arXiv:2306.09841v2 (2023).
18. XU Zhihao (徐智昊), DAI Qionghai (戴琼海), FANG Lu (方璐), “Large-scale Photonic Chiplet Taichi Empowers 160-TOPS/W Artificial General Intelligence,” *Science*, April 2024.
19. YUAN Luyao (袁路遥), ZHU Songchun (朱松纯), “Communicative Learning: a Unified Learning Formalism,” *Engineering*, March 2023.

20. ZHANG Chi (张驰), ZHU Yixin (朱毅鑫), ZHU Songchun (朱松纯), “Human-level Few-shot Concept Induction through Minimax Entropy Learning,” *Science Advances*, April 2024.
21. ZHANG Tielin (张铁林), XU Bo (徐波), “A Brain-inspired Algorithm that Mitigates Catastrophic Forgetting of Artificial and Spiking Neural Networks with Low Computational Cost,” Science Advances, August 2023.
22. ZHANG Yue (章岳), CUI Leyang (崔乐阳), SHI Shuming (史树明), “Siren’s Song in the AI Ocean: a Survey on Hallucination in Large Language Models,” arXiv preprint arXiv:2309.01219v2 (2023).
23. ZHAO Zhuoya (赵卓雅), ZENG Yi (曾毅), “A Brain-inspired Theory of Mind Spiking Neural Network Improves Multi-agent Cooperation and Competition.” *Patterns*, August 2023.
24. ZOU Xu (邹旭), YANG Zhilin (杨植麟), TANG Jie (唐杰), “Controllable Generation from Pre-trained Language Models via Inverse Prompting,” arXiv preprint arXiv:2103.10685v3 (2021).

The studies collectively address the litany of LLM deficits described in this paper’s sections 1 and 2, namely, those associated with theory of mind (ToM) failures, inductive, deductive, and abductive reasoning deficits, problems with learning new tasks through analogy to previous tasks, lack of grounding/embodiment, unpredictability of errors and hallucinations, lack of social intelligence, insufficient understanding of real-world input, in particular in video form, difficulty in dealing with larger contexts, challenges associated with the need to fine tune outputs, and cost of operation.

Proposed solutions to these problems range from adding modules, emulating brain structure and processes, rigorous standards and testing, and real-world embedding, to replacing the computing substrate outright with improved chip types.

Several prominent Chinese scientists cited in this study’s section 2, who made public statements supporting alternate GAI models, including Tang Jie, Zhang Yaqin, Xu Bo, Zhu Songchun, and Zeng Yi, are on the bylines of many of these papers, adding authenticity to their declarations.

In addition, virtually all of China’s top institutions and companies engaged in GAI research, including the Beijing Academy of Artificial Intelligence (北京智源人工智能研究院), the Beijing Institute for General Artificial Intelligence (北京通用人工智能研究院), the Chinese Academy of Sciences’ Institute of Automation (中国科学院自动化研究所), Peking University (北京大学), Tsinghua University (清华大学), University of Chinese

Academy of Sciences (中国科学院大学) and Alibaba, ByteDance, Huawei, and Tencent AI lab, are represented in the selected corpus, in most cases on multiple papers.[64]

The record of metadata adduced here, and conclusions drawn in prior CSET research[65] support the present study's contention that major elements in China's AI community question LLMs' potential to achieve GAI—through increases in scale or modalities—and are contemplating or pursuing alternative pathways.

Assessment: Do All Paths Lead to the Buddha?

When LLM-based chatbots first became available, early claims that LLMs might be sentient, i.e., experience feelings and sensations, or even show self-awareness,[66] were prevalent and much discussed. Since then, cooler heads have prevailed,[67] and the focus has shifted from philosophical speculations about the interior lives of LLMs to more concrete measurements of LLM abilities on key indicators of "intelligent" behavior and the strategically important question of whether LLMs might be capable of general artificial intelligence (GAI).

While it is far from clear whether consciousness and the capacity for emotions are critical to GAI, what is clear is that a GAI system must be able to reason and to separate facts from hallucinations. As things stand, LLMs have no explicit mechanisms that would enable them to perform these core requirements of intelligent behavior. Rather, the hope of LLM enthusiasts is that, somehow, reasoning abilities will "emerge" as LLMs are trained to become ever better at predicting the next word in a conversation. Yet, there is no theoretical basis for this belief. To the contrary, research has shown that LLMs' vast text memory has *masked* deficiencies in reasoning.[68]

Heuristic attempts to improve reasoning (e.g., chain-of-thought),[69] likely the basis for improved performance in OpenAI's new "o1" LLM, and more recent approaches such as "rephrase and respond,"[70] "tree-of-thoughts"[71] or "graph-of-thoughts"[72] have yielded improvements, but do not solve the underlying problem of the absence of a core "reasoning engine."

