Comments presented to the United States Senate AI Insight Forum
on
Risk, alignment, and guarding against doomsday scenarios

William C. Hannas
Research Professor and Lead Analyst
and
Huey-Meei Chang
Senior China S&T Specialist
Center for Security and Emerging Technology, Georgetown University

December 6, 2023

Introduction

We are privileged today to present our views on two of this forum's subtopics—risk mitigation and the potential long-term harm posed by AI. Our comments also address the forum's question about government's role in AI safety and national security.

We are both senior members of Georgetown's Center for Security and Emerging Technology (CSET), where we monitor threats posed by Chinese artificial intelligence. Prior to that we were employed by the Central Intelligence Agency to manage open source collection of Chinese scientific materials and provide instruction in document analysis. Our collaboration has led to several books and papers on Chinese AI and technology acquisition, including *China's Quest for Foreign Technology*,[1] which became a *de facto* community handbook, and the 2023 volume *Chinese Power and Artificial Intelligence*,[2] a comprehensive look at Chinese AI.

We focus primarily on Chinese sources. The ability to penetrate native language materials provides insights into China's S&T development that are hidden from analysts whose access is limited to public statements and sparse translations.

Summary of points

Our comments today offer a partial solution to a conundrum that has puzzled AI pundits for a decade and has seized the public's attention in recent months, namely, how to adjudicate AI safety with the need to outpace rivals.

[1] William C. Hannas and Didi Kirsten Tatlow, eds. *Beyond Espionage: China's Quest for Foreign Technology* (New York and London: Routledge, 2021).
[2] William C. Hannas and Huey-Meei Chang, eds., *Chinese Power and Artificial Intelligence* (New York and London: Routledge, 2023).

We believe the U.S. lead in artificial intelligence cannot be sustained without national policies aimed at safe and responsible innovation. Policies, however, are no better than the data on which they are based, which is the crux of the problem—we lack the information needed to make sound judgments. Without reliable data on China's infrastructure and plans, we do not know what pace of development is "too slow," "too fast," where to blow the whistle or at what points to intervene, because we have no way to track how quickly our competitor is moving.

Accordingly, we recommend a national center be stood up within the U.S. Government to monitor through open sources—where most of the information resides—the AI progress of our strategic adversaries. We see this as the best and only way to move beyond guesswork to data-driven policy for AI and all technologies in which the U.S. has a stake. Members of Congress have proposed similar concepts, including in legislation introduced this year by Senator Bennet.

China, for its part, has for six decades operated a technology tracking system that informs its policymakers exactly what the state-of-play is in the global science competition, so that the country's resources are directed appropriately. It understands the link between "safety" and "security" and, in fact, uses one word (安全) for both concepts.

The AI safety we seek is one in which the liberal democracies—not the Chinese Communist Party—determine how AI "aligns" with human values. The key to this outcome is data, so you do not run faster than needed; know when to engage if the adversary takes risks; and can control access to intellectual property so we do not empower our competitors.

Why our lead in artificial intelligence is tenuous.

Recently our respected colleagues have argued, reasonably, that the U.S. is far enough ahead in AI development that we can slow down temporarily in the interest of safety.[3] But how long will this last? Sadly, we don't know because we lack the means to know. There are at least four additional facts that must be considered before we can bank on this assumption.

1. The rapid follower is following rapidly.

A CSET study in July 2022 looked at China's mainstream approaches to advanced AI, defined as compute-intensive big data models based on machine learning, and highlighted ten major efforts to achieve artificial general intelligence through generative large language models.[4] Parameters (a rough measure of performance) numbered in the tens to hundreds of billions. A second, highly detailed study of Chinese LLMs published in July 2023 concluded that "the performance of top Chinese models is not far behind the state of the art (SOTA) in the West."[5]

2. China doesn't need to be at the cutting edge to win.

[3] Helen Toner, Jenny Xiao, Jeffrey Ding, "The Illusion of China's AI Prowess," *Foreign Affairs*, June 2, 2023.
[4] William Hannas, Huey-Meei Chang, Daniel Chou, Brian Fleeger, "China's Advanced AI Research," Center for Security and Emerging Technology, July 2022.
[5] Jeffrey Ding and Jenny Xiao, "Recent Trends in China's Large Language Model Landscape," Centre for the Governance of AI, April 21, 2023.

