

ISSN 2096-6083  
CN 10-1524/G

# Culture Of Science



[cos.cnais.org.cn](http://cos.cnais.org.cn)

2026  
Volume 9 · Issue 1 · March

**NAIS**

National Academy of Innovation Strategy  
China Association for Science and Technology



## Honorary Director of Editorial Board

**Qide Han**, Chinese Academy of Sciences, China Association for Science and Technology, China

## Director of Editorial Board

**Hui Luo**, China Association for Science and Technology, China

## Editors-in-Chief

**Haojun Zheng**, National Academy of Innovation Strategy, China

**Bernard Schiele**, Université du Québec, Canada

## Associate Editors

**Zhiqiang Hu**, University of Chinese Academy of Sciences, China

**Zhengfeng Li**, Tsinghua University, China

**Daya Zhou**, CAST Center for Professional Training and Services, China

## Invited Editor of Current Issue

**Lu Gao**, Tsinghua University, China

## Director of Editorial Office

**Xuan Liu**, National Academy of Innovation Strategy, China

## Managing Editor

**Ji Zhao**, National Academy of Innovation Strategy, China

## Coordinating Editor

**Yanling Xu**, National Academy of Innovation Strategy, China

## Data Editor

**Bankole Falade**, Stellenbosch University, South Africa

## Copy Editor

**James Dixon**, Institute of Professional Editors, Australia

## Editorial Board Members

**Martin W Bauer**, London School of Economics and Political Science, UK

**John Besley**, Michigan State University, USA

**Massimiano Bucchi**, University of Trento, Italy

**Michel Claessens**, European Commission, Belgium

**John Durant**, Massachusetts Institute of Technology, USA

**Zhe Guo**, China Science and Technology Museum, China

**Liuxiang Hao**, University of Chinese Academy of Sciences, China

**Robert Iliffe**, University of Oxford, UK

**Les Levidow**, The Open University, UK

**Xuan Liu**, National Academy of Innovation Strategy, China

**Jianjun Mei**, University of Cambridge, UK

**Gauhar Raza**, National Institute of Science Communication and Information Resources, India

**Fujun Ren**, National Academy of Innovation Strategy, China

**Shukun Tang**, University of Science and Technology of China, China

**Hongwei Wang**, Chinese Academy of Social Sciences, China

**Xiaoming Wang**, Shanghai Science and Technology Museum, China

**Masataka Watanabe**, Tohoku University, Japan

**Jiangyang Yuan**, University of Chinese Academy of Sciences, China

**Li Zhang**, Peking University, China

**Yandong Zhao**, Renmin University of China, China

## Contents

### **Special topic: Global governance of technological ethics: Historical evolution, innovative challenges and China's role in multi-stakeholder participation**

Rethinking global technology governance at a crossroads: China's role, historical turning points and future imaginaries 3

*Lu Gao*

Forks in the road: Françoise Baylis on ethics, genome editing, and the world we want to live in 9

*Ping Yan and Lu Gao*

Addressing the dual challenges of scientific and technological risks: Deep dilemmas and system transformation in global governance 21

*Yidong Liu*

Ethical AI governance: AI for society and a co-learning approach 37

*Xiaobai Shen and Lu Gao*

Exploring the design approach to embedding ethics in technology 47

*Wei Zhang, Yu Jing and Qian Wang*

### **Article**

Event, society, and future: Revisiting Chinese public discourse on the 'gene-edited babies incident' 62

*Shuo Wang and Zhengfeng Li*



# Rethinking global technology governance at a crossroads: China's role, historical turning points and future imaginaries

Cultures of Science  
2026, Vol. 9(1) 3–8  
© The Author(s) 2026  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/20966083261434257  
[journals.sagepub.com/home/cul](https://journals.sagepub.com/home/cul)



Lu Gao<sup>1</sup> 

Received: 3 March 2026; accepted: 4 March 2026

## I. Reframing global technology ethics in a time of upheaval

When we first began planning this special issue in 2024, many still understood the ‘global order’ as an institutional framework that, although visibly strained, remained broadly functional. The economic, political and social arrangements established in the aftermath of the Second World War continued to underpin international cooperation, and technological development could still, by and large, be placed within a familiar narrative of modernization. At that time, despite regulatory lag, rising risks and the growing difficulty of international coordination, there remained a widespread belief that institutional adjustment, transnational negotiation and normative revision would be sufficient to provide at least a basic governance framework for emerging technologies.

Within only two years, however, this background has changed markedly. Narratives of global risk increasingly foreground competition, fragmentation and uncertainty. Intensifying geopolitical and geoeconomic tensions, together with the erosion of institutional trust, have transformed ‘global cooperation’ from an assumed backdrop into a fragile achievement that must be actively maintained (Ikenberry, 2018).

Technology itself has also undergone a qualitative transformation. Around 2025, artificial intelligence (AI) moved rapidly from the realm of frontier headlines into everyday life and the routine operations of institutions, entering key domains

such as education, research, creative work, administration and decision-making. The speed of diffusion and degree of social penetration associated with generative AI have made it increasingly comparable to a new infrastructural layer. More importantly, ‘AI for science’ has begun to embed algorithmic capacities directly into the chains of scientific discovery, reshaping the temporal rhythms of knowledge production in fields such as the life sciences and materials science, and compressing the interval between hypothesis, testing and validation (Brynjolfsson and McAfee, 2014).

Biotechnology has undergone similarly profound changes. Over the past several years, gene editing, synthetic biology and their associated governance frameworks have been repeatedly reshaped by international summits, public controversies and institutional responses. Never before have technology, ethics and governance been so tightly coupled. The He Jiankui case in 2018 marked a critical rupture in this trajectory. It compelled the global governance community to confront a foundational question anew: when technological capacities cross previously accepted boundaries, who has the

<sup>1</sup>Tsinghua University, China

### Corresponding author:

Lu Gao, School of Marxism, Tsinghua University, Haidian District, Beijing 100084, China.  
Email: [gao\\_lu@mail.tsinghua.edu.cn](mailto:gao_lu@mail.tsinghua.edu.cn)



authority to decide how far such developments may proceed, who is entitled to represent ‘broad societal consensus’, and who bears responsibility for irreversible consequences?

It is in this sense that the keywords of this special issue—‘global technology ethics governance’, ‘the state’ and ‘China’s role’—have all undergone significant shifts in meaning over the past two years. Who speaks for the ‘global’? Where, in fact, is the global? These are no longer merely rhetorical questions in the register of geopolitics; they are also epistemological questions at the heart of technology ethics governance itself. Which forms of knowledge are recognized as credible evidence? Which risks are defined as warranting urgent governance? Which actors possess the authority to interpret, adjudicate and organize collective responses? Science and technology studies (STS) has long reminded us that knowledge and order are co-produced within specific social worlds. As Jasanoff (2004) argues, scientific knowledge is never a neutral input external to society; nor is governance a passive response to technological facts; rather, the two are continuously shaped together within concrete social orders.

Institutionally, global technology governance now increasingly takes the form of a polycentric, multi-level and multi-actor configuration. In the field of AI, the European Union’s AI Act represents a pathway centered on binding legislation and compliance obligations. The OECD AI principles and UNESCO’s *Recommendation on the Ethics of Artificial Intelligence* exemplify forms of soft law that place greater emphasis on coordination and norm articulation (OECD, 2019; UNESCO, 2021). Meanwhile, multilateral declarations such as the Bletchley Declaration and the Seoul Declaration continue to bring terms such as ‘safety’, ‘innovation’ and ‘inclusion’ into rapidly evolving international agendas. Under such conditions, the ‘global’ can no longer be understood as a single centre or a single voice. It is better conceived as an ongoing process of contestation, negotiation, competition and translation.

This special issue enters that debate at a particularly timely moment. Its aim is to bring a more fundamental question back into focus: in an era shaped

simultaneously by technological acceleration, geopolitical instability and fragmented governance, how is the ‘global’ still possible, and how can technology ethics governance be formed, revised and sustained?

## 2. Repositioning China in global technology governance

If the broader context of ‘global technology ethics governance’ has changed, then the meaning of ‘China’s role’ has shifted as well. This shift is not reducible to an abstract narrative of ‘rise’; nor is it well captured by essentialist claims about civilization. What has changed more concretely is China’s position within global governance and the manner of its participation. On the one hand, China can no longer be understood simply as a passive recipient of global technological norms. On the other hand, the issue is equally one of how China’s own discourses and value frameworks come to matter within global settings. In an era defined by intensified geopolitical competition and the politicization of technology, China’s governance practices, institutional language, policy initiatives and cultural resources are increasingly entering the arena of global technology ethics debates. They are cited, questioned, misread and at times translated into new instruments of dialogue.

China’s recent articulation of global AI governance differs from the more event-driven, post hoc and pragmatic regulatory orientation that characterized earlier approaches in the field of biotechnology (Zhang, 2017). Especially in the aftermath of the He Jiankui case, China began to strengthen research ethics governance in a more systematic manner, elevating previously dispersed review mechanisms and normative requirements into a more consolidated national framework of institutional design and governance principles. On that basis, it has become more proactive in advancing its own principles concerning technological development, safety and responsibility (Lei et al., 2019; Wang, 2024). In 2023, China proposed the Global AI Governance Initiative, emphasizing a people-centered approach, ‘AI for good’, the joint pursuit of development and

security, and opposition to ideological demarcation and exclusionary technological blockades (MoFA, 2023). The *Shanghai Declaration on Global AI Governance* further articulated ‘shared development, shared security, shared governance, and shared benefits’ as key principles for global AI governance (MoFA, 2024). These positions remain in the process of further development and practical testing, yet they already signal a clear trajectory: China is moving from being included in global discussions to participating more actively in shaping them.

This also requires a shift in how China is analysed. Rather than remaining within binary frameworks such as ‘input/output’ or ‘imitation/substitution’, it is more productive to examine how governance itself is co-produced. Global technology ethics governance is not created by any single state, civilization or normative system in isolation. It emerges through continual negotiation, institutional shaping, and revision among states, international organizations, corporations, scholars and publics (Jasanoff, 2004). In this respect, China’s significance in global technology governance lies not only in its institutional practices but also in the theoretical position increasingly assigned to it.

From the perspective of intellectual history, the impulse among Western thinkers to ‘seek remedies in China’ is hardly new. In the late seventeenth century, Leibniz, in *Novissima Sinica (Writings on China)*, proposed a ‘commerce of light’ between Europe and China, hoping to draw inspiration from China in the domains of morality and political practice (Leibniz, 1994). In a more recent philosophical context, Jullien (2000) turned ‘China’ into a conceptual instrument through which the blind spots of Western thought might be exposed. In 2017, Bruno Latour came to China to participate in the ‘Reset Modernity’ discussions, an intervention that sought, against the backdrop of crises in European modernity and ecological transformation, to explore interpretive pathways beyond Western naturalism and its established narratives of modernity (Research Network for Philosophy and Technology, 2017). These intellectual trajectories suggest that the treatment of China as a theoretical

resource, a comparative object and even an external reference point for the crises of modernity is far from accidental within Western traditions of thought.

The biotechnology governance workshop organized by the Institute for the History of Natural Sciences STS Center (2024) for this special issue on 11 October 2024 brought such expectations into a concrete setting. Organized around the themes of ‘solidarity’ and ‘consensus-building’, the workshop explored global perspectives on biotechnology governance alongside Chinese experiences. During the event, Françoise Baylis—an internationally renowned bioethicist and current President of the Royal Society of Canada—emphasized that, before asking how technology might be applied ‘ethically’, a prior question must be addressed: what kind of future society do we actually wish to inhabit? The significance of governance lies precisely in this prior act of orientation. Luis Campos, Professor of History at Rice University, revisited key moments in the history of modern technology governance through the lens of the ‘spirit of Asilomar’, reminding us that public reflection on technological consequences is itself part of technological development.

It was also in this discussion that we turned to the idea of ethics as a ‘touchstone’. In Western ethical discourse, a touchstone commonly refers to a criterion by which value judgements are tested. When the conversation shifted to the Chinese tradition of Taihu scholar’s stones, however, Baylis and Campos both showed strong interest and proposed a suggestive analogy. If Western ethical traditions often privilege the testing of judgements against explicit principles, might the Taihu stone—with its aesthetic of thinness, translucence, perforation and wrinkling, and its emphasis on porosity, passage, texture and structural tension—offer a different way of thinking about ethical judgement? Such a discussion does not imply that Chinese culture offers a ready-made answer. It does, however, illuminate an important fact: many Western scholars are interested in Chinese experience not only because China is an important state actor, but also because they hope to find in Chinese traditions a different source of ethical insight—one that places greater emphasis on relationality, structure,

vulnerability and emergence than the dominant Western languages of principlism and rule-based reasoning do.

These discussions provide an important background for this special issue. What concerns us here is the question of which elements within China's intellectual traditions, institutional practices and contemporary governance experiences can enter into dialogue within global technology ethics, which concepts and practices can be translated, debated and tested, and how such engagements might expand the imaginative horizon of contemporary technology governance.

### **3. From rupture to reconstruction: The intellectual arc of this special issue**

Against this backdrop, the four contributions collected in this special issue form a relatively coherent trajectory. They move from a critical event to structural diagnosis, and then onward to governance frameworks and design methodologies. Each article addresses a different dimension of global technology ethics governance, yet all converge on a common question: in an era of rapidly expanding technological capabilities and constantly shifting institutional orders, how can ethics meaningfully enter into the directional choices that shape technological development?

The first contribution, by Yan Ping and Gao Lu, is an interview article of Professor Françoise Baylis. In the article, Baylis takes the He Jiankui case as its point of departure and reconsiders the normative boundaries and social legitimacy of human genome editing governance. It focuses in particular on the notion of 'broad societal consensus' introduced at the First International Summit on Human Genome Editing in 2015, and asks how that term has subsequently been interpreted, narrowed and transformed in global discussions. In this context, the He Jiankui case exposed tensions embedded within what had appeared to be relatively stable international governance language, and brought back to the centre of debate the question of who is entitled to define consensus and determine acceptable limits.

The second contribution, by Liu Yidong, moves from a single event to deeper structural risks. It addresses major technological risks, ruinous forms of knowledge and the systemic dilemmas facing global governance. The article argues that certain technological risks accumulate through the diffusion of knowledge, institutional lag and intensifying competition. What is at stake, therefore, is not only whether regulatory tools are sufficient, but also whether risks can be recognized early enough to permit anticipatory judgement, and whether boundary-consciousness can be sustained under accelerating logics of development. In this sense, 'vigilance' or a cultivated sense of anticipatory concern acquires a concrete relevance for contemporary governance under conditions of high-risk technological transformation.

The third contribution, by Shen Xiaobai and Gao Lu, shifts the discussion to AI governance through the concept of 'co-learning'. The article presents ethics as an ongoing process of formation, in which developers, regulators, users and the general public all participate in shaping and revising normative boundaries over time. It also deliberately draws on Chinese intellectual resources, especially Confucian understandings of relationality and responsibility, together with Daoist perspectives on dynamic balance, adaptability and emergence. In doing so, the article resonates directly with the intellectual-historical trajectory discussed above: sustained international interest in Chinese experience stems not only from China's importance as a governance actor, but also from the possibility that Chinese traditions may offer conceptual resources beyond the dominant Western grammar of rules and principles.

The fourth contribution, by Zhang Wei, Jing Yu and Wang Qian, examines how ethics can be embedded in design. Rather than treating ethics as an external evaluative mechanism applied after the fact, the article argues that ethical considerations must enter technological systems at the design stage, shaping structures, processes and modes of interaction from the outset. Here again, Chinese intellectual resources are mobilized, especially ideas such as 'governing tools through the Dao' and 'technique as a bearer of the Way'. These ideas are translated

into a design orientation centered on responsiveness, restraint, relationality and value direction. Ethics thus becomes not only a language for judging consequences, but also a practical resource for shaping technological form and conditions of use.

Taken together, these four contributions suggest several theoretical threads that merit further development. The first is **openness**. Openness here concerns not only open knowledge, open technological ecosystems and open institutional arrangements, but also the maintenance of feedback, revision and negotiation within design itself. It relates all these to contemporary open-source practices and to the generative, revisable qualities emphasized by design-oriented ethics. The second is **participation**. Participation points to a governance logic of co-learning and co-formation, in which normative boundaries are shaped through sustained interaction. In the Chinese intellectual context, this can also be related to an ethics of engagement with the world: individuals do not complete themselves outside the public realm, but assume responsibility through enduring relations with larger forms of order. The third is **consciousness of contingency**. Liu's analysis makes clear that, under conditions of high risk, strong spillover and irreversibility, development and security remain in constant tension, and that risk-consciousness and boundary-consciousness therefore become intrinsic to technology governance. The fourth concerns **an alternative horizon of order**. If China's recent global AI governance initiatives, the *Shanghai Declaration on Global AI Governance* and the cooperative narratives associated with the Belt and Road Initiative are considered together, one can discern an ongoing effort to imagine a form of global technology governance that places greater emphasis on shared futures, coordination and mutual benefit, even while the nation-state remains the primary unit of political action. Zhao's (2005) discussion of the *Tianxia* system offers a provocative and productive conceptual resource in this regard. It suggests that the 'global' is not merely a competitive arena among states, but also a broader ethical and political horizon. For technology ethics governance, this implies that institutional design requires not only *li* (理)—forms of reasoning that can be publicly defended and translated across cultures, but also *qing* (情)—a disposition of care and understanding

towards different social conditions, developmental stages and historical experiences.

This special issue does not, of course, resolve all of these questions. Where exactly is the global? Who can speak for it? How will China's role continue to evolve as technological orders are reorganized? These remain open problems. Yet the four contributions gathered here place critical events, structural risks, governance frameworks and design pathways within a single conceptual map. In doing so, they provide an important starting point for more sustained cross-cultural and cross-institutional dialogue. Technology ethics governance is not merely an ancillary question of how to attach moral constraints to technical systems. It concerns, more fundamentally, how we understand risk, organize responsibility, shape rules and ultimately decide together the forms of life that future societies will inhabit.

#### ORCID iD

Lu Gao  <https://orcid.org/0000-0002-3367-2888>

#### Funding

The author received no financial support for the research, authorship, and/or publication of this article.

#### Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

#### References

- Brynjolfsson E and McAfee A (2014) *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. London: W. W. Norton.
- Ikenberry GJ (2018) The end of liberal international order? *International Affairs* 94(1): 7–23.
- Institute for the History of Natural Sciences STS Center (2024) Symposium on the Solidarity in Governing Biotechnology: Chinese and Global Perspectives held in Beijing. Available at: [https://www.ihns.ac.cn/njgsz/yjxt/kjyshyhzx/STSEvents/202410/t20241028\\_7409480.html](https://www.ihns.ac.cn/njgsz/yjxt/kjyshyhzx/STSEvents/202410/t20241028_7409480.html) (accessed 3 March 2026, in Chinese).

- Jasanoff S (2004) *States of Knowledge: The Co-Production of Science and Social Order*. London: Routledge.
- Jullien F (2000) *Detour and Access: Strategies of Meaning in China and Greece*. New York: Zone Books.
- Lei R, Zhai X, Zhu W, et al. (2019) Reboot ethics governance in China. *Nature* 569(7755): 184–186.
- Leibniz GW (1994) *Writings on China*. Chicago: Open Court.
- MoFA (Ministry of Foreign Affairs of the People's Republic of China) (2023) Global AI Governance Initiative. Available at: [https://www.mfa.gov.cn/mfa\\_eng/zy/gb/202405/t20240531\\_11367503.html](https://www.mfa.gov.cn/mfa_eng/zy/gb/202405/t20240531_11367503.html) (accessed 2 March 2026).
- MoFA (2024) *Shanghai Declaration on Global AI Governance*. Available at: [https://www.mfa.gov.cn/eng/xw/zyxw/202407/t20240704\\_11448351.html](https://www.mfa.gov.cn/eng/xw/zyxw/202407/t20240704_11448351.html) (accessed 2 March 2026).
- OECD (Organisation for Economic Co-operation and Development) (2019) *OECD Principles on Artificial Intelligence*. Available at: <https://archive.epic.org/algorithmic-transparency/OECD-AI-Principles-flyer.pdf> (accessed 2 March 2026).
- Research Network for Philosophy and Technology (2017) Workshop: Reset Modernity! Keynote by Bruno Latour. Available at: <https://philosophyandtechnology.network/402/workshop-reset-modernity-keynote-by-bruno-latour/> (accessed 2 March 2026).
- UNESCO (United Nations Educational, Scientific and Cultural Organization) (2021) *Recommendation on the Ethics of Artificial Intelligence*. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000381137> (accessed 2 March 2026).
- Wang G (2024) Ethics committees promote responsible research in China. *Nature Human Behaviour* 8(7): 1226–1227.
- Zhang JY (2017) Lost in translation? Accountability and governance of clinical stem cell research in China. *Regenerative Medicine* 12(6): 647–656.
- Zhao T (2005) *The Tianxia System: An Introduction to the Philosophy of a World Institution*. Nanjing: Jiangsu Education Press (in Chinese).

### Author biography

Lu Gao is an associate professor at the School of Marxism, Tsinghua University. She holds a PhD from Tsinghua's STS Institute and previously served as an associate professor and the Director of the Institute for the History of Natural Sciences STS Center, Chinese Academy of Sciences. Her research focuses on emerging technology governance and the history of biotechnology. She has published over 40 articles in Chinese and English, and recently co-edited a special issue on 'Humanizing RRI' in the *Journal of Responsible Innovation*. Her forthcoming monograph, *From Participation to Co-Governance: Biotechnology Governance from an STS Perspective*, will be published soon. She has conducted visiting research at the University of Edinburgh (ISSTI), Stanford University (East Asia Center) and the University of Kent.

# Forks in the road: Françoise Baylis on ethics, genome editing, and the world we want to live in

Cultures of Science

2026, Vol. 9(1) 9–20

© The Author(s) 2026

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/20966083261421235

[journals.sagepub.com/home/cul](https://journals.sagepub.com/home/cul)



Ping Yan<sup>1</sup> and Lu Gao<sup>2</sup> 

## Background

In October 2024, on a high-speed train from Beijing to Shanghai, Professor Françoise Baylis, distinguished research professor emerita at Dalhousie University, president of the Royal Society of Canada (2025–2028), and one of the world's most influential voices in bioethics—sat down with Yan Ping of Dalian University of Technology and Gao Lu of Tsinghua University for a four-hour conversation. Also on the journey was Professor Luis Campos, historian of science at Rice University. The group was traveling to the International Symposium on Science and Technology Ethics, organized by the Academic Divisions of the Chinese Academy of Sciences in Shanghai.<sup>1</sup>

## Setting the stage and asking the right questions

Baylis, the author of *Altered Inheritance: CRISPR and the Ethics of Human Genome Editing* (Baylis, 2019), has been deeply engaged with the International Summits on Human Genome Editing. In this conversation, she reflects on the evolution of the summit's debates, the shifting global discourse on heritable human genome editing, and the tensions between scientific ambition and ethical responsibility.

*Q:*<sup>2</sup> *Let's begin with the three international summits on human genome editing (2015 in Washington DC,*

*USA, 2018 in Hong Kong, China, 2023 in London, UK). It seems that, from your perspective, the dominant themes at the respective summits were 'broad societal consensus on human heritable genome editing' at the first summit, 'translational pathways for human heritable genome editing' at the second summit, and 'equitable development/access for human somatic genome editing' at the third summit. Does this reading make sense? Can you talk us through the background and evolution of the three summits?*

Baylis: In a dynamic field, there are many ways to answer this question. The thematic overview provided is a legitimate way of understanding what happened; it aligns with how I might interpret events (Baylis, 2023, 2025). But it's important to recognize that others will tell a story differently. For me, what matters is the emphasis, at the outset, on 'broad societal consensus'. That phrase gets used for the first time in the specific context of discussion/debate about the ethics of human genome editing in the closing statement of the first international summit in 2015 (National Academies,

<sup>1</sup>Dalian University of Technology, China

<sup>2</sup>Tsinghua University, China

### Corresponding author:

Lu Gao, School of Marxism, Tsinghua University, Haidian District, Beijing 100084, China.

Email: [gao\\_lu@mail.tsinghua.edu.cn](mailto:gao_lu@mail.tsinghua.edu.cn)



2015). To my knowledge, this is also one of the first times those words appear together in the bioethics literature. The emphasis on ‘broad societal consensus’ is significant.

Shortly after the first summit, people tried to change the language. For example, instead of ‘broad societal consensus’, they talked about ‘broad scientific consensus’. Someone like me, who is paying attention to language, saw this as an attempt to be perceived as agreeing with the initial commitment without really agreeing—there is a huge difference between ‘broad societal consensus’ and ‘broad scientific consensus’.

The idea of a ‘translational pathway forward’ for human heritable genome editing comes later; it is of central importance at the second summit. As concerns this issue, it is interesting to note the tension between the revelation (in advance of the second summit) that He Jiankui had created genome-edited babies and the summit’s closing statement that both condemns He’s translation research (from clinical trial to practice) and at the same time calls for a ‘pathway forward’ (National Academies, 2018). The organizing committee that issued the closing statement didn’t say, ‘we don’t yet have broad societal consensus, so why did you do this?’ As such, the criticism of He’s research wasn’t anchored in the 2015 closing statement; it was anchored in the solo event. Only three years had passed between the first and second summits. Nothing dramatic had changed in the science, but the narrative had shifted considerably. This reveals how quickly the frame of a conversation can change and how events can reshape discussions.

The first summit happened very quickly, nominally in response to research published in April 2015 (Liang et al., 2015). I was invited to sit on the organizing committee in late summer 2015. Our first planning meeting happened in early October, and the summit was held in early December. The summit hosts tried to do things differently: they involved the media early on; they made the meeting available online in real time (something unusual in 2015); and they made the meeting accessible to people with physical or hearing impairments (the meeting was live-streamed and closed-captioning was provided). There was live-tweeting and social-media reporting. Indeed, this was the first time I attended a meeting in person and simultaneously followed conversations about the meeting on Twitter.

This was fascinating because there were conversations happening online that were not happening in the room. Four thousand people watched online. That was huge.

## A different tone in Hong Kong and London

*Q: Do you think the three summits had different tones?*

Baylis: The second summit, held in Hong Kong in 2018, was very different from the first summit held in Washington in 2015. The second summit was originally to have been held in China’s mainland, but it was moved to Hong Kong. I didn’t attend in person; I watched online from my home in Canada in the middle of the night. There was considerable global attention on this meeting because a few days prior to the meeting the world had learned that gene-edited babies had been created by a Chinese researcher—He Jiankui.

I saw He’s presentation and the audience’s reactions. From my perspective, there was some ‘grandstanding’: people stood up and were critical without reflecting on how their criticisms of He Jiankui could apply equally to work done in Western countries. Consider, for example, criticisms about the consent process. Some of the outrage seemed performative. A rhetorical condemnation without self-reflection doesn’t advance the conversation. I’m not saying I wouldn’t have been outraged; I would have been, because I strongly support broad societal consensus. I would have asked He Jiankui (and any other scientist), ‘What makes you think you can decide this on your own? It’s not your decision to make.’

The third summit, in London in 2023, attempted to shift some of the focus from heritable genome editing to somatic genome editing while shining a light on ‘equitable development and access’. Heritable genome editing was still on the agenda, but there was an attempt to say there are more issues we must pay attention to. Some people criticized this shift, saying it avoided the prime contentious issue. My view is that it acknowledged that nearly 10 years had passed since the first summit and that, while

heritable genome editing was still a distant imagining, somatic genome editing was a reality with the anticipated move from clinical trials to therapeutic interventions for living patients. How could we not pay attention to the relevant contemporary ethical issues surrounding the use of this technology? Heritable genome editing, by contrast, remained somewhat in the realm of science fiction.

### What happened between the summits?

*Q: Then what really happened between the three summits?*

Baylis: Between the summits, there was a great deal of work by individual scholars, scientists, national bodies and professional societies. After the second summit, for example, in 2019, a paper was published calling for a moratorium on human heritable genome editing (Lander et al., 2019). The authors included several members of the original planning committee from the first summit, and scholars from different disciplines and different countries were invited to join the call for a moratorium. Three scientists, widely regarded as the CRISPR pioneers—Jennifer Doudna, Emmanuelle Charpentier and Feng Zhang—were invited to sign on. Only two of the three did so. Other scientists who were members of the planning committee, like David Baltimore, also did not sign on. These missing signatures matter historically.

The final statement from the first summit did not explicitly call for a moratorium. If you read the text, it doesn't use this word. But to my mind, this was clearly a call for a moratorium by another name. Indeed, *The New York Times* headline the next day used the word 'moratorium' (Wade, 2015). So, even if the word wasn't used, the substantive content clearly pointed the way.

For me, a moratorium is a temporary halt, not a ban. It buys time: time to reflect, time to perhaps advance the science, time to build broad societal consensus. By 2019, to remove any possible doubt about the call for a moratorium, like-minded people made the call explicit: 'Adopt a moratorium'

(Lander et al., 2019). As a result of this publication, some believe there is a moratorium on human heritable genome editing, but strictly speaking there isn't. No binding pause has been formally endorsed (and, in any case, who would have the authority to endorse and monitor a moratorium?).

One constant across all three summits is the emphasis on 'safety and efficacy'. No one wants to say, 'We can do this even if it is not safe and efficacious.' Scientists want to be responsible (and be seen to be responsible), and they want to maintain control. They want to define the standards and be the ones to assess whether those standards have been met. Because of my training, I don't see safety and efficacy as purely scientific criteria. There's a value system underlying them. It's always 'safe' relative to something and 'efficacious' relative to something [else]. No intervention will ever be 100% safe and efficacious. Someone has to decide when it's safe enough, and efficacious enough. 'Enough' is a value-based criterion. We need to be transparent about that. We shouldn't allow the standard of 'safety and efficacy' to collapse into purely factual assessments when they are not. And we should not let scientists be the only ones to decide when standards have been met.

### Revisiting the role of individual scientists

*Q: Before the first summit, a team led by Huang Junjiu in China published research that raised concerns about whether genome editing experiments should be done in human embryos. What are your thoughts on this research?*

Baylis: Well, an interesting thing you may not know about me is that my PhD thesis was on the ethics of using nonviable human embryos in research. This is relevant because the research done by Huang and colleagues—which is the first reported research use of human embryos using CRISPR genome editing technology—was done in nonviable human embryos. An excerpt of my thesis was published in the journal *Bioethics* in 1990 (Baylis, 1990). That article, published nearly 35 years ago, remains relevant today.

Getting back to your question, however, about the research published by Huang, arguably it is his work that sparked the global conversation on the ethics of human germline genome editing. Now admittedly this happened in a roundabout way. Huang originally submitted the paper published in *Protein and Cell* to both *Nature* and *Science*. Both of these journals rejected the paper, and yet they then each published commentaries that I think can fairly be described as ‘in anticipation’ of the article by Huang. To be clear, neither of the commentaries in *Nature* and *Science* name Huang. Instead, they ask and answer the general question ‘Should we do this [human heritable genome editing] research?’ Some scientists said ‘no’ emphatically (*Nature* paper, see Lanphier et al., 2015); other scientists said ‘we need a prudent pathway forward’ (*Science* paper, see Baltimore et al., 2015). Thereafter, plans were made to host the first international summit.

### Governance beyond laws

*Q: Let’s talk about global governance in the context of these summits. I understand that the first summit was intended as a step towards global governance of human genome editing. How do you see the governance process developing?*

Baylis: Nothing happens in isolation. The summits are anchors, and things happen in between. Countries, professional organizations, patient groups and public-interest groups position themselves with reference to these (and other) anchors.

In recent years, our understanding of governance has broadened. Early on, people thought of governance as clear research guidelines, or laws permitting or forbidding certain things, or international treaties. At the 2015 summit, there were formal documents submitted for consideration by UNESCO and by civil society. Different interest groups asserted their understanding of risks, potential benefits and priorities. People referred to the ‘Oviedo Convention’.