By the same token, multiple attempts to fix LLMs' hallucination problem[73] have run into dead ends because they fail to address the core problem that is inherent to LLMs' ability to generalize from training data to new contexts. Indeed, current efforts to improve reasoning abilities and fix hallucinations are a bit like playing "whack-a-mole" but with moles hiding in a billion-dimensional weight space and with a mallet that is uncertain to hit where intended. The resulting systems might be sufficient for situations where humans can assess the quality of LLM output, e.g., writing cover letters, designing travel itineraries or creating essays on topics that are perennial favorites of high school teachers. Yet, these capabilities are a far cry from GAI.

The public debates in the western world on the appropriate path to GAI tend to be drowned out by companies with financial interests in promoting their latest LLMs with claims of "human-like intelligence" or "sparks of artificial general intelligence,"[74] even in the face of ever more apparent shortcomings of LLMs, as detailed in section 1. The dominance of commercial interests that promote LLMs as sure paths to GAI has

already negatively affected the ability of academic research in the U.S. to pursue alternative approaches to GAI.[75]

The situation is different in China. While there are also companies in China developing LLMs for commercial purposes, leading Chinese AI scientists and government officials, as detailed in this paper, realize that LLMs have fundamental limitations that make it important to investigate other approaches to GAI or supplement LLM performance using "brain-like" algorithms. The latter strategy, of pursuing "brain-inspired" AI has led to major breakthroughs in the past, for example, by combining deep learning[76]—modeled on the brain's sensory processing hierarchy—and reinforcement learning[77]—modeling how the brain learns strategies from rewards—into "deep reinforcement learning,"[78] which, for instance, formed the basis of AlphaGo,[79] the first artificial neural network that beat human champions in the game of Go. This difference in research directions may give China an advantage in the race to achieve GAI.

It might be helpful to compare the current situation to how China came to dominate the global market for photovoltaic (PV) panels (or, more recently, battery technology and electric vehicles), based on Chinese government decisions made at the beginning of the millennium to become a world leader in PV. The ensuing policy decisions and investments to build up the domestic PV industry and increase the efficiency of PV panels led to innovation and economies of scale that now have China producing at least 75% of the world's solar panels. A decision by China to strategically invest in non-LLM-based approaches to GAI[80] may repeat this success, albeit in a field of even greater importance than photovoltaics.

Managing a China First-Mover Advantage

Geoffrey Hinton, recent Nobel Prize winner and recipient of a Turing Award for his work on multilayer neural networks—the first AI NN architecture that led to superhuman performance on a range of benchmark tasks in computer vision and other areas—acknowledges "a race, clearly, between China and the U.S., and neither is going to slow down."[81]

This race to general AI is typically characterized as a competition for data, chips, talent, and energy,[82] with success measured on benchmarks meant to assess "human-level intelligence." The assumption underlying these comparisons is that both sides are competing in the same arena.

This view is dangerously misleading. The present study shows major elements in China's AI community pursuing alternative paths to GAI in which model complexity—taken by many in the U.S. as a proxy for performance, conditioned by companies' focus on the number of parameters of their models as a distinguishing feature—plays only a subsidiary role. These non-traditional approaches, moreover, have Chinese state backing.

Utility aside, pragmatism likely also motivates PRC support for a general AI that avoids the inherent uncontrollability of large statistical models,[83] which along with their hallucinations and other pesky foibles also resist top-down government censorship, since their inner workings are and will likely remain a "black box."[84] The Chinese government's early concern with LLM "safety" (安全, which also means "security") should be understood in this context.[85]

So how do alternative models improve things from the state's perspective? BIGAI director Zhu Songchun, whose influence extends well past his Beijing circle of colleagues,[86] has an answer.

Zhu argues that for an AI to be general it must ingest principles that guide its exploration of the environment in which it is embedded. In Zhu's system, an AI: (1) must manage unlimited tasks including those not predefined; (2) have autonomy (自主性), including the ability to generate its own tasks; and (3) be "value-driven," not "data-driven" as in today's large models.[87]

Zhu correctly notes that current LLMs "do not have human cognitive and reasoning capabilities, and also lack human emotions and values."

"From the perspective of values, whether large models can understand human value orientation determines whether large models can be safely and reliably applied to important areas related to the national economy and people's livelihood."

Hence "values" are needed not only to drive the system as it learns but to ensure that what it does learn meets the needs of the nation and the people. As Zhu explains:

The core of people's concerns about the threat of artificial intelligence is their distrust of "big models." There are two levels of trust. The first is trust in the system's ability. The second is recognition of values. The core of trust between people is value identity.[88]

Zhu's test of a general AI system, on which BIGAI is founded, besides evaluating vision, language, cognition, movement, and learning also assesses its adherence to embedded *values* along five dimensions: elementary self-value, advanced self-value, "primary social value, advanced social value, and group value."[89]

The difficulty of guaranteeing that LLM outputs are consistent with a particular set of values has also been recognized in the western literature. Earlier, ham-fisted approaches to ensuring that LLM outputs were aligned with specific sets of values have been widely ridiculed.[90] At the core of the challenge of "alignment" is the absence of an explicit "moral engine" in LLMs. This has forced developers to resort to laborious "fine-tuning" of LLM parameters based on human feedback on problematic LLM-generated responses.[91]