Kai-Fu Lee (李开复), author of *AI Superpowers*,[6] argues—correctly in our view—that our obsession with building the world's best algorithms ignores a more important concern, namely, the need for derivative applications, an enterprise at which China excels. Lee demonstrated the practicality of China's rapid follower approach this year with the launch of his own large language model "Yi" that "draws on public results at the top level of the industry," namely, a rebadged version of Meta's "LLaMA" architecture.[7]

3. China can "beg, borrow, and steal" what it needs.

Lee's appropriation of an open source model is at the benign end of a phenomenon we have witnessed for decades, namely, China's unparalleled ability to acquire by legal and extralegal means technology created outside China.[8] The problem is acknowledged today and needs no elaboration here, other than to note that the very same techniques China has used since the PRC was established to relieve the world of its proprietary technology are being used in a state-run effort to close the gap in AI technology as well.[9]

4. There are many paths to the Buddha—not just LLMs.

The goal of AI since its inception has been "artificial general intelligence" (AGI) that equals or exceeds human cognitive skills across most tasks. The term—and the concept—once considered science fiction, have achieved respectability with the success of LLMs, which ironically are seen by some today as a dead end.[10] China is funding research into alternative "small data" paths to AGI that have mostly escaped scrutiny and it does not shy away from the idea of a literal merger of human and artificial intelligence.[11]

Why are we not tracking China's AGI developments?

Given these facts, and the importance of gauging China's AI progress for our own strategic calculations, one might expect doing so would be a high national priority. U.S. policymakers should have at their fingertips an understanding of China's AI aspirations, ability, infrastructure, and liabilities at least as good as any one person's in China. This holds for those concerned with "safety" in the usual sense and for "security" as it applies to our national interests.

So why don't we? As researchers who have worked the classified and academic sides of the spectrum, recently and over the long term, we can state unequivocally that the non-China world

[6] Kai-Fu Lee, *AI Superpowers: China, Silicon Valley, and the New World Order*, Harper Business, 2018.
[7] "Kai-Fu Lee's AI large language model Yi used Meta's Llama architecture without namechecking its source," TechNode Feed, November 15, 2023.
[8] William Hannas, James Mulvenon, Anna Puglisi, *Chinese Industrial Espionage*, Routledge, 2013.
[9] William Hannas and Huey-Meei Chang, "China's Access to Foreign AI Technology," Center for Security and Emerging Technology, September 2019.
[10] Nikolas Badminton, "Meta's Yann LeCun on autoregressive Large Language Models (LLMs),' *Insights*, February 13, 2023; Maximilian Riesenhuber, Professor of Neuroscience, Georgetown University, conversations with the authors 2022-2023; Zhu Songchun (朱松纯), "Promoting original and leading innovation with organized scientific research" (以有组织科研推进原创性、引领性创新), *Guang Ming Daily* (光明日报), February 18, 2023.
[11] Huey-Meei Chang and William Hannas, "Spotlight on Beijing Institute of General Artificial Intelligence," Center for Security and Emerging Technology, May 2023.

has no robust, dedicated effort to monitor China's science and technology advances, in AI or any other critical area, any more than we can track effectively China's efforts to shortcut growth by appropriating foreign IP. What efforts *are* made in the intelligence community and at think tanks such as our own are short-term, sporadic, and forgotten days after they are published.

We are not even sure this capability exists in a diffuse sense, given the neglect of open source collection inside and outside the "community" and the lack in this country of language qualified China science and technology analysts—probably fewer than 100 nationwide, compared to some 100,000 Chinese specialists monitoring foreign S&T.[12] The reasons for this disparity are clear and compelling:

1. Hubris blinds us to reality.

Success with large language models—one-type of AGI precursor—and our own national hubris has caused us mostly to brush off the efforts of competitors, who speak in a different language and are culturally disposed to view problems differently. We have been on top for so long—in science, wealth, and national power—that it is gauche to suggest American exceptionalism may be subject to the same cycle of decline experienced by every civilization in history.

2. We think we're looking but we're not.

In 2013, this august body had the foresight to commission a high-level review of the intelligence community's management of global S&T. Its top two findings were: (1) the USIC mostly ignores "the non-military R&D intentions and capabilities of our adversaries," and (2) its focus on traditional (classified) venues precludes it from adequately addressing S&T intelligence's most useful source, namely, publicly available information.[13] We assure you that neither of these conditions has improved and never will because we imagine the problem has been solved.

