

Towards a global community of shared future in AGI

迈向通用人工智能时代的人类命运共同体

Brian Tse 谢旻希

Asilomar Conference on Beneficial AGI

5th January 2019

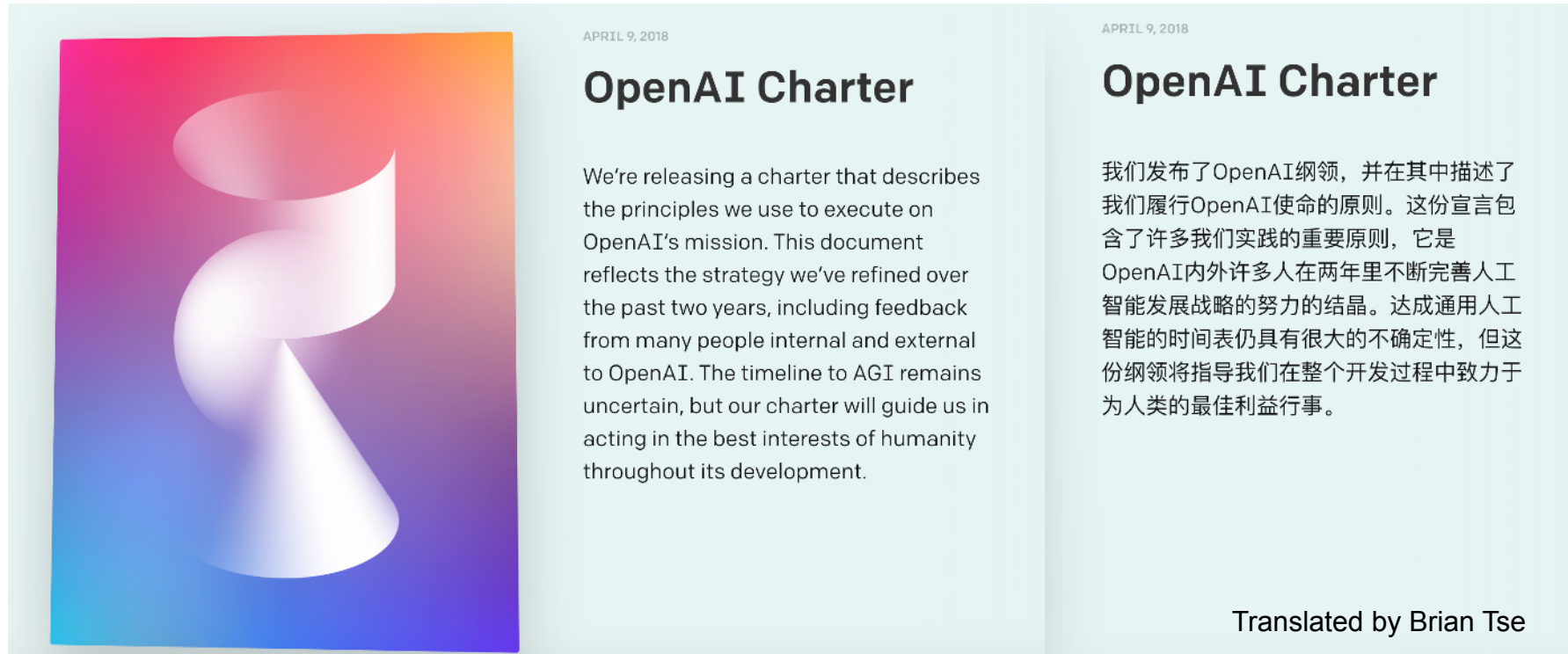




30
IRIS



Motivation for my presentation



“We are concerned about late-stage AGI development becoming a competitive race without time for adequate safety precautions.”