This approach based on tweaking undesirable or "non-aligned" answers and hoping for results that generalize to novel prompts has no guarantee of success.[92] A case in point is a recent study presenting the same ethical dilemmas in different languages to different LLMs.[93] That study found widely divergent behavior for ethical decisions not just across LLMs, but even for the same LLM when presented with the same ethical dilemma in different languages. New approaches such as training not based on human feedback but using different value models might be suitable for specific, well-defined scenarios,[94] but it is unclear how such a strategy could generalize to broader sets of ethical dilemmas, let alone lead to LLM responses that are consistent with a specific set of values. It is therefore currently far from clear if and how particular sets of values can be "trained into" an LLM, given that, for LLMs, "good" and "bad" are just words to be predicted, with no grounding in any kind of valuative framework.

In the end, Zhu's argument for an alternative GAI approach cuts three ways:

- Zhu claims that LLMs cannot achieve GAI because they lack human-like sensibilities that underlie the motivation to explore and learn. A cognitive architecture able to assimilate *values* is needed, in this view, for AI to achieve generality.
- He proposes to manage the AI safety problem by replacing largely unworkable "guardrails" and ad hoc fixes to LLM output with an internal requirement to behave according to first principles, i.e., a *value system* matching that of its users.
- He addresses the fear of China's ruling elite that large models will subvert the autocracy. Driven by socialist Party values, the GAI that China needs to stay competitive remains within comfortable bounds and reinforces State ideology, potentially forever.

Given these considerations, the AI "race" takes on a new dimension with challenges not only in the economic and military spheres but in human value orientation.

A final distinction between Chinese and western AI research evidenced throughout this paper needs to be made explicit, namely, the real possibility that China's directed, strategic approach may be more effective—all other things being equal—than the western profit-driven approach that focuses on quick wins at the possible expense of more successful strategies that require a longer time horizon.[95]

Our recommendations, accordingly, are twofold: (1) replace the monoculture of LLM research with government and institutional support for a multifaceted approach to general AI, and (2) take seriously the need to monitor Chinese technical developments through open sources.[96]

Authors

William C. Hannas is CSET's lead analyst and formerly the CIA's senior expert for China open-source analysis. He is currently focused on U.S.-China technology competition, community outreach, and data discovery methodologies.

Huey-Meei Chang is CSET's senior China S&T specialist, co-editor of *Chinese Power and Artificial Intelligence: Perspectives and Challenges* (Routledge, 2023), and co-author of several papers on China's AI development.

Maximilian Riesenhuber, PhD, is professor of neuroscience at Georgetown University and codirector of its Center for Neuroengineering. His research focuses on understanding brain function and how these insights can be translated to augmented cognition applications and neuromorphic AI.

Daniel H. Chou is a data scientist at CSET. He has collected, enhanced, and analyzed data for multiple studies on China AI and technology development while supporting government and private sector projects.

Acknowledgements

The authors are grateful to CSET's Helen Toner, who served as "red teamer," and to John Chen of the RAND Corporation and Dr. Mike Wolmetz of Johns Hopkins University's Applied Physics Laboratory for serving as outside reviewers. The authors are also grateful to CSET's Dr. Igor Mikolic-Torriera, Dr. Catherine Aiken, Matthew Mahoney, Shelton Fitch, and Ben Murphy for their generous support during the review and publication process.

Document Identifier: doi: 10.51593/20230048

Endnotes

[1] Wm. C. Hannas, Huey-Meei Chang, Catherine Aiken, Daniel Chou, "China AI-Brain Research," (Center for Security and Emerging Technology, September 2020), 50, https://cset.georgetown.edu/publication/china-ai-brain-research/.

[2] We avoid the term "artificial general intelligence" (AGI) here due to its historical association with features of biological brains and minds (such as emotional and social intelligence, affect, consciousness) that *may or may not* be necessary elements of a "general" AI. See Wm. C. Hannas, Huey-Meei Chang, Daniel Chou, Brian Fleeger, "China's Advanced AI Research," (CSET, July 2022), 1-3, for a description of the terminological ambiguities with "AGI" overall and as relates to China, https://cset.georgetown.edu/publication/chinas-advanced-ai-research/. The standard Chinese term 通用人工智能 equates literally to "general artificial intelligence." A summary of recent Chinese policy documents issued by the Centre for the Governance of AI notes editorially that "通用人工智能 is sometimes, depending on context, translated as 'artificial general intelligence' (AGI), sometimes as 'general artificial intelligence,' and sometimes as 'general-purpose artificial intelligence'." This description gels with our own observations. Fynn Heide, "Beijing Policy Interest in General Artificial Intelligence Is Growing," (Centre for the Governance of AI, June 8, 2023), https://www.governance.ai/post/beijing-policy-interest-in-general-artificial-intelligence-is-growing.