Taking a cue (for once) from China

By contrast, China has a world-class S&T information/intelligence (STI) network that allows it to minimize risk, avoid surprises, and identify opportunities for theft or collaboration.[14] It has been operating since 1956, encompasses multiple organizations, and functions as an appendage to the national (and local) governments *outside* China's clandestine intelligence network. Among its notable features the system boasts a cadre of some 100,000 workers,[15] state-of-the-art data

[12] William Hannas and Huey-Meei Chang, "China's STI Operations," January 2021.
[13] "Report of the National Commission for the Review of the Research and Development Programs of the United States Intelligence Community (unclassified version)," 2013. https://www.intelligence.senate.gov/sites/default/files/commission_report.pdf.
[14] Both concepts are covered in the one Chinese term 情报 (*qíngbào*).
[15] Zeng Jianxun (曾建勋), "基于国家科技管理平台的科技情报事业发展思考 (Reflection on the Development of the Scientific and Technical Information Industry Based on the National Science and Technology Management Platform)," 情报学报 (Journal of the China Society for Scientific and Technical Information), 2019: 38 (3), pp. 227-238.

processing equipment, industry journals, and direct access to China's S&T policymakers, who depend on it for major decisions.[16] And it is based entirely on open-source.

The present authors were immersed in this same discipline, managed China collection and analysis for the U.S. Government, and innovated where resources allowed. We have immense admiration for the potential of "publicly available information" to inform decision makers on technology policy, where classified sources contribute relatively little.[17] Our complaint is that the intelligence potential of open source is scarcely realized in the U.S., whereas in China it is the main staple of their global monitoring effort.

Our investment in open source intelligence (OSINT) is a fraction of China's and a thin line on the U.S. intelligence budget. Even this characterization understates the matter because Title-50 agencies treat open source as an "enabler" of classified intelligence—a servant of the house specialty—not something worth exploiting in itself. When used at all, it is directed mostly at traditional intelligence goals, of which S&T monitoring is last at a long table.

Calibrating speed and safety

Given China's goal to surpass the U.S. in all aspects of AI,[18] its embrace of artificial general intelligence,[19] and the aim of its top AI scientists to endow their creations with Chinese values and characteristics (中国特色),[20] there is as much danger to the United States—arguably to the whole of humanity—in ceding a "first mover advantage" (先发优势) to China as there is in the AI safety issue itself narrowly defined.

Although we have no solution to the problem of AI safety, it does seem obvious that calibrating the speed of our advance, and the risks we are willing to take, should depend in part on how quickly China is moving, which implies an ability to monitor Chinese AI development on a level for which we are not resourced. If the "safety" problem is important, the need for an in-depth understanding of China's capabilities in this quickly moving arena is no less important.

We lack the space here to describe in detail the resources needed to provide this capability.[21] It is enough to note that the idea for a National Science and Technology Analysis Center has been discussed with Senate and House committees, which found the concept attractive but difficult to execute owing chiefly to the issue of "ownership." Our goal today is simply to draw attention to what we call the "calibration problem"—how quickly to run and at what risk—which depends on an informed understanding of what our competitor is doing.

---

[16] Hannas and Chang, "China's STI Operations," 2021.
[17] The figure used in the U.S. and China for the contribution of open source to "S&T intelligence" is 85 percent.
[18] PRC State Council, "New Generation AI Development Plan" (国务院关于印发《新一代人工智能发展规划》的通知), PRC State Council, 2017.
[19] William Hannas, Huey-Meei Chang, Max Riesenhuber, Daniel Chou, "China's cognitive AI research: emulating human cognition on the way to general purpose AI," Center for Security and Emerging Technology, July 2023.
[20] Chang and Hannas, "Spotlight on Beijing institute of general artificial intelligence."
[21] See Tarun Chhabra, William Hannas, Dewey Murdick and Anna Puglisi, "Open-source intelligence for S&T analysis: Establishing a new open-source National Science and Technology Analysis Center." Center for Security and Emerging Technology, September 2020, https://cset.georgetown.edu/publication/open-source-intelligence-for-st-analysis/.