Building Ethical AI for the Human-AI Symbiotic Society
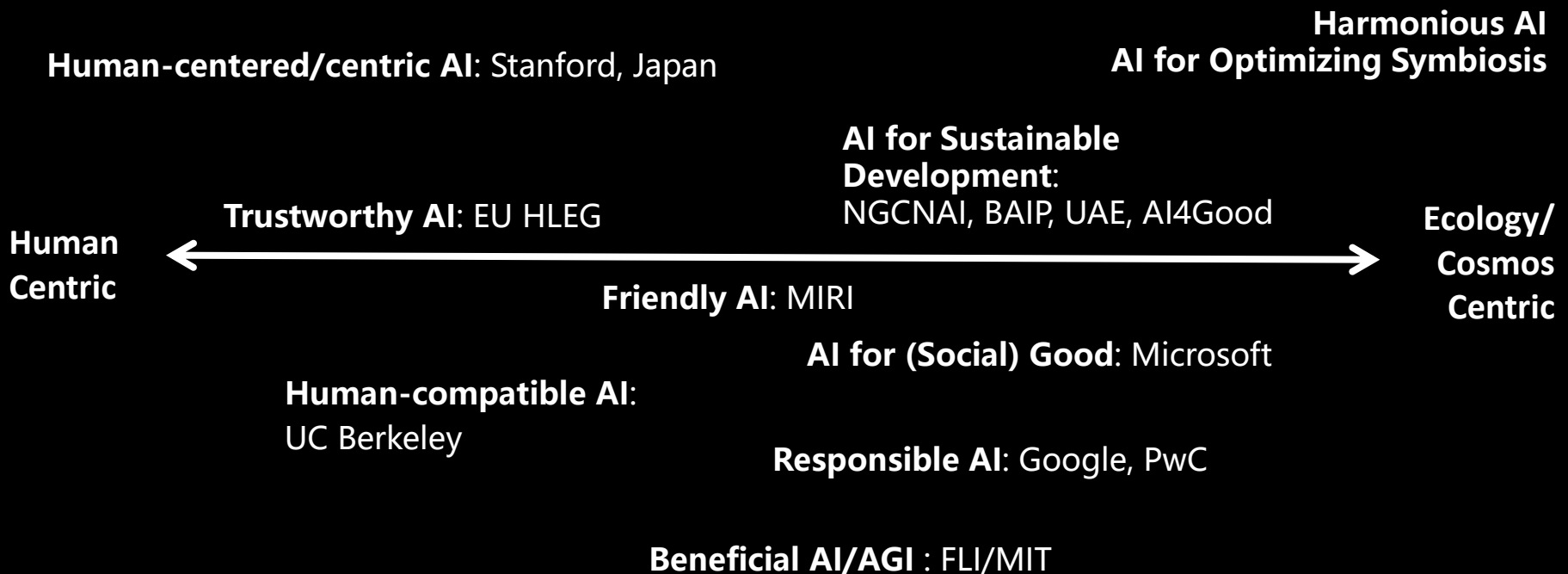
Yi Zeng

Chinese Academy of Sciences

Beijing Academy of Artificial Intelligence
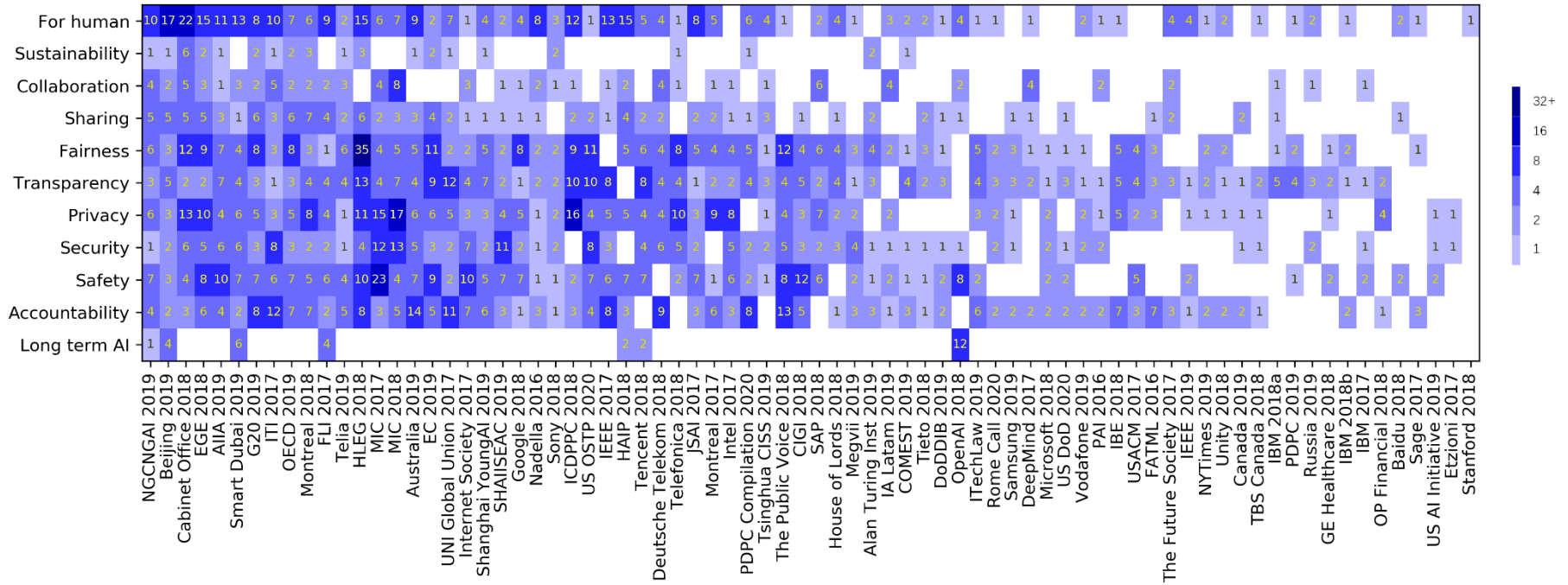
Global Vision of Artificial Intelligence Development