[3] Geoffrey Hinton, conversation with Joel Hellermark, April 2024, https://www.youtube.com/watch?v=tP-4njhyGvo&t=660s.

[4] Robert Hart, "Meta's AI Chief: AI Models Like ChatGPT Won't Reach Human Intelligence," *Forbes*, May 22, 2024, https://www.forbes.com/sites/roberthart/2024/05/22/metas-ai-chief-ai-models-like-chatgpt-wont-reach-human-intelligence/.

[5] Ben Goertzel, *The Consciousness Explosion* (prepublication copy, 2024), 12. "This sort of technology [large language models], on its own, seems clearly not capable of producing HLAGI [human-level AGI] but it does seem very promising as a component of integrated multi-module AGI systems." Also see, TWIML AI Podcast (09:03), https://www.youtube.com/watch?v=MVWzwIg4Adw&list=TLPQMDUwODIwMjPPUBk12t2hDg&index=8.

[6] "In 2023 venture-capital investors poured over $36bn into generative AI, more than twice as much as in 2022." Guy Scriven, "Generative AI Will Go Mainstream in 2024," *The Economist*, November 13, 2023, https://www.economist.com/the-world-ahead/2023/11/13/generative-ai-will-go-mainstream-in-2024.

[7] Jon Kleinberg and Manish Rashavan, "Algorithmic Monoculture and Social Welfare," *PNAS* 118, no. 22 (February 2021), https://www.pnas.org/doi/10.1073/pnas.2018340118.

[8] Hannas, Chang, Chou, and Fleeger, "China's Advanced AI Research."

[9] Huey-Meei Chang, "China's Bid to Lead the World in AI," *The Diplomat*, July 6, 2024, https://thediplomat.com/2024/07/chinas-bid-to-lead-the-world-in-ai/.

[10] "Big data, small task" refers to the design philosophy of LLMs, which are trained on large datasets (on the order of trillions of "tokens") with a simple task, viz. to predict the next word in a text. This is in contrast to traditional AI systems trained on specific tasks such as recognizing faces, with architectures optimized for the problem domain—which enabled these systems to learn from smaller datasets.

[11] Note that the scope of this report is limited to text-based models. Yet, all considerations regarding text-based LLMs as paths to GAI also directly apply to their multimodal extensions that process not just text but also images, video and audio, which are differences in input/output modalities that do not fundamentally augment the models' "intelligence."

[12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, et al., "Attention Is All You Need," arXiv preprint arXiv:1706.03762v7 (2023).

[13] Kyle Mahowald, Anna A. Ivanova, Idan A. Blank, Nancy Kanwisher, et al., "Dissociating Language and Thought in Language Models," *Trends in Cognitive Sciences* 28, no. 6 (March 2024); Boshi Wang, Xiang Yue, Huan Sun, "Can ChatGPT Defend its Belief in Truth? Evaluating LLM Reasoning via Debate," *Findings of the Association for Computational Linguistics*, EMNLP (2023).

[14] Nouha Dziri, Ximing Lu, Melanie Sclar, Xiang Lorraine Li, et al., "Faith and Fate: Limits of Transformers on Compositionality," 37th Conference on Neural Information Processing Systems, NeurIPS (2023).

[15] E.g., Wolfram GPT. See https://gpt.wolfram.com/.

[16] S.M. Rivera, A.L. Reiss, M.A. Eckert and V. Menon, "Developmental Changes in Mental Arithmetic: Evidence for Increased Functional Specialization in the Left Inferior Parietal Cortex," *Cerebral Cortex* 15 (November 2005).

[17] Geoffrey Hinton, conversation with Joel Hellermark.

[18] Winnie Street, John Oliver Siy, Geoff Keeling, Adrien Baranes, et al., "LLMs Achieve Adult Human Performance on High-order Theory of Mind Tasks," arXiv preprint arXiv:2405.18870v2 (2024).

[19] Tomer D. Ullman, "Large Language Models Fail on Trivial Alterations to Theory-of-Mind Tasks," arXiv preprint arXiv.2302.08399v5 (2023).

[20] Jie Huang, Xinyun Chen, Swaroop Mishra, Huaixiu Steven Zheng, et al., "Large Language Models Cannot Self-Correct Reasoning Yet," arXiv preprint arXiv:2310.01798v1 (2023).

[21] Lexin Zhou, Wout Schellaert, Fernando Martinez-Plumed, Yael Moros-Daval, et al., "Larger and More Instructable Language Models Become Less Reliable," *Nature*, September 25, 2024.

[22] Chaeffer, Brando Miranda, Sanmi Koyejo, "Are Emergent Abilities of Large Language Models a Mirage?" 37th Conference on Neural Information Processing Systems, NeurIPS (2023).

[23] Adrian de Wynter, "Awes, Laws, and Flaws from Today's LLM Research," arXiv preprint arXiv:2408.15409v2 (2024).

