LOFTY PRINCIPLES, CONFLICTING INTERESTS

AI ethics and governance in China

Rebecca Arcesati

KEY FINDINGS

- **Rather than being driven entirely from the top**, China’s AI ethics and governance landscape is shaped by multiple actors and their varied approaches, ranging from central and local governments to private companies, academia and the public. China’s regulatory approach to AI will emerge from the complex interactions of these stakeholders and their diverse interests.

- **Despite notable advances in tackling ethics issues in specific AI sectors and application areas**, a large gap remains between defining broad ethical principles and norms to guide AI development and putting these into practice through standards, laws and government or corporate regulation.

- **This gap is not unique to China**, but particularly pronounced in the Chinese context since AI is seen as a core means for fully achieving the governance vision of the Chinese Communist Party, which prioritizes state control and political security over individual rights. Genuine concern for AI ethics coexists with Beijing’s use of AI for mass surveillance and ethnic profiling.

- **Given China’s rapid AI advancements**, its expanding presence in global standards bodies and Chinese tech companies’ growing global reach, it will be critical for the EU to engage with Chinese actors. However, European policymakers must take the government’s rhetoric on AI ethics with great caution and push back against China’s use (and export) of AI for surveillance and other applications that threaten human rights and fundamental freedoms.
Various Chinese actors have tackled AI ethics issues
A timeline of seminal developments since 2017

Leading actors driving each development:
- Government or Government-Affiliated
- Academia
- Industry
- Civil Society

New Generation AI Development Plan (State Council) - July 2017

Four AI ethics principles (Baidu) - May 2018

AI Ethical Risks Analysis Report (National Artificial Intelligence Standardization General Group) - Sept. 2018

ARCC principles (Tencent) - April 2019

Beijing AI Principles (BAAI) - May 2019

Joint Pledge on AI Industry Self-Discipline (AIIA) - June 2019

AI for Children: Beijing Principles - Sept. 2019

Core Principles of our AI Practice (Megvii) - Nov. 2019

White Paper on AI Governance (CAICT, AIIA) - Sept. 2020

Public privacy backlash against deepfake face-swapping app ZAO - Oct. 2019

First civil lawsuit over the use of facial recognition technology - Nov. 2019

AI Security Standardization White Paper (TC260) - Sept. 2019

AIIA = Artificial Intelligence Industry Alliance
AITISA = Artificial Intelligence Industry Technology Innovation Strategic Alliance
BAAI = Beijing Academy of Artificial Intelligence
CAICT = China Academy of Information and Communications Technology
CESI = Chinese Electronics Standards Institute
MOST = Ministry of Science and Technology
NITS = National Information Technology Standardization Technical Committee
SAC = Standardization Administration of China
TC260 = National Information Security Standardization Technical Committee

Source: MERICS
1. INTRODUCTION

Countries around the world are harnessing the transformative impact of artificial intelligence (AI) on their economies and societies. There has been much focus on the competition and rivalry between countries with advanced AI research and development (R&D) capabilities, with talk of an “AI race” between the United States and China – and to a lesser extent Europe. However, the ethical and safety risks of not getting AI right are as great as its beneficial potential. From facial recognition and recruitment algorithms carrying biases to self-driving cars endangering lives, the challenges associated with AI governance failures are enormous and require joint solutions.

Artificial Intelligence (AI) refers to both a scientific field and a broad suite of technologies that accomplish tasks generally believed to require human intelligence, such as making decisions through the collection, processing and interpretation of data. The EU Commission defines AI as “systems that display intelligent behavior by analyzing their environment and taking actions – with some degree of autonomy – to achieve specific goals.”

Under the umbrella term “AI ethics,” experts are discussing questions such as what role AI systems should play in our societies, what risks they involve and how we should control them. In recent years, professional associations, companies, governments and international organizations have published a plethora of AI ethics principles and guidelines. Several European countries and organizations have played a pivotal role in these efforts, with the EU strongly advocating for the development of risk frameworks and legislation to ensure “trustworthy” AI, cemented in April 2021 in a proposal for the world’s first dedicated AI regulations.