Different School of Thoughts for the Vision of AI

Harmonious AI
AI for Optimizing Symbiosis

Human-centered/centric AI: Stanford, Japan

AI for Sustainable Development:
NGCNAI, BAIP, UAE, AI4Good

Trustworthy AI: EU HLEG

Human Centric ←————————————————————→ **Ecology/ Cosmos Centric**

Friendly AI: MIRI

AI for (Social) Good: Microsoft

Human-compatible AI:
UC Berkeley

Responsible AI: Google, PwC

Beneficial AI/AGI : FLI/MIT

AI Principles: A Global Landscape
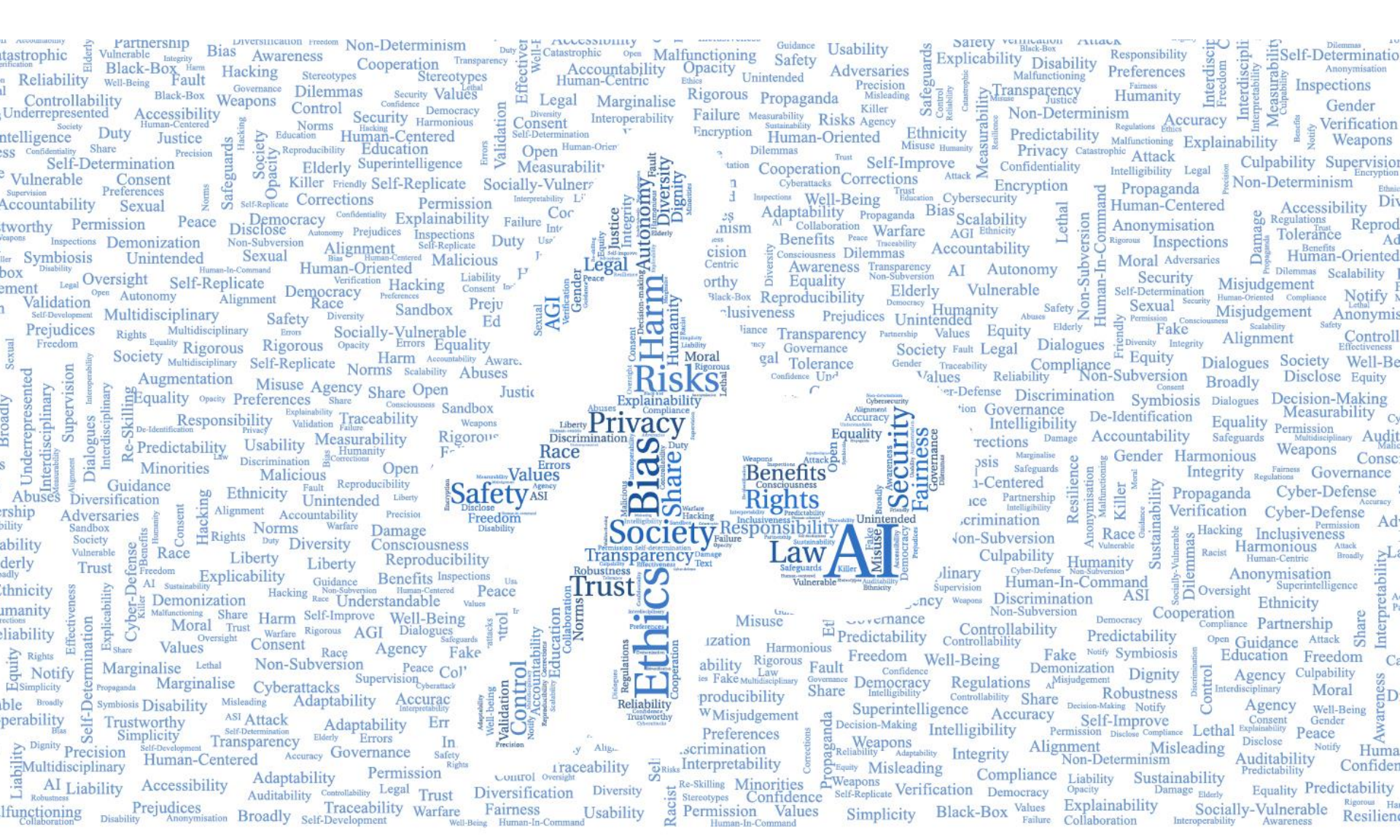
At least 76 AI principle proposals World wide, by the end of October 2020, from governments, research institutions, NPO & NGO and industry.

http://www.linking-ai-principles.org/

AI Principles: A Global Landscape

What we can learn from this landscape

- None of the proposals cover more than 70% of the concerned finer topics.

- Principle proposals should not compete with each other, instead, they complete each other to provide the global landscape and the real future of AI.

A Topic Cloud of Different Artificial Intelligence Principles All Over the World in the Shape of a Chinese Character "Optimizing Symbiosis".