[24] While LLMs excel on standardized tests that are similar to those in its massive training set (e.g., the SAT or the bar exam), recent research has shown that figuring out novel tasks based on one or two examples—something humans are quite good at—continues to present a challenge for LLMs. For instance, the influential "Abstraction and Reasoning Corpus" notes that humans can solve an average of 80% of ARC's pattern matching tasks, while the best AI systems perform at around 30%, https://lab42.global/arc/.

[25] Goertzel, "Generative AI vs. AGI: The Cognitive Strengths and Weaknesses of Modern LLMs," arXiv preprint arXiv:2309.10371v1 (2023), 63.

[26] Doug Lenat and Gary Marcus, "Getting from Generative AI to Trustworthy AI: What LLMs Might Learn from Cyc," arXiv preprint ArXiv:2308:04445 (2023).

[27] Julia Angwin, "Press Pause on the Silicon Valley Hype Machine," *The New York Times*, May 15, 2024, https://www.nytimes.com/2024/05/15/opinion/artificial-intelligence-ai-openai-chatgpt-overrated-hype.html.

[28] Dwarkesh Patel, "Francois Chollet, Mike Knoop – LLMs Won't Lead to AGI - $1,000,000 Prize to Find True Solution," Dwarkesh Podcast, June 11, 2024. Chollet added "I see LLMs as more of an off-ramp on the path to AGI actually. All these new resources are actually going to LLMs instead of everything else they could be going to." https://www.dwarkeshpatel.com/p/francois-chollet.

[29] Liu Yangnan (刘杨楠), "Before Achieving AGI, Global AI Leaders Are Arguing over These 4 Key Issues" (实现 AGI 之前，全球 AI 大佬在这 4 个关键问题上吵起来了), Sohu.com (搜狐), June 13, 2023, https://www.sohu.com/a/684748834_100016644.

[30] Zhipu (智谱) was founded in 2019, has a staff of 800, and a market valuation of $2.5 billion as of May 2024. It ranks at the top of China's 260 generative AI companies. Eleanor Olcott, "Four Start-ups Lead China's Race to Match OpenAI's Chat GPT," *Financial Times*, May 2, 2024, https://www.ft.com/content/4e6676c8-eaf9-4d4a-a3dc-71a09b220bf8.

[31] BAAI (北京智源人工智能研究) was founded in 2018 by Huang Tiejun (see below), vice-dean at Peking University's Institute for Artificial Intelligence (人工智能研究院), with a goal of building strong AI. https://www.aminer.cn/research_report/5f44cbf13c99ce0ab7bc8db9?download=false.

[32] See Tang Jie's CV at https://keg.cs.tsinghua.edu.cn/jietang/.

[33] Celeste Biever, "Chian's ChatGPT: Why China Is Building Its Own AI Chatbots," *Nature*, May 22, 2024, https://www.nature.com/articles/d41586-024-01495-6.

[34] A rule of thumb, similar to Moore's Law for semiconductor density, which holds that the power of a neural network increases in proportion to its size, training data, and computational resources.

[35] "Tsinghua University's Tang Jie: From GPT to GPT Zero Will Be a Major Milestone This Year" (清华大学唐杰: 从 GPT 到 GPT Zero 会是今年重大阶段性成果), Tencent Network (腾讯网), March 1, 2024, https://news.qq.com/rain/a/20240301A06U9500.

[36] Tang Jie (唐杰), "Big Models and Superintelligence" (大模型与超级智能), *Communications of CCF* (中国计算机学会通讯) 20, no. 6, (2024), also see https://hub.baai.ac.cn/view/37642.

[37] Chinese orthography lacks word division, i.e., white space between words. Accordingly, these distinctions must be made on the fly, complicating the process of nominating the "tokens" on which LLMs operate.

[38] Li Anqi (李安琪), "Kai-Fu Lee and Ya-Qin Zhang Fireside Chat: China's Big Models Have More Opportunities on the C-end, and Technology Will Not Bring Permanent Leadership" (李开复、张亚勤炉边谈话: 中国大模型在 C 端的机会更多, 技术不会带来永久领先), Tencent Technology (腾讯科技), June 14, 2024, https://wallstreetcn.com/articles/3717247.

[39] Liu Yangnan, "Before achieving AGI."

[40] "Take-Aways from WAIC Keynote Speeches by Leading Figures at the Science Frontier Conference" (WAIC 科学前沿会议大佬演讲干货!), zhidx.com (智东西), July 6, 2024, https://36kr.com/p/2849481785170817.

[41] "Huang Tiejun, the Earliest Promoter of China's Large-scale Model: The World May Only Need Three LLM Ecosystems" (中国大模型的最早推行者黄铁军：全球可能只需要三个大模型生态), Tencent Technology (腾讯科技), June 9, 2023, https://wallstreetcn.com/articles/3690752.