Understanding Chinese approaches to AI ethics and governance is vitally important for European stakeholders. China will be a fundamental force in shaping the trajectory of AI innovation and adoption as well as the way in which AI will be governed. It has embraced AI and aims to become the world’s primary AI innovation center by 2030. Chinese policymakers are paying increasing attention to ethics in the context of AI governance, having issued multiple related principles. Behind such initiatives there is a web of public and private players, interests and voices.

This MERICS Monitor provides an analysis of China’s emerging AI ethics and governance landscape. **It examines three issues:**

- The various approaches to AI ethics taken by government, corporate, academic and civil society actors in China
- Ethical issues related to specific applications (healthcare, autonomous driving and public security) and how they are being addressed
- China’s role in global AI ethics and governance efforts and its implications for European stakeholders
2. BEIJING’S STRATEGIC CONSIDERATIONS AND APPROACH TO AI ETHICS AND GOVERNANCE

The government’s ambition to lead the world in AI is accompanied by its growing attention to the technology’s governance. In 2018, President Xi Jinping called for the “healthy development” of AI through the establishment of laws, ethics, institutional mechanisms and regulations.1 In the leadership’s view, researching and preventing the short-term risks, such as privacy and intellectual property infringements, and long-term challenges AI systems could pose to the economy, social stability and national security, such as unemployment and changes to social ethics, is of utmost importance.

2.1 China’s policymakers pay increasing attention to ethics in the context of AI governance

Starting with the publication in 2017 of the State Council’s New Generation Artificial Intelligence Development Plan (AIDP), the government expressed its intention to tackle ethical issues arising from AI systems. The plan states that by 2025 China will set up an initial system of laws, regulations, ethical norms and policies as well as a security assessment framework to “ensure the safe, reliable and controllable development of AI.” A comprehensive system should be established by 2030. The AIDP calls for strengthening research on legal, ethical and social issues. It also urges measures like an ethical framework for human-machine collaboration and codes of conduct for personnel in AI product R&D and design.2

Since then, several principles and white papers have been issued to guide AI governance (see Exhibit 1). In a 2018 AI Standardization White Paper, the Chinese Electronics Standards Institute (CESI) recommended three overarching ethical considerations for AI: “human interest,” “liability” and “consistency of rights and responsibilities.” The document discusses safety, ethical and privacy issues and reflects the government’s wish to use technical standardization as a tool in domestic and global AI governance efforts.3

In 2019, the Ministry of Science and Technology (MOST) issued the Governance Principles for a New Generation of AI, which put forward eight principles for developing “responsible AI.”4 Drafted by a dedicated expert group, the Governance Principles are the most official formulation of China’s approach to AI ethics to date.

Understanding the terms the government uses is necessary to gauge its vision for AI governance. Reference to “human rights” in the Governance Principles does not imply endorsement of liberal democratic values, while “societal security” implies maintaining stability by prioritizing collective wellbeing, as defined by the Chinese Communist Party (CCP), over individual freedoms. Additionally, the concept of human-machine harmony, read alongside the AIDP’s call for strengthening “public opinion guidance,” may indicate the intent to prepare society for greater data-driven monitoring and governance through AI.

2.2 The government directs a multi-stakeholder conversation on AI ethics

While the debate on AI ethics is overseen by Beijing and takes place within the strict limits of the party-state’s goals and interests, it is a multi-stakeholder conversation.

MOST’s AI governance committee comprises experts from leading universities, the Chinese Academy of Sciences and private AI companies. The Beijing AI Principles, a key document that
The National New Generation Artificial Intelligence Innovation and Development Pilot Zones (国家新一代人工智能创新发展试验区) are experimental areas established under MOST's mandate to stimulate local AI innovation. 15 such zones have been announced thus far and five more are planned to be established by 2023.

Sources: MOST; provincial and municipal government documents; announcements as of March 2021 were considered.

**AI ethics and governance developments are not confined to Beijing**

Various AI pilot zones* are taking initiatives to implement high-level principles on the ground.