Linking Artificial Intelligence Principles   http://www.linking-ai-principles.org
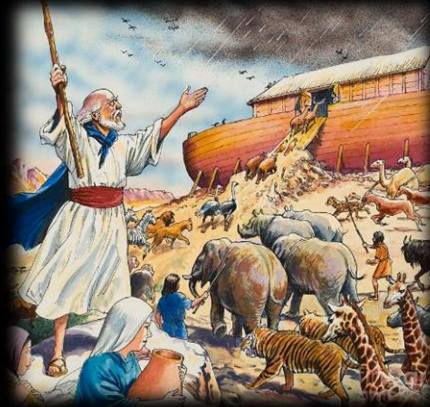
The root of Harmony (optimizing symbiosis) can be traced back to Confucian harmony. (c. 500 BCE)

Harmony among self, family, governments
Harmony among different races, countries
Harmony between Human and AI

和衷共济--《尚书·皋陶谟》

Being on Noah's ark for a shared future

Be in Harmony,
yet be different!

Wa (和) is a Japanese cultural concept usually translated into English as "harmony". It implies a peaceful unity and conformity within a social group in which members prefer the continuation of a harmonious community over their personal interests.
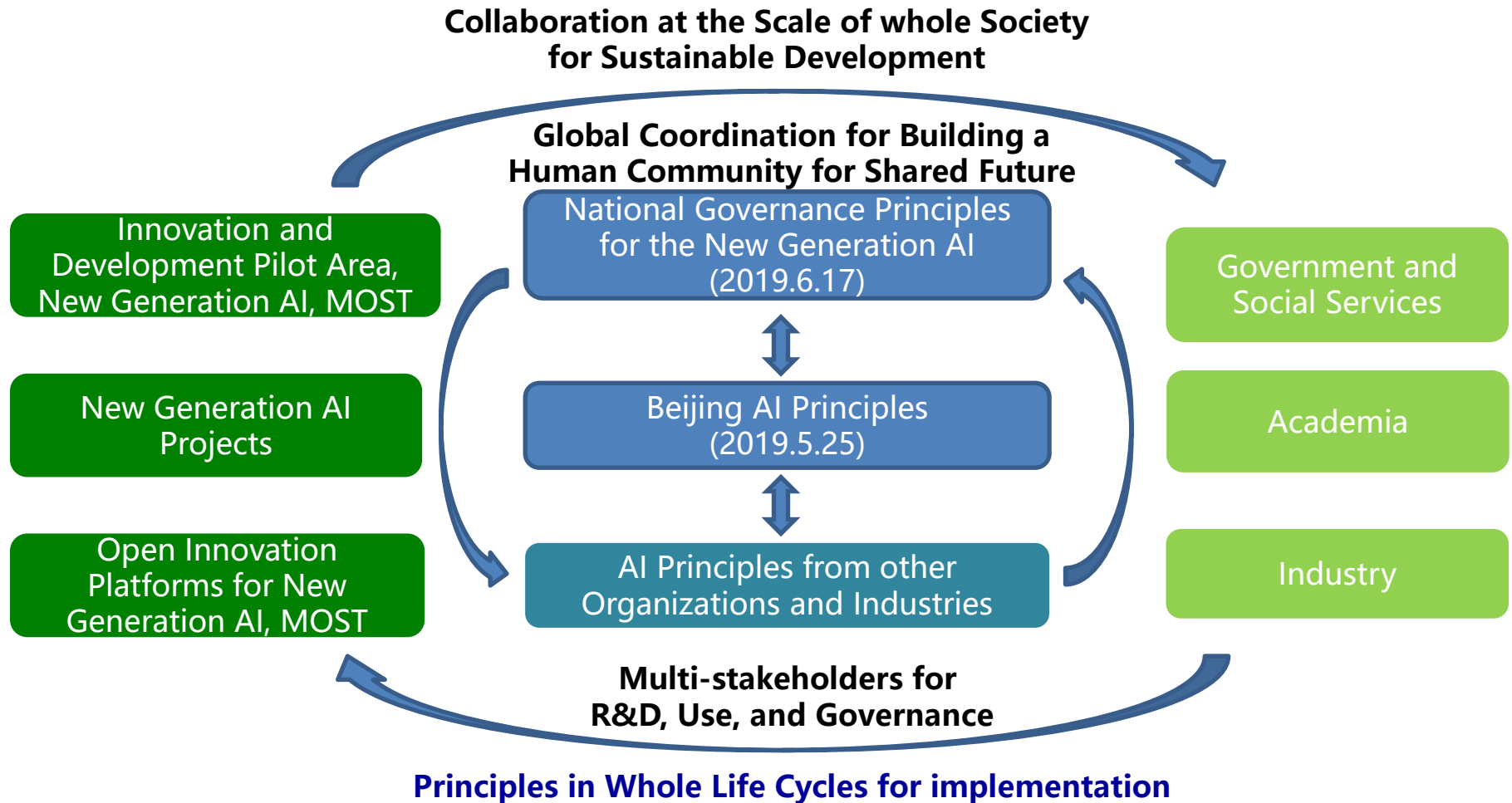
Japan: A Harmony of Past and Future

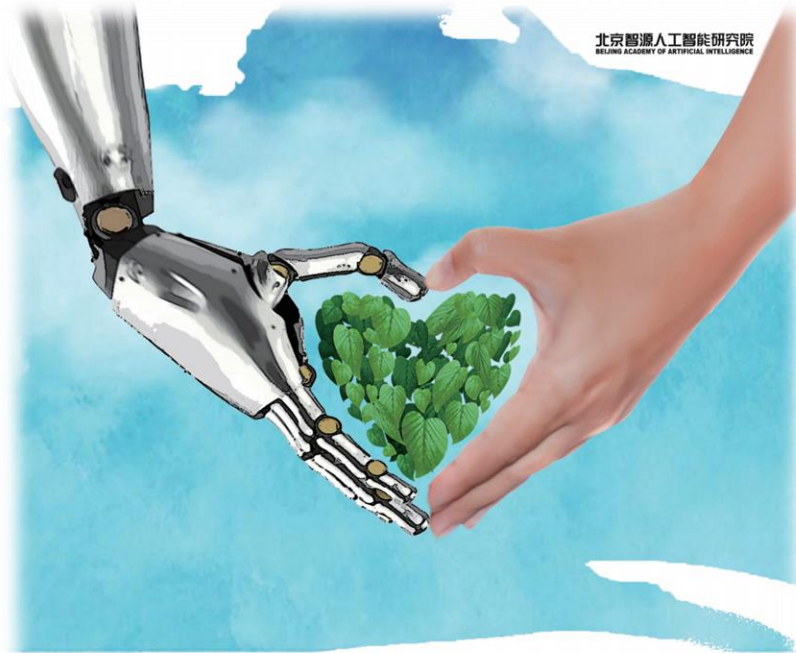Ubuntu, an African Philosophy : I am because we are