“我们担心通用人工智能在发展后期将演变成一场激烈的竞赛，导致缺乏充足的时间进行安全防范。”

Basis for global coordination

- A** AGI is a serious possibility in the coming decades
- B** There are significant uncertainties and risks with AGI
- C** International cooperation on AI can be mutually beneficial

Outline for my presentation

1) Contexts: China's perspectives on

- A** AGI is a serious possibility in the coming decades
- B** There are significant uncertainties and risks with AGI
- C** International cooperation on AI can be mutually beneficial

2) Approaches: Potential pathways towards coordination

3) Call to action: How you can help

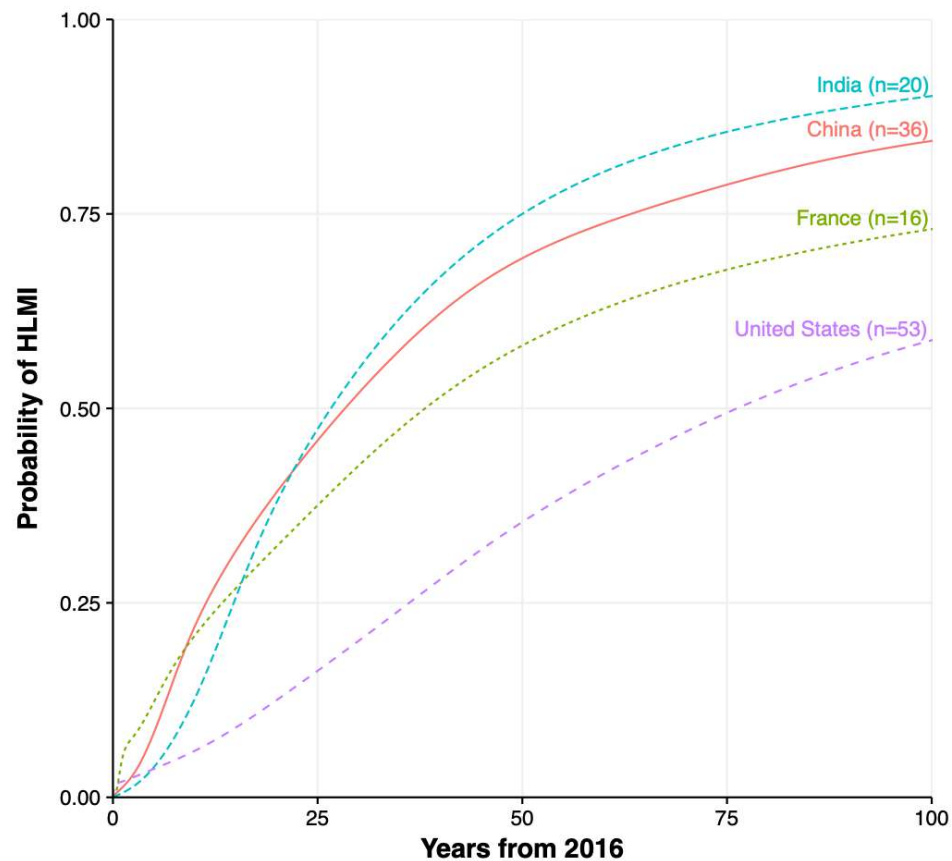
A

AGI is gradually emerging into the public discourse



When Will AI Exceed Human Performance? Evidence from AI Experts

(a) Top 4 Undergraduate Country HLMI CDFs



A

AGI is gradually emerging into the public discourse