[42] CAS's Institute of Automation is one of China's top institutes developing GAI through LLM and brain-inspired research. Its scientists are tied with Peking University for the highest number of GAI-related studies. Wm. C Hannas, Huey-Meei Chang, Max Riesenhuber, and Daniel H. Chou, "China's Cognitive AI Research: Emulating Human Cognition on the Way to General Purpose AI," (Center for Security and Emerging Technology, July 2023), 11, https://cset.georgetown.edu/publication/chinas-cognitive-ai-research/.

[43] CEBSIT is an umbrella organization for some 39 research institutes in China. See http://www.ion.ac.cn/yjsjj/zzjg/.

[44] Bo Xu and Muming Poo, "Large Language Models and Brain-inspired General Intelligence," National Science Review, October 2023, https://academic.oup.com/nsr/article/10/10/nwad267/7342449.

[45] Chang and Hannas, "Spotlight on Beijing Institute for General Artificial Intelligence."

[46] Zhu Songchun (朱松纯), "Zhu Songchun: Will the Rapid Development of Artificial Intelligence Definitely Pose a Threat?" (朱松纯: 人工智能高速发展一定会产生威胁吗?), ScienceNet (科学网), September 12, 2023, https://news.sciencenet.cn/htmlnews/2023/9/508316.shtm; Han Yangmei (韩扬眉), "Zhu Songchun: 20 Years of Exploration Has Given China an Advantage in Moving towards the Era of

General Artificial Intelligence" (朱松纯: 20 年探索, 为我国迈向通用人工智能时代赢得先机), *China Science Daily* (中国科学 报), July 25, 2024, https://news.sciencenet.cn/htmlnews/2024/7/527056.shtm.

[47] See Zeng Yi's home page at https://braincog.ai/~yizeng/. See also https://brain-cog.network/cn.

[48] "The Chinese Academy of Sciences Team Is Determined to Invest in the Next 20 Years or Even Longer to Build a General Brain-like Artificial Intelligence Infrastructure" (中科院团队: 决心投入未来 20 年乃至更长时间, 打造通用类脑人工智能基础设施), *Science & Technology Review* (科技导报), October 10, 2022, https://c.m.163.com/news/a/HJB51HH40511DC8A.html.

[49] "Shen Xiangyang: Rethinking the Human-machine Relationship in the Era of General Purpose Large Models" (沈向洋: 通用大模型时代 重新思考人机关系), EastMoney.com (东方财富网), March 23, 2024, https://finance.eastmoney.com/a/202403233021954617.html.

[50] "Tongji president Zheng Qinghua: Big Models Have Become the Pinnacle of Current Artificial Intelligence, but They Still Have Four Major Flaws" (同济校长郑庆华: 大模型已成当前人工智能巅峰, 但存四大缺陷), Sohu (搜狐), April 23, 2024, https://www.sohu.com/a/773698839_100016406.

[51] "Focus Comment: Big Models and General Artificial Intelligence | Turing Conference SIGAI Roundtable Forum held in Wuhan" (焦点评论: 大模型与通用人工智能 | 图灵大会 SIGAI 圆桌论坛在武汉举行), Beijing Institute for General Artificial Intelligence, July 30, 2024, https://www.bigai.ai/blog/news/焦点评论：大模型与通用人工智能 - 图灵大会 bigai 圆桌/.

[52] "Several Measures for Promoting the Innovation and Development of General Artificial Intelligence in Beijing" (北京市促进通用人工智能创新发展的若干措施), May 30, 2023. https://www.beijing.gov.cn/zhengce/zhengcefagui/202305/t20230530_3116869.html.

[53] "Wu Zhaohui Appointed as vice president of the Chinese Academy of Sciences" (吴朝晖任中国科学院副院长), April 11, 2024, http://www.cs.zju.edu.cn/csen/2024/0416/c38564a2902027/page.htm.

[54] Jiao Yifei (缴翼飞), "Wu Zhaohui, Vice Minister of the Ministry of Science and Technology: Big Models Push Artificial Intelligence toward the 3.0 Stage, and We Need to Explore the Development of General Artificial Intelligence through Multiple Paths" (科技部副部长吴朝晖：大模型推动人工智能迈向 3.0 阶段 要多路径探索通用人工智能发展), *21st Century Business Herald* (21 世纪经济报道), March 24, 2024, https://www.21jingji.com/article/20240324/herald/18cf65156b7d67322a4d0b3a9d98a47b.html.

[55] The figures omit universities and government-sponsored institutes engaged in AI research in Haidian.

[56] Sun Ying (孙颖) and Wang Haixin (王海欣), "Beijing Releases Three-year Plan and Lays Out Six Major Actions to Build a National Embodied Intelligence Innovation Highland" (发布三年计划、布局六大行动, 北京打造全国具身智能创新高地), *Beijing Daily* (北京日报), April 27, 2024, https://news.bjd.com.cn/2024/04/27/10758397.shtml.

[57] For example, LLMs by Zhipu, BAAI, iFlytek, Huawei, Baidu, Shanghai Artificial Intelligence Laboratory (上海人工智能实验室), Baichuan AI (百川智能), Moonshot AI (月之暗面), etc. with declared GAI goals. See "2023 H1 'China's Top 50 Most Valuable AGI Innovation Institutions' Officially Released" (2023 H1「中

国最具价值 AGI 创新机构 TOP 50」正式发布), Geek Park (极客公园), July 27, 2023, https://www.geekpark.net/news/322354.