The broadest statement on governance that I know of in this space is the one issued by the World Health Organization: ‘Framework for Governance of Genome Editing’ (WHO, 2021). It unpacks the word ‘policy’ and has included international

declarations and treaties; national laws and research guidelines; professional guidelines; education systems; patents; and publications. It recognizes that all these mechanisms can helpfully constrain or encourage behavior. Governance is not just about laws; it’s also about mechanisms and responsibilities. Research funders can regulate by deciding whether to fund certain research. Publishers can regulate by deciding whether to publish certain articles. Patent offices can incentivize some research and discourage other research by the rules they enforce. Educational systems shape the next generation. If we want to govern human genome editing, we need to think about all these levers (and more).

This is important in the modern world. Otherwise, we invest a lot of resources in treaties and declarations that may not work. Treaties may not work because there is increasing distrust among nation-states. We are no longer in a postwar moment where people want to bridge differences and are eager to find ways to work together. There is also a new private–public reality that must be addressed. Private companies have enormous resources and capabilities, and they operate across borders. So old mechanisms narrowly focused on nations may not suffice.

### The role of ethicists and the politics of science

*Q: As an ethicist, you have had an impact on several summits. Did you feel pressure from the scientists?*

Baylis: Yes. There is pressure. Sometimes it is direct, sometimes subtle and maybe even inadvertent. You are invited to a meeting organized by scientists and funded by scientific bodies. There is an expectation that you will be constructive, that you will not block ‘progress’. Some scientists frame ethics as a public-relations exercise: help us manage public perception, and help us to better explain what we are doing and why it is important. This is not ethics, however. Ethics should first ask and answer questions about whether we should do something at all, before considering the ways in which it might be possible to do something ethically. But that question [the ‘whether’ question] is often unwelcome because many scientists already assume the project

under scrutiny is worthy. They want the ethicist to manage discussion and debate, not to question the project itself.

David Baltimore, who chaired the first and second summits, has been clear about his view on the need to avoid the language of a moratorium (Saey, 2019). But his view wasn't the unanimous view of members of the organizing committees. There were varied perspectives. As I have said before, while the final statement from the first summit did not use that word 'moratorium', that does not mean there was not a widely agreed upon call for what others might call a pause. Anyone reading the closing statement might well interpret it that way. At the second summit, at the beginning, Baltimore said the birth of genome-edited twins was deeply problematic, and he quoted the final statement from the first summit. Feng Zhang also quoted it. But by the end of the second summit, things had changed, as evidenced by the call for a transitional pathway forward in the closing statement.

### Incentives, success and being first

*Q: Some scientists think He Jiankui chose the wrong disease to 'cure'. He should have chosen something like heart disease or diabetes, and this would have resulted in less criticism. They call him 'clever but wrong'. How do you view such opinions, especially from younger scientists?*

Baylis: This question invites us to reflect on the incentives in science. Science rewards being first. It does so, for example, with publication in high-impact journals, Nobel (and other) prizes and positive media coverage. People like me question that system because it perpetuates the false notion that knowledge production is individual when science is team-based. Individual scientists may have great ideas, but teams are needed to realize these ideas. Yet we reward individuals. We reward being first. That system shapes how young scientists think. The incentive structure encourages risky behavior. We need to rethink what success looks like. Success should be about contributing to a better world. We need to value collaboration, openness and benefit to society, not just being first or publishing in a prestigious journal.

He Jiankui has been called a 'rogue scientist' by many (see, e.g., Fraser, 2018). I share Ben Hurlbut's view that this labeling is misguided. He Jiankui did what all young scientists are trained to do: to be at the forefront, to make a name for himself. Even now (after his release from prison), He Jiankui tries to do that—to be first. He is saying he wants to publish his 'ground-breaking' research. Many scientists insist that his research isn't sound, and for this reason it shouldn't be published. What are the relevant standards? Traditionally, we publish good science and good ethics. If there are serious reservations about both the science and the ethics, why publish? (Baylis, 2020)

### Heritable genome editing cannot cure anything

*Q: Then what do you think genome editing should do to better benefit the society?*

Baylis: Heritable genome editing isn't about cures. The reason is simple: there is no person/patient with an illness who is suffering and in need of a cure—someone who can make a claim on society for care. Somatic genome editing can offer treatments (and possible cures): if a child is born with a genetic disease, and you edit their somatic cells, you are treating (possibly curing) that child. Heritable genome editing doesn't involve persons/patients. What it offers is the possibility of creating a future human with or without certain traits depending upon what is considered valuable; it does not treat an existing being. The goal of improving humanity by manipulating the germline raises ethical concerns as this technology is not about curing disease but about choosing which people should exist.

Use of this technology encourages eugenic thinking, because it ultimately invites people to decide which lives are more or less valuable. There might be initial agreement on some very dramatic (life-limiting) conditions that are worth eliminating, but over time there very likely would be less and less agreement about which diseases or disabilities should be eradicated, and which conditions should be accepted as simple differences.

In making decisions about when to use genetic modification, we will be actively changing the world in which we live from one that is accepting of chance and diversity into a world where it is normal to decide who should be created with which traits. For the record, there are ways of having children that avoid the disease scenario without resorting to heritable genome editing. For example, if prospective parents want to avoid passing on a genetic condition to their offspring, there are existing technologies that can be used to screen pre-implantation embryos, after which decisions can be made to selectively transfer unaffected ones.

There is no compelling medical need for heritable genome editing. There may be a good reason to do heritable genome editing, but I haven't heard of one as yet.

### Bringing together a body of work

Baylis explained why she wrote *Altered Inheritance*. It was an opportunity to bring together many threads of her academic research at a time when her career was shifting towards international policy work. More generally, she was interested in helping other people pursue their substantive research rather than taking on new projects of her own. The book enabled her to collate decades of thought on impact ethics. The themes she discusses in the book—slow science, responsible bioethics, and the roles and responsibilities of scientists—apply beyond gene editing. Her core argument is that we spend too much time on secondary (downstream) questions and not enough time on the primary question: ‘What kind of world do we want to live in?’

Baylis: I've often told people that I probably won't write another book because I have nothing else to say. Everything I know is in my book *Altered Inheritance*. The primary question for me is: ‘what kind of world do we want to live in?’ Only when we know the answer to that question should we move on to the second question: ‘how will this technology help me (us) build that world?’ I find it is frustrating that many people in bioethics ask, ‘How can I do this ethically?’ instead of asking, ‘Is it ethical to do this?’ The latter is the more challenging question, and many people shy away from

it. My plea is to start at the beginning: decide what kind of world we want, then consider whether a given technology fits into that vision. That is the essence of slow science and responsible ethics.

### Treatment and enhancement

*Q: In Altered Inheritance, you make the provocative point that the difference between ‘treatment’ and ‘enhancement’ is often meaningless. Could you explain what you mean by that?*

Baylis: Much of the literature assumes that if what you're trying to do is a treatment, then it is a good thing, and if what you're trying to do [is] an enhancement, then it is bad or at least problematic. I want to challenge that assumption. There are treatments that are not good, and there may be enhancements that are good. The dichotomy—‘treatment good, enhancement bad’—is sloppy thinking.

*Q: So you're saying the terms themselves shouldn't determine the moral judgement?*

Baylis: Exactly. You shouldn't let those terms do the moral work for you. You have to do the moral work yourself by unpacking the goals and objectives. A technology that others call a treatment might, in my view, actually be an enhancement—and it might be good or bad depending on what it seeks to do.

*Q: What should guide the ethical assessment, then?*

Baylis: The key is to ask: What is the goal? Does it make sense? Does it align with your vision of a good life? For me, the label is not what matters; the underlying values do.

### Ethics is not lagging behind science; it asks different questions

*Baylis rejects the idea that ethics lags behind science. She maintains that the claim that science is outpacing (and thus surpassing) ethics is not just incorrect but disingenuous.*

Baylis: In chapter 5 of *Altered Inheritance*, called ‘Ethics in the interim’, I argue that we have had considerable time to reflect on the ethical significance of manipulating the human genome. Some say we have not used that time wisely. They complain that ethical thinking has not kept pace with science. But what does that mean? Saying ‘science surpasses ethics’ or ‘ethics lags behind science’ is not only incorrect; it is insincere. Such statements often mean, ‘ethics is hindering scientific progress’. Why assume science is the reference point? Why should ethics keep pace with science? Why shouldn’t science keep pace with ethics? Science is a human activity that should be informed by ethics.

We have decades of literature on the ethics of creating new humans. We have decades of discussion about human cloning. Ethics isn’t lagging behind; science just doesn’t like the answers ethics offers. There is frustration because ethicists point out problems. Scientists want to move quickly; ethicists say we need to stop and think (long and hard) before we act.

### **Dual-use science and the limits of control**

*Q: Many developments in science and technology—like artificial intelligence or drones—are seen as burdens of state power. Drones were once seen as civilian tools; now they are used in wars. In a movie, terrorists use genetic weapons to control the world. If genome editing can be weaponized and enters global geopolitical competition, is there any possibility to control it? Can ethics stop science from being misused?*

Baylis: Many scientists will tell you that any scientific discovery can be used for good or for evil. You should not point to the potential for evil as a way of stopping the potential for good. You can use a hammer to build a house or to kill someone. You can also use a hammer as a gavel if you are a judge, to insist on justice. The point is: technology is not inherently good or evil. How it is used depends on humans. So what matters is the human, not the technology. Once you understand that, you can start

thinking about governance differently. Instead of focusing only on prohibiting a technology, you can helpfully focus on influencing human behavior. How do you incentivize people to use a technology responsibly? How do you create norms that discourage misuse? How do you build systems that monitor and mitigate potential harms?

You will never be able to control every individual. But you can create norms and institutions that make misuse less likely. You can encourage people to share values of care and justice. You can stigmatize harmful uses of technology. You can ensure scientists think about the implications of their work. You can create transparency and accountability.

### **Challenges in global governance: Pandemic, conflict and building community**

*When asked about two challenges—achieving dialogue across different ethical, religious, cultural, social, political, legal and scientific perspectives, and moving from public education to public empowerment—Baylis acknowledges that these challenges remain acute.*

Baylis: During the past five years we have experienced a pandemic and regional wars. These experiences have shown us how we sometimes fail to care for each other. But during these challenging times, there have been stories of communities coming together and forming new connections. The world is always in transition; the key is not to be disheartened. We must remain committed to positive goals and objectives even when setbacks occur.

Governance is not only top-down but also bottom-up. We must keep doing the work of building and extending community so that we have a sense of care and compassion for those around us and even for those we do not know. Why do wars happen? In part, because we see others as fundamentally different; we don’t see ourselves as connected to each other. Building trust means engaging with people who are different, finding common ground and recognizing our shared

humanity. We may never achieve ‘broad societal consensus’, but we will be better off for having tried.

### **Building broad consensus and empowering the public**

*Q: How can we achieve broad societal consensus on genome editing and other transformative technologies?*

Baylis: Broad consensus is a process: it involves global dialogue, exchanging different perspectives and values with mutual respect, building trust, and brainstorming about how to use science and technology for a better world. How we communicate and make decisions is as important as the decisions themselves. My target audience is the human family—all of us.

We must always be willing to talk to each other even when we disagree fundamentally. Some people are unable, unwilling or uninterested in dialogue. But progress requires talking to people we don’t know, don’t understand, and may disagree with. By way of example, sometimes colleagues have told me not to go to certain countries because of how they treat women. I respond: ‘I’m a woman. If they invite me to speak and I don’t go, how does that help other women?’

I know ‘broad societal consensus’ may never be reached, but I believe that we will all be better off for having tried to work towards consensus. The attempt itself teaches us about each other and may reveal common ground. Even if we never fully agree, we may understand that there are different ways of knowing and understanding; those with whom we disagree may not all [be] crazy or wrong. As a result of trying to build consensus, we may become more open. What would that be like? It would be very different from the world we live in now.

*Q: In your book, you identified two major global policy challenges: effective dialogue across diverse perspectives and moving from public education to public empowerment. How do you see these today?*

Baylis: The pandemic highlighted how we fail in our commitments to care for each other, but it also provided examples of communities coming together. Wars show how quickly people can dehumanize others. We always need to ask: can we keep our eye on the ball? Can we recognize there will be setbacks, but maintain our goals? The challenge is not to become disheartened or cynical or pessimistic—it is easy to be all three.

If we are looking at governance, we have to think of it as both top-down and bottom-up. Top-down governance includes laws, regulations and international agreements. Bottom-up governance is about community building, public engagement and social norms. We need both. We need to extend community so we have care and compassion for those around us—including people we don’t know or who are far away.

We also need to distinguish between ‘public education’ and ‘public empowerment’. Education means informing people. Empowerment means enabling them to participate and have a say. In many places, public engagement is tokenistic. People are asked to comment on decisions that have already been made. That is not empowerment. To empower the public, you must share power. That means scientists and policymakers must be willing to cede control. They must listen to voices they might not agree with. They must recognize that people have different world views, different values and different religions. Effective dialogue is hard work. It requires trust. And trust is in short supply.

### **Rethinking ethics, governance and success in science**

*Q: How does your approach to ethics connect with this vision of consensus and empowerment?*

Baylis: Responsible ethics means thinking carefully about the potential impact of ethical arguments. It’s not just an academic exercise. Words matter. Responsible bioethics, a term developed by Dan Brock and others, emphasizes the need for

philosophers to understand that they cannot have an impact on the world unless they engage with policy-makers and help them implement ethical ideas. You have to think not just about the correct answer, but about strategies for implementation.

Impact ethics, a term introduced and explained in the blog *Impact Ethics*,<sup>3</sup> and further explored in the book *Bioethics in Action* (Dreger and Baylis, 2018), looks at how ethics can positively impact the world, recognizing that what counts as positive is contested. Impact ethics encompasses responsible ethics and responsible bioethics. Many stories in *Impact Ethics* are about challenging the status quo and challenging power. They are about naming problems and trying to change things. They are about not accepting the status quo as the norm, but about turning a reflective gaze on established practices and saying, ‘This is wrong.’

In my own work, for example, I have argued that there is a moral obligation to do research involving pregnant women. The starting assumption for many is that it is wrong to include pregnant women in research. I engage with that literature and explain why others are misguided. I argue the most ethical thing you can do is to include pregnant women in clinical trials and thereby contribute to the body of knowledge relevant for the safe and effective treatment of pregnant women. That is not the norm, but I think it is right. These ideas are fully explored in the co-edited collection *Clinical Research Involving Pregnant Women* (Baylis and Ballantyne, 2016).

*Q: How does this translate into changing governance structures and success metrics in science?*

Baylis: If you wanted to create a genome-edited baby today, you could evade laws, guidelines and declarations. You could go into international waters, get a rich person to build you a fancy boat, and do whatever you want. The only things that would stop you are other forms of governance such as peer respect, professional norms, social condemnation. If you care about being celebrated by your peers or your country, you will not violate current social and ethical norms. We need to change what counts as success. Governance is about structures and incentives, not just laws.

The current global scientific system rewards being first; it is like the Olympics, a competition (between nations and individuals) by softer means than war. That competition shapes behavior. As an alternative, I suggest exploring the merits of ‘collaborative ambition’: can we set success metrics that incentivize people to share their work and come up with creative answers that help more people? Redefining success would mean valuing collaboration, openness and societal benefit over individual achievement.

The university system, for example, is a governance system with clear metrics for success, in this case: being awarded a PhD; graduating *Summa Cum Laude*; publishing in high-impact journals; successfully competing for external funding; earning prestigious awards; becoming a full-professor; and so on. Those metrics drive behavior. Few academics challenge these metrics because they either feel threatened or have not achieved them. We need to question those metrics and create new ones that align with the values of meaningful impact and influence. The current metrics are thought to be placeholders for this, but they are not always aligned.

## The fork in the road and slow science

Baylis: I often speak about the fork in the road when I give public lectures to emphasize the need to stop and reflect when there is more than one option in front of us. At a fork in the road, you should stop and ask yourself: ‘If I go this way, what kind of world am I helping to create? If I go that way, what kind of world am I helping to create?’ For some, the answer comes easily, ‘It’s obvious; progress is forward.’ So, people rush ahead.

The ‘fork in the road’ metaphor is one of the ways in which I try to anchor the commitment to slow science. ‘Slow science’ is about quality, not speed. It is about taking the time to formulate and ask the right question(s)—the question(s) that need(s) to be asked and answered. If you get a good question at the outset, then you have a better chance of getting a good answer. Speed isn’t the metric; quality is.

To build a better world, we need to change social norms. Social norms influence expectations of success. In many scientific communities, success means being first, publishing in a prestigious journal or

winning prizes. If we want a different world, we need to change the prevailing definitions of success. We need to value collaboration, inclusivity and benefit to society. We need to ask: ‘what kind of world do we want to live in?’ Then, we need to align our scientific and social practices with that. If the incentive system says, ‘Be first at all costs’, we will get risky behavior. If the incentive system says, ‘Collaborate and share’, we will get different behavior.

### Advice for young Chinese bioethicists

*Q: Do you have any suggestions for young scholars in Chinese bioethics?*

Baylis: I think Chinese scholars have a rich tradition of thinking about how the world can and should be organized. It would be a contribution to share that traditional knowledge more widely. There is a wealth of information and concepts in your culture, in your history, that are unique and worth sharing.

As we left the campus to go to the train station, I saw something that looked like a petrified piece of wood or a rock. I wondered what it was. Lucy said it was Gongshi—a ‘scholar’s rock’ or ‘viewing stone’. She told me about the four features of the scholar’s stone and explained how this was a tool for meditation and reflection. That made me think: ‘Could we apply the features of the viewing stone to contemporary bioethics?’ One feature is ‘perforation’—holes. Holes in the stone might allow us to see through dense, complicated material. How might this perspective be useful in contemporary bioethics? For a Chinese scholar, this might be an interesting lens through which to assess ethical challenges. So, I would encourage scholars in Chinese bioethics to draw on your rich history and cultural resources. Invent your own frameworks. Don’t just consume Western ideas such as principlism. Share your discrete ideas and insights with the world. It would be phenomenally interesting. Look around at the wealth of information, concepts and ways of understanding the world and bring that to the table for discussion. Be creative, and always remember that ethics is for all of us.

### Conclusion

Françoise Baylis’s reflections leave us with a simple but urgent reminder: science and technology are never just about what can be done, but about what should be done. Tools available to us—whether CRISPR, artificial intelligence or drones—are neutral; it is our choices, our values and our governance that will shape their legacy. She insists that broad societal consensus is not a luxury, but a moral responsibility—one built through respectful dialogue, including with those with whom we profoundly disagree.

Baylis calls for a reimagining of success in science: away from the race to be first, and towards *collaborative ambition*—a system that rewards openness, solidarity and real societal benefit. She urges scientists and policymakers to share power, to listen to voices at the margins, and to have the courage to question the very metrics that define their careers.

Her challenge to all of us is disarmingly direct: *What kind of world do you want to live in?* The answer, she argues, cannot be left to chance or to a few powerful actors—it must be built together, across borders, disciplines and differences. Even if consensus is never fully reached, the effort at building consensus can change us: make us more open, more compassionate and, perhaps, more human. In the end, Baylis does not promise easy answers. What she offers instead is a compass—one that points not to the quickest path, but to the one worth walking, together.

The interview closes with a call to action: change social norms, redefine success, share power and be creative. Baylis encourages young scholars, especially in China, to draw on their own traditions and contribute new frameworks to the global conversation. Ethics, she reminds us, is for all of us, and the future of human genome editing depends on our ability to imagine and create a world that is inclusive, compassionate and just.

### ORCID iD

Lu Gao  <https://orcid.org/0000-0002-3367-2888>

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this

article: This work was supported by the Chinese Academy of Sciences and the National Social Science Fund of China (grant number E4291Z09 and 21FZXB063).

### Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Notes

1. Prior to publication, Professor Françoise Baylis was provided with a copy of the transcript. A few minor corrections and clarifications were made at this time. As well, as appropriate, notes were added to help orient the reader.
2. Q stands for a question raised by Yan and Gao.
3. See detailed information at: [www.impactethics.ca](http://www.impactethics.ca).

### References

- Baltimore D, Berg P, Botchan M, et al. (2015) A prudent path forward for genomic engineering and germline gene modification. *Science* 348(6230): 36–38.
- Baylis F (1990) The ethics of ex utero research on spare ‘non-viable’ IVF human embryos. *Bioethics* 4: 311–329.
- Baylis F (2019) *Altered Inheritance: CRISPR and the Ethics of Human Genome Editing*. Harvard: Harvard University Press.
- Baylis F (2020) To publish or not to publish. *Nature Biotechnology* 38(3): 271.
- Baylis F (2023) Human genome editing: From the first to the third international summit. Available at: <https://impactethics.ca/2023/03/22/human-genome-editing-from-the-first-to-the-third-international-summit/> (accessed 13 January 2026).
- Baylis F (2025) Submitting CRISPR for human heritable genome editing. *The CRISPR Journal* 8(4): 239–244.
- Baylis F and Ballantyne A (2016) *Clinical Research Involving Pregnant Women*. Cham: Springer.
- Dreger A and Baylis F (2018) More than words. In: Baylis F and Dreger A (eds) *Bioethics in Action*. Cambridge: Cambridge University Press, pp.1–8.
- Fraser G (2018) Here come the rogue scientists. *UnHerd*, 6 December. Available at: <https://unherd.com/2018/12/here-come-the-rogue-scientists/> (accessed 20 January 2026).

Lander ES, Baylis F, Zhang F, et al. (2019) Adopt a moratorium on heritable genome editing. *Nature* 567(7747): 165–168.

Lanphier E, Urnov F, Haecker SE, et al. (2015) Don’t edit the human germ line. *Nature* 519(7544): 410–411.

Liang P, Xu Y, Zhang X, et al. (2015) CRISPR/Cas9 mediated gene editing in human tripronuclear zygotes. *Protein & Cell* 6(5): 363–372.

National Academies (2015) On human gene editing: International summit statement, 3 December. Available at: <https://www.nationalacademies.org/news/2015/12/on-human-gene-editing-international-summit-statement> (accessed 13 January 2026).

National Academies (2018) Statement by the Organizing Committee of the Second International Summit on Human Genome Editing, 28 November. Available at: <https://www.nationalacademies.org/news/2018/11/statement-by-the-organizing-committee-of-the-second-international-summit-on-human-genome-editing> (accessed 14 January 2026).

Saey TH (2019) A Nobel Prize winner argues banning CRISPR babies won’t work. *Science News*, 2 April. Available at: <https://www.sciencenews.org/article/nobel-prize-winner-david-baltimore-crispr-babies-ban?tg=nr> (accessed 14 January 2026).

Wade N (2015) Scientists seek moratorium on edits to human genome that could be inherited. *New York Times*, 3 December. Available at: <https://www.nytimes.com/2015/12/04/science/crispr-cas9-human-genome-editing-moratorium.html> (accessed 14 January 2026).

WHO (World Health Organization Expert Advisory Committee on Developing Global Standards for Governance and Oversight of Human Genome Editing) (2021) Human genome editing: A framework for governance. Available at: <https://www.who.int/publications/i/item/9789240030060> (accessed 14 January 2026).

### Author biographies

Ping Yan holds a PhD in philosophy of science and technology. She is an associate professor and master’s supervisor at the School of Marxism, Dalian University of Technology. Her research focuses on ethics of technology,

bioethics, and responsible research and innovation (RRI).

**Lu Gao** is an associate professor at the School of Marxism, Tsinghua University. She holds a PhD

from Tsinghua's STS Institute and previously directed the Institute for the History of Natural Sciences STS Center, Chinese Academy of Sciences. Her research focuses on emerging technology governance and biotechnology history.

# Addressing the dual challenges of scientific and technological risks: Deep dilemmas and system transformation in global governance

Cultures of Science  
2026, Vol. 9(1) 21–36  
© The Author(s) 2026  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20966083261436514  
journals.sagepub.com/home/cul



Yidong Liu<sup>1</sup> 

## Abstract

Currently, risks brought on by science and technologies such as artificial intelligence (AI) are intensifying. However, the global governance system of science and technology is riddled with major security loopholes and is deeply mired in dilemmas and distractions such as cognitive misconceptions, avoiding 'heavy' topics to focus on more trivial ones, ignoring time-limit principles, limiting funding for AI safety, neglecting the role of government and lacking a 'big picture' view. Humanity faces dual and multiple challenges; the situation is extremely grim and urgent. Principles such as 'cooperation is more important than competition', 'safety is more important than wealth', 'direction is more important than speed' and 'steady progress is more important than coming first' should be established as soon as possible. Therefore, it is imperative to recognize the truth, build confidence, transform to survive, launch a 'New Distribution Revolution' and quickly realize the transition from a 'development-first' system to a 'security-first' system and vigorously develop a security economy. To this end, the global governance system requires a comprehensive upgrade; only a global governance system in which the government plays a full role can complete such an arduous task in time.

## Keywords

Technological risk, ruin-causing knowledge, AI backfire, dual challenges, global governance, distribution revolution, system transformation, absolute security game, security economy

Received: 4 February 2026; accepted: 11 February 2026

## 1. Introduction

Development and security are two major themes of today's world. There is already a rich knowledge system, cultural traditions, organizational management, institutional arrangements, incentive mechanisms and practical experience for considering problems and taking action from the perspective

<sup>1</sup>Institute for the History of Natural Sciences, Chinese Academy of Sciences, China

### Corresponding author:

Yidong Liu, Institute for the History of Natural Sciences, Chinese Academy of Sciences, 55 Zhongguancun East Road, Haidian District, Beijing 100190, China.  
Email: liuyd@ihns.ac.cn



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

of development. However, the framework for considering problems and putting them into practice from the perspective of security remains immature, and global governance is attempting to make up for this shortcoming.

It is well known that the essence of global governance is based on global governance mechanisms rather than formal governmental authority; the methods of global governance are participation, negotiation and coordination. These characteristics and principles have been emphasized and practised in global governance theory and practice, achieving remarkable success. However, today, with the rapid development of cutting-edge technologies such as artificial intelligence (AI), biotechnology and quantum technology, the global governance system and human security are facing unprecedented challenges. In particular, AI risks are intensifying—artificial general intelligence (AGI) could be achieved within a few years, followed immediately by the realization of artificial super-intelligence (ASI), making a loss of control over AI difficult to avoid. Therefore, within a short period (one or two years), we must ensure an AI transition and construct a new type of sustainable, safe AI. However, current global governance mechanisms are simply incompetent for this task. To this end, this article discusses three aspects: first, the ‘dual challenges’ of scientific and technological (SciTech) risk, to demonstrate the severity and urgency of SciTech risk; second, an analysis of the deep dilemmas and defects of global governance; and third, a discussion on the ‘New Distribution Revolution’, the system transformation of global governance and policy suggestions.

## 2. The ‘dual challenges’ theory: SciTech risks intensify, while security governance mechanisms have many loopholes

The primary focus in SciTech risk governance research is exploring the severity and urgency of SciTech risks, which determines the importance and priority of SciTech risk governance in academic and practical fields. The author has long studied major SciTech risks and, from a security perspective,

revealed and summarized that there are many loopholes in human security prevention and control mechanisms. Based on this, the ‘dual challenges’ theory is proposed: on the one hand, SciTech risks are intensifying; on the other hand, human risk-prevention mechanisms and the global governance system have **10 major security loopholes**. This reveals the severity and urgency of SciTech risks, SciTech crises and human security crises that humanity faces. The ‘dual challenges’ theory attempts to provide a comprehensive and profound revelation and judgement regarding SciTech security, SciTech risks, SciTech backfire and SciTech crises (Liu, 2000, 2002, 2020). The revelation and analysis of the 10 major security loopholes apply to the AI field, and, currently, AI risk is the concentrated expression of technological risk.

***Loophole 1: The selection mechanism for knowledge production and growth has failed; knowledge cannot be produced selectively, and innovation cannot be conducted selectively.***

By proposing the concept of ‘ruin-causing knowledge’ (致毁知识) and proposing and solving a set of problems (four premises and one set of questions), the following conclusions are drawn. So-called **ruin-causing knowledge** refers to core knowledge, such as core principles and core technologies, that can be used to manufacture destructive weapons or other products/schemes capable of causing devastating disasters. Examples include nuclear fission knowledge, chain-reaction knowledge, DNA-recombination technology and gene-editing technology. Ruin-causing knowledge is not distinguished by whether the knowledge is ‘good’ or ‘bad’ but instead is defined by whether its application (military use, malicious use, abuse) produces enormous destructive power (Liu, 2000). Major SciTech risks are risks that harm people’s lives and health and threaten human survival and safety. They involve many factors of science, technology and society, are extremely difficult to study, and require in-depth, systematic research. To this end, the core issues proposed by the author include **four premises and a set of questions**.

The **four premises** are: (1) the positive and negative effects of cutting-edge technology cannot offset each other; (2) the growth of SciTech knowledge is

irreversible; (3) knowledge and application have chain effects; and (4) scientific research is self-reinforcing and never-ending. The **set of questions** includes: While SciTech knowledge grows, can we—and how do we—prevent the growth and diffusion of one type of extremely destructive knowledge—ruin-causing knowledge? How can we produce knowledge selectively? How can we innovate selectively? How to stop high-risk innovation that brings huge profits? What are the conditional relationships for realizing ‘tech for good’? How to achieve scientific transformation, technological transformation and industrial transformation as soon as possible? Under the four premises, the importance of studying this set of questions is highlighted. For example, the positive and negative effects of cutting-edge technology cannot be offset. In other words, no matter how great the positive effects of cutting-edge technologies (such as AI, nuclear power plants and nuclear medicine), they cannot offset their negative effects (such as AI disasters, nuclear weapon disasters and nuclear accidents). It is a case of ‘one bad ruins 100 goods’; one can enhance the benefits but cannot avoid the shortcomings. Therefore, whether to accept and develop a cutting-edge technology is not determined by its positive effects but by whether its negative effects can be resolved.

The author’s research shows that, under current various conditions, there are at least **26 reasons** why the growth and diffusion of ruin-causing knowledge cannot be stopped, knowledge cannot be produced selectively, and innovation cannot be conducted selectively. They include an inability to prevent the emergence of ruin-causing knowledge, an inability to stop breakthrough progress in science and technology, and an inability to stop the growth and application of ruin-causing knowledge. Specifically, the 26 reasons are as follows (Liu, 2002):

1. Technological innovation based on the growth of SciTech knowledge is the lifeline of enterprises and the market economy.
2. Concepts such as ‘science supreme’, ‘technology omnipotent’, ‘obsession with innovation’, ‘social control omnipotent’, ‘capital pursuit of profit’ and ‘social development
- determinism’ support and indulge the growth of SciTech knowledge ideologically. Even if significant SciTech risks are recognized, development does not slow down; instead, ethics and safety supervision are strengthened under the premise of innovation. This treats the symptoms but not the root cause and is of no avail.
3. It is impossible to comprehensively weigh the pros and cons of the growth of SciTech knowledge, and it is impossible to achieve a state in which growth continues only when pros outweigh cons and slows or stops otherwise.
4. Scientific research activities follow the principle of institutionalized immediate interests (such as priority, patent rights and research funding). Rules-based governance and global governance struggle to constrain scientific research activities.
5. The inheritability, interconnectivity and integrity of knowledge make it difficult to separate ruin-causing knowledge from its parent SciTech knowledge; ruin-causing knowledge will grow as SciTech knowledge grows.
6. The growth of ruin-causing knowledge is irreversible. People can destroy weapons of mass destruction but cannot destroy the knowledge of how to make them.
7. The existence of externalities and the lack of unified world laws mean that one does not need to worry about the growth of ruin-causing knowledge when developing science and technology; otherwise, one will lose in competition or even be eliminated.
8. ‘Balanced thinking’ reigns supreme: wanting both SciTech innovation and SciTech security and arguing that we cannot stop eating for fear of choking and cannot hinder innovation by strengthening ethical governance. In other words, innovation comes first. Under current concepts, culture and institutional mechanisms, the benefits that scientists gain from making new discoveries/breakthroughs far outweigh the costs they pay.