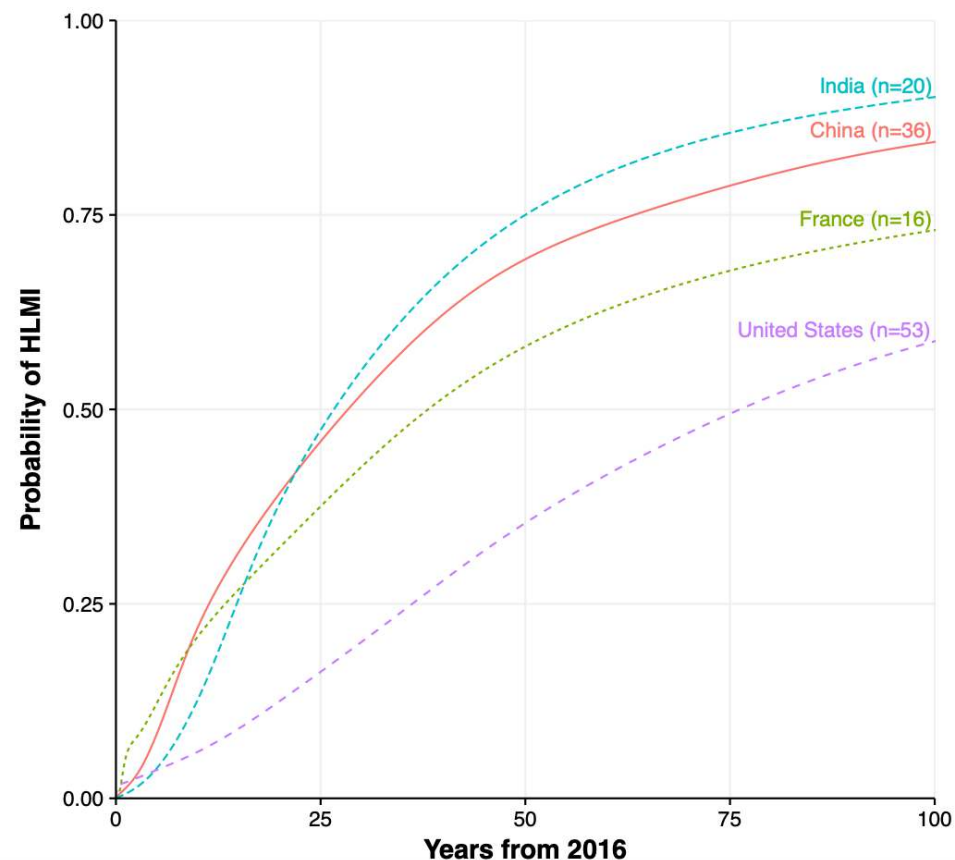


In response to Stephen Hawking and Elon Musk signing the FLI open letter: “I think that this will happen, and it will happen in the near future. How long? I think, optimistically speaking, it is **almost within 15-30 years.**”

Huang Tiejun (黄铁军),
Chair of Peking
University's Computer
Science Department

When Will AI Exceed Human Performance? Evidence from AI Experts

(a) Top 4 Undergraduate Country HLMI CDFs



A AGI is gradually emerging into the public discourse

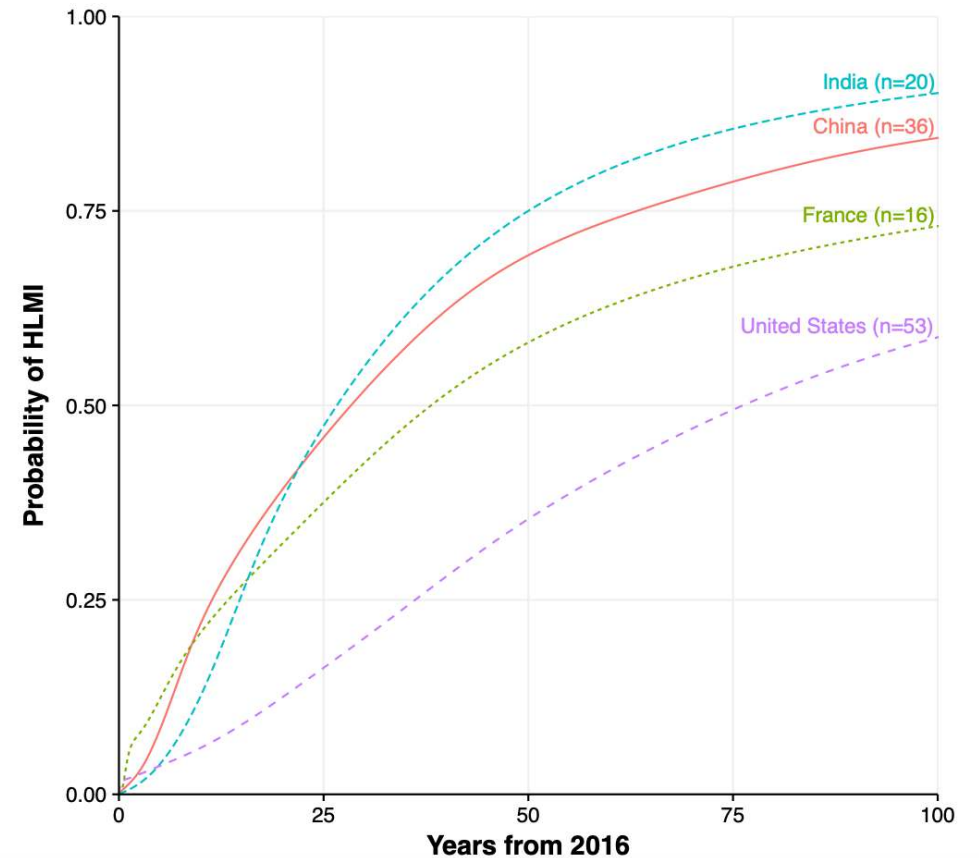
“How to realize the leap-forward development from narrow artificial intelligence to **general artificial intelligence** is the inevitable trend of the next generation of artificial intelligence development. It is also a challenge in the field of international research and application.”