[58] Hannas, Chang, Riesenhuber, and Chou, "China's Cognitive AI Research."

[59] William C Hannas, Huey-Meei Chang, Rishika Chauhan, Daniel H. Chou, et al., "Bibliometric Analysis of China's Non-Therapeutic Brain-Computer Interface Research," (Center for Security and Emerging Technology, March 2024), https://cset.georgetown.edu/publication/bibliometric-analysis-of-chinas-non-therapeutic-brain-computer-interface-research/.

[60] CSET merged corpus of scholarly literature including Web of Science, OpenAlex, Semantic Scholar, The Lens, arXiv, and Papers With Code. Searches included English and Chinese variants of the terms. The LLM keywords were: 大语言模型, 大型语言模型, large language model, LLM, GPT, LLaMA; the AGI keywords were: 通用人工智能, artificial general intelligence, 人工通用智慧, 强人工智能 strong artificial intelligence, strong AI, AGI, GAI.

[61] Online sources were primarily Chinese institute and scientific research (科研) websites.

[62] These papers' bylines typically list a half-dozen or more authors. For brevity we provide only the lead author, last author (often a senior advisor), the corresponding author if different from the first and last authors, and finally all authors claiming a non-China affiliation (marked in italics).

[63] YONG Silong's pinyin spelling ("si" for 子) is idiosyncratic. See https://www.bigai.ai/blog/news/联结场景理解和具身智能，首个智能体具身推理能/.

[64] Here are the scores: BAAI (3 papers), BIGAI (10), CASIA (5), other Beijing-located CAS institutes (10), Peking University (18) and Tsinghua University (11)—for a total of 57 appearances versus 18 for all other Chinese institutes combined, i.e., 76 percent of the cited affiliations. Prior CSET analysis of Chinese GAI research using a different (and much larger) corpus showed Beijing-based institutes on 70 percent of the papers. See Hannas, Chang, Riesenhuber, and Chou, "China's Cognitive AI Research," 12. Some 6 foreign organizations were among the corpus's cited affiliations, most prominently Carnegie Mellon University (3) and Microsoft Research Asia (2).

[65] Hannas, Chang, Aiken and Chou "China's AI-Brain Research;" Hannas, Chang, Chou and Fleeger "China's Advanced AI Research;" Hannas, Chang, Riesenhuber and Chou, "China's Cognitive AI Research;" Huey-Meei Chang and Wm. C. Hannas, "Spotlight on Beijing Institute for General Artificial Intelligence," (Center for Security and Emeriging Technology, May 2023), https://cset.georgetown.edu/publication/spotlight-on-beijing-institute-for-general-artificial-intelligence/; Wm. C. Hannas, Huey-Meei Chang, Rishika Chauhan, Daniel H. Chou, et al., "Bibliometric Analysis of China's Non-Therapeutic Brain-Computer Interface Research" (Center for Security and Emerging Technology, March 2024), https://cset.georgetown.edu/publication/bibliometric-analysis-of-chinas-non-therapeutic-brain-computer-interface-research/.

[66] Leonardo de Cosmo, "Google Engineer Claims AI Chatbot Is Sentient: Why That Matters," *Scientific American*, (July 2022), https://www.scientificamerican.com/article/google-engineer-claims-ai-chatbot-is-sentient-why-that-matters/.

[67] David J. Chalmers, "Could a Large Language Model Be Conscious?" arXiv preprint arXiv:2303.07103 (2024).

[68] Tang Xiaojuan, Zhu Songchun, Liang Yitao, Zhang Muhan, "Large Language Models Are In-context Semantic Reasoners Rather than Symbolic Reasoners," arXiv preprint arXiv:2305.14825v2 (2023).

[69] Jason Wei, Xuezhi Wang, Dale Shuurmans, Maarten Bosma, et al., "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," arXiv preprint arXiv:2201.11903 (2022).

[70] Yihe Deng, Weitong Zhang, Zixiang Chen, Quanquan Gu, "Rephrase and Respond: Let Large Language Models Ask Better Questions for Themselves," arXiv preprint arXiv:2311.04205 (2024).

[71] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, et al., "Tree of Thoughts: Deliberate Problem Solving with Large Language Models," arXiv preprint arXiv: 2305.10601 (2023).

[72] Maciej Besta, Nils Blach, Ales Kubicek, Robert Gerstenberger, et al., "Graph of Thoughts: Solving Elaborate Problems with Large Language Models," arXiv preprint arXiv:2308.09687 (2024).

[73] For a recent survey, see S.M. Towhidul Islam Tonmoy, S.M. Mehedi Zaman, Vinija Jain, Anku Rani, et al., "A Comprehensive Survey of Hallucination Mitigation Techniques in Large Language Models," arXiv preprint arXiv:2401.01313 (2024).