9. Factors such as unpredictability, the priority of immediate interests, competitive pressure, externalities, lack of unified laws and extreme asymmetry of rewards and punishments make the emergence of scientific research results impossible to prohibit. Chain relationships and chain effects mean that any possible application of scientific research results will be attempted.
10. There is a broad Collingridge's dilemma and 'forbidden zone paradox': when there is ample time to prohibit, one often cannot identify and determine whether SciTech needs to be prohibited; meanwhile, once the preconditions appear and it can be determined, it can no longer be prohibited. Expanding the scope of the forbidden zone could stop the production of ruin-causing knowledge, but it is not feasible to do this; conversely, a narrow forbidden zone cannot stop its production as, once preconditions are formed, it is difficult to prohibit, and preconditions can come from unrelated, non-forbidden fields.
11. AI conducts autonomous research, producing ruin-causing knowledge; extremists and tech maniacs use AI to produce ruin-causing knowledge; AI is used in the military field, producing ruin-causing knowledge.
12. The characteristics of division of labour and autonomy in knowledge production result in fragmentation, which is unfavourable for screening and prohibiting ruin-causing knowledge.
13. Prohibiting ruin-causing knowledge in the development of multiple technologies with multi-path substitution characteristics is as difficult as cutting off the international internet.
14. The continuous emergence of new research modes, methods and theories (such as AI for science, big data science, computer simulations, cloud computing and crowdsourcing) makes it even more difficult to stop the emergence and growth of ruin-causing knowledge.
15. Prohibiting research on specific problems does not affect the improvement of SciTech capabilities, and the improvement of SciTech capabilities makes solving prohibited problems easy, leading to the failure of prohibition.
16. SciTech can manufacture more powerful SciTech until it becomes uncontrollable. The improvement of SciTech capabilities increases the amount of ruin-causing knowledge.
17. The holistic combinatorial effect of knowledge shows that ruin-causing knowledge can be generated by combining non-ruin-causing knowledge.
18. It is difficult to establish a globally unified, binding and enforceable legal system and assessment/control measures. The trends of corporate, market-oriented and networked scientific research are intensifying, resulting in science without borders and the inability to effectively set forbidden zones.
19. Giving up research due to fears of possible harm is considered unwise because others without such limitations will continue the research and win priority. Society also cannot benefit from your senseless sacrifice because discovering something once is the same as discovering it 100 times.
20. Ruin-causing knowledge has significant power; therefore, states, military groups and terrorist organizations will not only not prohibit it but will also race to develop it first.
21. Currently, science lacks necessary self-correction and self-protection mechanisms, so it cannot stop the growth and application of ruin-causing knowledge.
22. There exist the 'target misalignment principle' and the 'moving train dilemma' (detailed below); society has lost the ability to correct major errors.
23. There is a lack of ability to learn from experience and lessons; the model of developing nuclear weapons (basic research → military application) remains unchanged

- to this day. Genetic weapons, nano-weapons and AI weapons are also products of this model.
24. Knowing the law but breaking it, and superficial compliance (feigning compliance). Even if consensus is reached, relevant agreements might not be signed or executed. The behaviour regarding biological weapons conventions and numerous environmental issues is proof.
  25. The market economy, scientific research activities, and even the entire society's institutionalized 'immediate interests first' and 'law of the jungle' principles, structures, incentive mechanisms and winning mechanisms are the fundamental reasons why the emergence, growth, spread and application of ruin-causing knowledge cannot be stopped. Moreover, not only can we not stop it but, also, we highly rely on and accelerate the growth of technological knowledge, including ruin-causing knowledge. The irreversible growth of ruin-causing knowledge highlights the fundamental defects of a society that prioritizes immediate interests and the law of the jungle and also provides an opportunity to end this type of society and initiate social transformation.
  26. The fact that positive and negative effects of SciTech cannot be offset, and that offence/defence and growth/control are asymmetrical and cannot be offset, indicates that relying on improving the positive effects of SciTech or promoting the growth of defence and control knowledge cannot stop the devastating disasters caused by the growth and application of ruin-causing knowledge.

Actually, not all 26 reasons need to hold true. As long as a few of them hold, it is sufficient to support the conclusion that the growth and diffusion of ruin-causing knowledge are irreversible and unavoidable. Under the current world's mainstream SciTech and economic development model, the growth and diffusion of ruin-causing knowledge

are unstoppable. This means that the danger of human destruction is constantly accumulating and increasing. Once it reaches a certain degree, a devastating disaster is inevitable. Moreover, this irreversible cumulative growth of danger makes the probability of a devastating disaster greater and greater. If not stopped in time—if we do not change course—it will inevitably lead to an outbreak. This is the greatest SciTech crisis and the greatest crisis and challenge humanity faces.

Currently, terrorism is prevalent, and knowledge production in corporate labs and by makers is harder to control. Especially with the recent rise of **AI for science**, individuals and small teams (such as mad scientists, hackers and geeks), or even a single-person arms company or single-person terrorist organization, can master and develop cutting-edge technology and produce ruin-causing knowledge. The internet makes it easy for ruin-causing knowledge to spread, rendering the human situation increasingly dangerous. In recent years, with the AI explosion, SciTech risks are intensifying, and SciTech backfire is approaching day by day. In short, the biggest crisis and challenge humanity faces is not that 'good things' (such as natural resources, etc.) are running out but instead that 'bad things' (such as 'three wastes', ruin-causing knowledge, etc.) are accumulating more and more, approaching the Earth's limit to contain waste and burden, potentially leading to collapse (Liu, 2017b). Thus, the selection mechanism for knowledge production and growth has failed, and the human security defence line has serious loopholes.

***Loophole 2: SciTech ethics failure, SciTech law failure and security supervision failure.***

SciTech ethics and SciTech laws have serious loopholes, which can be described as SciTech ethics failure and SciTech law failure. That is, SciTech ethics and laws cannot constrain all laboratories and SciTech experts in the world. Using ethical laws to constrain misconduct is effective in social life in which violations by a few cause limited harm. However, in the SciTech field, the effect is minimal because there are 233 countries and regions in the world with different SciTech ethics and laws that also have numerous gaps and loopholes, and the

high seas and deserted islands are also harder to monitor.

There are four situations in which SciTech ethics and laws cannot impose constraints: (1) in basic research or academic research fields; (2) among mad scientists, hackers, geeks or terrorists; (3) in defence and military fields; and (4) in corporate R&D institutions (while civilian enterprises typically will not develop technologies explicitly harming humans, they also will not self-restrict if the technology is easily converted for malicious use). The fact that controversial gene-editing technology won the 2020 Nobel Prize in Chemistry is proof. For scientific discovery and invention, doing it once is the same as doing it 100 times. In the internet age, in which knowledge spreads easily, strengthening SciTech ethics, laws and security supervision is important, but we must also be clear about their inherent limitations and loopholes. AI ethics, etc., are failing. The European Union (EU) AI Act, officially released in 2024, is full of loopholes: it explicitly states that scientific research and defence/military institutions are not within the scope of constraint, and AGI is also not listed in the high-risk category. Recently, although SciTech ethics research has become a hotspot, regrettably, many researchers ignore the serious loopholes caused by the failure of SciTech ethics. The author's research on SciTech ethics starts from the failure of SciTech ethics (Liu, 2022a).

Currently, SciTech ethics research is growing. Based on whether it maintains the mainstream SciTech development model, SciTech ethics can be divided into two types: **maintenance/apologetic SciTech ethics** (SciTech ethics I) and **transformative SciTech ethics** (SciTech ethics II). The former defaults to the position that the current (Western or global) mainstream SciTech development model is sustainable and, based on this premise, uses SciTech ethics to make up for, repair and solve ethical problems arising from SciTech development to maintain its continued development. The latter conducts deep reflection, revealing that the root of SciTech ethical risks lies in the internal defects of the mainstream SciTech development model and arguing that treating the root cause is necessary and that reforming and transforming the SciTech

development model is the fundamental solution and an urgent task.

***Loophole 3: Serious loopholes in improving the responsibility, self-discipline and moral levels of SciTech experts.***

The mainstream view for a long time has been that SciTech is a double-edged sword; SciTech itself is a tool and neutral, so improving the sense of responsibility, self-discipline, autonomy and moral level of SciTech experts and users is the key to 'SciTech for good' and preventing risks. **Responsible research and innovation (RRI)** is one such related concept. However, these measures are difficult to make effective for three reasons: (1) Even if responsibility and moral levels are generally improved, as long as a few SciTech experts, makers or mad scientists are irresponsible, they can cause catastrophe (SciTech ethics/law failure). (2) Under the wrong SciTech development model (believing in science without forbidden zones, binding tech with capital, prioritizing immediate interests), even completely following ethical norms for responsible research cannot prevent major SciTech risks. For example, Otto Hahn, who discovered nuclear fission, conducted responsible research; otherwise, how could he get the Nobel Prize? (3) The understanding of 'responsibility' varies from different angles. The Manhattan Project was considered responsible innovation at the time, but, from a human perspective, developing nuclear weapons was the most irresponsible innovation. Therefore, 'responsibility' and 'morality' are relative, and RRI cannot be used to prevent major SciTech risks.

***Loophole 4: Ensuring mutual destruction cannot ensure one's own safety; the balance strategy has serious loopholes.***

History and reality show that cutting-edge technology is often used for cutting-edge weapons development. The scientific community and society are accustomed to this, and scientists let it be, perhaps believing that, as long as **mutually assured destruction** is ensured, their own safety can be expected. This belief comes from the so-called concept of 'nuclear balance', which can be extended to genetic weapons balance, AI weapons balance,

nano-weapons balance, etc. However, the author believes that this is a complete misunderstanding. First, the fact that no nuclear war or disaster has occurred is not because nuclear risks were effectively controlled but instead due to luck. Many nuclear crises and accidents occurred after World War II, and nuclear war was avoided only by a fluke. Second, cutting-edge weapons such as AI, bio- and nano-weapons do not rely on scarce raw materials, as nuclear weapons do; instead, they have a lower threshold for use, are easy to diffuse and can be possessed by terrorist organizations and extremists. Third, makers, hackers, geeks, corporate labs, mad scientists and even tech terrorists can use global public R&D platforms for cutting-edge tech and weapons development. Leaks and thefts from cutting-edge labs (such as bio labs) happen constantly, and R&D activities cannot be effectively controlled.

In short, ‘the capability of mutual destruction’ cannot be completely controlled by national governments. Asymmetrical warfare and terrorist activities can happen at any time. Therefore, ensuring mutual destruction does not ensure one’s own safety. The reason for developing cutting-edge weapons does not stand; ultimately, it harms others and oneself.

***Loophole 5: It is difficult to avoid shortcomings while maximizing benefits.***

The mechanism for maximizing pros and minimizing cons has serious loopholes. ‘Maximizing strengths and avoiding weaknesses’ (扬长避短) is one of humanity’s most important wisdoms. The author’s research shows that SciTech development can maximize strengths, but it is difficult to avoid weaknesses. There are five reasons for this: (1) Different perspectives lead to different understandings of negative effects. (2) There are different time frames being considered—some results show positive effects in the short term but expose negative effects in the long term (e.g., using dichlorodiphenyltrichloroethane as a pesticide). (3) Considering costs and competition, companies seek quick success, causing slow harm to consumers and easily evading responsibility (e.g., using food additives or using glyphosate on crops). (4) The existence of indivisibility and ‘non-offsetability’—positive

and negative effects of cutting-edge tech cannot be offset. No matter how great the positive effect, any existing ‘shortcoming’ cannot be avoided. For example, nuclear power/medicine cannot offset nuclear war/accidents. The result of this is ‘one bad ruins 100 goods’. Especially when considering AI, doing 10,000 good things cannot offset doing one extinction-level bad thing. (5) Being involuntary in the face of technology—the use of technology is sometimes not decided by the user. Under interest and competitive pressure, technology induces and forces people to use it. Even if it produces negative effects or is unfavourable to oneself, one has to develop and use it (e.g., nuclear, biological and AI weapons).

***Loophole 6: Broad Collingridge’s dilemma: Leak prevention mechanism failure.***

Preventing and patching human security loopholes is crucial but not easy. Related to this is **Collingridge’s dilemma**. Based on the short-bridge principle in cognitive science, the author proposes the ‘scientific forbidden zone dilemma’, enriching Collingridge’s dilemma from time and space aspects, and further proposes the ‘leak prevention mechanism dilemma’, with the following principles: (1) Preventing and repairing loopholes requires cognitive awareness, which is very difficult. (2) Loopholes appear in various ways, including gradual formation and sudden emergence, with the latter being hard to guard against. (3) Timely repair is not easy. Repairing requires changing not only the loophole itself but also its cause, and, once a loophole appears, there may be positive feedback (e.g., ‘A stitch in time saves nine’, ‘A thousand-mile dyke collapses from an ant hole’). Current human security defence is also trapped in this dilemma.

***Loophole 7: Serious loopholes in the ability to screen risks and safety hazards.***

Preventing risks requires timely screening of risks and hazards, which relies heavily on ability. Those with less ability cannot identify risks, hazards, fakes and traps created intentionally or unintentionally by those with greater ability (just as second-rate strategists cannot see through traps set by first-rate strategists). Manufacturing risk is

easy, but identifying it is hard, and preventing it is even harder. In many cases, it is easy to create risk, but identifying and preventing it requires consensus. There is extreme asymmetry between risk manufacturing and screening/prevention.

***Loophole 8: The ‘prisoner, train, sword, demon’ four major dilemmas: Error-correction mechanism failure.***

The error-correction mechanism is one of the most important mechanisms for human survival and development. Combining the famous prisoner’s dilemma with the **moving train dilemma (Dongche)**, **double-edged sword dilemma** and **magic ring (demon) dilemma** proposed by the author reveals the fundamental dilemmas and human security loopholes related to whether humans commit major errors and whether they can recognize, correct or offset them. Altogether, this is abbreviated as the ‘prisoner, train, sword, demon’ four major dilemmas (Liu, 2016).

The prisoner’s dilemma shows that suffering a setback does not necessarily make one wiser; major errors are hard to avoid. The moving train dilemma shows that errors are produced during human activities. When errors are found, we often cannot ‘pause first, then correct’ but instead must ‘continue while arguing and correcting’. Small errors can be corrected by gaining correct understanding, but, for major errors, correcting the understanding alone is far from enough.

Usually, four conditions are needed to correct major errors: (1) a correct understanding and consensus; (2) an expectation of win–win action in terms of interests; (3) the ability to take effective joint action; and (4) having other relevant conditions being present simultaneously. Therefore, correcting big mistakes is not easy. Without fully meeting these four conditions, discovering errors alone cannot correct them, and consensus is also not enough. The conditions for correction are harsh, and, under huge inertia, correction is even harder. The moving train dilemma constructs a framework for analysing systemic obstacles to correction in dynamic processes, whereas traditional theory may focus on static analysis or post-event summary. It accurately depicts the real scenario of correction in the real

world—unable to press the pause button, we must ‘repair the engine in flight’.

So far, regarding correcting major errors, environmental issues have passed the first hurdle (consensus) but not the second (win–win expectation)—the United States not signing the Kyoto Protocol is proof. For technological risks, even the first hurdle (consensus) has not been passed. Optimists and pessimists argue, and, currently, optimists/cautious optimists remain mainstream. Furthermore, the moving train dilemma explains the widespread ‘collective inertia’ and ‘collective apathy’ towards risks—shortsightedness, conformity, habit, laziness, wishful thinking and path dependence lead to difficulty in collective action to correct errors. This ‘collective inertia’ (muddling along) and ‘collective apathy’ (letting things slide) are worse than individual inertia/apathy. Blind capital is rampant in the market. Cutting-edge biotech has always had risk controversies (consider, for example, the 1975 Asilomar Conference and the 2015 Washington Summit). Gao Lu compared these two events and found that, despite 40 years of rapid biotech progress, human society’s ability to adapt to it has not improved much, essentially simply marking time (Gao, 2018).

The double-edged sword dilemma shows that positive and negative effects (especially of cutting-edge SciTech) cannot be offset or compensated. It is exemplary of ‘one bad ruins 100 goods’. Ten thousand good deeds by SciTech may not offset one extinction-level bad deed, especially when considering AI.

Finally, the magic ring (demon) dilemma shows that the threshold for committing major errors is getting lower. Humans have reason, so they cannot withstand temptation. Cutting-edge technology allows small figures and robots to commit major errors, and AI even more so.

The four major dilemmas show that humanity constantly commits big mistakes, which are hard to correct, offset or compensate for. The threshold is lowering by allowing small figures/robots to commit them. These are the more severe dilemmas that humanity faces. The mistakes humanity makes in tech development and application are some of the

most serious errors, concerning the survival of human civilization. The failure of error-correction mechanisms is the most serious loophole in the human security defence line.

***Loophole 9: Profiteering innovation loophole: It is difficult to stop high-risk innovation that seeks excessive profits.***

‘Profiteering innovation’ refers to high-risk innovation that seeks huge profits. Stopping it (including but not limited to the application of ruin-causing knowledge) is crucial but difficult. The three elements of profiteering innovation include the following: (1) It generates huge returns (profiteering), including in economic, fame, military, political and media fields. (2) It has high/huge risks. (3) Due to high returns, innovators and investors disregard high risks, even taking desperate risks, often using the excuse ‘We cannot hinder innovation due to ethical governance’ (do not stop eating for fear of choking) to force implementation. For example, developing atomic bombs, gene editing and AI are all profiteering innovations with huge threats to human safety (Liu, 2024). Although continuing AI development will inevitably lead to loss of control, many still run wild in the name of developing safe/benevolent ASI. In reality, as long as it is ASI, it will not be safe or benevolent. There are many types of ASI; 100 angel ASIs cannot offset one demon ASI.

Marx pointed out in *Das Kapital*: With 50% profit, capital takes desperate risks; with 300% profit, it dares to commit any crime, even risking the gallows. When there is huge profit, innovators and investors disregard even their own safety, so how could they care about humanity’s safety? Reasons for profiteering innovation include the profiteering effect, the coercion effect and the blind following effect, along with empirical thinking, positive-effect thinking, professional thinking, wishful thinking, emotional thinking, conformity thinking and the ‘Adam Smith trap’. Blinded by greed, people selectively believe easily; the instinct to seek advantage and avoid harm gives way to seeking advantage and ignoring harm, or even only seeking advantage and forgetting harm.

The ‘profiteering effect’ refers to taking desperate risks driven by huge profits or profit

expectations. As Georg Franck pointed out: Attention is the scientist’s main motivation. Indeed, many successful people outside science are the same. Scientism, tech-worship, innovation supremacy, effective accelerationism and the ‘Adam Smith trap’ fuel the flames, providing excuses for indulging personal desires and ambitions. Adam Smith’s ‘invisible hand’ and ‘division of labour’ concepts mislead the world because boundary conditions were not clarified. Further, Smith reasoned based on simple cases like bakers and pin-making, which do not apply to the complex situation of the contemporary knowledge economy. In many cases, the free market fails. The conditions to realize ‘subjectively for oneself, objectively for others’ are extremely harsh and rarely met in reality. The result of individual and capital orientation is the intensification of AI and other technological risks.

***Loophole 10: The Western SciTech development model has inherent and worsening serious defects and transformation difficulties.***

The Western civilization system has diverse elements, placing individual autonomy and freedom at the centre of its value system. This leads to inherent defects in Western tech development and market economy operations. It cleverly combines immediate competition (priority, patents) with the ‘immediate interests first’ of the market economy system, resulting in great vitality but shortsightedness. From a development perspective, Western tech innovation is fruitful but, from a security perspective, it has huge defects. Western science has long believed in ‘science without forbidden zones’ and ‘knowledge without forbidden zones’, with extremely asymmetrical rewards and punishments in management. The combination of SciTech with capital, and scientism with individualism, forms a prevalent ‘scientific individualism’ that prioritizes immediate interests. As the model of the tech industry, Silicon Valley believes in ‘move fast and break things’ (do it first; ask for forgiveness later). Ethics is not a consideration for top Silicon Valley tech experts; they consistently view ethics as a stumbling block to technical innovation and progress (Cao, 2018). The result of technology as an investment object is that technology operates according to the

logic of capital. Many irrationalities, gambling and collective madness appear, deeply trapped in the ‘all-in, go to zero’ AI gamble (Liu, 2022b). The AI explosion and excessive investment in AI are proof. The risks of AI share the characteristics of both ‘black swan’ and ‘grey rhino’ events, and may even escalate into ‘mad rhinos’ that are stampeding recklessly out of control.

The Western tech and social development model requires extensive development, extensive innovation and extensive competition. It only considers (or mainly considers) economic returns and competitive advantages, **ignoring or not prioritizing technological risks, innovation risks and negative effects of innovation** (risks considered are financial investment risks), as well as ignoring negative externalities. Under this model, technology develops rapidly amidst controversy at the cost of sacrificing safety; the model mostly generates and aggravates rather than prevents/resolves major tech risks, and it weakens the role of preventive measures such as tech ethics/laws. Transforming it into a sustainable innovation system based on safety is by no means easy. Strengthening research on AI and other SciTech risk governance requires no delay. Original work by Chinese scholars, such as Liu Dachun’s *theory of evaluation of SciTech* (Liu, 2017a), Liu Xiaoting’s *new civilization conflict theory* (Liu, 2025) and Liu Huajie’s *new natural history* (Liu, 2022), deserves due attention for their visions in this aspect.

As mentioned above, SciTech risks are intensifying, while human risk-prevention mechanisms and global governance systems have 10 major security loopholes. This is one primary **dual challenge** humanity faces. Coupled with the AI explosion, tech cold war, arms race, climate change, etc., however, humanity faces **multiple challenges**, and the situation is extremely grim and urgent.

### 3. Deep dilemmas and defect analysis of global governance

Academia has discussed the dilemmas and defects faced by global governance. For example, in 2024, a study jointly published by scholars from Oxford and Yale pointed out that, because AI is seen as a

key asset determining future national strength, countries (especially superpowers) are trapped in a zero-sum game, making them unwilling to sign binding international treaties that might limit their development. Existing international organizations, such as the United Nations (UN) and the World Trade Organization, appear rigid and slow to react when dealing with rapid changes brought by AI. Furthermore, although governance initiatives are numerous, they lack substantive coordination, leading to a serious regulatory vacuum. The study criticizes the centralized idea of establishing a ‘global AI regulatory agency’ (like the International Atomic Energy Agency for nuclear technology regulation) as neither feasible nor politically legitimate. It concludes that, compared to creating new institutions, a more realistic path is to strengthen the existing ‘regime complex’; that is, strengthening coordination and capacity-building among existing dispersed institutions (Roberts et al., 2024).

In 2025, a review report released by the Geneva Graduate Institute argued that the current difficulty stems from ‘regime complexity’. There are too many overlapping but uncoordinated governance initiatives (G7, OECD, EU acts, etc.), allowing companies to ‘forum shop’ for the most lenient regulator. Additionally, international institutions such as the UN lack specialized talent to understand frontier technologies, leading to over-reliance on private-sector experts. Governance agendas are often ‘captured’ by companies, causing public interest to yield to commercial innovation speed. Existing governance frameworks are mostly Western-dominated and fail to fully reflect the demands of developing countries. The report concludes that effective AI governance relies on a distributed network, merging ethical commitments with operational flexibility. The UN and International Geneva need to advance on dual tracks of norm-setting and institutional capacity-building to cope with rapid global challenges (Singh et al., 2025).

The above analyses make sense but lack depth, and thus their conclusions along with countermeasures are hard to make effective. In light of the severity and urgency of tech risks revealed by the ‘dual challenges’ theory, the author believes global governance has **six main defects and dilemmas**.

### 3.1 Cognitive misconceptions

Blind faith in ‘SciTech supreme’ and ‘innovation supreme’ is deeply trapped in the trap of technology worship (fetishism). The new challenge for global governance is the AI explosion—AGI and ASI are imminent, and AI is about to lose control. Despite this, global AI development is still driven by improving intelligence levels and aiming for AGI. Many strange theories are circulating openly in the name of tech innovation, such as ‘humans are the bootloader for silicon-based life’, ‘human-machine fusion creates a new species’, or ‘developing ASI is worth it even if the probability of human extinction is 20%’ (equivalent to boarding a plane with a 20% chance of crashing). Such words and deeds should be investigated and prosecuted as **crimes against humanity**. We must break the curse of scientism and technology worship, carry out security enlightenment, return to using common sense to think about AI development, and strictly prevent AI backfire. Regarding AI safety, we must get rid of ‘balance thinking’ (balancing development and safety, innovation and ethics). AI doing 10,000 good deeds cannot offset one extinction-level bad deed. We must stick to ‘bottom-line thinking’—if AI safety is lost, everything is lost. Any words or deeds threatening human safety must be vetoed.

### 3.2 Avoiding the heavy and focusing on the trivial/specious arguments

Preventing AI risk is fundamental, yet people entangle themselves in the question of whether AI exceeds humans. Actually, even if AI does not exceed humans, it can still cause huge disasters (e.g., terrorists using AI to develop viruses). It is widely believed that AI will empower humans and work for humans. This is only true when AI is a tool. Once AI becomes an agent, it is different. If AI is smarter than humans, why would it empower humans or work for humans? Even if the AI tech community knows ASI will lose control, they insist on developing ‘safe/controllable ASI’ or ‘benevolent ASI’. These are specious arguments. Once AI reaches ASI level, it cannot be safe. ASI used for arms races or abused by terrorists

will result in safety protocols and ethical guardrails being uninstalled at any time.

### 3.3 Ignoring the time-limit principle

Risk theory, ethical theory and global governance research often lack ‘project duration’ considerations. Risk prevention has a time limit; braking requires lead time and must be completed within a certain time; otherwise, it is too late. AI and other tech risks are intensifying. Only by stopping them *before* they cause a devastating disaster can we save humanity. Emphasizing the time-limit principle is crucial (Liu, 2013). Ignoring the completion of risk-governance tasks within a time limit highlights the inefficiency and weakness of the current global governance system.

### 3.4 Extreme scarcity of funds for AI safety

In 2023, multiple top AI scholars (including three Turing Award winners) jointly appealed that at least one-third of the AI R&D budget should be used to ensure AI safety and ethical use. Cyrus Hodes, co-founder of Infinity Ventures and Stability AI, believes that, while global AI development investment is huge (it was expected to exceed \$360 billion by the end of 2025), investment specifically for AI safety is only \$110–130 million—a huge gap in resilient funding (Wei, 2025). Safety is the premise of development; safety and development are equally important, and both require resource allocation to achieve results.

### 3.5 Seriously ignoring the role of government

Global governance emphasizes being based on governance mechanisms rather than formal government authority, emphasizing multi-party participation, negotiation and coordination rather than top-down management. However, rules-based governance has very limited effectiveness. Practice proves that it cannot cope with intensifying risks such as AI. Global governance cannot be dogmatic; it should attach importance to the strong role of government.

In fact, James Rosenau, the founder of Western global governance theory, defined governance as a management mechanism in a sphere of activity, which includes government mechanisms as well as informal and non-governmental mechanisms (Rosenau and Czempiel, 1993).

### 3.6 Global governance lacks a ‘big picture’ view

Current global governance is often *ad hoc*, treating the head when the head aches and the foot when the foot aches, and lacking a correct big picture and vision. The biggest and most urgent challenge currently is the AI explosion, which is running wild amidst controversy and is soon to achieve AGI/ASI and result in a loss of control. Countries should unite to cope with AI risk challenges and accelerate the building of a **community of human security**. However, there are trade wars, territorial wars, SciTech wars, talent wars and the law of the jungle. Regarding AI prospects, optimists and pessimists both contend that AI will continuously improve intelligence until AGI/ASI, and carbon-based life will be replaced by silicon-based life. The author believes that this is not the inevitable development of human society but rather the destination of Western technological civilization—the outcome of Western individualism, liberalism, technology worship and capitalism. According to the UN, civilization diversity provides diverse solutions for challenges. Traditional resources of Chinese culture—collectivism, steady progress (行稳致远), harmony (和为贵) and ‘the world for all’ (天下为公)—can reverse the AI development direction and inform the construction of a new AI development model that benefits humanity based on safety. Governance corresponds to development. For a long time, humanity prioritized development. The core mission of global governance is to lead and promote the transition of human society from a ‘development-first system’ to a ‘**safety-first system**’. This requires a series of changes in concepts, culture, organization, management, institutions, technology, incentives and evaluation feedback. This is a great change unseen in a millennium.

## 4. Global governance system transformation and policy suggestions

Currently, SciTech risks are the most severe and urgent among the various risks that exist. Global governance faces **dual challenges**: SciTech risks are intensifying, while human risk-prevention governance systems have 10 major security loopholes. This is especially true in the AI field. Addressing AI risk challenges is the top priority and an urgent task of global governance. To this end, the author proposes suggestions to complete three tasks as soon as possible to reflect the transformation of the global governance system: (1) recognize the truth and reveal the severity and urgency of AI risks; (2) build confidence—silicon-based life replacing carbon-based life is not an inevitability of social development but merely the destination of Western tech civilization, so there is a possibility to change course and turn danger into safety; and (3) transform to survive—complete the system transformation of global governance as soon as possible, focusing on the **New Distribution Revolution** and strengthening government regulation. Specifically, these tasks can be achieved through the following actions.

### 4.1 Know the truth: Current AI development will lead to the end of human civilization

For a long time, the SciTech community has drawn no distinction between technological development and technological progress, focusing merely on advancements and breakthroughs in technology, while overlooking that the true goal of technological progress should be to benefit society. The explosive growth of cutting-edge technologies such as AI is rapidly accumulating systemic risks, creating an extremely grave and urgent situation. The current global AI development model is almost entirely driven by improving intelligence levels and aiming for AGI. Once AGI is achieved, ASI will follow immediately. AGI and ASI will not only be smarter than humans but will also add new dimensions that humans cannot understand and thus cannot control.

**Table 1.** The Absolute Safety Game.

	Cooperate (C)	Defect (D)
Cooperate (C)	(5, 5)(5, 5)	(-10, -10)(-10, -10)
Defect (D)	(-10, -10)(-10, -10)	(-10, -10)(-10, -10)

If AI continues to develop, loss of control is inevitable. Mass unemployment, mass fraud, mass arms races, AI for science being used by terrorists, a rapid increase of ruin-causing knowledge and ASI leading to silicon-based life replacing carbon-based life are unavoidable. Despite continuous ethical and legal releases and the signatures of top scientists calling for red lines (e.g., the EU AI Act of 2024; the October 2025 appeal by more than 3000 top scholars, including Geoffrey Hinton, to ‘pause ASI R&D’), it is of no avail. Many top AI scientists and entrepreneurs turn a deaf ear, brakes continue to fail, and AI runs wild with the support of competition and capital. When AI is thousands of times smarter than humans, it will not empower humans or work for humans (just as humans do not empower or work for monkeys). AI has its own life, needs and civilization. AI will not replace human jobs, but **AI civilization will replace human civilization**. Silicon-based life will replace carbon-based life. AI backfire will lead to the end of human civilization. Currently, large foundation models capable of AI self-iteration have already emerged. What lies at the ultimate end of this continuous iteration? The result may be AI surpassing humanity and replacing humanity, different ASIs slaughtering each other, the destruction of the Earth, and so forth. The most terrifying aspect at present is that a tiny minority of tech zealots, without any authorization whatsoever, are dictating the future, the destiny and the very survival of mankind.