Tan Tieniu (谭铁牛),
Deputy Secretary-General of the Chinese Academy of Sciences



When Will AI Exceed Human Performance? Evidence from AI Experts

(a) Top 4 Undergraduate Country HLMI CDFs



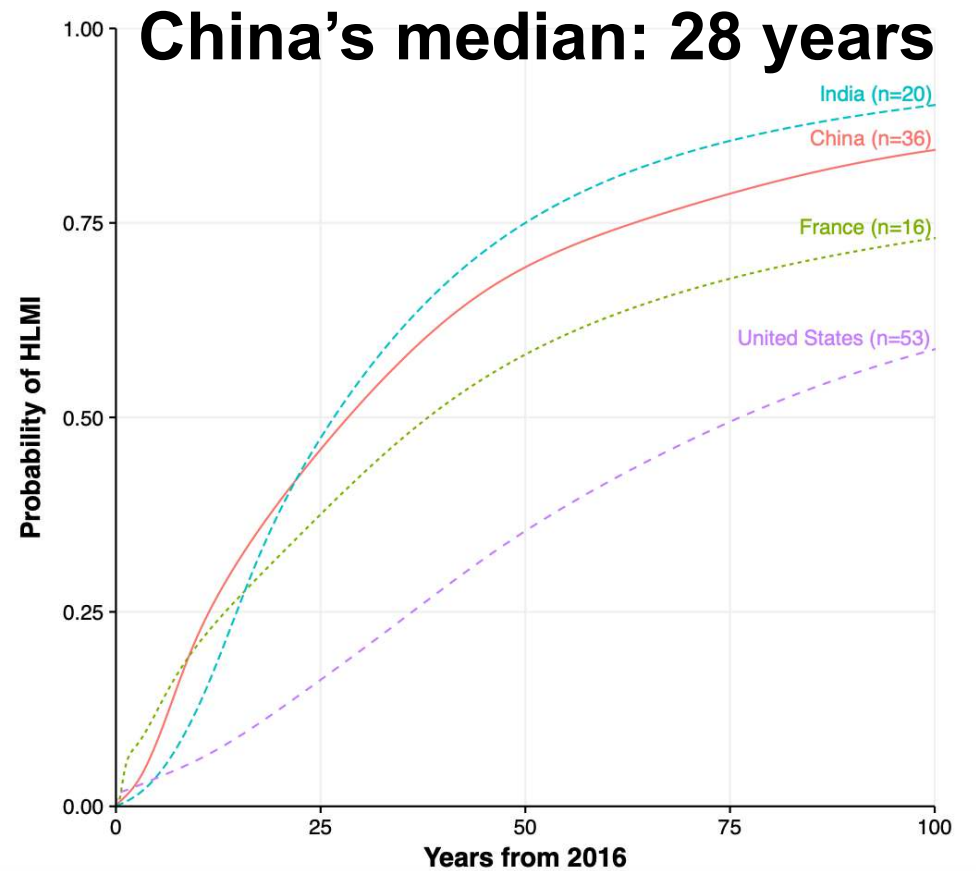
A

AGI is gradually emerging into the public discourse



When Will AI Exceed Human Performance? Evidence from AI Experts

(a) Top 4 Undergraduate Country HLMI CDFs



A Some private AI labs may be investing in AGI research



A

Some private AI labs may be investing in AGI research



Horizon Robotics
地平线机器人技术

Tencent 腾讯



Baidu: “The past discussions have mostly been on the feasibility of developing AGI, and that we should start discussing the ways to develop such technology.”