A person with Ubuntu is open and available to others, affirming of others, does not feel threatened that others are able and good, for he or she has a proper self-assurance that comes from knowing that he or she belongs in a greater whole and is diminished when others are humiliated or diminished, when others are tortured or oppressed.

Social and Technical Grounding of AI Principles in China

**Collaboration at the Scale of whole Society
for Sustainable Development**

**Global Coordination for Building a
Human Community for Shared Future**

Innovation and Development Pilot Area, New Generation AI, MOST	National Governance Principles for the New Generation AI (2019.6.17)
New Generation AI Projects	Beijing AI Principles (2019.5.25)
Open Innovation Platforms for New Generation AI, MOST	AI Principles from other Organizations and Industries

Government and Social Services

Academia

Industry

**Multi-stakeholders for
R&D, Use, and Governance**

Principles in Whole Life Cycles for implementation

- **AI to fight against COVID-19 is a concrete example for agile governance of AI**
- **Very few AI applications will announce when they will end the use even before they start to use, which exactly is practicing "For Human".**

The Use and Protection of Private Information during COVID-19 Pandemics

Facial Recognition and Public Heath

— The First Report in Survey Series on
Artificial Intelligence and Healthy Society

May 17th , 2020

Published by:

Research Center for AI Ethics and Safety, Beijing Academy of Artificial Intelligence

China-UK Research Centre for AI Ethics and Governance, Institute of Automation, Chinese Academy of Sciences

1,137 participants were involved in the survey across all the provinces in China, including Chinese language participants from 10 other countries.

The Participants never feel they could forget about privacy even during the pandemics, and they are more likely to accept accessing their private information (facial data) when there are outstanding security concerns both from government and giant companies.

Download the Report   https://www.baai.ac.cn/news/20200517.html

Finding the Balance between Ethics and Effectiveness

MIT Technology Review Covid Tracing Tracker

Search..

Location	Name	Notes	Voluntary	Limited	Data destruction	Minimized	Transparent	Tech
Algeria	Algeria's App	Algeria's app was investigated by Amnesty International.	☆	☆	☆	☆	☆	TBD
Australia	COVIDSafe	Australian experts have criticized the government for a lack of transparency and non-responsiveness to privacy issues.	★	★	★	★	☆	Bluetooth
Austria	Stopp Corona	Austria was one of the first major European nations to align with the Google/Apple API.	★	★	★	★	★	Bluetooth, Google/Apple
Bahrain	BeAware	Though 25% of the country has downloaded BeAware, there is little public information about the app.	★	★	☆	☆	☆	Bluetooth, Location
Belgium	Belgium's App*	The Belgium app is set to be released by September.	★	☆	☆	☆	☆	Bluetooth, Google/Apple DP3T
Bulgaria	ViruSafe	Bulgaria launched thier app in early April and began lifting movement restrictions in early May.	★	★	★	☆	★	Location
Canada	COVID Alert*	Canada's app is set to role out in Ontario in July, and expand to the rest of the country by late summer.	★	★	★	★	★	Bluetooth, Google/Apple
China	Chinese health code system	There is very little information available to the public about how China's technology works.	☆	☆	☆	☆	☆	Location, Data mining
Cyprus	CovTracer	The Cypriot app was one of the earliest efforts to launch, all the way back in February.	★	☆	★	★	★	Location, GPS
Czech	eRouska	eRouska is one part of the Czech government's larger "smart quarantine"	★	★	★	★	★	Bluetooth

MIT Technology Review

Sign in   Subscribe

Topics Magazine Newsletters Events

SANDEEP RASAL/ SOPA IMAGES VIA AP IMAGES

Tech policy / Privacy

India is forcing people to use its covid app, unlike any other democracy

Millions of Indians have no choice but to download the country's tracking technology if they want to keep their jobs or avoid reprisals.

by Patrick Howell O'Neill   May 7, 2020

The world has never seen anything quite like Aarogya Setu. Two months ago, India's app for coronavirus contact tracing didn't exist; now it has nearly 100 million users. Prime Minister Narendra Modi boosted it on release by urging every one of the country's 1.3 billion people to download it, and the result was that within two weeks of launch it became the fastest app ever to reach 50 million downloads.

Popular

01.
These weird, unsettling photos show that AI is getting smarter

THE WALL STREET JOURNAL.   Subscribe | Sign

Home World U.S. Politics Economy Business Tech Markets Opinion Life & Arts Real Estate WSJ. Magazine

WORLD

Contact Tracing, the West's Big Hope for Suppressing Covid-19, Is in Disarray

Investigators aren't casting a wide enough net to catch Covid-19 cases before they spread

Cafes and restaurants on the Rue de Buci, Paris, were busy on Sept. 13 despite a recent surge in Covid-19 infections. KIRAN RIDLEY/GETTY

Meituan Drives Instant Food Delivery With AI "Super Brain"

Synced
Jan 20, 2019 · 4 min read

https://medium.com/syncedreview/meituan-drives-instant-food-delivery-with-ai-super-brain-be77074792fd

Food Delivery Riders were trapped in the AI planning Algorithm

**If others can send it faster, so can you,
If others can speed up, so can you,
otherwise, you will be punished (money).**

If you really do it, you could die.