**Planned or ongoing AI ethics initiatives**
- General pledge
- Specific ethics research and/or regulatory tasks
- Issuing of local principles and/or implementation guidelines

**Tianjin:** Focus on building ethical guidelines; explicit mandate for think tanks to build an AI Governance Platform (announced Aug. 2020), research to cover ‘norm setting’ and governance technologies

**Beijing:** Beijing Academy of AI; Beijing AI Principles; experimental research plan focused on public welfare, sustainable development and social impacts of AI

**Jinan:** Planned AI Governance Committee to focus on conducting research and advising on responsible AI, incl. ethical risks (announced Jun. 2020)

**Chongqing:** First project batch covering institutional reforms, incl. in the areas of privacy protection and the construction of preliminary ethical guidelines, policies and regulations

**Shanghai:** AI Pilot Zone Expert Advisory Committee Governance Working Group (est. Jul. 2020); Action Proposals on Collaborative Implementation of AI Governance Principles; major AI governance conferences

**Hangzhou:** Focus on technical standards and security frameworks for healthy AI development, esp. for autonomous vehicles and smart city governance

*The National New Generation Artificial Intelligence Innovation and Development Pilot Zones是国家新一代人工智能创新发展试验区, are experimental areas established under MOST's mandate to stimulate local AI innovation. 15 such zones have been announced thus far and five more are planned to be established by 2023.

Sources: MOST; provincial and municipal government documents; announcements as of March 2021 were considered.
A central feature of all these discussions is their applied approach. To drive implementation at the local level, MOST is encouraging municipal governments to step up relevant work in AI pilot zones (see Exhibit 2). Additionally, both the Beijing AI Principles and the Joint Pledge focus on applicable and action-oriented goals and measures to ensure that the trajectory of AI development throughout the lifecycle of systems, from R&D to commercialization, is beneficial for society.7

2.3 Safeguarding stability is a key objective of China’s AI strategy

The government’s rhetoric and attention to ethics can appear hypocritical given its use of AI for mass surveillance, repression and ethnic profiling (see section 4.3). However, from the perspective of China’s leadership and of its moral and ethical frameworks, this poses no contradiction. National security and stability are the highest collective goods, taking priority over personal privacy, transparency, accountability and individual human rights.

The CCP sees security and stability as preconditions as well as products of economic development, a key objective of China’s AI strategy. A major goal of the AIDP is the modernization of social governance, which entails not only the optimized provision of public services but also the construction of a modernized socialist society through, for example, the use of AI to “grasp group cognition and psychological changes.”8

Additionally, the party justifies its control over the legal system by arguing for the need to ward off threats from internal and external enemies to meet the superior goal of preserving political security.9 Thus, from the CCP’s perspective, the use of AI against a part of the population it sees as a terrorist threat to society, as is the case with Uighurs in Xinjiang, can coexist with efforts to ensure that AI systems do not cause harm to the majority.

Ethical questions about algorithmic decision-making are framed around the interests of the collective – of which the party-state claims to be the sole legitimate representative – rather than the individual.10 This logic also explains why the emerging data protection regime aims to impose restrictions on companies’ ability to collect personal information but leaves the government with nearly unrestrained power to harvest and use citizens’ data for public security and law enforcement.11

3. HOW INDUSTRY, ACADEMIA AND CIVIL SOCIETY DRIVE FORWARD ETHICAL AI

3.1 Industry plays a pivotal role in shaping Chinese discussions

China’s leadership sees industry as a key driver in coordinating self-regulation, research and education on AI ethics, though regulators ultimately set governance rules. It has highlighted the importance of corporate self-regulation, with a recent white paper published by the CAICT identifying companies as the main AI governance entities in the near-term.12 Many leading tech companies and startups have issued calls to address governance and ethics issues related to the development and commercialization of AI applications. They are also joining multi-stakeholder efforts to develop ethics principles and industry standards for responsible AI development, while initiating their own research and principles to tackle ethics issues.

Many companies were directly or indirectly involved in each of China’s three seminal AI documents, of which the joint pledge is an industry commitment to self-regulation. The seven members of MOST’s AI governance committee, for example, include two executives
from e-commerce giant JD.com and facial recognition unicorn Megvii, demonstrating that companies are directly involved in the formulation of policy recommendations and guiding documents such as the Governance Principles.

Tech giants and AI startups are founding members of the previously mentioned BAAI and other key industry alliances behind AI principles and white papers. Baidu and Tencent have also submitted proposals on AI ethics directly to China’s leadership. Many companies are meanwhile active participants in domestic standard-setting activities related to AI.