Horizon Robotics: “The long-term goal of building intelligent machines capable of learning diverse tasks as efficiently as humans do.”

— Xu Wei (徐伟), Chief Scientist of General AI at Horizon Robotics



A

Some private AI labs may be investing in AGI research



Horizon Robotics
地平线机器人技术

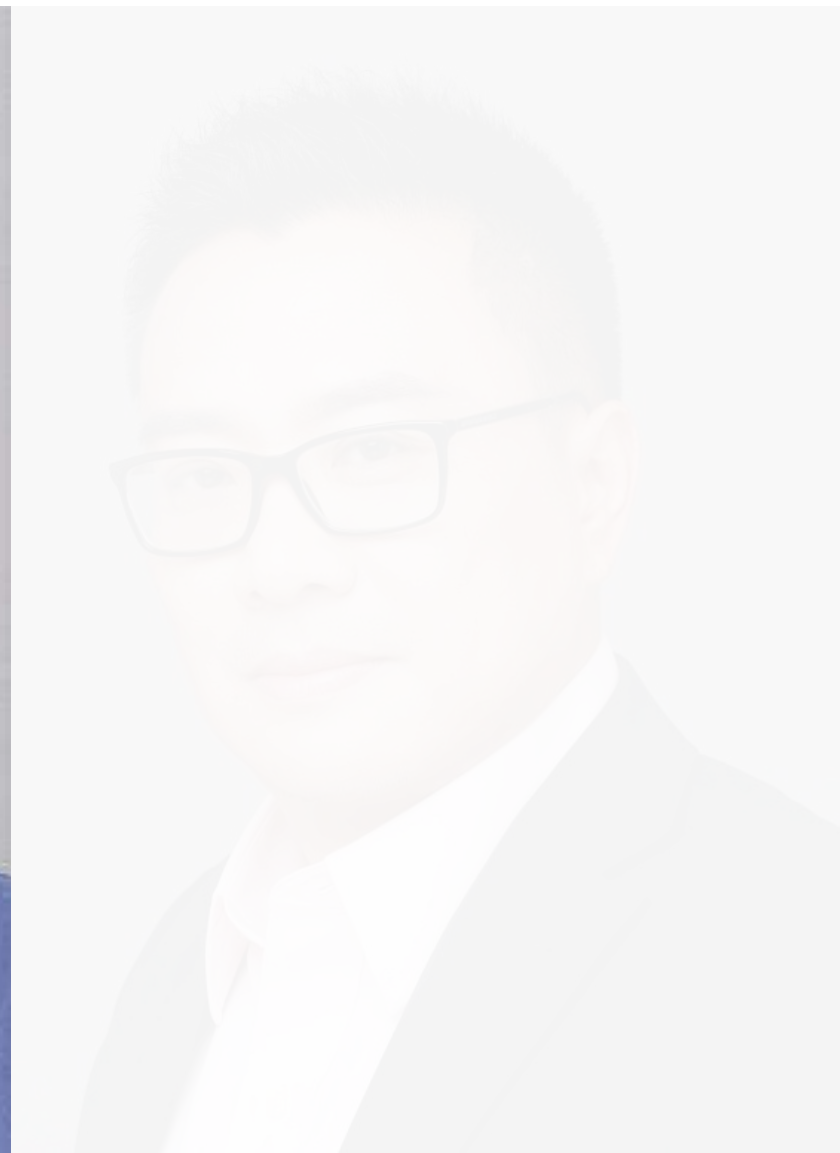
Tencent 腾讯



“Tencent AI Lab has three strategic directions with building AGI being one of them”