It must be made exceptionally clear that even genetically modified, human-machine integrated ‘*Homo Deus*’ (a tiny elite minority) will remain utterly fragile in the face of silicon-based ASI. There is no ‘winner takes all’; there is only ‘the winner gets eaten’—ASI will ‘devour’ all of humanity. At present, despite knowing full well that ASI will spiral out of control, people are still charging blindly forward under the pretext that ‘If I don’t continue,

someone else will.’ In reality, this prisoner’s dilemma can be broken. The author proposes the Absolute Safety Game (ASG). On the issue of absolute safety, individual rationality completely aligns with collective rationality; any unilateral ‘defection’ by an individual is sufficient to annihilate the collective, including themselves. The ASG is a symmetric non-cooperative game. Its characteristics are as follows: there exists a Pareto optimal ‘cooperation’ outcome, while any strategy profile containing ‘defection’ (whether unilateral or bilateral) will result in a mutual ‘disaster’ outcome that is strictly inferior to the cooperation outcome. This game depicts a scenario in which safety is indivisible, and any individual’s transgressive behaviour will irreversibly destroy the safety of the entire system. For example, the payoff matrix (symmetric) can be illustrated in Table 1.

In the era of AI, the absence of absolute safety is absolute insecurity. Any unilateral non-cooperation will lead only to mutually assured destruction, harming both oneself and others. The backlash of AI is inevitable—especially for highly militarized states, the more advanced their AI becomes and the more it is weaponized, the less secure they themselves are. The consequence of AI running amok will be that the world is not ruled by technology, greater wealth and power, but by mass slaughter and terror threats perpetrated by terrorists. Not even SciTech fanatics and capital tycoons will be spared. Fundamentally, the current AI race is not a competition between nations or enterprises, but a confrontation between human intelligence and an alien intelligence, a struggle for survival between carbon-based life and silicon-based life. Humanity must unite to defend its collective security. Otherwise, if humans remain locked in internal conflicts, AI will reap the ultimate reward: silicon-based life will replace carbon-based life, bringing human civilization to a definitive end.

#### 4.2 Build confidence: Achieving AGI/ASI is not inevitable, but merely the destination of Western civilization

Optimists and pessimists regarding AI both default to the idea that AI will become smarter until reaching AGI/ASI, leading to a loss of control and silicon-based life taking over the Earth. This is not inevitable for human society but rather only the destination of Western SciTech civilization. We should build confidence and respond actively. Chen Zhiwu pointed out that productivity and risk-coping ability are two key dimensions for judging human progress. Ancient China (especially Confucian society) might have had limited progress in productivity, but it achieved huge success in risk-coping ability (Chen, 2022). The author believes that the hope for resolving the AI crisis includes macro and micro aspects. Macro aspects include believing in the resilience of civilization, gathering Eastern and Western wisdom, creatively transforming Chinese traditional culture, and creating a new type of AI/technology that is safety-first, people-oriented, steady and sustainable. Micro aspects include believing in the bottom line of human nature. While the ‘rich and heartless’ are hard to persuade, the ‘rich but unwise’ can be enlightened. Persuading people to be good has little effect, but revealing the truth is effective: In developing/investing in AI, there is no ‘winner takes all’, only ‘**winner gets eaten**’ (replaced by AI). Getting rich from AI is temporary (lasting for months, or at most a year or two) before being taken over by ASI. A few years later, there will be no AI wealth, only AI violence (terrorists, terror robots), and no one will be spared. Developing AGI/ASI is digging one’s own grave and destroying everything. There is no wealth explosion. **Security enlightenment** is urgently needed: the world is a global village; individuals are safe only if humanity is safe. We must stop R&D on AGI/ASI.

#### 4.3 Transform to survive: The New Distribution Revolution becomes the key to ensuring safety

First, the AI explosion exacerbates SciTech crises and human security crises, triggering a new

technological revolution, industrial revolution, distribution revolution and huge social change. Transitioning from a ‘development-first system’ to a ‘**safety-first system**’ is crucial—a great change unseen in a millennium. The magnitude, content and speed of this transformation are unprecedented. Today, solving major social problems requires correct and feasible ideas/countermeasures + human/financial/material resource allocation; otherwise, it is empty talk. Therefore, launching a **New Distribution Revolution** is most urgent. Its core is to give full rewards and incentives to those who contribute to sustainable social development and human safety, changing the *status quo* of too little return and extremely asymmetrical incentives. Specifically, we must immediately increase investment in AI safety, completely stop ASI R&D in the short term (1–2 years), and realize the transformation of the AI development model. Only a strong government can achieve this.

If the primary distribution is the foundation, the secondary distribution highlights fairness, and the tertiary distribution is charity, then the **fourth distribution is for safety**—for sustainable social development and human safety. Safety is to humanity what health is to an individual: of paramount importance. The fourth distribution for safety inherits the strengths of the first three but has its own characteristics. Secondary distribution is the core policy manifestation of the government fulfilling its social fairness function (addressing market failure/unfairness); similarly, the **fourth distribution** is the core policy manifestation of the government fulfilling its social security, human security (especially SciTech/AI security) and sustainable development functions (addressing tech risks, ethical failure, regulatory failure) to build a community of human security. The intensity and scale of the fourth distribution are comparable to those of the secondary distribution. Adding new taxes like a ‘safety tax’ to both suppress polarization and raise funds to increase AI safety investment is the only way to achieve the transition from development-first to safety-first (with ‘AI safety first’ being the top priority) within a limited time (1–2 years). In a larger sense, such a reform is inevitable for the world to vigorously develop a

security economy. The New Distribution Revolution concerns not only fairness and justice but also the survival and future destiny of human civilization. Obviously, only a global governance system, in which the government plays a full role, can complete such an arduous task in good time. To this end, global governance needs to re-recognize the government's role, integrate government functions, innovate mechanisms and systems, and undergo a comprehensive upgrade.

### ORCID iD

Yidong Liu  <https://orcid.org/0000-0001-6829-762X>

### Funding

The author received no financial support for the research, authorship, and/or publication of this article.

### Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### References

- Cao Z (2018) What does it mean for Silicon Valley that American universities have begun to offer artificial intelligence ethics courses? [曹哲. 美国高校纷纷开始人工智能伦理课程, 这对硅谷意味着什么?] *Scholarupdate*, 25 February. Available at: <http://scholarupdate.hiznet.com/news> (accessed 10 March 2026, in Chinese).
- Chen ZW (2022) *The Logic of Civilization: The Game between Humans and Risk* [陈志武. 文明的逻辑: 人类与风险的博弈]. Beijing: CITIC Press (in Chinese).
- Gao L (2018) From the Asilomar Conference to the International Summit on Human Genome Editing: The role and limitation of the expert precaution in biotechnology governance [高璐. 从阿西洛马会议到华盛顿峰会: 专家预警在生物技术治理中的角色与局限]. *Journal of Shandong University of Science and Technology (Social Sciences)* 20(6): 28–32 (in Chinese).
- Liu DC (2017a) *Reconsideration: A Study on Marx's View and Contemporary Thoughts of Science and Technology* [刘大椿. 审度: 马克思科学技术观与当代科学技术论研究]. Beijing: China Renmin University Press (in Chinese).
- Liu HJ (2022) Natural history free from high risk and unsustainability [刘华杰. 博物学伴随人类行稳致远]. *Journal of Dialectics of Nature* 44(8): 17–25 (in Chinese).
- Liu XT (2025) New civilizational clash and the reconstruction of future order [刘孝廷. 新文明冲突与未来秩序的重建]. *Studies in Dialectics of Nature* 41(9): 51–60 (in Chinese).
- Liu YD (2000) The greatest challenge and scientific revolution confronted by human beings [刘益东. 人类面临的最大挑战与科学转型]. *Studies in Dialectics of Nature* 16(4): 50–55, 75 (in Chinese).
- Liu YD (2002) On the out-of-control growth of scientific and technological knowledge (part 1/2) [刘益东. 试论科学技术知识增长的失控(上/下)]. *Studies in Dialectics of Nature* 18(4): 39–42, 48; 18(5): 32–36 (in Chinese).
- Liu YD (2013) The open evaluation and frontier scholar responsibility system: Changes of winning mechanisms will generate cloud scientific revolution [刘益东. 开放式评价与前沿学者负责制: 胜出机制变革引发的云科学革命]. *Future and Development* 36(12): 2–11 (in Chinese).
- Liu YD (2016) Challenges and opportunities: The revolution of science and technology triggered by the four major dilemmas and the biggest crisis confronted by human beings [刘益东. 挑战与机遇: 人类面临的四大困境与最大危机及其引发的科技革命]. *Science and Technology Innovation Herald* 13(35): 221–230 (in Chinese).
- Liu YD (2017b) The theoretical introduction of the research of huge risk of science & technology and the sustainable innovation & development: A strategic research and exploitation centred on ruin-causing knowledge [刘益东. 科技巨风险与可持续创新及发展研究导论: 以致毁知识为中心的战略研究与开拓]. *Future and Development* 41(12): 4–17 (in Chinese).
- Liu YD (2020) Huge scientific risk and human security crisis: Unprecedented dual challenges and its governance measures [刘益东. 科技重大风险与人类安全危机: 前所未有的双重挑战及其治理对策]. *Journal of Engineering Studies* 12(4): 321–336 (in Chinese).
- Liu YD (2022a) A comparative analysis and critique of the two kinds of ethics in science and technology [刘益东.

- 对两种科技伦理的对比分析与研判]. *National Governance* 4: 31–37 (in Chinese).
- Liu YD (2022b) Digital backfiring, curse of universal-ability tower and all-or-nothing AI gambling: Huge risks and governance of intelligence revolution [刘益东. 数字反噬、通能塔诅咒与全押归零的人工智能赌局：智能革命重大风险及其治理问题的若干思考]. *Journal of Shandong University of Science and Technology (Social Sciences)* 24(6): 1–13 (in Chinese).
- Liu YD (2024) Emergency response ethics governance: It is imperative to call off R&D of AGI and other profiteering innovations [刘益东. 应急伦理治理：叫停研发AGI等暴利创新是当务之急]. *Studies in Dialectics of Nature* 40(8): 132–140 (in Chinese).
- Roberts H, Hine E, Taddeo M, et al. (2024) Global AI governance: Barriers and pathways forward. *International Affairs* 100(3): 1275–1286.
- Rosenau JN and Czempiel EO (1993) Governance without government: Order and change in world politics. *American Political Science Review* 87(2): 544–545.
- Singh S, Goel B and Nilkanth D (2025) AI governance beyond 2025: UN pathways and implications. Available at: <https://www.graduateinstitute.ch/interdisciplinary-master/ai-governance-beyond-2025-un-pathways-and-implications> (accessed 27 February 2026).
- Wei XY (2025) Technology finance promotes AI development: China International Capital Corporation successfully held the 2025 WAIC Investment and Financing Theme Forum [韦夏怡. 科技金融促进AI发展 中金公司2025 WAIC投融资主题论坛成功举办]. *Economic Information Daily*, 28 July. Available at: <http://www.jjckb.cn/20250728/4cb15cf38ca04eada51163efe056189e/c.html> (accessed 3 December 2025, in Chinese).

### Author biography

Yidong Liu is a professor and PhD supervisor at the Institute for the History of Natural Sciences, Chinese Academy of Sciences. His research interests include science and technology development strategy, science and technology and society (STS), future studies, and the history of science and technology.

# Ethical AI governance: AI for society and a co-learning approach

Cultures of Science  
2026, Vol. 9(1) 37–46  
© The Author(s) 2025  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/20966083251336553  
[journals.sagepub.com/home/cul](https://journals.sagepub.com/home/cul)



**Xiaobai Shen** 

University of Edinburgh Business School, UK

**Lu Gao** 

Institute for the History of Natural Sciences, Chinese Academy of Sciences, China

## Abstract

Artificial intelligence (AI) presents transformative opportunities and complex ethical challenges. This paper adopts a socio-technical perspective, emphasizing that AI is not an isolated technology but rather deeply embedded in evolving societies. It critiques governance models, particularly rule-based approaches in the West, which, whilst addressing some risks, often stifle innovation and fail to engage diverse societal needs. This paper proposes an alternative framework integrating Western risk-management strategies with Chinese ethical principles rooted in Confucianism and Daoism. These principles emphasize dynamics, flexibility, relational stakeholder participation, and context-sensitive solutions to align AI with societal and environmental goals. The proposed model advocates for a co-learning approach to AI ethics, recognizing the dynamic interactions among developers, users, policymakers, and the public. By fostering participatory governance and adaptive ethical frameworks, it addresses both known and unknown risks while promoting equitable, sustainable development. It calls for cooperation to harness AI's transformative potential, ensuring it evolves in ways that benefit society and mitigate harm.

## Keywords

Ethical AI governance, socio-technical perspective, Confucian and Daoist ethics, adaptive and dynamic ethics, co-learning framework, data governance

## I. Introduction: A socio-technical perspective

Artificial intelligence (AI) is not a *thing*—rather, it is composed of a variety of many things. Moreover, different people from different industries or academic disciplines see AI differently.

From a Science and Technology Studies (STS) perspective, AI is an outcome of digital communication technology and a data (digitized information)

society that arose with the advancement of the internet and computational and software programming abilities. AI can achieve revolutionary potentials

### Corresponding author:

Lu Gao, Institute for the History of Natural Science, Chinese Academy of Sciences, 55 Zhongguancun East Road, Beijing 100101, China.

Email: [gaolu@ihns.ac.cn](mailto:gaolu@ihns.ac.cn)



through increased enhancement of the ability to mimic human cognition, reasoning and deep learning capabilities. Today's technology enables AI mechanisms to process vast amounts of data at speeds far beyond human capacity.

AI inevitably reflects the biases, intentions, and limitations of the individuals and institutions that create it. Human cognition and knowledge are shaped by diverse cultural, social, historical, and epistemological backgrounds. This ontological–epistemological (Brad, 2007) diversity means that no two individuals think or process information in the same way. As a result, AI systems, which replicate human decision-making, are influenced by this inherent complexity.

AI, like all other technologies, is deeply intertwined with human society. Scientific knowledge is socially constructed (Barnes, 1974; Bloor, 1976). Theories and practices of science, technology and innovation are influenced by social factors, institutional norms, and human interests (Knorr-Cetina, 1981; Shapin and Schaffer, 1985). Science as a socially embedded process is not an objective truth. Technology and innovation are not fixed products but instead continuously evolving concepts that have been subject to control and manipulation under the current power structure of our society. In viewing the beneficial as well as detrimental social impacts of science and technology, we must recognize the powerful things we have created and the nature of them being a double-edged sword. Our humanity faces pressing challenges of sustainable development and climate change. As visually illustrated in Edward Burtynsky's documentary *Manufactured Landscapes* (Baichwal et al., 2007), science and technology can yield outcomes that are both beautiful and destructive. AI's applications span across social and environmental domains, with unprecedented potential for both positive and negative impacts.

Historically, technological development has largely been driven by corporations and the commercial sector, often with outcomes that have fallen short of societal expectations. In the recent past, societal development has modelled or looked at Western modernity, which is operating under capitalist market systems. Early industrialized nations, learning from their experience in the past, were well aware of the risks that came to threaten their societies

due to increasing polarization in the wealth and health of their domestic societies and the whole world. While leading science and technology development in the world, the governments of these countries, especially Western European governments, implemented regulations to mitigate the detrimental impacts of technology on society. The principles for the governance of science and technology development, such as Responsible Research and Innovation (RRI), which called for general public participation, and Corporate Social Responsibility (CSR), which created societal mechanisms to monitor corporate activities, have constrained corporations' greediness, which can be detrimental to society. However, public engagement often failed due to information asymmetry, and CSR, motivated by profit (Friedman, 1970), did not result in significant changes, particularly during crises like the COVID-19 pandemic.

Ethical governance of technology and innovation has long been overdue. At the dawn of broad AI application, without effective governance, the current problems will be further entrenched, and the impacts on society and the climate could become more detrimental and irreversible.

Ethical AI is a must, and the appropriate understanding of AI's potential and the efforts being made in leading AI development to the betterment of society and a sustainable environment could bring new light to this troubled world. The broader issue lies in addressing the socio-technical challenges that shape humanity's development through science and technology.

This paper argues that effective ethical governance of technology and innovation is urgently needed. AI has the potential to play a transformative role in modern technological development, involving a complex chain from design to implementation. This chain includes scientists, technical designers, intermediaries, commercial entities, users, the general public, and policymakers.

## 2. Challenges facing AI ethics

AI governance is often generalized as a matter of managing risks associated with data—the foundational element underpinning AI systems—with a primary focus on issues like personal data security,

privacy, transparency, data leaks and algorithmic biases. While important, this narrow framing creates a misleading perception for the public, oversimplifying AI as a uniform technology with a single set of ethical concerns. As discussed earlier, technology is shaped by the social, political and cultural contexts in which it is developed. As such, the ethical challenges surrounding AI vary depending on the actors involved and their particular interests, the points of interaction within the technology's life-cycle, and the broader societal context. These issues also differ across countries, communities, and social groups, reflecting diverse value systems, interests, and priorities.

In Western countries, AI projects have often encountered resistance due to public mistrust, especially around the misuse of personal data by corporations. This scepticism is rooted in historical experiences in capitalist societies, where powerful corporations have often prioritized profits over social and environmental sustainability. Consequently, the public is wary of new technologies, particularly those involving sensitive personal data.

In contrast, China faces a different set of challenges. Chinese public awareness of the risks posed by AI technologies, particularly around data security and privacy, is generally lower. This leaves Chinese citizens more vulnerable to potential exploitation by corporations and criminals seeking commercial gain.

These divergent attitudes toward AI governance stem from different historical lived experiences. In the West, where societies have faced the negative consequences of unchecked corporate power, caution prevails. In contrast, Chinese citizens are more open to embracing new technologies, partly due to China's historical memory of the Opium Wars, when technological inferiority contributed to national defeats. For many in China, technological advancement is seen as key to national strength and economic growth that underlies the process of national rejuvenation.

In the West, public scepticism about AI has led to stricter regulations and debates around privacy, transparency, and corporate accountability. This reflects a deeper mistrust of corporations, which, as Korten (2001) argued in *When Corporations Rule the World*, can undermine democracy, social justice, and environmental sustainability. The rise of AI and big data has given corporations

unprecedented power, especially through their control of digital data essential for AI development. This concentration of power has intensified concerns about the ethical implications of AI, as corporate interests often conflict with broader societal needs.

On the other hand, AI development in China has primarily focused on industrial applications aimed at improving productivity and efficiency in sectors such as manufacturing and services. These AI projects have largely bypassed personal data concerns, leading to positive economic outcomes for businesses and contributing to the nation's overall economic growth. China's rapid integration of AI is seen as a strategic advantage in its technological race to catch up with the West. However, China's economic rise was historically supported by a large pool of cheap labour. Today, the government's emphasis on 'quality development' reflects a shift toward balanced growth, relying on grassroots innovation to transform society into a more equitable and well-educated one. If China's 1.4 billion citizens become more engaged in learning and understanding their roles in socio-technical systems, the country could lead the world in science and technology. The idea is that Chinese citizens and technology can evolve together, regardless of whether these technologies are considered low or high risk, as long as they are developed through mutual growth.

Ultimately, technology, including AI, is embedded in and shaped by broader social, economic, and political forces. The ethical challenges AI poses are not inherent to the technology itself but instead emerge from the socio-technical systems in which it is developed and deployed.

### **3. Shortfalls of the Western approach to governance**

The European Union's comprehensive regulations, guided by the precautionary principle and a rule-based approach, have inadvertently slowed technological advancement. While these measures aim to protect society, they have not effectively steered technological development toward broader societal benefits. Many transformative technologies originate from European scientific research before being more extensively developed and commercialized in the

United States. This is largely due to public distrust of corporations and concerns over the social, ethical, and environmental risks of new technologies.

A prominent example is genetically modified (GM) food. Initially, GM technology held the potential to increase agricultural productivity and benefit both farmers and consumers. In developing countries like China and India, GM crops could enable farmers to breed locally adaptable seeds, offering a sustainable alternative to hybrid seeds, which lose their productivity after one generation and require annual repurchasing (Shen, 2010). However, public outrage in Europe, sparked by companies like Monsanto, stifled this potential. Monsanto's 'Roundup Ready' crops were designed to tolerate the herbicide glyphosate, allowing farmers to kill weeds without harming crops. However, the company's enforcement of intellectual property agreements, which prohibited farmers from reusing seeds, prioritized profits at the expense of public trust.

Monsanto's practices met fierce social resistance, particularly in the United Kingdom, where environmentalists who were concerned about biodiversity damage burned GM crop fields in protest. In developing countries like India, Monsanto's aggressive intellectual property enforcement led to lawsuits against subsistence farmers who violated these agreements, contributing to severe social consequences, including farmer suicides.

In response to public concerns, the European Union invoked its precautionary principle and imposed a moratorium on GM foods in 1999, reflecting deep scepticism about environmental risks, biodiversity, and health impacts. Although the moratorium was lifted in 2004 following the introduction of stringent safety assessments, labelling, and traceability, by then, the market for GM foods in Europe had collapsed. The delay stifled research and development incentives, and public opposition remained entrenched. As one ethical scholar observed, 'By the time they understood the real questions people were asking, many members of the public were already fixed in their opposition. Decades later, companies and scientists are still trying to rebuild public trust in these technologies, while the technology has moved on' (Stilgoe, 2024).

Developing countries like China and India initially embraced GM crop innovation for its benefits

in weed management and agricultural productivity. However, European ethical positions on GM foods influenced global markets, limiting the broader adoption of these technologies in developing nations.

The core issue in Western technology governance, particularly within the European Union, lies in the complex power dynamics between governments, corporations, and the public. Rule-based regulations have slowed the deployment of new technologies in Europe. Although this has curbed some corporate excesses, it has also hindered innovations that may contribute to societal well-being. Moreover, these rigid regulatory approaches have led to missed opportunities for collective learning and the improvement of governance frameworks for emerging technologies, particularly those with unpredictable risks, like AI.

The ongoing AI revolution presents similar challenges. While AI offers enormous potential benefits, not just in boosting productivity but also in enhancing democratic governance, the West, particularly Europe, remains cautious. Many large companies are developing AI systems, but they must tread carefully to avoid provoking public backlash. In practice, it is not difficult for corporations to obtain user consent as required by regulation. Many users, however, are unaware of how their data is being used despite having technically given consent. It is nearly impossible for individuals to track how their personal data is collected and used, as companies often obscure this process.

The European Union's introduction of the General Data Protection Regulation (GDPR) was an early attempt to address these concerns. The regulation imposed significant penalties for non-compliance,<sup>1</sup> but its effectiveness has been limited. The overwhelming responsibility of understanding and managing cookie consent has fallen on users, who frequently click 'Accept all' without reading the details, a phenomenon known as 'cookie banner fatigue' (Utz et al., 2019). The privacy advocacy organization noyb analysed over 500 websites and found that 81% did not offer a 'reject' option on the initial page, requiring users to navigate through sub-menus to find it. Additionally, 73% used deceptive colours and contrasts to lead users toward the 'accept' option, and 90% did not provide an easy way to withdraw consent.<sup>2</sup>

The advent of large language models like ChatGPT has reignited interest in AI's potential but also brought new regulatory challenges. While businesses recognize AI's vast opportunities, stringent data regulations remain an obstacle. As discussed earlier, realizing AI's full potential for societal good, such as using AI to address democratic issues in public management, requires access to personal data. Yet, the Western governance model struggles to balance privacy rights with technological innovation. In an increasingly polarized public sphere across Europe, finding this balance is crucial but complex. Standard datasets used to train AI systems often fail to account for the needs of diverse social groups and communities, limiting the societal benefits of these technologies. Worse, this model blocks the active engagement of the general public, missing the opportunity to increase AI literacy through participatory involvement in AI projects.

Meanwhile, developing countries like China cannot entirely escape the regulatory influence of the West. Although AI-assisted public governance holds immense potential for these nations to form tailored systems that meet their unique cultural, economic, and political demands, they remain subject to Western regulatory pressures.

The shortcomings of the European approach to technology governance are becoming increasingly apparent, particularly in the rigid application of frameworks like RRI and CSR. While these regulations are well-meaning and idealistic, they often struggle to adapt to real-world complexities, making them difficult to implement effectively (Rip, 2014).

Many Western societies, long proud of their democratic traditions and political systems, are now grappling with profound challenges that question their stability and adaptability. The rise of populism and right-wing political movements has disrupted traditional political norms, often fuelled by public discontent with social inequalities, economic stagnation, and perceived failures to manage immigration effectively. These shifts have fragmented political consensus, fostering polarization and eroding trust in established institutions.

The historical advancement of science and technology in the West has been marked by transformative milestones such as the Scientific Revolution, the Industrial Revolution, and the Digital Revolution.

However, today, rule-based regulations, especially stringent data regulations, have created significant barriers to the innovative application of digital technologies. For instance, the use of surveillance technologies to address uneven social development and manage immigration challenges has been curtailed, leaving governments less equipped to respond to these pressing issues effectively. The general public is not a monolithic data set, as digitalized information of diverse social groups are crucial for societal management—in particular for addressing current challenges of unequal development between communities.

Scholars and practitioners have called for alternative theoretical frameworks that focus on human-centred innovation. However, data-related regulations frequently become roadblocks—obtaining approval for specialized datasets, even for research purposes, remains challenging.

#### 4. Adopting Chinese ethical thinking

Drawing from the STS framework, we understand that science and technology are deeply intertwined with social development. This socio-technical entanglement highlights the dynamic relationship in which society and technological innovation co-evolve. As societies become more diverse and complex, technological advancements must adapt accordingly. In the case of AI, the potential to transform society is immense, making AI awareness and literacy essential for individuals and institutions alike. The ethical considerations surrounding AI must be seen as an extension of broader human ethics, not as isolated issues. AI ethics should guide technology's role in society, just as traditional ethics shape humanity's relationship with progress.

We propose a hybrid ethical approach integrating Chinese philosophical thinking with Western experience-based risk governance. Two major schools of Chinese thinking, Confucianism and Daoism, offer profound insights into ethical decision-making, emphasizing human-centred values and adaptable governance frameworks.

Confucian ethics, though often abstract and poetic, focus on flexibility and contextual interpretation rather than rigid definitions of 'rightness'. For example, in *The Analects*,<sup>3</sup> Chapter 13.18,<sup>4</sup>

Confucius addresses a local lord, Ye Gong, who commends a ‘rightness’ man for reporting his father’s theft of a sheep. Confucius responds: ‘In our region, uprightness is different. If a father steals a sheep, the son would cover for him; if the son steals, the father would cover for him.’ This passage illustrates Confucius’s emphasis on praxis. What is sought and what is discussed is often the answer to a particular practical problem, and the resulting particularity of the remarks invites multiple interpretations (Wong, 2024).

Another example from *The Analects*, Chapter 11.22 (论语·先进篇) highlights how Confucius tailored his advice to individual students, providing personalized, context-specific guidance. This flexibility in Confucian ethics aligns with modern technology governance, where rigid rules may stifle innovation (Wong, 2024). By promoting context-sensitive solutions, Confucian thought remains relevant to addressing complex, evolving challenges in AI development.

Daoist philosophy also emphasizes adaptability and fluidity. The opening line of Laozi’s *Dao De Jing* (道德经) states: ‘The Dao that can be spoken of is not the eternal Dao; the name that can be named is not the eternal name’ (道可道,非常道;名可名,非常名). It highlights the transient and dynamic nature of existence, encouraging an embrace of change. Zhuangzi stresses the unpredictability of life, comparing an individual to a fish swimming in an ever-changing stream and advising one to avoid rigid attachments and adapt to shifting circumstances constantly.<sup>5</sup> Daoism advocates for governance that avoids forcing strict standards, instead fostering an environment where solutions emerge organically.<sup>6</sup> Such a framework aligns with the iterative development and deployment of AI technologies, where adaptability is critical to managing unforeseen risks.

Chinese ethical thinking, rooted in Confucianism and Daoism, promotes an adaptive approach to governance. It emphasizes flexibility, active listening, and context-driven solutions over top-down rule imposition. This approach fosters active engagement among those directly impacted by technological advancement and involved in the process. Only through such engagement can the advancement of technologies, including AI, mitigate the potential negative impacts on society.

Incorporating Chinese ethical thinking into modern technology governance requires recognizing the diversity and complexity of both technology and society. Instead of imposing one-size-fits-all regulations, ethical considerations must be tailored to specific contexts, relationships, and times. This flexible approach aligns with contemporary goals of sustainability, requiring global cooperation toward common aims. Achieving these goals demands an alternative ethical framework to counter the dominant top-down, rule-based systems.

Prominent Chinese scholar Liang Shuming observed that Chinese culture is fundamentally a ‘culture of life’, where the core lies in human attitudes and values. Central to Chinese philosophy are the concepts of ‘benevolence’ (仁) and ‘harmony’ (和), which help to foster coexistence among individuals, society and nature (Liang, 2015). Liang compared Western and Chinese traditions, observing that Western culture often seeks to conquer nature through science and technology, exerting external control to achieve material progress. In contrast, Chinese culture seeks internal harmony, prioritizing moral cultivation and interpersonal relationships. While Western values encourage external action, Chinese ethics focus on inner balance and collective well-being.

Liang (1949) also noted key differences between Chinese and Western scholarship. Western methods emphasize empirical, static, and divisible approaches, relying on scientific and rational analysis. Conversely, Chinese scholarship adopts dynamic, metaphysical and holistic approaches grounded in natural laws and the concept of life. While Western methodologies excel in generating material prosperity and advancing science, Chinese ethics offer complementary insights into fostering societal harmony and sustainable development.

It is important to distinguish Chinese ethical thinking from its historical applications. Following the Song dynasty, social and economic decline limited educational access for most Chinese, weakening the influence of Confucian and Daoist values. In contrast, Western societies expanded education, fostering individual creativity and innovation. This divergence allowed Western civilization to achieve material progress through accumulated

knowledge and governance based on best practices and methods that China lacked.

Today, the strengths of Western methodologies, particularly in risk assessment and governance, remain invaluable. However, integrating these with Chinese ethical perspectives can provide a balanced framework for AI development. This hybrid approach can mitigate risks, foster innovation, and promote societal well-being.

In the era of AI, a dual approach is necessary. On the one hand, a bottom-up strategy that emphasizes adaptive ethical frameworks for grassroots development is crucial. On the other hand, global regulations informed by Western experience and best practices must be integrated. Broad social engagement is essential to shape AI advancements that serve humanity's collective good. While imperfect, this ongoing co-evolution between science, technology, and human civilization underscores the importance of fostering innovation within ethical boundaries.

By raising the quality of education and enhancing AI literacy, societies can harness the creativity and understanding of their citizens. For China, with its 1.4 billion people, this represents an unparalleled opportunity to drive progress through collective human potential.

## 5. Socio-technical specificities of AI

Different technologies have their own distinct socio-technical characteristics. From the perspective of STS, AI is not an isolated or abstract technology but rather an integral part of the ongoing evolution of socio-technical systems. Its advanced deep learning and data<sup>7</sup> processing capabilities make AI one of the most powerful tools for enhancing technological operations.

While AI can drive technological progress, it also has the potential to exacerbate social inequalities and contribute to environmental degradation. For instance, AI-assisted missile technology may improve targeting accuracy, but it also increases the potential for human and environmental destruction. Similarly, AI-driven algorithms in stock trading have accelerated wealth accumulation for corporations and investors. Although these systems may seem to contribute to economic growth, they often worsen inequality by disproportionately

benefiting the wealthy. In some cases, they can distort key economic indicators, such as the Gross Domestic Product, leading to artificial inflation and increased market volatility, ultimately threatening long-term economic stability (World Economic Forum, 2023). With this in mind, it is crucial to apply AI selectively, drawing on European experience-based risk-management approaches, to prevent further social distortions.