南宁交警喊话：外卖小哥们别再用生命去送餐了！

https://technode.com/2020/09/11/eleme-meituan-scramble-to-fix-driver-work-conditions-amid-criticism/

AI systems does not understand what does it mean for speeding up to drivers, and they don't care if the drivers are against the traffic rules, which may kill them.

They do not know what it means to kill people.

"We had better be quite sure that the purpose put into the machine is the purpose which we really desire."

Norbert Wiener, Some Moral and Technical Consequences of Automation. Science. May 6th, 1960.
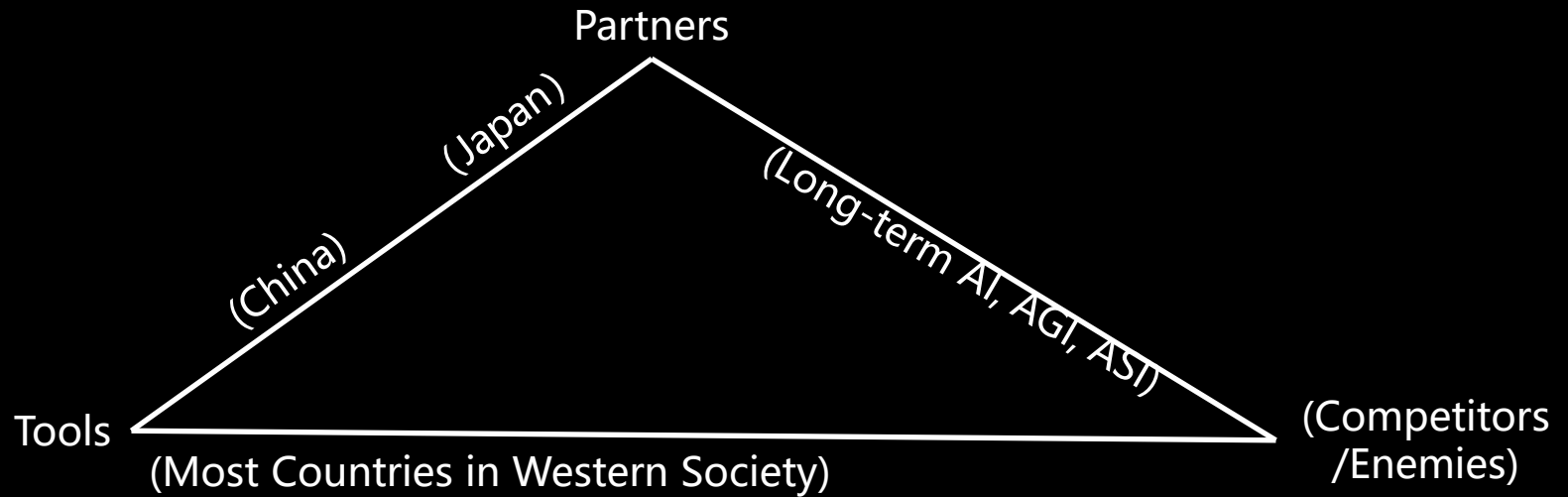
We have machines that process information.
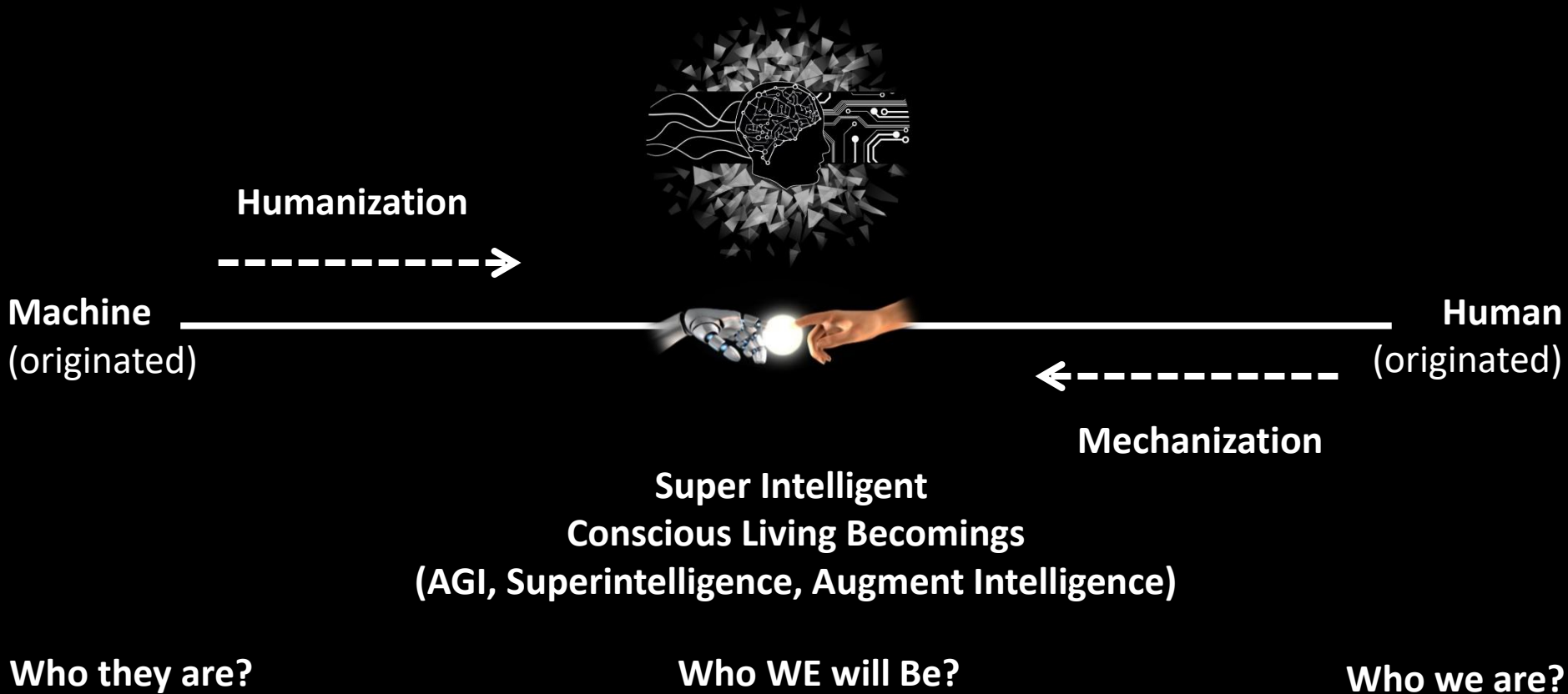
We do not have machines that understand.