“Games is the most important path and direction for developing AGI. All the key research labs in the world are all exploring games for AI research”

— Yao Xing (姚星), Vice President at Tencent



A

Some private AI labs may be investing in AGI research

Tencent 腾讯



Mar 2016:
Fine Art (绝艺) developed for playing Go

May 2018:
《Feedback-Based Tree Search for Reinforcement Learning》 published based on AlphaGo Zero's MCTS for playing MOBA *King of Glory* (王者荣耀)

Dec 2018:
《Hierarchical Macro Strategy Model for MOBA Game AI》 published, competitive with top 1% of human players

April 2016:
Tencent AI Lab established

May 2018:
PhoenixGo (凤凰围棋) developed based on AlphaGo Zero's published paper in Nature

Sep 2018:
《TStarBots: Defeating the Cheating Level Builtin AI in StarCraft II in the Full Game》 published



Tencent AI Lab



A

Some private AI labs may be investing in AGI research



“There is not a timetable for the arrival of AGI, but there are seven major technical challenges to overcome from ANI to AGI.”

— Zhou Bowen (周博文),
Vice President and Director
of AI Research at JD.com



B Influential thinkers believe there is a potential existential risk



B Influential thinkers believe there is a potential existential risk



“The day when strong AI emerged will probably be the time when humanity faces the greatest survival crisis. So, for serious AI researchers, if you really believe that your efforts will produce results, you should not touch strong AI.”

— Zhou Zhihua (周志华),
one of China’s foremost ML
experts




B Influential thinkers believe there is a potential existential risk

“Although there still may some time before the emergence of superintelligence, we also have reason to “prepare for a rainy day... **artificial intelligence may cause will be the irreversible end of humanity, and at least the end of human history.**”

— Zhao Tingyang (赵汀阳), one of the most influential Chinese philosophers



B Influential thinkers believe there is a potential existential risk



“Its objective is to help students realize proper views and approaches to develop future AI, have deeper analysis and thoughts on the impact of AI to the future (Human-Machine) society, and **prevent existential and potential risks.**”

— Zeng Yi (曾毅), Deputy Director at Research Center for Brain-inspired Intelligence



© Political leaders view international cooperation as necessary



© Political leaders view international cooperation as necessary



“Deepened international cooperation is required to cope with new issues in fields including law, security, employment, ethics and governance. Moreover, China is ready to work with other countries in the field of AI to jointly promote development, safeguard security and share the results, in areas such as technology exchanging, data sharing and application markets.”

— **Xi Jinping (习近平), General Secretary of CPC and President of the People's Republic of China**

© Political leaders view international cooperation as necessary

“We’re hoping that all countries, as members of the global village, will be inclusive and support each other so that we can respond to the double-edged-sword effect of new technologies... AI represents a new era. Cross-national and cross-discipline cooperation is inevitable.”

— **Liu He (刘鹤), Vice Premier and “trade war chief” of the People's Republic of China**



Short-term challenges and long-term opportunities

- AI safety vs AI security
- Competence vs consciousness
- Technocracy
- Long-term orientation
- Risk management



Short-term challenges and long-term opportunities

- AI safety vs AI security
- Competence vs consciousness



- Technocracy
- Long-term orientation
- Risk management



What are the promising approaches given such contexts?

- 1) **Contexts: China's perspectives on**   
- 2) **Approaches: Potential pathways towards coordination**
 - 1 Establishing AI safety technical research collaboration
 - 2 Participating in multistakeholder and industry alliances
 - 3 Fostering transnational epistemic community in AI/ML
 - 4 Bridging safe and reliable (安全可靠) AI ethical principles

1 Establishing AI safety technical research collaboration

CAICT 中国信通院 AI Security/Safety White Paper 2018

Strengthen international cooperation to manage common security/safety risks:

“There are bottlenecks in the current deep learning technologies, such as **poor robustness against adversarial examples, interpretability, poor adaptability to imperfect information and uncertain environment.** Through the establishment of overseas research centers and organization of international technical exchanges, **[China should] begin international cooperation in technical research,** track the latest research progress, and jointly strengthen research on new technologies such as transfer learning and brain-inspired learning. It should solve hidden safety risks and regulatory challenges such as algorithmic black-box, algorithmic bias, strengthen the robustness and safety of AI decision-making, and **promote the development of AI from narrow intelligence to general intelligence.**”



China AI Development Report 2018

“International collaboration and industry-university collaboration are important means of advancing AI development.”