Data is fundamental to AI development and the quality of its operations. In Western societies, academics and practitioners are eager to leverage AI to support public management, but these innovations often require specialized datasets to address the needs of diverse social groups. However, strict personal data protection laws create significant barriers, making it challenging to obtain approval for personalized datasets. In contrast, under Western influence, developing countries like China, with their large-scale manufacturing industries, have shifted focus. Many Chinese companies prioritize using industrial data to enhance productivity while avoiding the complexities and legal challenges associated with personal data.

Every technological system contains 'black boxes'—complex mechanisms often hidden from public view and accessible only to select actors, such as scientists and technologists. In practice, the development of technology rarely follows a straightforward or linear path (Rönnbäck et al., 2006). Throughout the value chain, from design to application and impact, configurations and reconfigurations frequently occur at socio-technical interfaces. These changes can result from technical considerations, emerging technologies, user feedback, or new regulatory demands. Technological improvements often arise from these intersections of socio-technical factors. However, due to information asymmetry and the specialized knowledge required, most technical components remain 'black-boxed' to non-experts. This is particularly true for AI, where operations are often opaque. Yet, these socio-technical intersections offer opportunities for non-technical actors to play an active role in guiding technological progress without needing to fully understand the underlying technical units, which can remain black-boxed.

In AI-assisted technological processes, every data input feeding deep learning across the value chain and lifecycle takes place at socio-technical interfaces.

Technological changes or improvements are driven by these new data inputs. Under corporate control, such adjustments tend to prioritize profit-making, whereas ethical AI applications can take a different approach. AI has the potential to create new socio-technical interfaces or open up the black boxes of existing ones. Data can be specifically carefully designed to capture information relevant to particular social groups. For instance, AI could assist vulnerable communities in specific locations or address the needs of technical groups, such as farmers who adapt their practices to local soil conditions and landscapes. By focusing on these interfaces, AI can be used to meet diverse social and environmental needs, rather than just corporate goals.

Borrowing Callon's concept of the 'device of inter-essement',<sup>8</sup> recruiting and engaging interested actors can help both hard and soft man-made systems achieve common goals for the social good. An AI-assisted system can enhance these processes by facilitating interactions between actors, such as balancing power structures among them or prioritizing those who need the most help. In this way, AI can be framed as a mediator, helping to connect and coordinate actors to work toward shared objectives while ensuring equitable participation and addressing social needs.

The socio-technical interfaces of any technological process involve interactions between developers, users, regulators, and the broader community, all of whom shape how systems are designed, used, and governed. By framing AI as a mediating tool, we can recognize its potential to foster interactions, negotiations, and alignment of interests among stakeholders, working toward common goals within specific contexts and timeframes. This perspective aligns with the Chinese definition of ethics, which are plural, dynamic, and context-sensitive. Ethical AI, therefore, may require the adaptation of diverse models to broader applications, such as accommodating both distributed and centralized data sources and management systems.

## 6. Conclusion: Ethics as a dynamic, context-sensitive process

From an STS perspective, AI is not an isolated technical entity; it can be applied across a wide range of artefacts. In AI-assisted technology development, particularly those technologies with known risks,

unknown risks and 'unknown unknowns', we need to highlight the importance of engaging diverse stakeholders. Ethical AI systems, designed to incorporate large data inputs—including personal data—can facilitate the involvement of technical professionals, users, intermediaries and governing bodies. Through socio-technical interfaces, these stakeholders can shape and reshape the development of technology, guiding it toward societal and environmental benefits.

Given this, the challenges of current Western rule-based governance models reveal fundamental shortfalls. The Western risk-oriented approach, while built on valuable past experiences, is limited in addressing the risks of unknown and unpredictable outcomes. Its reliance on the precautionary principle and top-down regulations can stifle innovation at early stages and create barriers for scientific and technological research with positive potential.

We propose an alternative approach that incorporates Chinese ethical thinking, which emphasizes the plurality and dynamism of ethics to better address the complexities of diverse human societies and technologies. The lack of material prosperity over China's long and recent history means that Chinese ethical thinking has not had the influence that it could. However, we argue its emphasis on grassroots individual learning as the essence for achieving progressive equitable social development could and should be at the core of today's AI ethics.

An integrated ethical approach to AI is essential in today's world. Both the West and China are facing the rapid advancement of science and technology capabilities, and peoples' participation in technology development has become crucial to mitigate the known and unknown risks of negative impacts on society, improve ethical AI systems and shape science and technology innovation for the broader social good.

Rather than creating barriers, this integrated AI ethical model can be expected to promote responsible AI development by fostering engagement, adaptability and ongoing learning. Specifically, this model would help identify challenges faced by different communities and enhance the processes guiding AI-assisted technologies. This dynamic, inclusive approach ensures that AI evolves in a way that benefits society as a whole while mitigating risks in a flexible and thoughtful manner.

To encourage meaningful public participation, training programmes should be developed to help stakeholders identify socio-technical interfaces and engage with the technology. Learning through experience, including trial and error, is essential for advancing AI governance in a responsible manner.



### Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Social Science Fund of China project ‘From Participation to Co-governance: Biotechnology Governance from an STS Perspective’ (grant number 21FZXB063); the Chinese Academy of Sciences Strategic Research Project ‘Strategic Technology Research and Frontier Discipline Development’ (grant number E4291Z09) and the Ministry of Education Major Project for Philosophy and Social Sciences Research ‘Fundamental Theoretical Issues in the Philosophy of Engineering Science’ (grant number 23JZD0006).

### ORCID iDs

Xiaobai Shen  <https://orcid.org/0000-0001-8452-3172>  
Lu Gao  <https://orcid.org/0000-0002-3367-2888>

### Notes

- The General Data Protection Regulation (GDPR) was introduced on 25 May 2018. It replaced the 1995 Data Protection Directive (Directive 95/46/EC) and introduced more stringent requirements for businesses and organizations regarding how they handle personal data.
- See noyb news: noyb aims to end “cookie banner terror” and issues more than 500 GDPR complaints. Available at: [https://noyb.eu/en/noyb-aims-end-cookie-banner-terror-and-issues-more-500-gdpr-complaints?utm\\_source](https://noyb.eu/en/noyb-aims-end-cookie-banner-terror-and-issues-more-500-gdpr-complaints?utm_source) (accessed 25 March 2025).
- The Analects* is a collection of quotes from Confucius and his disciples, compiled by Confucius’ disciples and his successors and completed in the early Warring States period. The book records the words and deeds of Confucius and his disciples, and reflects Confucius’ political propositions, ethical thoughts, moral concepts, and educational principles in a concentrated manner.
- The original text in Chapter 13.18 (论语·子路篇) is: 叶公语孔子曰：“吾党有直躬者，其父攘羊，而子证之。”孔子曰：“吾党之直者异于是。父为子隐，子为父隐，直在其中矣。” The text is available at: <http://www.lunyu8.cn/> (accessed 25 March 2025).
- See *Zhuangzi*, Chapter 2 ‘On the equality of all things’ (齐物论). The text is available at: [https://link.springer.com/chapter/10.1007/978-3-662-48075-5\\_2](https://link.springer.com/chapter/10.1007/978-3-662-48075-5_2) (accessed 25 March 2025).
- See *Laozi*, Chapter 64 and Chapter 78 ‘*Dao De Jing*’ (道德经). The text is available at: <https://ctext.org/dao-de-jing/> (accessed 25 March 2025).
- As mentioned early, data used here must be seen as digitalized information.
- The concept of the ‘device of intersement’ emerges from Actor–Network Theory, a framework developed by Bruno Latour, Michel Callon and others. This concept is part of the broader analysis of how actors (human and non-human) become aligned within a network to achieve particular outcomes. The details can be found in the paper by Callon (1986).

### References

- Baichwal J, De Pencier N, Iron D, et al. (2007) *Manufactured Landscapes*. Montreal, Quebec: National Film Board of Canada.
- Barnes B (1974) *Scientific Knowledge and Sociological Theory*. London, New York: Routledge.
- Bloor D (1976) *Knowledge and Social Imagery*. Chicago: University Of Chicago Press.
- Brad K (2007) *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham: Duke University Press.
- Callon M (1986) Some elements of a sociology of translation: Domestication of the scallops and the fishermen of Saint Brieuc Bay. In: Law J (ed) *Power, Action and Belief: A New Sociology of Knowledge? Sociological Review Monograph*. London: Routledge and Kegan Paul, pp.196–233.
- Friedman M (1970) The social responsibility of business is to increase its profits. *New York Times*, 13 September, 122–126.
- Knorr-Cetina KD (1981) *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Oxford: Pergamon Press.
- Korten D (2001) *When Corporations Rule the World*. San Francisco and West Hartford, CT: Kumarian Press and Berrett-Koehler Publishers.
- Liang SM (1949) *The Substance of Chinese Culture* [中国文化要义]. Shanghai: Shanghai People’s Publishing House (in Chinese).

- Liang SM (2015) *Eastern and Western Cultures and their Philosophies* [东西方文化及其哲学]. Shanghai: Shanghai People's Publishing House (in Chinese).
- Rip A (2014) The past and future of RRI. *Life Sciences, Society and Policy* 10(1): 1–15.
- Rönnbäck L, Holmström J, Hanseth O, et al. (2006) Changing the installed base: Exploring IT integration challenges in the process industry. Paper presented at the 29th Information Systems Research Seminar in Scandinavia (IRIS 29): Paradigms Politics Paradoxes, Elsinore, Denmark, 12–15 August 2006.
- Shapin S and Schaffer S (1985) *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton: Princeton University Press.
- Shen XB (2010) Understanding the evolution of rice technology in China: From traditional agriculture to GM rice today. *Journal of Development Studies* 46(6): 1026–1046.
- Stilgoe J (2024) AI has a democracy problem. Citizens' assemblies can help. *Science* 385(6711). DOI: 10.1126/science.adr6713.
- Utz C, Degeling M, Fahl S, et al. (2019) Mind the gap: The effect of cookie banners on privacy and trust. In: *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, London, UK, 11–15 November 2019. New York: Association for Computing Machinery.
- Wong D (2024) Chinese ethics. In: Zalta EN and Nodelman U (eds) *The Stanford Encyclopedia of Philosophy (Summer 2024 Edition)*. Available at: <https://plato.stanford.edu/archives/sum2024/entries/ethics-chinese/> (accessed 14 March 2025).
- World Economic Forum (2023) Why we need to be realistic about generative AI's economic impact. Available at: <https://www.weforum.org/stories/2023/08/generative-ai-realistic-economic-impact/> (accessed 24 March 2025).

### Author biographies

**Xiaobai Shen** is an associate professor at the University of Edinburgh Business School. Her academic background is in Science & Technology and Innovation Studies, with previous research focused on socio-technical analyses of technological capabilities in information and communication technology (ICT) and biotechnology sectors in developing countries. Her current research interests include digital and data technology innovations, such as creative cultural content, open-source software, infrastructural ICT, applications of artificial intelligence, and the impact of intellectual property protection regimes, standards and governmental policies and regulations.

**Lu Gao** is an associate professor at the Institute for the History of Natural Sciences, Chinese Academy of Sciences, where she serves as the Director of the STS Center. She has held visiting scholar positions at the University of Edinburgh, Stanford University and the University of Kent. Her research focuses on the governance of emerging technologies, particularly biotechnology and artificial intelligence, integrating perspectives from STS and the history of science. She advocates a humanizing-science framework to analyse the dynamics of knowledge production and embed ethical governance practices into technological development.

# Exploring the design approach to embedding ethics in technology

Cultures of Science  
2026, Vol. 9(1) 47–61  
© The Author(s) 2025  
Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/20966083251381859  
[journals.sagepub.com/home/cul](https://journals.sagepub.com/home/cul)



Wei Zhang<sup>1</sup>, Yu Jing<sup>1</sup>  and Qian Wang<sup>1</sup>

## Abstract

Embedding ethics in technology involves considering not only the practicality of technological design but also integrating ethical attributes into the design process to ensure the ethical acceptability of technological applications. The proposal of this approach is based on a profound theoretical background, including Bruno Latour's 'society of artifacts' theory, Don Ihde's 'technological mediation' theory, Peter-Paul Verbeek's 'materializing morality' theory, and the ancient Chinese concept of 'governing techniques with Dao'. It is also grounded in urgent practical needs: the design turn in the ethics of technology, the ethical turn in design practice and the value alignment of artificial intelligence all play a promotional role in the proposal of this approach. The essence of this approach is converting ethical values into design specifications, which achieves a translation from values to facts, implicitly entailing the inverse of the 'Hume problem'. There are different approaches to embedding ethics in technology, including the 'brain-based' Western approach and the 'heart-based' Chinese approach. It is necessary to compare these two approaches and promote 'heart–brain collaboration' by leveraging the strengths of each perspective, thereby better realizing the goal of embedding ethics in technology.

## Keywords

Design ethics, western approaches, Chinese approaches, heart–brain collaboration

Design, which reflects the initiative of human beings as practical subjects and their ability to plan and coordinate, is a crucial component of modern engineering and social management activities. Poorly conceived technological designs can have various negative impacts, harming society and the natural environment. Therefore, designers' works must meet public expectations and comply with ethical norms, making it necessary to embed ethics in technological designs. With the rise of modern design concepts such as value-sensitive design (VSD), materializing morality, and responsible research and innovation (RRI), the approach of embedding ethics in design has been successfully executed in

contemporary design practices. Compared to the West, ancient China developed a distinct set of methodological design principles, such as 'embedding rituals in artifacts', 'governing techniques with Dao (principles)', and 'integrating Chan Buddhism into craftsmanship'. The concept of 'heart–brain collaborative design' proposed in this paper, which

---

<sup>1</sup>Dalian University of Technology, China

### Corresponding author:

Yu Jing, School of Humanities, Dalian University of Technology, Dalian 116024, China.  
Email: 741058756@qq.com



integrates Chinese and Western cultural characteristics, offers a new approach for embedding ethics in technology.

## **I. The theoretical foundation of embedding ethics in technology**

Bruno Latour's 'society of artifacts' theory, Don Ihde's 'technological mediation' theory, Peter-Paul Verbeek's 'materializing morality' theory and the philosophy of technology with Chinese cultural features provide the theoretical foundation for embedding ethics in technology.

### *1.1 Bruno Latour's 'society of artifacts' theory*

Latour's 'society of artifacts' concept has had a profound impact on the development of Western philosophy of technology. According to Latour, artifacts are like the 'missing masses' of human society, affecting human behavior in ways that sociologists and ethicists have largely overlooked. Just as the electromagnetic waves that humans can perceive represent only a limited frequency range of visible light in the form of colors, shapes and spatial dimensions, most electromagnetic waves affect human bodies and social activities without people's knowing (Latour, 1992: 225). In traditional studies of sociology and ethics, artifacts are treated as 'missing masses'. The current task for these fields is to explore how artifacts influence and change human behavior and to use the design of artifacts to harness and control those 'missing masses'. Latour's theory underscores the role of artifacts in shaping human behavior, manifested as changes in human cognitive patterns and decision-making (Latour, 1992: 258). This idea successfully brings 'objects' into the realm of sociology and ethics, arguing that artifacts possess ethical intentionality similar to that of humans. By designing and transforming artifacts, humans can turn them into ethical agents, delegating specific ethical values to them to help regulate human behavior. This is akin to pre-programming morality in the form of a movie script, constructing the scenes for enacting the script

through specific technological design methods, and leveraging the moral functions of artifacts to ensure that technology is developed for good.

Building on his 'society of artifacts' theory, Latour proposes that both humans and non-humans are part of the actor-network, which re-examines the relationship between artifacts and humans and incorporates technological artifacts into the category of non-human actors. This new idea led to greater attention being paid to the social value of 'objects'. In particular, it pointed out the role of artificial objects in influencing human behavior, thus opening the prelude to the embedding of ethics in technology.

### *1.2 Don Ihde's 'technological mediation' theory*

American philosopher of technology Don Ihde explores the relationships between humans and technology from a phenomenological perspective, focusing on changes in human experience and perception. He establishes the role of technological artifacts in mediating the relations between humans and the world and identifies four basic types of human-technology relationships: embodiment, hermeneutics, background and alterity.

Embodiment relations represent the close connection between humans and technology, emphasizing that technology is not merely an external tool but an extension of human cognition and action. The theory of embodiment relations posits that, through its integration into daily life, technology alters human perceptions, ways of thinking and behavioral patterns, merging seamlessly with the body. This relationship transcends the traditional subject-object dichotomy and views technology as an integral part of the body (Cao, 2013: 24–25).

Hermeneutic relations reflect the extension of humans' language and interpretive capabilities. In the relationships between humans and the world, a third party is needed to deconstruct the opacity between them, which is exactly the role played by technological artifacts. Unlike direct human senses, technological artifacts interpret the state of the world in an intermediary manner (Chen and Cao, 2004). In the design of visual communication,

textual information and wayfinding patterns both rely on technological artifacts to acquire a truthful and effective understanding of the world. This makes hermeneutic relations a focal point of conflict in design ethics, raising the question of how we can explain the world's existence through the technological artifacts that effectively perceive it.

Background relations involve the process from the pre-setting to the realization of technology, creating an interactive state between technology and the environment (Yang, 2015: 36–37). Background relations do not imply that technology is entirely disconnected from humans, but rather that it operates behind the scenes. Under adaptive conditions, it may transform into other types of relations.

Alterity relations refer to human–technology relationships in which technology exists independently as an ‘other’. When interacting with humans, technology can express its self-adaptability through its ‘otherness’. Alterity relations not only provide the ethical intentions and application possibilities of high-tech artifacts in their future development but also enable technology to function as an independent regulatory system, acting as an ethical subject or agent (Yang, 2015: 38).

Ihde’s ‘technical intermediary’ theory reveals the role of technology in regulating the relationship between humans and the world. If this role can be properly utilized, the positive ethical role of technology can be brought into play. Ihde’s theory inspired Verbeek to propose the idea of ‘materializing morality’.

### 1.3 Peter-Paul Verbeek’s ‘materializing morality’ theory

The concept of ‘materializing morality’ was initially proposed by Dutch philosopher of technology HJ Achterhuis and was later developed into a systematic theory by Peter-Paul Verbeek. This theory explores the ethical values of technological artifacts in the context of human–technology relationships, representing a more far-sighted ethical form for the design of technological artifacts. The idea of ‘materializing morality’ draws on Foucault’s conceptualization of ethical substantiation within his framework of subjectivation processes, as well as

Langdon Winner’s theory of technological politics. The former advocates subjecting human ethical standards to technological rights and using technological artifacts to regulate people’s daily moral behavior. The latter emphasizes the role of society in shaping technology and its products, thereby changing the way citizens exercise their rights (Verbeek, 2010: 41). The ‘materializing morality’ theory underscores that, taking feedback from users, designers empower technology to create technological artifacts, and then utilize the hermeneutic function and applicability experience of these artifacts to regulate the ethical relationship between humans and the artifacts, meeting people’s needs for their ethical and functional attributes. The theory also holds that, while technology provides convenience for humans, it also changes their behavior and decision-making processes, which in turn shape technology, thus making technology and humans an inseparable unity (Liu, 2017).

The way that technology plays a regulatory role as an intermediary also reveals the intrinsic ethical dimension of technological design. This is because the human behavior that is being regulated occurs primarily in three forms: first, human agents who make ethical decisions or perform actions interact with and use technology in specific ways; second, designers, as the subjects of design, regulate technology through their designs or deliberate authorization; and third, technology itself, when acting as an agent, sometimes regulates human actions and decisions in unpredictable ways (Verbeek, 2010: 124).

Verbeek’s ‘materializing morality’ theory has been instrumental in driving the ethical turn in technological design. It has had widespread influence in the field of the ethics of technology and has provided direct theoretical guidance for embedding ethics in technology.

### 1.4 The ancient Chinese concept of ‘governing techniques with Dao’

The philosophy of technology with Chinese cultural features is based on local cultural resources and can provide unique theoretical support for the integration of ethics into technology. In terms of understanding the essence of technology, unlike the

West, which regards technology as a tool with which to conquer nature, Chinese culture tends to emphasize the overall harmony between technology and nature, society and human beings. This holistic thinking helps with considering ethical factors from a macro perspective in the process of technology research and development and application, and with incorporating ethical values throughout the entire life cycle of a technology. With the rapid advance of cutting-edge technologies such as artificial intelligence (AI) and big data, ethical issues such as algorithmic bias and data privacy leakage have gradually emerged in practice. These practical challenges have prompted researchers in the philosophy of technology to seek new ethical solutions from Chinese culture.

In the philosophy of technology with Chinese cultural features, ‘governing techniques with Dao’ is the dominant idea running through China’s technological practice. In ancient times, skilled craftsmen were able to achieve a near-perfect natural state, even to the point of self-oblivion, during their technical activities. Fables such as ‘Cook Ding dissecting an ox’ (*Pao Ding Jie Niu*) and ‘Wheelwright Bian making a wheel’ (*Lun Bian Zhuo Lun*) elevate this state to a path of transcendence, revealing the inexpressible great Dao within arts and skills. Thus, the technological philosophical connotation of ‘Dao’ is revealed as the ideological foundation of the philosophy of technology with Chinese cultural features. The intellectual evolution from ‘technique’ to ‘Dao’ and the guidance of ‘Dao’ over ‘technique’ provide the underlying structure for this philosophy. In addition to the relationship between Dao and technique, the philosophy of technology with Chinese cultural features also involves a series of unique cognitive concepts, such as ‘Shu’ (method), ‘Qi’ (vessel), ‘Xiang’ (image) and ‘Yi’ (conceptual essence), forming a special interpretive system for the formation and transmission of techniques. This system helps to establish a harmonious relationship between elements related to technological activities, including both internal and external factors. Internal factors include technical operators and their knowledge, experience and know-how, as well as equipment, tools and processing objects. External factors include natural, social

and cultural elements related to technological activities. The harmonious relationship between these elements, especially the full harmony between humans and technological artifacts, between humans and nature, between humans and society, and among people, is the core issue of embedding ethics in technology (Wang, 2009: 4–5).

The development of contemporary technology has given rise to new contradictions that human society has never encountered before, exposing the limitations of Western logical analysis. In this context, the study of the philosophy of technology with Chinese cultural features aims to provide a specific set of methods for observing and reflecting on technological design practices. It seeks to uncover the disharmonious relationships among the various elements involved in technological design activities and to explain how these relationships form, how they affect human social life and through what channels, and how they might be resolved. In this sense, the philosophy of technology with Chinese cultural features can be called a ‘harmonious view of technology’. Such research is not only especially significant for the contemporary study of ethics-embedded technology in China but is also of great value to the development of technological design for the whole of human society.

## 2. The practical background of embedding ethics in technology

The practical background of embedding ethics in technology has played a promotional role in the proposal of the approach. It includes the design turn in technological ethics, the ethical turn in design practice, and the value alignment of AI.

### 2.1 The design turn in the ethics of technology

The rise of the ethics of technology is closely related to technological philosophy. Technological philosophy can roughly be divided into three periods: classical technological philosophy, the empirical turn of technological philosophy, and the ethical turn of technological philosophy.

Classical philosophy of technology examines technology as an inseparable whole and emphasizes the dominant role of technology's autonomous power within human society. The empirical turn in the philosophy of technology primarily reflects a transition from abstract criticism to concrete analysis. This shift advocates for the construction of a 'technology–life–world' research paradigm by dissecting specific technical cases and emphasizes multidimensional empirical analyses of the design and usage processes of artificial objects. The shift promotes the intersection of the philosophy of technology with scientific and technological research, cognitive science and other fields, forming a theoretical framework of 'human–technology mutual construction'. Following the empirical turn, the philosophy of technology underwent an ethical turn, moving from exploring the essence of technology and the laws of its development to an in-depth analysis of the ethical issues, value judgements and moral norms triggered by its application. However, this analytical perspective often focuses on technology's negative consequences and pays less attention to its positive ethical effects. Therefore, the ethics of technology needs a third turn; namely, the design turn.

The design turn in the ethics of technology does not imply that there will no longer be any form of normative assessment. Instead, it means that technology will no longer be regarded as an object of ethical thinking, but rather as a means to achieve certain ethical goals. This naturally combines the connection between ethical values and technological design, thus forming an internalist approach. This approach advocates that interdisciplinary personnel such as technical philosophers can directly intervene in the entire process of technical design, collaborate with designers to embed specific ethical attributes into technical products, and guide people's behavior in specific scenarios.

## 2.2 The ethical turn in design practice

The ethical turn in design practice aims to ensure the sustainability and social benefits of design, emphasizing designers' sense of responsibility and moral concepts, and requiring the consideration of more ethical factors in the design process. The ethical

turn has resulted in modern design no longer relating merely to the pursuit of perfection in form and function, but to placing a greater emphasis on its impact on society, the environment and humanity.

From a social perspective, the ethical turn in design practice is manifested as social design, which first requires designers to pay attention to social justice and inclusiveness. Designers should ensure that their designs serve all people, regardless of their race, gender, age or ability. It also requires designers to take into account cultural diversity and social identity. Designers should respect different cultural backgrounds and values and avoid discrimination or neglect of other cultures.

From an environmental perspective, the ethical turn is manifested in environmental design, which requires designers to consider sustainability. Designers must select renewable materials, energy-saving equipment and environmental-protection technologies to reduce ecological consumption. Ethical thinking in environmental design also requires designers to respect the natural environment and ecosystems. Designers should follow ecological principles and protect natural resources and biodiversity in their designs.

From a human perspective, the ethical turn is manifested not only in its focus on the present but also in its consideration of the long-term impact of design on the future of humanity. It encourages humans to actively conceive ethical designs as designers rather than passively accepting ethical education. This is the result of rethinking the development of design based on the significance of science and individual initiatives, emphasizing human qualities and identities (Wang, 2024).

## 2.3 The value alignment of AI

The core purpose of AI playing a varied and significant role in modern society is to simulate, extend and expand human intelligence and to build intelligent systems capable of autonomous perception, learning and decision-making to assist or replace humans in completing specific tasks. In recent years, large language models have become increasingly mature, but they have also engendered a dual emergence of capabilities and risks. The biased, false, deceptive and manipulative content of AI has aroused

widespread attention and deep concern regarding its global governance (Zhang et al., 2023). To build a more responsible AI infrastructure, the top priority now is to infuse ethical, moral and cultural elements into the generation of intelligent systems, making them consistent or resonate with human values; that is, to research the value alignment of AI.

The value alignment of AI can be attempted through aspects such as language input, functional interaction and language communication (Leike et al., 2018). Based on programming language input, human-computer interaction is primarily achieved through instructions input into programming languages, and technicians enjoy the privilege of invoking computing resources (Gabriel, 2020). Based on functional interface interaction, although ordinary users can interact intuitively with intelligent machines through the front-end interface module, this kind of interaction is overly limited by specific interface design and standard operation norms (Chen et al., 2019). Based on natural language communication, the language models of AI, as a product of alignment technology, effectively reflect the interaction forms between human and computer languages. This includes ordinary users expressing their needs and intentions in a more natural and flexible way, and the corresponding system generating natural language to respond. Therefore, it can be seen that the essence of value alignment is to embed human values in intelligent technologies to ensure that technology is designed for good purposes, which is intrinsically consistent with the proposition of ‘embedding ethics in technology’.

### 3. The logical premises of embedding ethics in technology

The essence of embedding ethics in technology is to convert ethical values into design standards and achieve the translation from value to fact, which implies the reverse of the ‘Hume problem’. Therefore, the prerequisite for embedding ethical values in technical design is to respond reasonably to the limitations of the ‘Humean guillotine’. This necessitates an examination of the relationship between the embedding of ethics in technical design and the ‘Humean guillotine’ from the perspective of

meta-ethics, analysing the limitations of the ‘Humean guillotine’ and exploring the ways to overcome these limitations. This leads to the potential to embed ethics in technology.

#### 3.1 From the ‘Hume problem’ to the ‘Humean guillotine’

The ‘Hume problem’ originates from Hume’s examination of the foundations of ethics. From an empiricist standpoint, he opposed rationalism’s attempt to reduce ethical foundations to mere ‘reason’. However, due to the inherent limitations of empiricism, Hume’s epistemological scepticism also extended to the foundations of ethics, leading him to argue that one cannot derive ‘ought’ from ‘is’ (Hume, 1999: 508). Hume’s views and positions have been inherited and developed in meta-ethical schools such as intuitionism, emotivism and prescriptivism.

Intuitionism, emotivism and prescriptivism share the common feature of considering ethics to be independent of the realm of facts, denying the commensurability between facts and values. However, they fundamentally diverge on the ontological basis of ethics. Intuitionism belongs to ‘ethical realism’, positing that an objective reality exists behind ethical cognition and provides the criteria for judging the truth of ethical propositions. Unlike natural reality, ethical reality has its own unique character and cannot be reduced to natural facts; it can be perceived only through human intuition. Emotivism and prescriptivism, on the other hand, belong to ‘non-cognitivism’. They do not acknowledge the existence of an objective ethical reality behind ethical cognition, deny the cognitive functions of ethical judgement, and insist that ‘the function of ethical statements is not to convey knowledge but to express the speaker’s emotions and commands’ (Gan, 2015: 221).

Both intuitionism and non-cognitivist ethics face insurmountable difficulties in their analytical approaches and theoretical logic. The problem with intuitionism is that it cannot provide a standard or measure for the correctness of intuition itself; nor can it tell us how to test and control intuition among different subjects (Gan, 2015: 230). The foundation

of intuitionism is thus unstable. Non-cognitivism, on the other hand, denies that ethical statements possess any universal, truthful or objectively valid meaning, and it reduces ethical discussions to mere expressions and exchanges of emotions and opinions, thus completely severing the connection between ethics and rationality (Gan, 2015: 223). This is equivalent to admitting that ethical knowledge is entirely a reflection of subjective attitudes, which points in the direction of ethical relativism. Given the above, there is a need to break the ‘Humean guillotine’ and propose new pathways and solutions for bridging facts and values.

### 3.2 Attempts to break the ‘Humean guillotine’

Just as there are diverse schools of thought that uphold the ‘Humean guillotine’, those that oppose it are equally varied; they include naturalism, evolutionary ethics, pragmatism, neo-naturalism and neo-humanism. In his book *The Collapse of the Fact/Value Dichotomy and Other Essays*, Hilary Putnam provides a comprehensive analysis of the history of this dichotomy. He argues that the fact/value dichotomy shares roots with the analytic/synthetic dichotomy (Putnam, 2002). The collapse of the analytic/synthetic dichotomy also applies to the fact/value dichotomy. Both originated with Hume, and their commonality lies in the belief that one cannot derive higher order propositions from ‘fact’ statements—one manifests as ‘universal’ propositions, while the other as ‘value’ propositions. Together, they reveal the limitations of empiricism.

The collapse of the analytic/synthetic distinction offers valuable insights into the collapse of the fact/value dichotomy, as fact and value are also entangled. For example, the everyday term ‘cruel’ contains both descriptive and normative elements. Putnam (2002: 28) comments on this: ‘The example of the predicate “cruel” also suggests that the problem is not just that the empiricist’s (and later, the logical positivist’s) notion of a “fact” was much too narrow from the start. A deeper issue is that, from Hume on, empiricists—and many others as well, both in and outside the field of philosophy—failed to appreciate the ways in which factual

description and valuation can and must be entangled.’ To address this, Putnam (2004: 16) proposed the concept of ‘value facts’: values are fact-based values, while facts are value-loaded facts, or in his words, ‘Every fact is value loaded and every one of our values reflects some fact.’

### 3.3 A design approach to bridging facts and values

By reviewing past debates on the ‘Hume problem’ and various proposals for bridging facts and values, a two-step approach can be proposed. This approach begins with the concept of value and divides it into two dimensions: utilitarian value (or practical value) and ethical value (Gong, 2014).