Corporate self-regulation has thus far primarily taken the shape of high-level ethics codes. Most notably, Baidu, Tencent and Megvii have issued documents that put forward ethics principles to guide their own and the industry’s development of AI. All three highlight similar notions such as the importance of technical robustness and safety, human oversight, data privacy and accountability. Tencent’s AI principles are the most detailed principles developed by a Chinese company so far. Issued in 2018, they urge for AI to be available, reliable, comprehensible and controllable, and highlight specific issues such as algorithmic transparency.

Companies also conduct extensive research into governance and ethics issues through dedicated departments. Their research, much of which predates the government’s increased attention to AI ethics, ranges from techniques for preserving privacy in machine learning to methods for protecting against adversarial attacks on deep-learning systems.

CEOs and AI executives also advocate for interdisciplinary exchanges and collaborative action on AI ethics, while positioning themselves as thought leaders on AI governance issues at key industry forums such as Shanghai’s annual World AI Conference. Some also raise public awareness of the risks of AI applications in everyday life, through campaigns such as AI for Good.

While many tech companies and AI clearly recognize the importance of governing the societal and ethical impact of AI, few have institutionalized steps that turn high-level commitments into concrete procedures. Their AI ethics research and principles, while representing good-faith intentions, mostly lack concrete implementation measures that address the specific issues they identify, from algorithmic fairness to data privacy.

Megvii is one of the few companies to create internal structures such as an AI Ethics Committee to oversee the implementation of its AI principles. This committee is said to make recommendations to the board based on internal investigations and a whistleblowing procedure. However, one listed international member says he never joined the committee and it remains unclear what kinds of changes – if any – it has effected.

It seems logical for companies to be at the forefront of identifying and addressing the harmful impacts of AI applications, given that they research, develop and deploy AI in real-life situations. They are also incentivized to anticipate and address the risks of their AI products and services to avoid backlash from regulators or the general public.

However, for now it is still unclear whether corporate AI ethics declarations are leading to meaningful changes in internal research and development processes, or whether they are ultimately empty commitments that serve only to enhance companies’ reputation. Companies are also commonly reluctant to implement potentially costly and time-intensive mechanisms to ensure safe and ethical AI products.
The close relationship of tech and AI companies with the government adds an additional layer of complication since the government not only provides extensive policy support but is also often a major client for corporates. Companies’ pledges on AI ethics thus often stand in stark contrast to their sale of AI products such as facial recognition or ethnic minority analytics tools to the public security apparatus (see section 4.3).

3.2 Chinese academic research also shapes AI ethics discussions

Academic research on the social and ethical implications of AI is increasingly informing discussions about AI governance in China. A review of relevant publications since 2017 reveals that although research efforts approach the issue from various angles, most are still limited to conceptualizing the changes brought about by AI and suggesting normative and regulatory frameworks. Critical research on specific applications is mostly lacking, although there are notable exceptions.²⁰

Ethics research is conducted through state-sponsored projects and individual scholars’ initiatives. China’s two leading research institutes under the aegis of the State Council – the Chinese Academy of Sciences (CAS) and the Chinese Academy of Social Sciences (CASS) – undertake relevant work, some of which is sponsored by China’s largest public research fund for social sciences, the National Social Science Fund of China. One project led by the Institute of Automation at CAS explores issues like the relationship between humans and AI and challenges associated with determining liability. CAS-sponsored researchers also apply social science research to practical problems, such as social ethics issues caused by the introduction of robots into families.²¹

Several prominent scholars are particularly influential in driving forward ethics research. At CASS, Duan Weiwen (段伟文) – one of China’s most prominent thinkers on philosophical, ethical and social issues surrounding AI and Big Data – leads a Science, Technology and Society Research Center. Duan frequently emphasizes that innovation runs faster than ethics, which requires targeted work to tackle ethical risks in specific technology application scenarios rather than abstract prescriptions. He also advocates for public participation and oversight in ethics matters.²²

Some researchers approach AI ethics from the perspective of traditional Chinese philosophy. CAS-affiliated Zeng Yi (曾毅) spearheaded the formulation of Harmonious Artificial Intelligence Principles, which are based on the concept of “harmony” in Chinese philosophy. These principles emphasize harmony between humans and machines, a concept that is also present in the Beijing AI Principles, and advocates for a positive symbiosis between the two. In addition to playing a leading role in drafting several seminal documents mentioned in this report, Zeng drives major applied ethics research efforts in areas like brain-inspired neural network architectures.²³