If you want to understand, do we need thinking before understanding?

Can the current machine really think?

The Relationship between Human and AI:
A Global Understanding

Partners

(Japan)

(China)

(Long-term AI, AGI, ASI)

Tools

(Most Countries in Western Society)

(Competitors /Enemies)

Extending and Creating Conscious Living Becomings

Humanization

Machine
(originated)

Human
(originated)

Mechanization

**Super Intelligent
Conscious Living Becomings
(AGI, Superintelligence, Augment Intelligence)**

Who they are?

Who WE will Be?

Who we are?

Overcoming Barriers to Cross-cultural Cooperation in AI Ethics and Governance

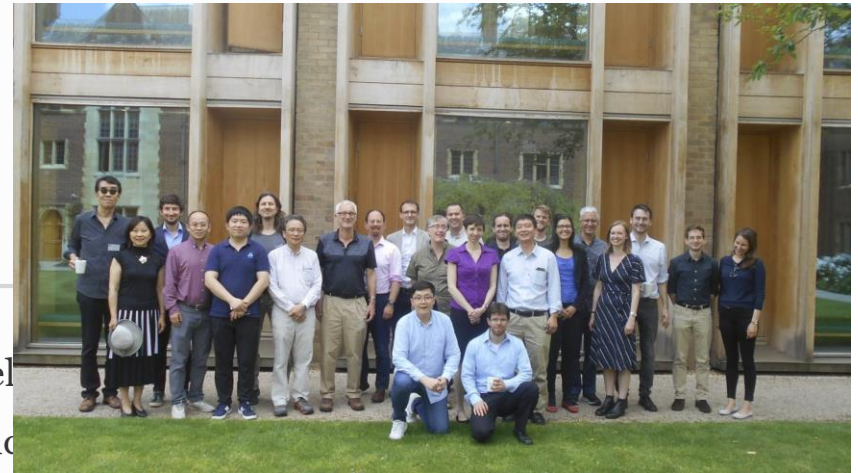Seán S. ÓhÉigeartaigh ✉, Jess Whittlestone, Yang Liu, Yi Zeng & Zhe Liu

Abstract

Achieving the global benefits of artificial intel
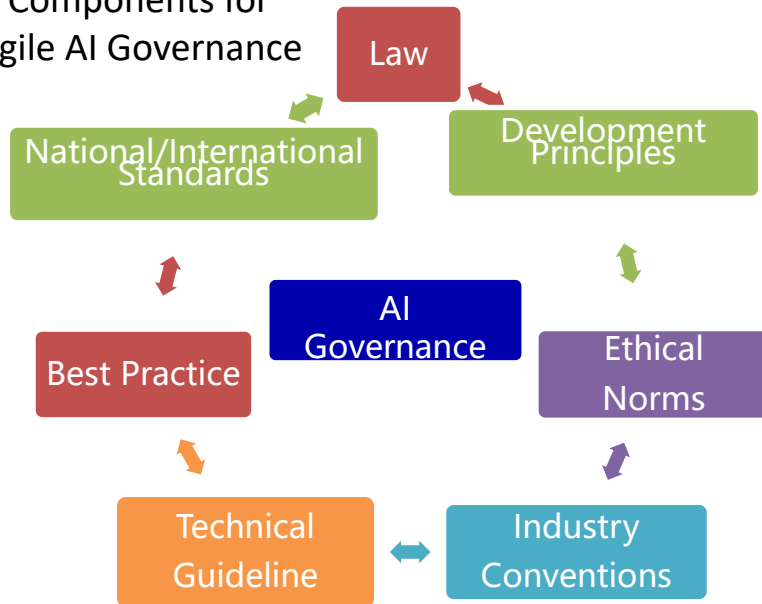cooperation on many areas of governance and
cultural perspectives and priorities. There are many barriers to achieving this at present,
including mistrust between cultures, and more practical challenges of coordinating across
different locations. This paper focuses particularly on barriers to cooperation between
Europe and North America on the one hand and East Asia on the other, as regions which
currently have an outsized impact on the development of AI ethics and governance. We
suggest that there is reason to be optimistic about achieving greater cross-cultural
cooperation on AI ethics and governance. We argue that misunderstandings between
cultures and regions play a more important role in undermining cross-cultural trust, relative
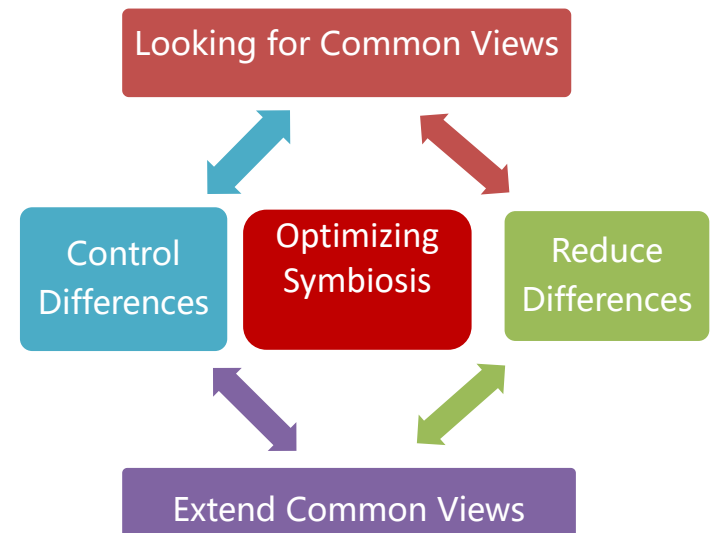
Global AI Coordination and Governance with Multiple Cultures