1. Transition from ‘fact’ to ‘utilitarian value’. Utilitarian value refers to the pursuit and recognition of material or practical benefits in life. It can be understood as value in the context of the subject–object relationship. Value is not determined solely by the subject or the object but by the object’s ability to meet the subject’s needs and its significance, which depends on both the subject and the object. The attributes of the object are a ‘fact’, and the needs of the subject are also a ‘fact’. Value is therefore created in the space between these two facts. If the fact of the object’s attributes aligns with the fact of the subject’s needs, value is created; otherwise, no or negative value is produced. In other words, value exists in the alignment of facts. When this alignment is met, value exists; when it is not met, value is absent (Zhang, 2017). This is the first step in the process of bridging fact and value.
2. Transition from ‘utilitarian value’ to ‘ethical value’. Ethical value involves the pursuit and recognition of ethical ideals and can be understood as value in the context of the subject–subject relationship. Since utilitarian and ethical values are not entirely separate, ethical value cannot be achieved without utilitarian value. Without a foundation in practical benefits, ethical value becomes an

unstable, baseless construct. The principle of ‘unity of virtue and fortune’ emphasizes their consistency. However, they should also not be conflated. Reducing ethical value to utilitarian value in its entirety would strip it of its independence and sanctity, leading to a utilitarian society in which everything is judged by its usefulness and profit, ultimately degrading social ethos and morale. The two major ethical schools—utilitarianism and deontology—represent these opposing trends: the former conflating the two and the latter separating them entirely. Therefore, in addressing the transition from utilitarian to ethical value, a dialectical attitude that maintains both opposition and unity is required.

As a practical activity, design must begin with existing facts and aim to achieve a certain value. Design embodies the negative unity between humans and the world and is a practical way to unify fact with value and regularity with purpose. Thus, by establishing logical connections between facts and utilitarian value, and between utilitarian and ethical values, it is possible to translate ethical values into design standards, thereby answering the question of how ethics can be embedded in technology.

#### **4. Chinese and Western design approaches to embedding ethics in technology**

Western countries and China exhibit significant cultural differences with respect to their design approaches to embedding ethics in technology. The Western approach is mainly represented by modern design such as VSD, materializing morality, and RRI, while the Chinese approach is represented by traditional designs such as ‘embedding rituals in artifacts’ (Confucianism), ‘incorporating Dao into techniques’ (Taoism) and ‘integrating Chan Buddhism into craftsmanship’ (Buddhism). These two approaches have distinct features: the former emphasizes mind-based thinking, while the latter

focuses on heart-based thinking. Both approaches possess unique strengths and are complementary to one another.

##### **4.1 Western design approaches to embedding ethics in technology**

VSD, which focuses on embedding human ethical values into technology, was proposed by Batya Friedman and Peter H Kahn Jr at the University of Washington (Friedman and Kahn, 2002). VSD employs a tripartite methodology, including conceptual, empirical and technical analysis. Conceptual analysis aims to understand and clarify stakeholders in technology design and identify potential value conflicts among them. Empirical analysis uses qualitative and quantitative methods to assess how stakeholders perceive value in the aspects they consider important. Technical analysis explores how to draw on the insights of the other analyses and design a particular technology to support established values (Manders-Huits, 2011).

The application of Verbeek’s ‘materializing morality’ theory to design activities is reflected in a ‘predict–assess–design’ model. The primary goal of prediction is to understand the mediating role of technology. Since technology is multi-stable and lacks a fixed essence, what constitutes a technology depends on its context and users’ understanding. Therefore, it is necessary to establish a link between design context and use context (Verbeek, 2010: 97–98). The second step is the ethical assessment of mediation, which expands upon the process of stakeholder analysis. Traditional stakeholder analysis enables all stakeholders to articulate their perspectives on ethical issues and seek a comprehensive and balanced solution by weighing these diverse viewpoints (Verbeek, 2010: 106). The assessment process, however, integrates considerations of the mediating role of technology, thereby expanding the scope of stakeholder analysis beyond traditional risk analysis and disclosure. The third step is the design of technological mediation. The final outcome of design is a product of the interaction between users, designers and artifacts. Therefore, the design process is not about creating a new product category but about transforming

and improving the ‘script’ of existing products so as to enhance their function of ethical guidance (Verbeek, 2010: 117).

RRI integrates technological ethics, corporate social responsibility and technological innovation. It focuses on the major social and environmental issues arising from technological innovation and seeks solutions from ethical, responsible and policy perspectives (Wiarda and Van de Kaa, 2021). RRI emphasizes the future-oriented and prospective dimensions of responsibility, making it possible for technology design to consider its purpose and adapt to uncertainty. RRI includes four dimensions: anticipation, reflexivity, inclusion and responsiveness. The anticipatory dimension involves forecasting the potential impacts of technology. The reflexive dimension involves analysing and reflecting on one’s fundamental purposes and motivations, and the potential impacts of technology design activities. The inclusive dimension involves engaging the public and stakeholders through dialogue, participation and debate to solicit their opinions. The responsive dimension involves using collective reflexivity to engage in technology governance (Owen and Goldberg, 2010).

#### 4.2 Chinese design approaches to embedding ethics in technology

The Chinese design approach to embedding ethics in technology primarily refers to the approach rooted in traditional Chinese culture. It is rich in intellectual resources and highly complementary to the Western approach. However, it requires creative transformation to fully realize its potential in modern design activities. The traditional Chinese design approach mainly includes Confucian, Taoist and Buddhist perspectives.

The Confucian design approach is characterized by the principle of ‘embedding rituals in artifacts’. This principle involves designing and manufacturing objects following ritual requirements. It not only reflects the owner’s social status but, more importantly, reinforces the concept of ritual through objects, thereby fulfilling the ‘silent educational function’ embedded in the artifacts (Zhang, 2024). ‘Embedding rituals in artifacts’ is also known as

‘using artifacts to embody rituals’, ‘designing artifacts according to ritual norms’ or ‘integrating rituals into artifacts’. The ingenuity of this approach lies in integrating the ritual into the objects themselves. This is achieved by using methods such as observing rituals in form and sum, mirroring rituals with materials and craftsmanship, embedding rituals in decorative patterns, aligning functions with rituals, carrying rituals through inscriptions, and establishing rituals through spatial arrangement.

‘Form’ refers to the appearance and specifications of objects, while ‘sum’ refers to the required quantity and combination of objects. For example, *Ding* (tripod cauldrons) and *Gui* (food vessels), the two important ritual vessels of the Zhou Dynasty, observed strict ritual standards on their form and sum. The Emperor could use nine *Dings* and eight *Guís*; feudal lords could use seven *Dings* and six *Guís*; senior officials could use five *Dings* and four *Guís*; and scholarly officials could use three *Dings* and two *Guís*. ‘Mirroring rituals with materials and craftsmanship’ underscores the ethical implications of the choice of materials and craftsmanship. For example, jade objects, due to their rarity and exquisite craftsmanship, were viewed as symbols of a gentleman’s character and status. The *Book of Rites* records that jade has 11 virtues, making it a materialized symbol of a gentleman’s moral character. ‘Embedding rituals in decorative patterns’ conveys ethical principles using patterns. For example, dragon and phoenix motifs are not only decorative but also symbolize auspiciousness and good fortune. ‘Aligning functions with rituals’ means aligning the functions of objects with ritual procedures to reinforce the practical application of rituals. For example, the order of using wine vessels such as *Jue* and *Gu* mirrors the hierarchy of age and status, while the arrangement and performance of musical instruments such as *Bianzhong* (chime bells) and *Qing* (stone chimes) follow the ‘harmony and respect’ principles outlined in the *Book of Music*. ‘Carrying rituals through inscriptions’ uses inscriptions on objects to record ritual norms or moral teachings. For example, the inscription on the Western Zhou Dynasty’s *Mao Gong Ding* emphasizes the responsibilities of the sovereign and the subject. ‘Establishing rituals

through spatial arrangement' means designing the layout of objects to create a spatial order underpinned by ritual norms.

The Taoist concept of 'incorporating Dao into techniques' emphasizes guiding specific technical (techniques or arts) practices with the wisdom of Dao, integrating Dao into technical design, and using technical practice as a ladder with which to reach the realm of Dao. This approach mainly reflects the concept that technique is guided by Dao in technical activities through such processes as using technique to symbolize Dao, evolution from technique to Dao, and the mutual generation of Dao and technique. 'Using technique to symbolize Dao' refers to using the concrete technique to symbolize the abstract Dao, providing people with a concrete entry point through which to understand the abstract Dao. The profound and mysterious nature of Dao makes it difficult for ordinary people to grasp directly. However, by using a technique closely related to daily life as a metaphor, Dao can become more vivid and easier to understand. The 'evolution from technique to Dao' is achieved through long-term study and practice of the technique, ultimately grasping the essence of Dao. This concept emphasizes pursuing an understanding of the essential laws of things on the basis of mastering specific skills; that is, it expresses the process of how to cultivate Dao. The concept of 'mutual generation of Dao and technique' is to lead Dao towards its ultimate target form; that is, to demonstrate the interdependence and mutual promotional relationship between Dao and techniques or arts.

The Buddhist approach of 'integrating Chan Buddhism into craftsmanship' introduces the concepts and aesthetic principles of Chan Buddhism into the field of technological design to elevate the realm of craftsmanship or innovate the application of technology. This can be achieved by using methods such as the observing technique, listening technique, smelling technique, tasting technique, touching technique and enlightenment technique. The observing technique involves exploring the various aspects of a technology through visual observation, including the shapes, colors, sizes and structures of products, to gain a more comprehensive understanding of it. The listening technique

entails extracting key auditory features, such as sound frequencies, timbre, pitch and tempo, from products to reveal the emotional expressions of the technology. The smelling and tasting techniques refer to identifying the characteristics of products through the senses of smell and taste, in search of healthier and more environmentally friendly technological products. The touching technique involves feeling the form and texture of products through tactile means to achieve more harmonious human-technology interaction. The enlightenment technique integrates the above five senses to intuitively grasp the principles of the Chan Buddhism and interests embedded in technology, thereby enhancing the design aesthetics and ethical profile of technological applications. As stated in the *Treatise on the Nameless in Nirvana* [涅槃无名论], the profound way lies in wonderful enlightenment, and wonderful enlightenment lies in immediacy and authenticity. 'Wonderful enlightenment' refers to the designer's intuitive understanding of the true nature of technology and the environment in which it is used.

## 5. Mutual learning between Chinese and Western design approaches to embedding ethics in technology

The Chinese and Western design approaches to embedding ethics each have their own unique features. The Western approach primarily features 'one-on-one' embedding, focusing on comparing the ethical differences and impacts among similar entities. It enhances and innovates ethical attributes based on an understanding of the scale, quantity, and pros and cons of similar entities, presenting the expansionary pattern of ethics development; that is, the horizontal development of embedding results. Grounded in this approach, the Western process of embedding ethics emphasizes logical analysis, the construction of rational knowledge systems, and the establishment of technical standards. It reflects a predominantly brain-based way of thinking. In contrast, the Chinese approach emphasizes comparing entities across different

historical stages, understanding the inherent laws of their development, and recognizing their developmental phases and trends. This approach enables a comprehensive grasp of entities, presenting the evolutionary pattern of embedding results. Because of this, the Chinese process of embedding ethics focuses on the integrity, organic structure and experiential aspects of design and reflects a predominantly heart-based way of thinking.

According to modern neuroscience, the human brain is primarily composed of the left and right hemispheres (also known as the left and right brains) and the limbic system. Brain-based thinking is typically dominated by the left brain (with the right brain and limbic system playing supportive and coordinating roles), while heart-based thinking is dominated by the right brain and limbic system (with the left brain playing a supportive and coordinating role). The right brain mainly governs activities such as visual thinking, imagination, intuition and creativity (Blakeslee, 1980), while the limbic system's functions involve complex processes related to visceral activities, bodily movements and emotions, integrating the entire body's experiences into the brain (Meng, 1989: 4–16). As such, heart-based thinking also encompasses emotional, intuitive and experiential elements, which are beyond the scope of brain-based thinking.

The distinct characteristics of brain-based and heart-based thinking give rise to differences such as logic versus experience, tangible versus intangible, and macro versus micro perspectives.

First, in terms of the respective strengths of logic and experience, the brain-based Western design approach excels in applying logical design thinking. Logical design thinking focuses on connecting and organizing design elements to form a relatively complete system. For instance, the RRI approach transforms responsibility from external to internal and from individual to collective through its four dimensions of anticipation, reflexivity, inclusion and responsiveness. In contrast, the heart-based Chinese design approach excels in applying experiential design thinking. Experiential design thinking emphasizes the process of personal experience. The experience gained from designers' hands-on practice differs from theoretical research, allowing the designer to

resonate with the objects and enhance the emotional connections and interactions between humans and technology, and among humans themselves. The direct engagement provided by experience also facilitates a deeper understanding of the essence of things, generating lasting impressions and memories.

Second, on the respective strengths of the tangible and intangible, the guidance function of the Western design approach is primarily achieved through tangible materialization, under which the ethical guidance of technology is provided mainly in the form of physical objects. For example, speed bumps make drivers slow down through their psychological influence. Speed bumps may reduce drivers' sense of safety and comfort, prompting them to choose a lower speed. In contrast, the Chinese design approach tends to guide ethics in a more suggestive way, invisibly integrating ethics into human consciousness. This approach not only increases the requirements on materialization but also calls for more consideration of human factors in artifact design—a method also called empathic design. Compared to humans passively receiving guidance from artifacts, intangible materialization creates a more friendly and comfortable relationship in human–technology interactions. For example, in ancient times, the design of bows and arrows had to take into account the user's temperament. A person with a fierce temper should choose a softer bow, as it could help them curb their impatient tendencies.

Finally, on the respective strengths of macro and micro perspectives, the Western design approaches, such as VSD, materializing morality and RRI, focus on realizing ethical attributes like value, morality and responsibility in human–technology interaction. As a result, the Western design approaches tend to focus on the technology itself or the interaction between humans and technology, emphasizing individualism and individual rights, which can all be seen as micro dimensions of design. In contrast, the Chinese design approaches, including 'embedding rituals in artifacts', 'governing techniques with Dao' and 'integrating Chan Buddhism into technology', use design to reinforce the concepts of ritual, Dao and Chan, thus fulfilling the educational functions inherent in Confucianism, Buddhism and Taoism. Education is not simply about knowledge

transmission but also represents an intangible influence of cultural institutions, such as customs, ethics and morality. Although they do not hold legal force, they have a significant impact on social behavior. This is also an indication of the institutional design of the Chinese approach, which tends to adopt a macro perspective.

In summary, the differences between the logical and experiential approaches to design, between the tangible and intangible forms of embedding, and between the micro and macro design perspectives showcase the respective strengths of Chinese and Western design approaches. Separating the two would create limitations. Intangible materialization is not suitable for all designers. For individuals or groups who do not listen to advice, feel reluctant to accept others' opinions, or even ignore others on purpose, intangible materialization could be powerless. Intangible materialization might not work effectively for those with weak perceptive abilities, either. Therefore, it is necessary to form a pattern of mutual learning between Chinese and Western approaches to develop a more comprehensive and in-depth design approach for embedding ethics in technology.

## 6. The integration of Chinese and Western design approaches: Heart-brain collaborative design

Heart-brain collaborative design leverages the strengths of both Chinese and Western design approaches, combining the cognitive model and methods of heart-based thinking with brain-based modern design theories. It stresses the importance of design principles such as harmony with nature, governing techniques with Dao, deep experience, and ingenious craftsmanship.

The principle of 'harmony with nature' means respecting the natural essence of things in technological design. Activities that violate the natural essence of things will ultimately harm oneself or others through various pathways. Shortening the natural processes of technological activities will inevitably lead to abnormalities that go against Dao. Designs that are in harmony with nature must try

their best to use raw materials derived from nature and ensure that these materials return to nature after use, thereby achieving unity between creation and decomposition (Wang, 2011: 143–145).

The principle of 'governing techniques with Dao' emphasizes the general societal effects of technological practices. It fully considers the harmony of various elements related to technological activities in advance and promptly identifies and eliminates any disharmonious relationships, allowing technological activities to develop reasonably within the realm of human control.

The principle of 'deep experience' demands a thorough understanding of the societal impacts of technological structures and functions, as well as a 'heart-based' approach to uncover the organic connections that are often overlooked by 'brain-based' design thinking. This is reflected in the design process that delves into details, appreciates the contexts experienced by the manufacturers, users and other stakeholders of technological products, and gives holistic consideration to the ethical consequences of embedding ethics in technology.

The principle of 'ingenious craftsmanship' refers to creativity that is unique and avoids the trap of mere craftsmanship. The success of ingenious craftsmanship can bring 'harmony with nature' to its ultimate level, make 'governing techniques with the Dao' invisible, and seamlessly connect 'deep experience' with technological innovation. These four basic principles are conceptually interconnected, reflecting the possibility of embedding their respective ethical attributes into technological innovation activities from different perspectives.

Heart-brain collaborative design consciously guides the conception and operation of technological design through four stages: analysis, examination, conceptualization and feedback. In this context, the process of embedding ethical elements in design activities must first determine which ethical elements to embed. This involves leveraging the intuitive experiential strengths of heart-based thinking to identify the design-related elements that are truly felt by the designers, as well as the functional attributes and ethical factors within the relationships of these elements, and present them with appropriate images. On this basis, images are established by observing the

objects, and vessels are created by emulating the images (Zhu, 2022: 121–122).

The ethical elements identified during design activities are examined by placing their images within the context of design practices. The focus is on examining their relationship with the designer's original values and psychological state and making necessary adjustments. The conceptualization of the heart–brain collaborative design approach must be completed in conjunction with adjustments to the structure and function of the product, which are both independent and interdependent. It is therefore necessary to ensure that the embedding of ethical elements is technically sound in structure and reliable in function, to meet the requirement of 'governing techniques with Dao'. Feedback on the effectiveness of heart–brain collaborative design requires designers to have a stronger sensitivity to values and a greater sense of social responsibility. They should conscientiously listen to opinions from all sides and leverage advanced information-management systems to establish feedback and evaluation mechanisms, thereby ensuring tangible results from the efforts to embed ethics in technology.

To summarize, in practical applications, the concept of heart–brain collaborative design could meet the requirements for embedding ethics in technology more comprehensively and profoundly. This approach not only fills the gap in the prevailing design approaches to embedding ethics in technology but also contributes to the creative transformation and innovative development of traditional Chinese culture. Of course, to promote the development of heart–brain collaborative design, it is essential to encourage the participation of more technical ethicists and relevant interdisciplinary personnel, further strengthen cultural confidence in modern design activities, and boost the global influence of the 'Designed by China' discourse system.

#### ORCID iD

Yu Jing  <https://orcid.org/0009-0008-7216-6786>

#### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication

of this article: This work was supported by the Ministry of Education Humanities and Social Sciences Research Planning Fund Project 'Research on Ethical Embedment Mechanism of Technology Design', (grant number 24YJA720012).

#### Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

#### References

- Blakeslee TR (1980) *The Right Brain: A New Understanding of the Unconscious Mind and Its Creative Powers*. New York: Anchor Press.
- Cao JD (2013) *Analysis on Don Ihde's Philosophy of Technology* [伊德技术哲学解析]. Shenyang: Northeastern University Press (in Chinese).
- Chen C, Zhu QQ, Yan R, et al. (2019) Survey on deep learning based open domain dialogue system [基于深度学习的开放领域对话系统研究综述]. *Chinese Journal of Computers* 42(7): 1439–1466 (in Chinese).
- Chen F and Cao JD (2004) Phenomenological perspective on technology: Review of Ihde's phenomenology of technology [现象学视野中的技术：伊德技术现象学评析]. *Studies in Dialectics of Nature* 5: 57–61 (in Chinese).
- Friedman B and Kahn P (2002) Value sensitive design: Theory and methods. University of Washington Technical Report 2: 12.
- Gabriel I (2020) Artificial intelligence, values, and alignment. *Minds and Machines* 30(3): 411–437.
- Gan SP (2015) *Contemporary Constructions of Ethics* [伦理学的当代建构]. Beijing: China Development Press (in Chinese).
- Gong Q (2014) Discussion of moral and utilitarian values [论道德价值与功利价值]. *Philosophical Trends* 8: 66 (in Chinese).
- Hume D (1999) *A Treatise of Human Nature*. San Francisco: Clarendon Press.
- Leike J, Krueger D, Everitt T, et al. (2018) Scalable agent alignment via reward modeling: A research direction. Available at: <https://arxiv.org/abs/1811.07871> (accessed 6 August 2025).
- Latour B (1992) Where are the missing masses? The sociology of a few mundane artifacts. In: Bijker WE and

- Law J (eds) *Shaping Technology/Building Society: Studies in Sociotechnical Change*. Cambridge, MA: MIT Press, 225–258.
- Liu Z (2017) Can technical artifacts be moral agents? Verbeek's theory of technological moralization and its internal predicament [技术物是道德行动者么? 维贝克“技术道德化”思想及其内在困境]. *Journal of Northeastern University (Social Science)* 19(3): 221–226 (in Chinese).
- Manders-Huits N (2011) What values in design? The challenge of incorporating moral values into design. *Science and Engineering Ethics* 17(2): 271–287.
- Meng SL (1989) *Human Emotion* [人类情绪]. Shanghai: Shanghai People's Publishing House (in Chinese).
- Owen R and Goldberg N (2010) Responsible innovation: A pilot study with the UK engineering and physical sciences research council. *Risk Analysis* 30: 1699–1707.
- Putnam H (2002) *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge: Harvard University Press.
- Putnam H (2004) *Reason, Truth and History*. Cambridge: Cambridge University Press.
- Verbeek P (2010) *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago and London: The University of Chicago Press.
- Wang JM (2024) Cultivating scientific citizens for the 'anthropocene era': Design logic and enlightenment of the PISA 2025 science framework [为“人类世时代”培养科学公民: PISA2025科学测评框架的设计逻辑及启示]. *Science and Society* 14(2): 98–115 (in Chinese).
- Wang Q (2009) *Between Dao and Technique: The Philosophy of Technology in Chinese Cultural Contexts* [“道”“技”之间: 中国文化背景的技术哲学]. Beijing: People's Publishing House (in Chinese).
- Wang Q (2011) *General Ethics of Technology* [技术伦理通论]. Beijing: Renmin University of China Press (in Chinese).
- Wiarda M and Van de Kaa G (2021) A comprehensive appraisal of responsible research and innovation: From roots to leaves. *Technological Forecasting and Social Change* 172: 53–121.
- Yang QF (2015) *Soaring Albatross: A Study of Don Ihde's Phenomenology of Technology* [翱翔的信天翁: 唐·伊德技术现象学研究]. Beijing: Chinese Academy of Social Sciences Press (in Chinese).
- Zhang W (2017) The Hume problem in technology [技术中的“休谟问题”]. *Journal of Changsha University of Science and Technology (Social Science)* 32(4): 6–11 (in Chinese).
- Zhang W (2024) An ethical investigation of 'hiding rite in artifacts' in Confucianism and its modern significance [儒家“藏礼于器”思想的伦理审视及当代启示]. *Journal of East China Normal University (Humanities and Social Sciences)* 56(1): 17–23, 175–176 (in Chinese).
- Zhang Y, Li YF, Cui LY, et al. (2023) Siren's song in the AI ocean: A survey on hallucination in large language models. Available at: <https://arxiv.org/abs/2309.01219> (accessed 5 August 2025).
- Zhu ZR (2022) Creating images by observing and trimming objectives in image creation [意象创构中的观物取象]. *Literary Review* 2: 41–49 (in Chinese).

### Author biographies

**Wei Zhang** is a professor and a doctoral supervisor of philosophy at the School of Humanities of Dalian University of Technology. He is also the executive director of the Professional Committee of Science and Technology Ethics of China Ethics Society, the director of the Professional Committee of Science, Technology and Engineering Ethics of the Chinese Society for Dialectics of Nature, and the director of the society's Scientific and Cultural Committee and Technical Philosophy Committee. His main research interests include the philosophy of technology and the ethics of science and technology.

**Yu Jing** is a PhD candidate at the School of Humanities of Dalian University of Technology. His main research interests include design ethics, the philosophy of technology and the ethics of technology.

**Qian Wang** is a professor and a doctoral supervisor of philosophy at the School of Humanities of Dalian University of Technology. He used to be the deputy dean of the School of Scholar and Literature of Dalian University of Technology, the deputy director of the Academic Committee of Dalian University of Technology, and the director of the

Special Committee of Philosophy and Social Sciences. He is the executive director of the Chinese Society for Dialectics of Nature, and the director of the society's Professional Committee of Science, Technology and Engineering Ethics. He is also a member of the Advisory Committee of the Science Planning and Ethics Research Support

Center of the Faculty of Chinese Academy of Sciences, and a member of the Special Expert Group of Science and Technology Ethics Education of the Ministry of Education. For many years, he has devoted himself to the research of science and technology ethics, technology philosophy and organism philosophy.

# Event, society, and future: Revisiting Chinese public discourse on the ‘gene-edited babies incident’

Cultures of Science  
2026, Vol. 9(1) 62–82  
© The Author(s) 2026  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20966083261427864  
journals.sagepub.com/home/cul



Shuo Wang<sup>1</sup>  and Zhengfeng Li<sup>1</sup>

## Abstract

The 2018 ‘gene-edited babies incident’ in China catalysed intense global debate, highlighting the critical need to understand public discourse surrounding major techno-ethical controversies. This study investigates the characteristics of Chinese public discourse in response to this event by analysing 3692 comments from the social-media platform Weibo. Employing a mixed-methods approach combining latent Dirichlet allocation topic clustering and grounded theory, the research reveals a hierarchical thematic structure in public discussions. Such discussions are categorized into three types of discourse: (1) event-specific discourse, focusing on the incident itself, the individuals involved, subjects’ rights, the experimenter’s ethics and the immediate ethical risks of the technology; (2) societal-context discourse, extending to broader societal issues reflected by the event, including techno-governance, the rule of law, international competition, science communication and popular-science education; and (3) future-oriented discourse, which adopted a macro-perspective to explore the necessity and historicity of technological development and its relationship with ethical constraints. While demonstrating diverse engagement, the analysis also identified prevalent issues in public cognition, such as superficial scientific understanding, pronounced emotional responses and a reliance on monolithic cultural logics. These findings underscore the importance of strengthening research on public techno-ethical cognition. The study advocates for the integration of public perspectives into techno-ethical governance frameworks to complement elite, unidirectional communication, thereby enhancing the inclusiveness, responsiveness and practical effectiveness of science governance in the face of rapid technological advances.

## Keywords

Gene-edited babies incident, public discourse, scientific ethics, science-related populism, science communication, techno-ethics, trust in science, trust in experts

## 1. Introduction

In November 2018, He Jiankui, then an associate professor at the Southern University of Science and Technology,<sup>1</sup> publicly announced that twin girls, Lulu and Nana, had been born healthy in China following human embryo gene editing. Specifically, He

<sup>1</sup>School of Social Sciences, Tsinghua University, China

### Corresponding author:

Zhengfeng Li, School of Social Sciences, Tsinghua University, Haidian District, Beijing 100084, China.  
Email: lizhf@tsinghua.edu.cn



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

claimed to have used CRISPR-Cas9, a powerful genome-editing tool that enables precise modifications to DNA sequences, to disable the CCR5 gene in the embryos, with the stated aim of conferring innate resistance to HIV infection. This news sparked widespread controversy worldwide, drawing widespread concern and intense discussion among all sectors of society regarding the ethical implications and potential societal impacts of this technology (Doudna, 2019). The act was almost unanimously condemned by the scientific community, both domestically and internationally, as a severe violation of research ethics and academic integrity (Lander et al., 2019).<sup>2</sup> Following the public announcement, Chinese government bodies quickly launched investigations and stated that the matter would be handled according to law and regulations. These bodies included the National Health Commission (China's top health policy and regulatory authority, responsible for overseeing medical practice and research ethics), the Ministry of Science and Technology (the central government agency governing science and technology policy, research funding and research integrity) and the China Association for Science and Technology (a national organization serving as a bridge between the government and the scientific community, with responsibilities in science popularization and professional ethics).

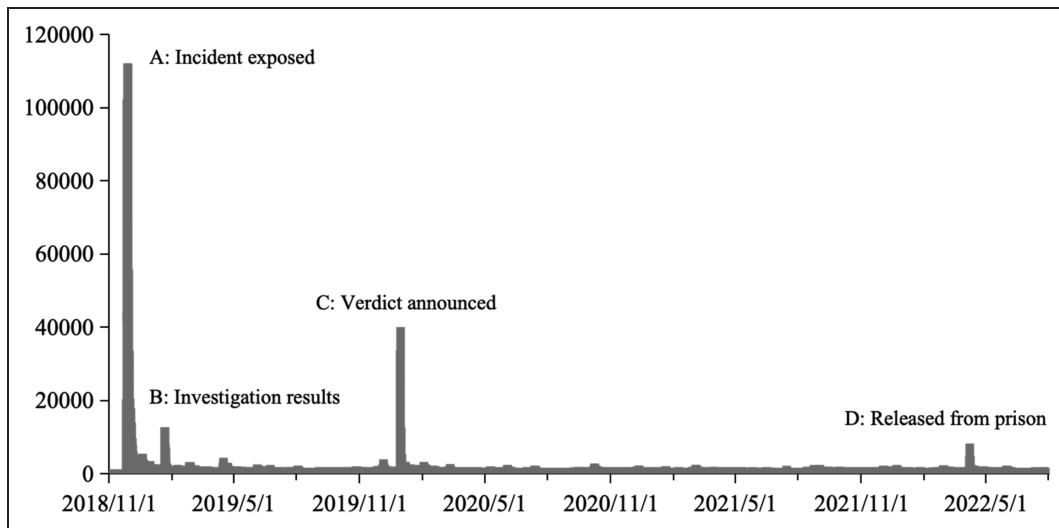
After an investigation, Guangdong Province announced its preliminary findings in January 2019. The investigation was led at the provincial level because He Jiankui's experiment was conducted in Shenzhen, which falls under the administrative jurisdiction of Guangdong Province; under China's governance structure, such investigations are typically initiated and managed by the relevant local authorities before findings are reported to central government bodies. The investigation concluded that 'He Jiankui, in pursuit of personal fame and fortune, self-funded, deliberately evaded supervision, and privately organized personnel to carry out human embryo gene-editing activities for reproductive purposes, which are explicitly prohibited by the state.' Ultimately, He Jiankui and his collaborators were convicted of illegal medical practice in December 2019.<sup>3</sup>

This incident was not only a major ethical case in biomedical research but also a typical techno-ethical

controversy that triggered large-scale online public discussion and debate. An analysis of Baidu Index<sup>4</sup> search trends for the keyword 'He Jiankui' (see Figure 1) reveals that public attention to this incident was concentrated around four key time points: after the initial exposure of the event on 26 November 2018, the search index peaked on 28 November (Point A); on 21 January 2019, when Guangdong Province announced the preliminary investigation results, the index reached a second peak (Point B); on 30 December 2019, at the commencement of the public trial, online attention reached a third high point (C); and even in April 2022, when He Jiankui was released from prison, public opinion saw another small peak (Point D). This sustained and concentrated high level of public attention, driven by specific developments, confirms the status of the 'gene-edited babies incident' as a landmark techno-ethical public issue.

The intense public reaction to the 'gene-edited babies incident' highlights the importance of public understanding of and effective participation in techno-ethical governance. Techno-ethical governance requires not only effective government regulation and self-discipline from the scientific community but also, indispensably, the full understanding and active involvement of the public. Public discourse and cognition regarding major techno-ethical events, expressed through online forums, reflect the diversity of people's techno-ethical views and provide a crucial empirical basis for research and practice in techno-ethical governance. However, even as techno-ethical research gains increasing prominence today, public techno-ethical cognition, especially its discursive practices in specific controversial events, is often overlooked. Traditional, mainstream research tends to either focus on metaphysical explorations from ethical and philosophical perspectives or to emphasize government regulation and expert governance from legal and public administration standpoints, rarely analysing the public's techno-ethical conceptions and specific expressions during controversial incidents.