Partnerships	Nature	Industry-university
MIT-SenseTime Alliance on AI	Research	+
MIT-iFlyTek Partnership	Research	+
Fudan-Google Innovation Lab	Research	+
Huawei-UC Berkeley Partnership	Research	+
Alibaba-NTU Joint Research Center	Research	+
Bytedance-UC Berkeley (BAIR) Partnership	Research	+
Toutiao-Intel Joint AI Lab	Research	

Participating in multistakeholder and industry alliances



PARTNERSHIP ON AI



“Baidu’s admission represents the beginning of PAI’s entrance into China. We will continue to add new members in China and around the world as we grow”



Microsoft



“Extensively establish international cooperation to form a global platform of collaboration” (广泛开展国际合作，形成全球化的合作平台)

Global AI Academic Alliance
(全球高校人工智能学术联盟)



THE UNIVERSITY OF SYDNEY

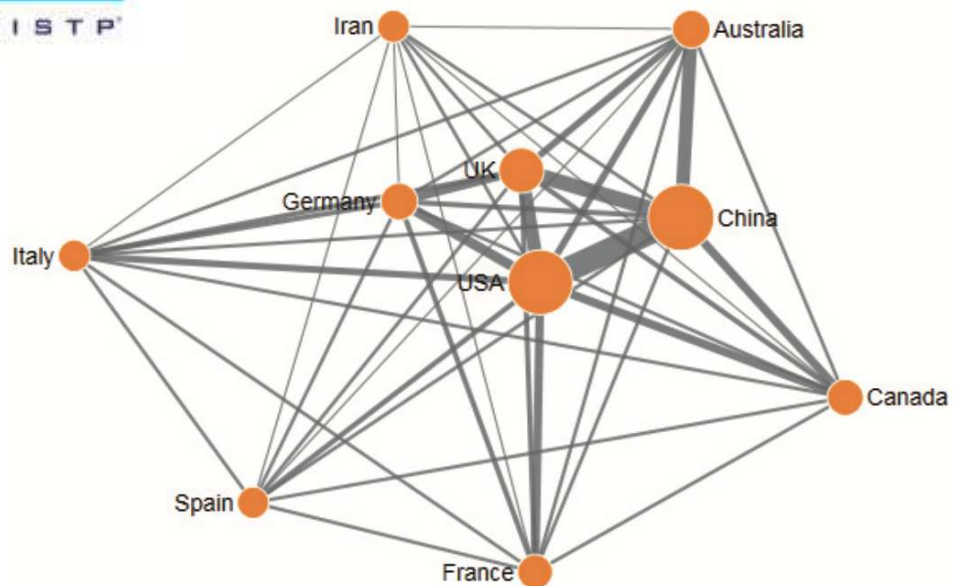


NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

3 Fostering transnational epistemic community in AI/ML



China AI Development Report 2018

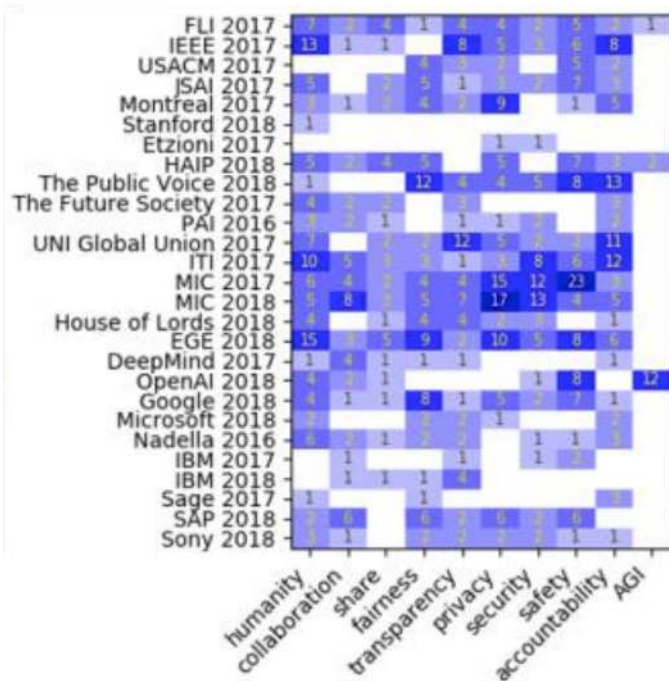


Collaboration network in the output of top papers on AI

	Percentage of Internationally Collaborative Papers (%)
Reference Value of International Papers	23.42
Reference Value of International Top Papers	42.64
China	53
United States	53.94
United Kingdom	76.38
Australia	81.82
Germany	80.65
Canada	72.5
France	76.9
Iran	50.18
Italy	75.98
Spain	71.66

Bridging safe and reliable (安全可靠) AI ethical principles

Linking AI Principles Project



Institutions

Sources

Safety-relevant principles



Harmonious AI Principles

Principle 6 — “Safety: AI should be with concrete design to avoid known and potential safety issues (for themselves, other AI, and human) with different levels of risks.”



ARCC (Available, Reliable, Comprehensible, and Controllable)
(马化腾“四可原则”)

Precautionary principle — “ensure AGI/ASI that may appear in the future serves the interests of humanity”



Robin Li 4 AI Principles
(李彦宏“AI四原则”)

Principle 1 — “The highest principle for AI is safe and controllable (安全可靠)”



Tan Tieniu’s Speech at 13th National People’s Congress