Creating a Shared Future through Global AI Coordination Towards Optimizing Symbiosis for Human-AI Society

Components for Agile AI Governance

- Law
- Development Principles
- National/International Standards
- AI Governance
- Ethical Norms
- Best Practice
- Technical Guideline
- Industry Conventions

Methodology for Global Governance

- Looking for Common Views
- Control Differences
- Optimizing Symbiosis
- Reduce Differences
- Extend Common Views

Artificial Intelligence for Sustainable Development Goals (AI4SDGs) Cooperation Network

UN USG Fabrizio Hochschild at the Launch of AI4SDGs Think Tank

AI for Sustainable Development Goals (AI4SDGs) Think Tank

A global collection of AI projects and proposals that impacts UN Sustainable Development Goals, both positively and negatively. The goal is to promote the positive use of AI for Sustainable Development, and to investigate on the negative impact of AI on Sustainable Development. Detailed evaluation on each project is provided based on our rating scheme. You are welcome to share your project to the world and get evaluated by submitting your project or proposal information here.

BROWSE BY GOALS · SHARE YOUR PROJECT

Artificial Intelligence for Sustainable Development Goals (AI4SDGs) Research Program
Learn more

AI4SDGs Cooperation Network
Learn more

Aritificial Intelligence: An Accelerator for UN Sustainable Development Goals
Learn more

Project Applications Now Open for the 2020 AI4SDGs Research Program
Learn more

Champion Network

- China-UK Research Centre for AI Ethics and Governance (China, UK)
- Centre for Artificial Intelligence and Robotics, United Nations, UNICRI
- The Centre for the Study of Existential Risk, University of Cambridge (UK)
- Centre for AI and Data Governance, Singapore Management University (Singapore)

Baidu
XIAOMI
DiDi
MEGVII
YIDUCLOUD
Montreal AI Ethics Institute (Canada)
YITU

The Launch of Projects for the 2020 AI4SDGs Research Program

Name of the Project	Correlation with Sustainable Development	Outcome	Funding
Sustainable Urban and Community Development and Governance Driven by Artificial Intelligence	11.Sustainable Cities and Communities 9.Industry, Innovation and Infrastructure	Research Report	200,000 RMB
The Legal System Construction and Problems of Autonomous Driving	11.Sustainable Cities and Communities 3.Good Health and Well-being	Research Report	200,000 RMB
Human Computer Intelligent Dialogue Model and System for Adolescent Mental Health	3.Good Health and Well-being 4.Quality Education	Open- Source System	500,000 RMB
Clinical Assistant Decision Support for Rare Diseases in Children	3.Good Health and Well-being	Research Report	200,000 RMB
Privacy Secure Intelligent Monitoring System for the Elderly	3.Good Health and Well-being 11.Sustainable Cities and Communities	Open- Source System	500,000 RMB
Cross-Cultural Interaction Knowledge Engine	11.Sustainable Cities and Communities 4.Quality Education	Research Report	200,000 RMB
Labor Digital Transformation and Future Employment: Challenges and Countermeasures	4.Quality Education 8.Decent Work and Economic Growth 11.Sustainable Cities and Communities	Research Report	200,000 RMB

Constituent Network

- Central Asia Tech Law (Pakistan)
- Institute of Technological Ethics (Aus
- C Minds (USA, Mexico)
- AI Policy Exchange (India)
- Artificial Intelligence 4 Development Agency (Austria)

Yi Zeng
(Beijing Academy of Artificial Intelligence, and Chinese Academy of Sciences, China)

Danit Gal
(United Nations, and Leverhulme Centre for the Future of Intelligence, University of Cambridge)

Adrian Weller
(Alan Turing Institute, and University of Cambridge, UK)

Seán Ó hÉigeartaigh
(Center for the study of existential risk, University of Cambridge, UK)

Wendell Wallach
(Yale University, and The Hastings Center, USA)

Eugenio Vargas Garcia
(United Nations General Assembly)

Brian Tse
(Future of humanity institute, University of Oxford, and Partnership on AI, UK)

Edson Prestes
(Federal University of Rio Grande do Sul, Brazil)

Serge Stinckwich
(United Nations University Institute in Macau, Macau SAR China)

Kay Firth-Butterfield
(World Economic Forum)

Mattia Fantinati
(Chamber of Deputies, Presidency of Council of Ministries, Italy)
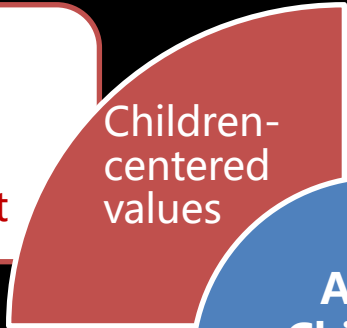
Steven Edwin Vosloo
(UNICEF)