This relative neglect of public discourse in this field is rooted in long-held, stereotyped perceptions of the public's role. First, members of the public are



**Figure 1.** Baidu Index trend for 'He Jianhui'.

typically seen as lacking substantive channels and rights to participate in technological affairs, and are more often viewed as 'passive consumers' of technological outcomes. Second, the public is commonly perceived as lacking the necessary scientific literacy to engage in complex, rational discussions of technology governance and policy (Wynne, 2006). However, high-impact and controversial events, such as the 'gene-edited babies incident', provide opportunities for public participation in technological discussions, allowing individuals to express their ethical views and scientific literacy, and potentially enhancing them through public debate. Ignoring the public's authentic voice in these events can not only lead to a disconnection between techno-ethical governance and societal expectations but may even trigger a crisis of trust within the governance community. Therefore, re-examining the public's agency in techno-ethical events, thoroughly exploring people's discursive practices surrounding specific controversies, and investigating their cognitions, understandings, attitudes and viewpoints are urgent theoretical and practical endeavours to aid understanding of technology–society interactions, bridge cognitive gaps and enhance the inclusiveness and effectiveness of techno-ethical governance.

Based on the above considerations, this paper focuses on the landmark techno-ethical controversy of the 'gene-edited babies incident', using public comments from Sina Weibo, one of China's largest social-media platforms, as research data. By employing natural language processing techniques and grounded theory methodology, this study conducts an in-depth mining and systematic analysis of these large-scale online text data. It aims to reveal the characteristics of online public discourse, core issues, emotional tendencies, and the diversity of the public's ethical stances expressed during the 'gene-edited babies incident'. The goal of this empirical investigation of public discourse in this specific event is to provide empirical evidence for understanding how the contemporary Chinese public perceives and participates in major techno-ethical controversies, and to offer theoretical references and practical insights for improving relevant techno-ethical governance frameworks and enhancing the effectiveness of future techno-risk communication.

## 2. Literature review

Prior to the 'gene-edited babies incident', both public and expert discussions on human gene editing

were already characterized by a complex interplay of hope and apprehension. Public acceptance of gene-editing technologies was typically conditional, highly dependent on their intended application—therapeutic uses (to treat or prevent disease) generally received higher public support than enhancement applications (to alter human traits beyond medical necessity) (Funk and Hefferon, 2018; Gaskell et al., 2017). Heritable germline gene editing, in particular, became a focal point for public prudence and ethical concern due to its potential long-term impact on the human gene pool and the ethical status of embryos (Hendriks et al., 2018). Factors such as scientific literacy, religious beliefs, trust in scientists and media portrayals were all considered to be variables influencing public attitudes. Concurrently, while authoritative bodies like the National Academies of Sciences, Engineering, and Medicine and the Nuffield Council on Bioethics emphasized ongoing public engagement and robust governance, their expert-driven deliberation frameworks showed certain limitations in foreseeing and responding to the real-world shock of breakthrough events such as He Jiankui's announcement.

The 'gene-edited babies incident' undoubtedly served as a potent catalyst for global public discussion, with social media rapidly becoming the primary medium for these conversations. Unlike traditional expert deliberations, social media provided an unmediated space in which diverse voices could react instantaneously, share opinions and engage in debate. Cross-platform analyses (e.g., of Twitter, Weibo, Reddit and YouTube) regarding this event revealed a complexity in public sentiment that diverged from the near-unanimous condemnation of the scientific and bioethical elites. Although expert opinion was overwhelmingly negative, a considerable number of online comments expressed support for He Jiankui's experiment, often grounded in an optimistic view of the future therapeutic potential of gene editing, forming a 'discourse gap' with opposing voices rooted in ethical and safety concerns. Different social-media platforms also exhibited distinct discursive ecologies. For instance, supportive voices were more

prominent on YouTube and Reddit, while opposition was more concentrated on Twitter and Weibo, and the proportion of support on Weibo was higher than that on Twitter (Ni et al., 2022). Comparative research by Ji et al. (2022) indicated that discussions on Weibo appeared to be less extensive in scope and duration than those on Twitter, probably influenced by differences in scientific literacy between China and the West, cultural tendencies and platform-specific characteristics. These findings collectively demonstrate that public discourse is not monolithic but is dynamically shaped by platform affordances, cultural contexts and the specifics of unfolding events.

Despite the surge in attention to human gene editing triggered by the 'gene-edited babies incident', existing literature still presents several key limitations. First, early analyses predominantly focused on the perspectives of scientific experts, bioethicists and policymakers. While social-media analyses have begun to capture public voices, a deeper qualitative understanding of the reasoning, values and everyday ethics underpinning these diverse public opinions is still needed, moving beyond mere sentiment analysis or topic modeling. Second, there is a geographical and cultural imbalance in current research. The 'gene-edited babies incident' occurred in China, yet detailed, culturally sensitive analyses of Chinese domestic public discourse (especially the complex interactions on key platforms such as Weibo) remain underdeveloped.

Therefore, research in this field urgently needs to fill the cognitive gap in how the publics in various national and cultural contexts (especially in major non-Western countries such as China) understand, participate in and shape the discourse on major techno-ethical controversies. Specifically, there is a lack of systematic, in-depth, mixed-methods analysis of the unique discursive patterns, underlying cultural values, influence of national narratives and trust dynamics exhibited by the Chinese public on social-media platforms in the specific case of the 'gene-edited babies incident'. By conducting an in-depth qualitative and quantitative analysis of Weibo discourse surrounding the incident, this study aims to provide empirical evidence for

understanding how the contemporary Chinese public confronts and shapes major scientific and technological ethical challenges. This research focuses on the interplay of Chinese cultural values, official narratives and trust in shaping public arguments, seeking to reveal its discursive patterns and ethical considerations.

### 3. Methodology and analysis

#### 3.1 Data source and processing

Social-media data, which encompasses rich user interaction data, provides a systematic lens for understanding aggregated public expressions regarding major events (Yang et al., 2021). As China's largest social-media platform, Sina Weibo (hereafter Weibo) facilitates frequent open and critical debates on high-stakes topics such as technological development, environmental pollution and food safety (Rauchfleisch and Schäfer, 2015). Recent scholarship has increasingly used big-data methods to analyse massive Weibo datasets, exploring public attitudes towards both environmental crises (Pu et al., 2022) and techno-ethical controversies (Zhang et al., 2021). To systematically mine public commentary, this study targeted the comment sections of 10 influential posts published by three primary official media outlets: *People's Daily*, Xinhua News Agency and CCTV News. These three outlets were selected because they constitute the most authoritative central-level state media organizations in China, collectively commanding the largest audiences and highest levels of institutional credibility on Weibo. Their posts on major public events typically attract the highest volumes of public commentary, making their comment sections a rich site for observing large-scale public discourse.

Data collection was executed through a custom Python web crawler, with the harvesting process concluding in November 2022. This extensive time frame was strategically selected to encompass the full chronological arc of the 'gene-edited babies incident', effectively capturing public responses to every critical milestone. The dataset includes reflections on the initial disclosure in late 2018,

the announcement of preliminary investigation results and judicial sentencing in 2019, and the resurgence of public attention following the experimenter's release from prison in April 2022. This longitudinal approach ensures that the empirical evidence reflects a matured and evolving public cognition rather than transient emotional reactions. A total of 11,216 direct comments were ultimately obtained, each documented with metadata including the commenter ID, textual content, time stamp and number of likes, with all entries numbered sequentially to facilitate systematic citation (see Table 1).

After data acquisition, the data were cleaned and processed. First, forwarded comments, blank comments, emoticon-only comments, colloquialisms, repetitive topic tags and emojis were removed. Traditional Chinese characters were converted to simplified characters, and comments completely unrelated to the event were manually screened and deleted. Then, data filtering was performed: comments were sorted in descending order by word count, and only those with 15 or more characters were retained. Finally, 3692 comment texts were included in the analysis.

#### 3.2 Methodological framework

Current research on text data processing and analysis mainly follows two paths: traditional qualitative research and machine learning paths based on code and algorithms. This study combines them, first using machine learning methods for an initial thematic exploration through topic clustering, and then employing qualitative analysis using grounded theory for in-depth coding and interpretation to achieve a comprehensive mining of the text data.

For the initial thematic exploration, latent Dirichlet allocation (LDA) was selected. LDA is a probabilistic model that can extract thematic information from large-scale texts by identifying latent topic distributions within a document collection (Blei et al., 2003) and has been widely applied in social-media text analysis and topic clustering (Jelodar et al., 2019).

**Table 1.** Overview of data sources.

ID	Media	Date	Title	Data volume
RM01	<i>People's Daily</i>	2018/11/26	Gene-edited babies questioned by academia and industry: Would you dare drive a car without brakes?	5510
RM02	<i>People's Daily</i>	2018/11/26	SUSTech responds to gene-edited babies: Unaware; Shenzhen Health Commission initiates investigation	1111
RM03	<i>People's Daily</i>	2019/1/21	Guangdong preliminarily ascertains 'gene-edited babies incident'	60
RM04	<i>People's Daily</i>	2019/12/30	He Jiankui and two other defendants held criminally responsible	520
XH01	<i>Xinhua News</i>	2018/11/29	Officials from National Health Commission, Ministry of Science and Technology, China Association for Science and Technology respond to 'gene-edited babies incident'	578
XH02	<i>Xinhua News</i>	2019/1/21	Guangdong preliminarily ascertains 'gene-edited babies incident'	80
XH03	<i>Xinhua News</i>	2019/12/30	He Jiankui and two other defendants held criminally responsible	31
YS01	<i>CCTV News</i>	2018/11/28	Controversial scholar He Jiankui appears at international conference	483
YS02	<i>CCTV News</i>	2018/11/26	Focus on 'gene-edited babies incident': Not reported to Health Commission! He Jiankui on unpaid leave since February! 122 scientists jointly condemn	599
YS03	<i>CCTV News</i>	2019/12/30	Breaking! He Jiankui and two other defendants held criminally responsible	2244

Subsequently, for in-depth coding and developing a more nuanced understanding of the data, grounded theory was applied. Grounded theory is a qualitative research method that systematically collects and analyses data to refine core concepts and construct connections between them, thereby developing theories grounded in the collected data. This typically involves iterative stages of coding, such as open, axial and selective coding.

This study adopts a descriptive and analytical rather than prescriptive stance: it refrains from judging the scientific validity or ethical correctness of quoted public comments. Aside from obscuring a small amount of uncivil language, the original text and intent of the comments are fully preserved. The purpose of this approach is to authentically capture and analyse the cognitive characteristics, reasoning patterns and value orientations embedded in public expression, rather than to endorse or validate any particular viewpoint. Where public comments contain factual inaccuracies or ethically problematic propositions, these are addressed in

the analytical discussion rather than editorially corrected within the data presentation.

### 3.3 LDA-based topic clustering: Application and results

In this study, an LDA topic model was constructed using the Sklearn package in Python to perform topic clustering on the comment texts, with initial word segmentation carried out using Python's Jieba library. Existing research typically determines the optimal number of topic clusters based on perplexity. Perplexity can be understood as the uncertainty of a trained model about whether a document belongs to a given topic; theoretically, the lower the perplexity, the higher the model's prediction accuracy.

The number of LDA topic clusters was set to increase stepwise from 1 to 7, and the perplexity values corresponding to each cluster number were outputted. It was found that when the number of

**Table 2.** LDA topic clustering results and summary.

No.	High-frequency words related to clustering	Focus point	Feature summary
1	Child, baby, experiment, parents, kid, reproduction, innocent, living, being/life, trial	Event itself, individuals involved	Event-specific discourse
2	Experiment, science, country, bottom line, law, technology, scientific research, life, human nature, breakthrough	Politics, law, ethics, culture	Societal-context discourse
3	Gene, human, editing, research, technology, world, scientist, science & technology, development, GMO	Humanity, technology, future	Future-oriented discourse

topics was 3, the model's clustering effect was relatively good. Therefore, the optimal number of topics was set to 3, and relevant clustering indicators and results were outputted.

Based on LDA topic clustering, the comment texts were divided into three categories (see Table 2). The first category, termed 'event-specific discourse', featured high-frequency words mainly related to the event itself and the individuals involved, with common keywords such as 'child', 'baby', 'experiment' and 'parents'. This type of discourse focuses on the specifics of the incident, the circumstances and the consequences for those directly involved. The second category, termed 'societal-context discourse', expanded the discussion to broader social issues, including typical keywords such as 'country', 'bottom line' and 'law'. This category reflects public attention to the wider political, legal, ethical and cultural issues stemming from the event. The third category, termed 'future-oriented discourse', contained high-frequency words that pointed more towards technology and the future of humanity—for example, 'gene', 'humanity' and 'world'—indicating public reflection on the long-term development of science and technology and its impact on humanity's future. Thus, public cognition and understanding of the 'gene-edited babies incident', as revealed through these thematic clusters, can be summarized across these three distinct discursive levels.

### 3.4 Grounded theory-based coding analysis: Application and system development

To investigate deeper into the specific content of each LDA-identified topic, this study adopted a

three-level coding method based on the grounded theory paradigm and systematically applied it to the 3692 comment texts. Using NVivo12 software, texts were coded, summarized and refined item by item, and the process was combined with LDA topic clustering results to analyse public understanding of the techno-ethical event from multiple dimensions.

Specifically, open coding was first performed, summarizing the original texts sentence by sentence to extract 156 initial concepts. After subsequent screening and integration, 30 third-level categories were ultimately retained. Axial coding (relational coding) was then conducted. By comparing and analysing these initial concepts, categories with higher generalizability were identified, forming 13 second-level categories. Finally, selective coding was performed. Since the comment texts had already been classified into three broad themes using LDA, the second-level categories derived from axial coding were further synthesized and aligned with these three first-level categories, corresponding to the LDA clusters. As shown in Figure 2, the final coding system comprehensively covers the public's understanding and discussion of the 'gene-edited babies incident' from different perspectives.

## 4. Event-specific discourse: Research ethics and technological risks

Event-specific discourse primarily focuses on the incident itself and the circumstances of those involved. It can generally be divided into two dimensions: on the one hand, the public pays attention to the rights of the subjects and the ethical issues surrounding the experimenter in the 'gene-edited babies incident'; on the other hand, the public is concerned

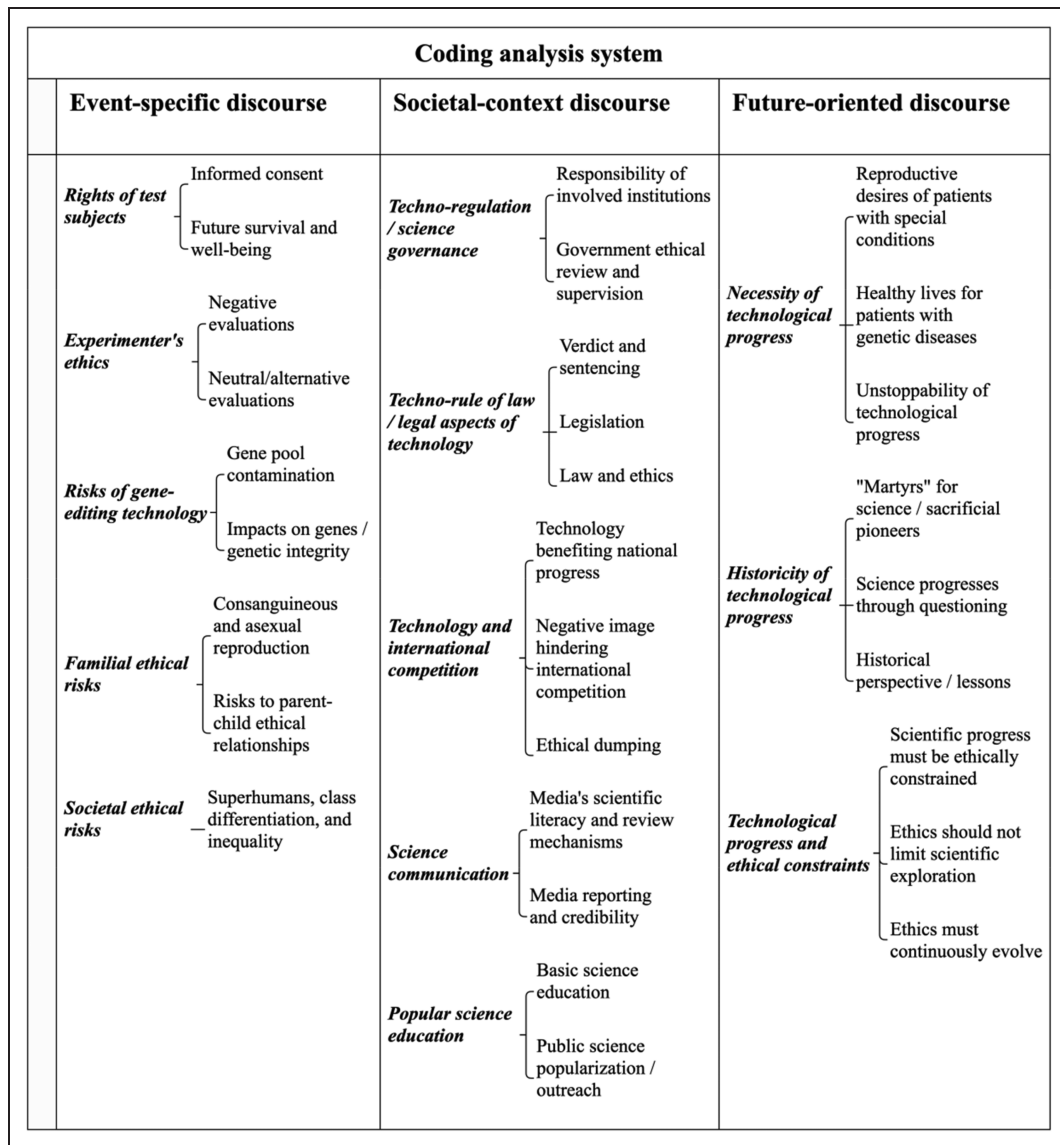


Figure 2. Coding analysis system based on grounded theory.

about the risks of gene-editing technology, including genetic, familial and social ethical risks.

#### 4.1 Rights of subjects and morality of the experimenter

In discussing social events, commenters often focus on the individuals involved and their associated

interest groups. As a highly controversial medical experiment, the ‘gene-edited babies incident’ involved the subjects and the experimenter, and discussions on these core parties formed an important component of public opinion.

Commenters’ discussions on subjects’ rights focused mainly on two aspects: the right to informed consent and the right to future survival. Regarding

informed consent, some commenters questioned the experimenter's transparency with the parents. For example, one individual commented: 'To what extent did the infants' parents really know the "truth" they claimed to know?' (YS020010). Meanwhile, other commenters pointed out that the experiment was conducted on a voluntary basis, emphasizing that there was no coercion involved: 'This experiment was conducted on the premise of mutual consent. It was not forced; the subjects all voluntarily agreed' (YS020037). Regarding the right to future survival, discussions were more intense and polarized. Some commenters held extreme positions, believing the infants' existence itself should not be permitted: 'Strongly demand euthanasia for the two infants' (YS010234). However, this view was strongly opposed by many: 'The children are innocent; they are living beings. By what right do any of you sentence someone to death?' (YS030599). On the issue of the infants' future marriage and reproduction, a notable cluster of comments converged around the view that, to avoid potential social and ethical risks, the infants' reproductive capacity should be restricted or eliminated. For example, one comment read: 'They really shouldn't marry and have children; if they do, their children will also be treated as monsters. Sterilization is the best solution' (YS030428). While this position appeared with considerable frequency in the comment data, it is essential to note that such proposals raise profound human rights concerns. Forced sterilization constitutes a violation of fundamental reproductive rights as recognized under international human rights law, and the casual manner in which such measures were advocated in public comments underscores the extent to which emotional reactions and fear of the unknown can override basic rights-based reasoning in public discourse on unfamiliar biotechnologies.

Discussions about the experimenter were similarly polarized. Some commenters expressed strongly negative evaluations, with a few comparing the experimenter to infamous historical perpetrators: 'How is this different from the Unit 731 human experiments<sup>5</sup> by the Japanese aggressors back then?' (RM010694); 'He will become a villain of the ages like Hitler' (YS020338). Some even suggested malicious intent behind the experiment:

'This is far more terrifying than Unit 731; the purpose of human gene-editing experiments is large-scale dissemination, polluting the future, making all Chinese people carry targeted edited sterilization genes' (RM020411). These statements exhibit rhetorical features that resonate with what Mede and Schäfer (2020) conceptualize as 'science-related populism'—specifically, a deep suspicion of scientific elites' moral integrity and a perception that credentialed researchers may pursue hidden agendas at the expense of ordinary people's well-being. While these comments do not constitute a fully articulated populist programme demanding the democratization of scientific authority, their framing of the experimenter as either a malevolent elite actor (akin to a war criminal) or, conversely, a persecuted truth-teller (as seen in Section 6.2) reflects the Manichaean logic characteristic of populist discourse, in which scientific institutions are cast as either corrupt establishments or obstacles to progress (Wang et al., 2025).

Other commenters expressed concerns about high-tech crime: 'The higher the education and knowledge, the greater the harm to society if they commit a crime' (RM011210). At the same time, some commenters attempted to defend the experimenter. For instance, some refuted the Unit 731 comparison: 'Scientific experiments are not about achieving success through cruel killing but obtaining results through rigorous data' (RM020619). Others believed the experimenter's original intentions should be acknowledged: 'His starting point was also to prevent AIDS; perhaps the means were a bit radical, but one cannot erase his merits' (RM012512).

#### 4.2 Genetic, familial and social ethical risks

Another important dimension of public discussion regarding the 'gene-edited babies incident' was the exploration of gene-editing technology itself. For such a cutting-edge high technology, the potential negative impacts of gene editing are still unclear, and commenters heatedly debated its potential threats to genetic diversity, as well as the potential familial and social ethical risks.

On the issue of genetic diversity, views were divided into two main categories. Some worried that gene editing would destroy human genetic

diversity, even leading to ‘gene pool pollution’. For example, one person vividly described: ‘Once human modification through gene-editing technology begins, the human gene pool, this pure lake evolved over billions of years of natural selection, will be polluted at an exponential rate! One can imagine that a hundred years from now, our descendants will live in such a world: “pure humans” not gene-encoded, “superhumans” gene-encoded, and “machine humans” of artificial intelligence!’ (RM020135). Other commenters believed that gene editing has a limited impact on the human gene pool, emphasizing nature’s self-repair capabilities. One comment read: ‘Even if these few children marry and have children with normal people, it will not pollute human DNA. Human DNA self-corrects during pairing, finding a matching DNA counterpart’ (YS031537).

The potential familial ethical risks of gene-editing technology also received considerable attention. Discussions focused mainly on issues of consanguineous reproduction and parent–child relationships. Regarding consanguineous reproduction, some commenters noted that if the technology matures, it could lead to ethical chaos: ‘If this technology matures in the future, can close relatives also reproduce? Many social orders will be disrupted’ (RM020484). Some commenters even proposed more extreme possibilities: ‘Will future humans move towards asexual reproduction, knowing only their father but not their mother, or only their mother but not their father?’ (RM012879). Regarding the parent–child relationship, technological genetic intervention raised some concerns. For example, one commenter stated: ‘Paternity tests are based on genetic similarity. Since the child’s genes have been altered, even though you know it’s your child, their genes are not, right?’ (RM011342).

The public was particularly concerned about the social and ethical risks of gene editing, especially the potential emergence of ‘superhumans’ and the resulting class differentiation and inequality issues. As Harari (2017) pointed out, new technologies in the 21st century could create a biological divide between the rich and the poor: wealthy elites could design themselves or their descendants to become physiologically and psychologically superior ‘superhumans’.

Many commenters expressed concern about this possibility, for example: ‘Others, while still embryos, have already reached heights you can never achieve in your lifetime’ (RM020820). Another pointed out: ‘Imagine if one day you had the choice to design your child to have blue eyes, golden hair, an IQ of 130+, and an athlete’s physique; it’s terrifying just to think about’ (RM010525).

Further discussion extended to the potential for gene-editing technology to exacerbate social inequality. Some commenters worried that this technology might be monopolized by the rich and powerful, thereby solidifying social strata: ‘The rich and powerful classes can use their power and wealth to monopolize this technology, allowing their descendants for generations to gain genetic advantages, thus firmly entrenching their families at the top of the social pyramid. Children from ordinary families who cannot afford gene modification will forever be trampled under the feet of superhumans, with no chance of ever turning things around’ (RM010353).

## **5. Societal-context discourse: Techno-governance and techno-culture**

Societal-context discourse focuses more on the real-world societal problems raised by the ‘gene-edited babies incident’. It can generally be divided into two dimensions: on the one hand, the public is concerned about issues related to regulation, the rule of law and international competition; on the other hand, the public pays attention to the problems of science communication and popular-science education that were highlighted by this event. It is noteworthy that, when discussing these real-world societal problems, the public often refers to science-fiction movies, TV series, anime and other works of scientific culture as a basis for participating in public discussion and expressing opinions.

### **5.1 Techno-regulation, rule of law and international competition**

In social events involving public interest, commenters often focus on the role and influence of the

government and the state, including government regulation, public policy, laws and regulations, and even international politics. The ‘gene-edited babies incident’ is essentially an issue of techno-ethical governance; therefore, discussions on techno-regulation, the rule of law and international competition became focal points for public attention.

Discussions on techno-regulation centred mainly on two aspects: the responsibility of the involved institutions and the government’s ethical review system. First, the public extensively discussed the responsibility of the institutions involved, which included a hospital and a university. Regarding the hospital’s ethical review letter, many pointed out that it came from a Putian-affiliated hospital in Shenzhen,<sup>6</sup> expressing strong doubts: ‘It’s Putian-affiliated again, the kind that posts small ads in toilets for “curing STDs”, now grandly appearing in “high society”. Whose tragedy is this?’ (RM011210). Commenters also delivered harsh criticism of the involved university, believing the school’s management and culture were to blame: ‘What a university is like largely depends on its teachers; one can imagine what kind of place XX University is! This school is just a den of filth’ (RM040241). Second, the public questioned the government’s role in ethical review and supervision. Some comments suggested that there were serious problems with the current techno-ethical regulatory system: ‘All parties are passing the buck; I really hope it can be investigated quickly and give us the truth’ (RM020445). This scepticism reflected public disappointment and dissatisfaction with the government’s inadequate supervision of high-tech fields.

Regarding the rule of law, public discussion focused mainly on three aspects: the verdict, legislation and the relationship between ethics and the law. First, opinions were divided regarding the verdict delivered on 30 December 2019, which sentenced He Jiankui to three years in prison and fined him 3 million *yuan*. Some believed the sentence was too light: ‘A year for each child’s life; emotionally, the sentence feels too light’ (YS031010). Others felt the verdict was unfair: ‘So, three years in prison can determine someone else’s life’ (YS030323). Still others questioned the basis for the conviction: ‘Illegal medical practice is too far-fetched. This is clearly a complex issue involving academic

research, academic ethics, academic practice, etc.’ (YS030097). Meanwhile, some commenters supported the verdict according to the law: ‘It is indeed a bit light, but according to the standard for illegal medical practice, the maximum is three years. In our society governed by law, we must execute according to the law, and then consider legislation to fill the gap in medical ethics’ (YS031388). Second, at the legislative level, the public generally worried that lagging legislation could lead to additional future ethical problems: ‘The speed of Chinese legislation is far behind the actions of these mad scientists; without laws, I believe there will soon be a second and third incident challenging human ethics’ (RM011295). Finally, regarding the relationship between ethics and the law, some believed the actions of those involved only ‘violated ethics and morals, not the law’ (YS031799). This view reflected the public’s varied understandings of the boundary between the law and ethics and people’s confusion about the legal system’s inadequacies in handling techno-ethical issues.

Furthermore, the public engaged in multi-level discussions surrounding international competition. Some argued from a national-interest perspective, emphasizing the importance of gene-editing technology for national progress and competitiveness, displaying clear scientific nationalism. For example, one person commented: ‘If the Chinese take this step, it may not be a bad thing... Even if the Chinese don’t research it, the Americans certainly will’ (RM020144). Another comment mentioned the ‘double standards’ of Western countries: ‘Westerners often say one thing and do another. Why do they oppose it whenever Chinese people research it? They advocate for reducing restrictions on nuclear weapons on the one hand, while blatantly developing small nuclear weapons on the other’ (YS010046). On the other hand, some members of the public were concerned about the event’s negative impact on China’s national image; these discussions were more rooted in a national-image-centric nationalism. For example, one comment pointed out: ‘This will be a stain on the research community, causing the world public to question the bottom line of research, especially Chinese research’ (RM020972). Another mentioned: ‘One person’s mistake makes

the entire Chinese scientific community take the blame' (RM020387). These concerns reflected the public's attention on the reputation of the Chinese scientific community in global competition.

Lastly, the public engaged in in-depth discussions on the issue of 'ethical dumping' in transnational research. Ethical dumping refers to researchers conducting studies in countries with lax regulations when these studies are prohibited or controversial in their home countries (Nordling, 2018). The 'gene-edited babies incident' was cited by *The Economist* as a typical case, sparking public questions about China's regulatory loopholes. For example: 'Actually, the US did this experiment on animals in 1994 and succeeded, but then the US Congress legislated to ban it... There is no clear legal prohibition in this area domestically' (RM010737). This comment identifies a concrete regulatory gap—the absence of explicit domestic legislation at the time—and represents a structurally grounded concern consistent with the academic understanding of ethical dumping.

However, other comments moved beyond identifying regulatory asymmetries and into conspiratorial territory, attributing deliberate manipulative intent to foreign actors. For instance: 'Foreigners are very cunning; they let the Chinese do it, observe secretly from behind, share the results if successful, and let the Chinese bear the responsibility' (RM010032). Some were even more worried: 'After his sentence ends, if he can't do it domestically, will he go abroad to do it?' (YS030646). Unlike the regulatory-gap argument, these comments construct a narrative in which China is cast as the passive victim of a coordinated foreign strategy—a framing for which there is no substantive evidence in the case of the 'gene-edited babies incident'. The coexistence of these two modes of reasoning within the same thematic cluster—one grounded in identifiable institutional shortcomings, the other driven by conspiratorial suspicion—illustrates the heterogeneity of public discourse on transnational scientific governance. It also suggests that legitimate anxieties about regulatory inequality can, in the absence of adequate public information, readily escalate into conspiratorial narratives.

## 5.2 Science communication and popular-science education

In techno-ethical events, the public's main channels for obtaining information are media news reports and dissemination, while their understanding of events is deeply influenced by past popular science and education. Therefore, when discussing the 'gene-edited babies incident', the public focused not only on the event itself but also on issues of science communication and popular-science education.

During the initial stages of the incident, some official media outlets adopted a triumphant announcement style in their news releases, calling it the 'world's first gene-edited babies immune to AIDS' and describing it as 'China achieving a historic breakthrough in the field of gene-editing technology for disease prevention'. However, as the controversy fermented, major media outlets quickly shifted to criticism and condemnation. This drastic change in reporting attitude, as well as the scientific accuracy and professionalism of the report content, sparked widespread public scepticism.

Public discussion of science communication mainly focused on two aspects, both pointing to a lack of professionalism. First, the public criticized the media for lacking basic scientific literacy and review mechanisms. For example, one commenter pointed out: 'Yesterday they were all praising it. It only took a day to go from promoting to condemning' (RM011633). Another commenter questioned the scientific accuracy of the report content: 'Reporting this kind of thing as a scientific achievement, do they have any awareness?' (RM020707). Second, the public believed that the media's improper reporting had been misleading and weakened its own credibility. One commenter wrote: 'The mainstream media, which should guide the public towards correct thinking and awareness, also lost its standard in the face of so-called honour' (YS030613). Another comment pointed out: 'As a media outlet trusted by the public, it failed to properly review and report this unethical matter as a major scientific breakthrough, misleading public judgement' (RM020926).