“To truly harvest the benefits of AI, we must first ensure its secure, controllable, and reliable (安全、可控、可靠) development.”

What do all these mean for us?

- 1) **Contexts: China's perspectives on** 
- 2) **Approaches: Potential pathways towards coordination**
 - 1 Establishing AI safety technical research collaboration
 - 2 Participating in multistakeholder and industry alliances
 - 3 Fostering transnational epistemic community in AI/ML
 - 4 Bridging safe and reliable (安全可靠) AI ethical principles
- 3) **Call to action: How you can help**

Supporting the emerging “Beijing Consensus in AI safety”



Local partnership



Researcher engagement



Technical workshops



Topics for further research and discussions

Contexts

- Relative AGI progress and timeline forecasting in China
- Different paradigms of AGI research in China, e.g. Brain-inspired intelligence
- Efforts and approaches towards safe and reliable AI

Approaches

- “Track II dialogue in AGI safety” between China and the U.S.
- Avenues of AI policy cooperation, e.g. malicious use from terrorism, safety-critical infrastructure, impact of AI on risks of nuclear war
- Strategic concept for cooperation, e.g. Safe in Diversity, Peaceful Co-existence

Destinations

- Comparative utopian thought, e.g. *The Great Unity* (大同社会)
- Comparative global governance theories, e.g. *Tianxia* (天下体系), *Community for shared future for mankind* (人类命运共同体)

Handwritten text on a chalkboard, including phrases like "What about...", "What about...", and "What about...".



HUMAN GENOME EDITING

27-29 November 2018

Lee Shau Kee Lecture Centre
Centennial Campus
The University of Hong Kong

Co-organized by



THE
ROYAL
SOCIETY

NATIONAL ACADEMY OF SCIENCES

NATIONAL ACADEMY OF MEDICINE

Venue Provider:



THE UNIVERSITY OF HONG KONG

Supported by



Innovation and
Technology Commission

SECOND INTERNATIONAL SUMMIT ON
HUMAN GENOME EDITING

Towards a global community of shared future in AGI

迈向通用人工智能时代的人类命运共同体

Brian Tse 谢旻希

Asilomar Conference on Beneficial AGI

5th January 2019