[74] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, et al., "Sparks of Artificial General Intelligence: Early Experiments with GPT-4," arXiv preprint arXiv:2303.12712 (2023).

[75] Brian Owens, "Rage against Machine Learning Driven by Profit," *Nature*, September 18, 2024, https://www.nature.com/articles/d41586-024-02985-3.

[76] Yann LeCun, Yoshua Bengio and Geoffrey Hinton, "Deep Learning," *Nature*, May 28, 2015.

[77] Emre O. Neftci and Bruno B. Averbeck, "Reinforcement Learning in Artificial and Biological Systems," *Nature Machine Intelligence*, March 2019.

[78] Matthew Botvinick, Jane X. Wang, Will Dabney, Kevin J. Miller, et al., "Deep Reinforcement Learning and Its Neuroscientific Implications," *Neuron*, August 19, 2020.

[79] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, et al., "Mastering the Game of Go with Deep Neural Networks and Tree Search," *Nature*, January 28, 2016.

[80] It was noteworthy that a number of papers surveyed emphasized the potential of "human-like" or "brain-like" algorithms to overcome shortcomings of current AI systems, given the brain's status as the past and current "gold standard" for intelligent systems.

[81] Geoffrey Hinton, conversation with Joel Hellermark, April 2024, https://www.youtube.com/watch?v=tP-4njhyGvo&t=660s.

[82] Joyce Guo, Zoey Duan, Chance Flanigen, Ethan Hsu, et al., "Energy Consumption Ramifications of U.S.-China AI Competition," *SSRN* (May 2024), http://dx.doi.org/10.2139/ssrn.4839772.

[83] As discussed, for example, in Zou, et al., "Controllable Generation from Pre-trained Language Models via Inverse Prompting" (item 24 in this paper's Section 3).

[84] "I honestly don't think we will get to an AI that we can trust if we stay on the current path… [LLMs] are recalcitrant, and opaque by nature—so-called "black boxes" that we can never fully rein in." Gary Marcus, "OpenAI's Sam Altman Is Becoming One of the Most Powerful People on Earth. We Should Be Very Afraid," The Guardian, August 3, 2024, https://www.theguardian.com/technology/article/2024/aug/03/open-ai-sam-altman-chatgpt-gary-marcus-taming-silicon-valley.

[85] Cyberspace Administration of China, et al., "Interim Measures for the Administration of Generative Artificial Intelligence Services" (生成式人工智能服务管理暂行办法), July 10, 2023, https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm; National Technical Committee 260 on Cybersecurity of Standardization Administration of China (SAC/TC260, 全国信息安全标准化技术委员会), "Technical Documentation of National Technical Committee 260 on Cybersecurity of Standardization Administration of China: Basic Security Requirements for Generative Artificial Intelligence Services," (全国网络安全标准化技术委员会技术文件：生成式人工智能服务安全基本要求), October 11, 2023, translated by CSET, https://cset.georgetown.edu/publication/china-safety-requirements-for-generative-ai-final/.

[86] Chang and Hannas, "Spotlight on Beijing Institute."

[87] Zhu Songchun, "Will the Rapid?"

[88] Zhu Songchun, "Will the Rapid?" Our emphasis.

[89] Yujia Peng, Jiaheng Han, Zhenliang Zhang, Lifeng Fan, et al., "The Tong Test: Evaluating Artificial General Intelligence through Dynamic Embodied Physical and Social Interactions," *Engineering* 34, (2024): 12-22.

[90] Nico Grant, "Google Chatbot's A.I. Images Put People of Color in Nazi-Era Uniforms," *The New York Times*, February 26, 2024, https://www.nytimes.com/2024/02/22/technology/google-gemini-german-uniforms.html.

[91] The process is called "reinforcement learning from human feedback" (RLHF).

[92] Cf. the "whack-a-mole" challenge referred to in this paper's section 4 in the context of efforts to reduce hallucinations or improve reasoning abilities of LLMs.

[93] Utkarsh Agarwal, Kumar Tanmay, Aditi Khandelwal and Monojit Choudhury, "Ethical Reasoning and Moral Value Alignment of LLMs Depend on the Language We Prompt Them in," arXiv preprint arXiv:2404.18460v1 (2024).

[94] Elizaveta Tennant, Stephen Hailes, and Mirco Musolesi, "Moral Alignment for LLM Agents," arXiv preprint arXiv:2410.01639v1 (2024).

[95] The authors are grateful to RAND's John Chen for this insight.

[96] Wang Yue (王悦), "After Nearly 1,500 Days of Wrestling with Large Models, Beijing Academy of Artificial Intelligence Is Still Insisting on Original Innovation" (与大模型交手近 1500 天，智源仍在坚持原始创新), Leiphone (雷峰网), June 20, 2024, https://www.leiphone.com/category/ai/LqUMGN3BTN3sZLyQ.html, also see https://scout.eto.tech/?id=3558.