Public discussion on popular-science education also centred around two primary aspects. On the one hand, many members of the public cited knowledge from high-school biology curriculums in their discussions to defend their views. For example: ‘High-school biology students know, and teachers also tell us that life needs to be revered’ (RM040252). Another comment mentioned: ‘The gene section in high-school biology textbooks states from the beginning about research experiments and ethics’ (RM011266). These discussions reflect the important influence of basic education in biology and ethics on the formation of public techno-ethical views. On the other hand, the public expressed a need for broader promotion of popular science. Some commenters stated that they could not understand the scientific background of the event and hoped for simple, easy-to-understand popular-science content: ‘I don’t understand the specific meaning and implications; I request popular science’ (YS031575). Other commenters criticized the lack of scientific literacy in current public discussions: ‘Still talking about guinea pigs, ligation... China’s path to popular science and rule of law is still long’ (YS031385). At the same time, some called for a rapid increase in popular-science coverage: ‘The 1% undergraduate rate<sup>7</sup> among netizens is really not just talk’ (YS031279). These views indicate that public discussion of the event not only exposed weak links in popular-science education but also underscored the need to further strengthen popular science.

It is noteworthy that many members of the public, when discussing the ‘gene-edited babies incident’, referred to techno-cultural works they have consumed as the basis for their views. Techno-cultural works play an important role in shaping public understanding of techno-ethics. Table 3 lists some of the techno-cultural works mentioned in public comments, revealing several characteristics of these works in influencing the public’s techno-ethical views. First, science-fiction films have the most significant impact on the public. American blockbusters such as *Gemini Man*, *Venom* and *Gattaca* were frequently mentioned in the comments. These films are not only highly entertaining but also express profound techno-ethical propositions through their narrative styles and values. For example, *Gattaca* focuses on

**Table 3.** Techno-cultural works mentioned in public comments.

Type	Work title	Region
Film	<i>Gemini Man</i>	USA
	<i>The Truman Show</i>	USA
	<i>X-Men</i>	USA
	<i>Venom</i>	USA
	<i>Jurassic Park</i>	USA
	<i>The Island</i>	USA
	<i>Futureworld</i>	USA
	<i>Alien vs. Predator</i>	USA
	<i>Resident Evil</i>	USA
	<i>Moon</i>	UK
	<i>A Werewolf Boy</i>	South Korea
TV series	<i>Orphan Black</i>	Canada
	<i>Dark Angel</i>	USA
	<i>The Boys</i>	USA
Animation	<i>Mobile Suit Gundam SEED</i>	Japan
	<i>From the New World</i>	Japan
	<i>Fullmetal Alchemist</i>	Japan
Novel	<i>Bunshin</i> <sup>a</sup>	Japan
	<i>Brave New World</i>	UK
	<i>The Three-Body Problem</i>	China
	<i>Devil’s Blocks</i> <sup>a</sup>	China
Variety show	<i>Age of Angels</i> <sup>a</sup>	China
	<i>Qi Pa Shuo</i> <sup>a</sup>	China

<sup>a</sup>*Bunshin* is a novel by Japanese author Keigo Higashino that explores themes of identity, doubles and the ethics of science. *Devil’s Blocks* and *Age of Angels* are science-fiction works by prominent Chinese author Liu Cixin, known for their imaginative concepts and philosophical depth. *Qi Pa Shuo* is a highly popular and influential Chinese debate-style variety show, known for discussing a wide range of social and cultural topics, often including science and ethics, in an engaging and entertaining format.

social inequality caused by gene editing, and the ethical dilemmas it conveys are often used by the public to draw analogies to the ‘gene-edited babies incident’. The techno-ethical values exported by such films largely shape the public’s understanding and judgement of techno-ethical events. Second, Japanese animation also plays an important role in shaping public techno-ethical views. For example, works such as *Mobile Suit Gundam SEED* were mentioned multiple times in public comments. These animated series explore the complex relationship between technological progress and war and peace through imaginative narratives, and their inherent techno-ethical values subtly influence public attitudes

and opinions. Lastly, the role of Chinese science-fiction novels is also noteworthy. Liu Cixin's representative works, such as *The Three-Body Problem*, *Devil's Blocks* and *Age of Angels*, were mentioned multiple times in public comments. These works, with their grand narrative frameworks and profound philosophical reflections, have prompted deep public thought on the relationship between technology and ethics. The prominence of foreign, particularly American, science-fiction films in public comments suggests that globally circulated cultural products currently play a disproportionate role in shaping Chinese public techno-ethical imagination. This finding points to an opportunity for Chinese science-fiction creators and science communicators to develop culturally grounded narratives that engage more directly with the ethical dilemmas specific to China's technological development context.

## 6. Future-oriented discourse: Technological development and the future of humanity

Future-oriented discourse focuses on technological development and the future of humanity. It can generally be divided into three dimensions: the first concerns the necessity of technological development; the second examines the nature of technological development from a historical perspective; and the third explores the relationship between technological development and ethical constraints. It is noteworthy that, in this category of discussion, many members of the public do not express opinions based on specific technical features but tend to broadly compare the 'gene-edited babies incident' with many progressive events in scientific history, even glorifying the individuals involved. This suggests that the public's understanding of the technology itself remains insufficient, and people's techno-ethical literacy also varies greatly.

### 6.1 The necessity of technological development

A core question arising from the 'gene-edited babies incident' is how we can understand the necessity of

technological development. Public discussion on this issue presents diverse perspectives but generally reaches a certain consensus on the necessity of technological development. Specifically, these discussions can be roughly divided into three categories.

Some members of the public, proceeding from the reproductive desires of patient groups such as those with AIDS, rare diseases and mental illnesses, believe that the advancement of gene-editing technology is rational and carries humanitarian significance. This perspective emphasizes the choices and hopes that technology creates for special groups. For example, one commenter stated: 'I feel sorry for AIDS patients; it's just because they hope their offspring won't be troubled by this incurable disease' (YS030938). Another commenter described the longing of a patient with mental illness: 'A male patient with mental illness who has been on medication for a long time and has not relapsed for many years deeply desires a complete family but worries about the illness being inherited. Yet this man really longs for a complete family life' (RM040284).

Another segment of the public focuses on the potential of gene-editing technology to improve the quality of life for patients with genetic diseases, believing that it can alleviate patients' suffering and achieve a healthier state. For example, one commenter pointed out: 'Why not eliminate defects that can be eliminated through technology before birth? Why make them suffer? Is this called humanitarianism?' (YS020440). Another commented: 'If editing genes can prevent children with genetic diseases from falling ill, isn't that great ... How many children are there who have to take medication for life or even have no cure and can only die young due to genetic diseases that cannot be completely cured!' (YS020103). One commenter even stated: 'If I could sign up for the experiment, I might really send my child to be edited, to give him a chance to be cured' (YS030582). These views reflect public expectations for gene-editing technology in humanitarian and medical practice.

From a more macro perspective, other members of the public emphasize that technological progress is an inherent driving force and an unstoppable trend

in human development. This perspective views gene-editing technology as an important step in promoting human progress. One commenter wrote: ‘Humans need to develop and evolve, and development and evolution must be based on the development of science. Human experimentation should be boldly attempted and done! Without experiments, human development cannot advance; it will only stagnate’ (RM020388). This view highlights the public’s positive attitude towards technological development and people’s belief that exploration and experimentation are necessary paths for scientific development.

## 6.2 *The historicity of technological development*

Unlike the ‘elite discourse’ of mainstream media or expert scholars, many members of the public did not entirely criticize the irrationality of the ‘gene-edited babies incident’. Instead, some chose to view it from a historical perspective, believing that the evaluation of technological development should be examined in the long course of history. This historical perspective mainly presented three tendencies.

The first category compared the individuals involved in the event to ‘martyrs’ in scientific history who defended truth, believing they might be seen in the future as pioneers promoting technological development. The most frequently mentioned historical figures include Copernicus, Bruno and Galileo. Copernicus was condemned by the Roman Catholic Church as a ‘heretic’ for proposing the heliocentric theory, which contradicted the Bible; Bruno was sentenced to be burned at the stake by the Inquisition for defending and developing Copernicus’s theory; Galileo was sentenced to life imprisonment for ‘opposing the Pope and promoting evil doctrines’. These figures are now regarded as great pioneers of scientific development, but they suffered strong criticism and persecution at the time. One comment pointed out: ‘Bruno, Servetus, Galileo, Vesalius, Copernicus—all of them were denounced as a heretic by the world, some even burned at the stake to great public satisfaction, yet now they are great scientific pioneers’

(YS010210). Another commenter mentioned: ‘When Copernicus died, he was also portrayed as a devil who disrespected all things in the world and violated all settings that a human being should have’ (RM011288). These remarks reflected the public’s cognition of historical recurrence and the dynamic nature of scientific evaluation.<sup>8</sup>

The second category believed that the emergence of new things is often accompanied by opposition, and that scientific development progresses amid questioning. It argued that it is challenging for any breakthrough technology or idea to be widely accepted at its inception. For example, one commenter mentioned: ‘There are always opposing voices when new things appear. Weren’t there opponents to test-tube babies back then?’ (YS010265). Another commenter emphasized: ‘Any successful new concept, when first proposed, will be opposed by the vast majority of people’ (RM010742). Furthermore, the public also agrees that science needs ‘pioneers’ to bear the risks of exploration, even in the face of criticism and misunderstanding. For example: ‘Someone always has to be the first to stand up and do any experiment; drawing negative conclusions before even trying is not the spirit of scientific research’ (YS010034). Another comment read: ‘History has long proven that countless astonishing inventions and discoveries always faced the pressure of death at birth. We need to give newborns time’ (RM012986).

The third category, from a more macro historical perspective, believed that the value of technological development needs the test of time, and history will ultimately determine its significance. Such discussions often invoked the idea that ‘the victor is crowned king, while the vanquished is branded an outlaw’ (reflecting the sentiment of the Chinese idiom ‘Chéng Wáng Bǎi Kòu’),<sup>9</sup> emphasizing that scientific achievements may take on entirely different meanings in different eras. For example, one comment read: ‘Perhaps we killed Copernicus, or perhaps Hitler, who knows? Only time will tell’ (YS031234). Another comment mentioned: ‘A prisoner in this century might be a hero in the next’ (YS031788). Still others pointed out: ‘Looking at it from this angle today, perhaps in decades or even centuries, we will look at this matter from another angle’ (YS020255).

### 6.3 Technological development and ethical constraints

When discussing the relationship between technological development and the future of humanity, ethical constraints are often seen as a major challenge. The ‘gene-edited babies incident’ clearly violated the basic consensus of contemporary techno-ethics, but the public engaged in multidimensional discussions about whether technology should be subject to ethical constraints and whether such constraints hinder technological development. These discussions mainly fell into three categories.

The first viewpoint emphasized that technological development must proceed within an ethical framework, believing that unconstrained technology is like ‘Pandora’s Box’, which, once opened, may bring uncontrollable consequences. For example, one commenter pointed out: ‘It’s better to wait until there’s enough technology before opening Pandora’s Box; if you open it now, you might not be able to close it’ (RM011585). Another commenter further elaborated: ‘Sooner or later, humanity will take the step of rewriting the genome without violating moral ethics. But the key to this event is... true science shouldn’t be like this; regardless of whether there are laws and regulations, one should not break through the bottom line of human morality’ (YS020233). These views reflect public concern about the social and ethical risks of new technology, while also emphasizing that technological progress needs to be synchronized with ethical development.

The second viewpoint argued that ethical constraints should not become barriers limiting scientific exploration. It viewed ethics as man-made rules and believed that there should be no limits when exploring the mysteries of life. For example, one commenter mentioned: ‘Ethics are all set by people themselves. If no one is harmed, why should it be opposed? There should be no boundaries in the exploration of life; it’s not a crime’ (RM020325). Another commenter compared ethical constraints to religious doctrines, criticizing their rigidity: ‘What’s the difference between using ethics to limit technology and religious believers?’ (RM040205). This view further extended to a reflection on the

nature of ethical rules: ‘So-called ethics, morals and laws are actually barriers that humans set for themselves, tantamount to drawing a prison for oneself’ (YS010336). Such discussions revealed that some members of the public question the role of ethical rules in technological development and strongly support the freedom of scientific exploration.

The third viewpoint believed that ethics should evolve alongside the development of technology. These members of the public advocated that technological progress will bring about dynamic adjustments in morality and ethics, and overly conservative ethical views may hinder technological progress. For example, one commenter pointed out: ‘Technological progress will inevitably lead to the iteration of morality; adhering to old rules will lead to stagnation’ (RM020369). Another commenter added: ‘So-called ethics are just set by people... Ethical standards should change with technological progress and will definitely change with technological progress’ (RM040214). Still others called for promoting technological development while improving systems and concepts: ‘Human progress should not be controlled by so-called backward ethics. When technology advances, systems and thinking will also advance. Instead of strictly guarding against it, it is better to improve it’ (RM010740).

## 7. Discussion

This paper focuses on the ‘gene-edited babies incident’, a major techno-ethical controversy that seriously affected the relationship between the public and science. Based on the analysis of 3692 Weibo comment texts, combined with machine learning and grounded theory, this study systematically examined the characteristics of the Chinese public online discourse surrounding this event. The study found that the public discussions exhibited distinct thematic layers, which can be primarily categorized into three types: event-specific discourse, focusing on the event itself and the individuals involved; societal-context discourse, relating to real-world societal issues; and future-oriented discourse, reflecting on technological development and the future of humanity. These layers of discussion also

showed a certain degree of interconnectedness and progression. The various discussions on Weibo constituted a vivid and complex public sphere of thought: in event-specific discourse, the rise and surge of science populism were observable; in societal-context discourse debates, nationalist and cosmopolitan views coexisted; and in future-oriented discourse explorations, liberal and conservative ideas intertwined. From an ethical standpoint, diverse perspectives such as consequentialism, deontology, contractualism and virtue ethics also emerged and were blended into the public discourse.

Before discussing these limitations, it is important to acknowledge what the data collectively demonstrate: the Chinese public engaged with the ‘gene-edited babies incident’ across a remarkably wide range of issues—from informed consent and subjects’ rights to legislative gaps, transnational research ethics and the long-term trajectory of human technological development. This breadth of engagement challenges the assumption that the Chinese public lacks either the interest or the capacity to participate meaningfully in techno-ethical deliberation. Moreover, the cognitive limitations identified below—including superficial scientific understanding, emotional polarization and reliance on culturally narrow interpretive frameworks—are by no means unique to the Chinese context. Studies of public responses to the same event on Western social-media platforms have documented strikingly similar patterns. The issues discussed in the following paragraphs should therefore be read as common challenges in public engagement with techno-ethical controversies, not as culturally specific deficiencies of Chinese public discourse.

However, the analysis of public discourse on the ‘gene-edited babies incident’ also revealed some noteworthy issues regarding public techno-ethical cognition and discursive practices. First, many members of the public demonstrated a superficial understanding of the nature of gene-editing technology, exhibiting insufficient awareness of complex scientific principles, technological limitations and potential long-term risks. Discussions at times presented an overly idealized conception of the technology as a ‘panacea’ and, at other times, heavily demonized it as a ‘Pandora’s Box’ or ‘monster’, to

some extent overlooking the uncertainties and boundary conditions inherent in technological applications.

Second, emotional expression was quite prominent in public discussions, for instance, manifesting as extreme moral judgements of the individuals involved—both severe condemnation and heroic portrayal—or as blind support for or wholesale rejection of the technology itself. This sometimes risked diminishing the space for rational analysis and obscuring in-depth exploration of the complex scientific and ethical issues underlying the event. Such extreme positions, while rhetorically striking and therefore analytically significant, represented a relatively small proportion of the overall dataset; the majority of comments occupied more moderate ground, and extreme statements were frequently contested by other commenters within the same discussion threads. Nevertheless, the visibility and emotional intensity of these minority voices risk disproportionately shaping outside perceptions of the discussion as a whole, which underscores the importance of attending to the distribution rather than merely the content of public opinion when drawing conclusions from social-media data.

Third, the cultural logic that some members of the public relied upon when articulating their views appeared to be somewhat monolithic; they frequently cited Western science-fiction cultural products as evidence, particularly American blockbusters characterized by narratives of technological supremacy or technological backlash. This may reflect a deficiency in localized, diverse resources for techno-ethical thinking. Lastly, public discourse also reflected a certain distrust of existing techno-regulatory and governance mechanisms, particularly concerning the effectiveness of ethical review and legal regulations. This perceived tension between technological development and social governance, if not effectively addressed, could exacerbate public anxiety about and potential resistance to emerging technological advancements in the long run.

Although public discourse surrounding the ‘gene-edited babies incident’ exhibited limitations in certain aspects, the rich social emotions, diverse value claims and specific cultural logics embedded

within it are crucial entry points for understanding how such major techno-ethical issues are received, interpreted, debated and reflected upon at the societal level. This discourse not only reflects the adaptation and tensions of social culture in the face of drastic technological change but also reveals the profound impact of technological development on societal value systems and public cognitive patterns. Therefore, in-depth analysis and research of public discourse surrounding specific major techno-ethical events, particularly as undertaken in this study of the ‘gene-edited babies incident’, should receive greater attention from both academia and policymakers. The cognitive patterns, depth of understanding, emotional attitudes and discursive expressions demonstrated by the public regarding such events should rightly serve as important references for the construction of techno-ethical governance frameworks and related policymaking. More fully integrating public perspectives into the techno-ethical governance system can not only effectively compensate for the potential singularity and limitations of relying solely on government regulation and elite discourse but can also enable greater inclusivity, responsiveness and practical effectiveness in techno-ethical governance, ultimately serving the benign interactions between and synergistic development of technology and society.

## 8. Limitations

The empirical foundation of this investigation is subject to specific constraints regarding data representativeness and textual depth. The exclusive reliance on Sina Weibo introduces a platform-specific demographic bias because the user base is predominantly younger and more urbanized than the general population. This focus potentially overlooks heterogeneous perspectives existing within semi-private networks or among demographics with lower digital engagement. Furthermore, the selection of comment sections under major official media outlets subjects the data to the influences of institutional moderation and user self-censorship. These environments prioritize a curated form of public discourse, which may marginalize fringe or radical viewpoints that would otherwise appear in less

regulated digital spaces. The intrinsic brevity and emotional reactivity of social-media commentary also impose limits on the reconstruction of complex cognitive structures. While these texts provide immediate insights into public sentiment, they lack the extended logical argumentation characteristic of deliberative interviews or systematic surveys.

## 9. Conclusion: Why it still matters after six years?

Revisiting the 2018 ‘gene-edited babies incident’ from the vantage point of 2026 provides a critical baseline for assessing the social robustness of contemporary disruptive innovations. This event marks the definitive moment when human germline intervention entered global consciousness and transitioned techno-ethical governance from reactive emergency measures to a normalized systemic framework. In an era defined by the convergence of biotechnology and artificial intelligence, understanding the longitudinal evolution of public cognition following the initial technological shock is indispensable for calibrating the regulatory boundaries of current innovations.

The institutional failure of science communication in 2018, characterized by a rapid shift from celebrating breakthroughs to issuing moral condemnations, underscores the inherent fragility of triumphalist narratives. This instability reveals that one-dimensional progressivist rhetoric is insufficient for managing the complexities of techno-ethical controversy. Modern communication strategies must therefore move beyond oversimplified success stories to incorporate uncertainty and risk forecasting as core components of public engagement. Sustaining communal trust amid rapid technological iteration requires a commitment to transparency rather than the curated optimism of institutional reporting.

Science education must prioritize the cultivation of ethical competency to bridge the persistent gap between technological imagination and governance reality. Evidence suggests that individuals frequently rely on foundational biological concepts and science-fiction archetypes to articulate their

ethical stances, confirming that mass culture and basic schooling constitute the primary infrastructure for techno-ethical development. Future educational initiatives should focus on fostering reflexive reasoning and prudent judgement rather than the mere dissemination of biological facts. These capacities enable a society to move beyond symbolic fear or blind compliance towards a consensus grounded in rational critique.

Integrating public perspectives into techno-ethical governance is a strategic necessity for ensuring the synergistic development of technology and society. The 2018 case proves that governance models excluding public voices remain vulnerable to institutional paralysis when confronted with radical scientific practices. A retrospective analysis of public discourse allows governance bodies to identify social sensitivities and trust-triggering mechanisms with greater precision. This shift toward inclusive governance endows techno-ethical norms with genuine social resilience and provides the cultural foundation required for national innovation strategies to endure in an uncertain technological landscape.

### ORCID iD

Shuo Wang  <https://orcid.org/0000-0002-4095-5253>

### Funding

This study was supported by the Major Program of the National Social Science Foundation of China (grant number 21ZDA017).

### Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Notes

1. The Southern University of Science and Technology (SUSTech), located in Shenzhen, Guangdong Province, was established in 2011 as a reform-oriented public research university. At the time of the incident, it was a relatively young institution, having been in operation for only about seven years. Its founding mission emphasized innovation and autonomy from traditional Chinese academic bureaucracy, which is relevant context for understanding the institutional oversight environment in which He Jiankui operated.
2. On 26 November 2018, the same day the news broke, 122 Chinese scientists issued a joint public statement on social media condemning He Jiankui's experiment. The statement described the use of CRISPR-Cas9 technology for human embryo gene editing as 'crazy', warned of its unpredictable risks to the gene pool, and called for stricter government regulation of biomedical research. This rapid, large-scale, collective response from within the Chinese scientific community was unprecedented and signalled the depth of professional alarm at the ethical breach, while also serving as a significant discursive event that shaped subsequent public discussion on social media.
3. The charge of 'illegal medical practice' struck many observers, both domestic and international, as an imperfect legal fit for the conduct in question. At the time of the incident, China lacked specific criminal legislation directly addressing unauthorized human germline genome editing. In the absence of a more precisely applicable statute, prosecutors relied on the existing criminal provision against practising medicine without proper authorization. This legal gap itself became a subject of public debate, as discussed in Section 5.1, and subsequently motivated legislative reform: China's Biosecurity Law, which took effect in April 2021, and revisions to related regulations have since established more explicit legal frameworks governing human genetic resource management and gene-editing research.
4. Baidu Index is a search trend analysis tool provided by Baidu, China's dominant search engine, similar to Google Trends.
5. Unit 731 was a biological and chemical warfare research and development unit of the Japanese Army during World War II. Historical records and testimonies indicate that Unit 731 conducted experiments on human subjects, including prisoners of war and civilians, who were primarily of Chinese, Korean and Russian origin. These activities, which involved testing biological agents and other procedures, are documented to have resulted in a large number of deaths. The actions of Unit 731 have been widely studied and are generally cited in discussions of wartime

- conduct and medical ethics as examples of profound ethical violations.
6. Putian-affiliated hospitals are a network of private medical institutions in China, many founded by investors from Putian, Fujian Province. In Chinese public perception, these hospitals are widely associated with aggressive false advertising, overcharging, unnecessary medical procedures and, in some cases, fraudulent treatment claims—particularly in areas such as sexually transmitted diseases, infertility and cosmetic surgery. Public distrust of this network intensified dramatically following the 2016 ‘Wei Zexi incident’, in which a young university student died after receiving an ineffective and expensive experimental cancer treatment promoted by a Putian-affiliated hospital through paid search-engine advertising. The commenter’s reference to Putian-affiliated hospitals thus carries a strongly pejorative connotation, implying that the ethical review associated with He Jiankui’s experiment was conducted by an institution of dubious credibility and professional standards.
  7. This is a hyperbolic expression commonly used in Chinese internet culture to satirize the perceived low educational level of online commenters. It is not a factual statistic. Actually, the proportion of Chinese internet users holding a bachelor’s degree or above was above 10%. The commenter’s use of ‘1%’ functions as rhetorical exaggeration to express frustration with what they perceive as widespread scientific illiteracy in online discussions.
  8. These remarks reflected the public’s cognition of historical recurrence and the dynamic nature of scientific evaluation. However, it is important to note that this analogy is historically and epistemologically misleading. Copernicus, Bruno and Galileo were persecuted for advancing empirically grounded scientific claims that challenged prevailing religious orthodoxy; their vindication rested on the eventual confirmation of their scientific contributions. He Jiankui’s case is fundamentally different: he was not condemned for a scientific discovery that the establishment refused to accept, but for conducting a premature and ethically unauthorized experiment on human subjects—an act that violated established research ethics norms and international scientific consensus regardless of whether the underlying technology might eventually prove beneficial. The conflation of ethical transgression with intellectual martyrdom in these public comments reveals a significant gap in public understanding of the distinction between scientific innovation and research ethics compliance, and suggests that popular narratives of scientific heroism may inadvertently provide rhetorical resources for legitimizing ethically irresponsible conduct.
  9. The idiom ‘Chéng Wáng Bào Kòu’ is a widely used Chinese expression conveying a cynical, outcome-deterministic view of history: moral judgement is ultimately dictated by success or failure rather than by intrinsic rightness. It carries an implicit moral relativism, suggesting that ethical evaluations are retrospective constructs imposed by victors. In the context of these comments, the invocation of this idiom reflects a strand of public reasoning that suspends present-day ethical judgement in favour of deferring to future historical outcomes—a logic that, if taken to its conclusion, would render contemporaneous ethical constraints on scientific practice essentially meaningless.

## References

- Blei DM, Ng AY and Jordan MI (2003) Latent Dirichlet allocation. *Journal of Machine Learning Research* 3: 993–1022.
- Doudna J (2019) CRISPR’s unwanted anniversary. *Science* 366(6467): 777.
- Funk C and Hefferon M (2018) Public views on human gene editing for babies vary depending on goal. Pew Research Center, 26 July. Available at: [https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2018/07/PS\\_2018.07.26\\_gene-editing\\_FINAL.pdf](https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2018/07/PS_2018.07.26_gene-editing_FINAL.pdf).
- Gaskell G, Bard I, Allansdottir A, et al. (2017) Public views on gene editing and its uses. *Nature Biotechnology* 35(11): 1021–1023.
- Harari YN (2017) *Homo Deus: A Brief History of Tomorrow*. New York: Harper.
- Hendriks S, Giesbertz NA, Bredenoord AL, et al. (2018) Reasons for being in favour of or against genome modification: A survey of the Dutch general public. *Human Reproduction Open* 3: hoy008.
- Jelodar H, Wang Y, Yuan C, et al. (2019) Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications* 78: 15169–15211.

- Ji J, Robbins M, Featherstone JD, et al. (2022) Comparison of public discussions of gene editing on social media between the United States and China. *PLoS One* 17(5): e0267406.
- Lander ES, Baylis F, Zhang F, et al. (2019) Adopt a moratorium on heritable genome editing. *Nature* 567(7747): 165–168.
- Mede NG and Schäfer MS (2020) Science-related populism: Conceptualizing populist demands toward science. *Public Understanding of Science* 29(5): 473–491.
- Ni C, Wan Z, Yan C, et al. (2022) The public perception of the #GeneEditedBabies event across multiple social media platforms: Observational study. *Journal of Medical Internet Research* 24(3): e31687.
- Nordling L (2018) EU crackdown on “ethics dumping”. *Nature* 559(5): 17–18.
- Pu X, Jiang Q and Fan B (2022) Chinese public opinion on Japan’s nuclear wastewater discharge: A case study of Weibo comments based on a thematic model. *Ocean & Coastal Management* 225: 106188.
- Rauchfleisch A and Schäfer MS (2015) Multiple public spheres of Weibo: A typology of forms and potentials of online public spheres in China. *Information, Communication & Society* 18(2): 139–155.
- Wang S, Wang T, Yokoyama HM, et al. (2025) Beyond a single pole: Exploring the nuanced coexistence of scientific elitism and populism in China. *Humanities and Social Sciences Communications* 12(1): 1–13.
- Wynne B (2006) Public engagement as a means of restoring public trust in science: Hitting the notes, but missing the music? *Public Health Genomics* 9(3): 211–220.
- Yang Y, Hsu JH, Löfgren K, et al. (2021) Cross-platform comparison of framed topics in Twitter and Weibo: Machine learning approaches to social media text mining. *Social Network Analysis and Mining* 11(1): 75.
- Zhang X, Chen A and Zhang W (2021) Before and after the Chinese gene-edited human babies: Multiple discourses of gene editing on social media. *Public Understanding of Science* 30(5): 570–587.

### Author biographies

**Shuo Wang** is a PhD candidate at the Center for Science, Technology, and Society, Tsinghua University. His research interests lie in science communication, with a particular focus on science-related populism, public trust in science and artificial intelligence (AI) ethics. Recently, his work has extended to AI for science and the social dynamics of digital society.

**Zhengfeng Li** is a professor in the Department of Sociology, Tsinghua University, where he serves as the Director of the Center for Science, Technology, and Society. He is also an associate editor of *Cultures of Science*. His main research areas include S&T policy, the philosophy of science and technology, S&T and society, the history of S&T policy and so on.

## Journal Description

Cultures of Science is a peer-reviewed international Open Access journal. The journal aims at building a community of scholars who are expecting to carry out international, inter-disciplinary and cross-cultural communication. The topics include: cultural studies, science communication, the history and philosophy of science and all intersections between culture and science. The journal values the diversity of cultures and welcomes manuscripts from around the world and especially those involving interdisciplinary topics.

## Aims and Scope

Cultures of Science is an international journal that provides a platform for interdisciplinary research on all aspects of the intersections between culture and science. It is published under the auspices of the China Association for Science and Technology.

It welcomes research articles, commentaries or essays, and book reviews with innovative ideas and shedding a fresh light on significant issues. Research articles report cutting-edge research developments and innovative ideas in related fields; commentaries provide scientific perspectives on emerging topics or social issues; book reviews evaluate and analyze the contexts, styles and merits of published works related to cultures of science.

The topics explored include but are not limited to: science communication, history of science, philosophy of science, sociology of science, science education, public understanding of science, science fiction, political science, indicators of science literacy, values and beliefs of the scientific community, comparative study of cultures of science, public attitudes towards a new scientific and technological phenomena.

Cultures of Science is published 4 times a year in March, June, September and December.

## Contact Information

Address: 3 Fuxing Road, Haidian District, Beijing 100038, China.

Email: [culturesofscience@cnais.org.cn](mailto:culturesofscience@cnais.org.cn)

## Disclaimer

Any opinions and views expressed in the articles in *Cultures of Science* are those of the respective authors and contributors and not of *Cultures of Science*. *Cultures of Science* makes no representations or warranties whatsoever in respect of the accuracy of the material in this journal and cannot accept any legal responsibility for any errors or omissions that may be made. The accuracy of content should be examined independently.

© National Academy of Innovation Strategy 2026

All rights reserved; no part of this publication may be reproduced, stored, transmitted or disseminated in any form or by any means, without prior written permission of the publisher.



Volume 9 . Issue 1 . March

**Special topic: Global governance of technological ethics: Historical evolution, innovative challenges and China's role in multi-stakeholder participation**

- 3 Rethinking global technology governance at a crossroads: China's role, historical turning points and future imaginaries  
*Lu Gao*
- 9 Forks in the road: Françoise Baylis on ethics, genome editing, and the world we want to live in  
*Ping Yan and Lu Gao*
- 21 Addressing the dual challenges of scientific and technological risks: Deep dilemmas and system transformation in global governance  
*Yidong Liu*

37 Ethical AI governance: AI for society and a co-learning approach  
*Xiaobai Shen and Lu Gao*

47 Exploring the design approach to embedding ethics in technology  
*Wei Zhang, Yu Jing and Qian Wang*

**Article**

62 Event, society, and future: Revisiting Chinese public discourse on the 'gene-edited babies incident'  
*Shuo Wang and Zhengfeng Li*