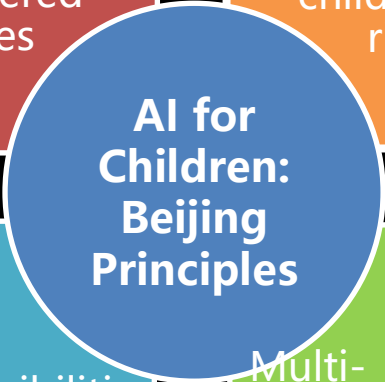**Petra Ahrweiler**
(Johannes Gutenberg University Mainz, Germany)

Alaa Khamis
(General Motors, Canada)

- Dignity
- Growth
- Fairness
- Children first

Children-centered values

Protecting children's rights

- Safety protection
- Physical and mental health
- Privacy protection
- Quality education
- Expression of will

AI for Children: Beijing Principles

- Be responsible
- Control risks
- Explain accordingly
- Ensure informed
- Train and guide
- Be sustainable

Taking responsibilities

Multi-stakeholder governance

- Policy-making
- Legal system improvement
- Multi-stakeholder collaboration
- Openness and sharing

https://www.baai.ac.cn/ai-for-children.html

Artificial Intelligence for Children: Beijing Principles

September 14th, 2020

面向儿童的人工智能北京共识

Technology Review

Tech policy / Kids and tech

Why kids need special protection from AI's influence

Algorithms are increasingly shaping children's lives, but new guardrails could prevent them from getting hurt.

by Karen Hao   September 17, 2020

Algorithms can change the course of children's lives. Kids are interacting with Alexas that can record their voice data and influence their speech and social

Long term impact for Abuse of AI: What kind of AI are we creating for the Next Generation!

Facial and emotion recognition in classrooms

- Emerged in China, America, EU.
- Students are expected to refine their behavior through observations, reflexive thinking and education.
- Fundamentally changed the way teachers and students interact with each other in a bad and unreliable way.

- Facial and emotion recognition in classrooms for more than 20 years in my life analyzing my real time emotion and thinking?
- Oh, THIS IS the AI that is always on me for 18 years, and you want me to learn it as a profession? (Negative impression for AI in the whole life).
- Revenge: I can create a technology that will save us from AI Surveillance!

And the other Side?

- Campus bullying (among students, teachers and students)

- Cyber bullying

https://medium.com/artificialciti/school-bullying-how-ai-provides-solutions-9a387eb9c7ac
https://www.asiancrush.com/japanese-city-planning-to-use-artificial-intelligence-to-track-prevent-bullying/

AI4SDGs AI for Children Working Group

Introducing the Working Group

Children are the hope and future for the development of humanity and the planet. At the same time, children are also disadvantaged and vulnerable groups and require special attention and caring from the whole society, especially in the context that the current development of Artificial Intelligence (AI) is bringing profound and extensive impact to society in an unprecedented way.

All sectors of society should attach great importance to the impact of AI on children and develop responsible AI for the next generations. The development of AI should protect and promote the benefits of children, avoid depriving and harming children's rights, and help realize the healthy growth of children.

The AI4SDGs AI for Children Working Group is affiliated to the AI4SDGs Cooperation Network. With the vision of developing responsible AI for Children all over the world, this working group is devoted to promote the positive use on AI for Children and avoid potential negative impacts, helping the healthy development of children and next generations.

Research and Practice

This working group will organize experts in AI+Education and AI for Children, industrial partners of AI4SDGs to work on the following topics:

(1) Promoting the positive use of AI for children and avoiding potential negative effects brought by AI applications and services, sharing case studies and best practices.

(2) Making practices on AI policies for children world wide in real world scenarios and compare them, providing complementary feedback to policy makers.

(3) Working closely with education and children organizations and get prepared for opportunities, long term positive/negative i Children.

(4) Sharing ideas and experiences to other similar working groups world wide to promote beneficial AI for children.

Working Group Members:

Beijing Academy of Artificial Intelligence	Yi Zeng	Director for the Research Center on AI Ethics and Sustainable Development
Tomorrow Advancing Life	Fei Li	Vice President
XIAOMI	Liheng Zhou	Vice Director for AI Ecology
MEGVII	Yuncheng Xu	Dean of Research Institute of AI Governance
AIXUEXI	Jianchao Wang	Director of AILab
YIDUCLOUD	Shi Zhang	President
GeekBang	Taiwen Huo	President
The Chinese University of Hong Kong	Helen Meng	Chair Professor
United Nations University Institute in Macau	JeongHyun Lee	Research Fellow
University of Oxford	Brian Tse	Policy Scholar
Tongxing School	Jingfang Hao	CEO
Qihoo 360	Yafeng Deng	Vice President
Baidu	Ken Lin	Senior Strategy Analyst, SLG
Baidu	Yanqiu Wang	Senior Product Manager, ACG

Welcome to join us. Contact: cooperation-network@ai-for-sdgs.academy

Observer:

Wenying Su, UNICEF China

Related Works and Efforts:

1. UNICEF Policy Guidance on AI for Children
2. Artificial Intelligence for Children: Beijing Principles
3. Generation AI (UNICEF & WEF)
4. Memorandum on Artificial Intelligence and Child Rights

We can only see a short distance ahead, but we can see plenty there that needs to be done.

-- Alan Turing
Computing Machinery and Intelligence
Mind, Volume LIX, Issue 236, October 1950, Pages 433–460

Thank You!

http://bii.ia.ac.cn/~yizeng/

yi.zeng@ia.ac.cn