Renmin University’s Guo Rui (郭锐), another prominent scholar and government advisor, focuses on translating ethical guidelines into an actionable governance system. Guo has advocated for companies to set up ethics committees, and in his latest book examines the ethical risks of specific AI applications, from precision marketing and content recommendation algorithms to sex robots and smart courts.²⁴

Chinese academia actively engages in global exchanges on AI ethics. This aligns with the government’s call to increase the country’s “discourse power” (话语权) in the field. A prominent example of the interplay between scholarly exchanges and the state’ soft
power ambitions is the Berggruen China Center, established by Peking University and the Berggruen Institute in 2018 with the stated goal of engaging Chinese thinkers to “examine, share and develop ideas to address global challenges.” AI ethics is one of the center’s main research areas. Additionally, in 2020 Tsinghua University established the Institute of Artificial Intelligence Global Governance to “actively contribute Chinese wisdom” and shape the field.

While promoting official Chinese global governance concepts is an important goal behind these initiatives, it would be wrong to view all academic research and collaborations as being driven by the state’s aims. The diverse range of individual research initiatives reflects scholars’ genuine aspiration to make AI beneficial for mankind, as well as to overcome political tensions and cultural barriers between China and the West to advance cooperation. Xue Lan (薛澜), the Director of Tsinghua’s abovementioned institute, has warned that geopolitical tensions between China and the United States are having a chilling effect on industry and policy exchanges in the AI field, which may hinder cooperation on global AI governance.

The BAAI has emerged as China’s leading AI research institute and a hub for multi-stakeholder and international collaboration. The institute has a research center, led by Zeng, which is dedicated to investigating AI ethics, governance and solutions for sustainable development. To foster international dialogue, a recently published study by BAAI and researchers at Cambridge University urges academia to play a greater role in overcoming cultural barriers to collaboration on AI ethics and governance.

Chinese academia seems to be gaining influence in official government efforts to govern AI. Xue and Zeng, for instance, are also members of MOST’s AI Governance Committee. Yet it remains to be seen to what extent scholars will be able to directly influence government policy, corporate practices and regulation towards higher ethical standards.

3.3 Public pushback on AI risks has led to some regulatory changes

While, generally, the public is not seen as the decisive force in China’s AI development, Chinese citizens are pushing for ethical constraints on some use cases. Despite the common perception in the West that Chinese people are particularly trusting of new technologies, there is growing awareness, debate and occasionally pushback related to the risks of AI. In some cases, this has led to policy changes and corporate self-regulation.

Chinese consumers care about the protection of their personal information. When in 2018 Baidu’s CEO Robin Li said Chinese people were less sensitive about privacy and more willing to trade it for convenience, he faced intense opposition on social media. During the Covid-19 outbreak, the use of monitoring apps that collect health information and location data also provoked public criticism due to concerns over discrimination and the erosion of privacy.

In recent years, consumer backlash has played a key role in holding Chinese tech companies accountable for data privacy violations and spurring on regulators to create more stringent regulations. The resulting data regime, which notably includes a Personal Information Security Specification and the soon-to-be finalized Personal Information Protection Law, imposes wide-ranging restrictions on companies’ ability to handle personal information.

The prevalence of AI-powered surveillance technologies also worries citizens. In one 2019 survey, over 70 percent of respondents expressed concerns over privacy violations in the rollout of facial recognition systems.
A series of civil lawsuits have also drawn attention to citizens’ privacy concerns with China’s growing use of facial recognition in public spaces. The first was filed in 2019 against a wildlife park in Hangzhou for introducing a facial recognition access-control system. The individual won the landmark case, which sparked intense online discussions about the excessive collection of facial data. Several other high-profile cases followed. Meanwhile, prominent scholars have recommended the use of tech measures and processes such as ethics by design in addition to regulation to ensure the responsible use of biometric recognition.32

It seems no coincidence that facial recognition regulation has received increasing attention from China’s top lawmakers. Both the recently enacted Civil Code and the abovementioned personal information regulations tighten restrictions over biometric data collection. More recently, standard-setting authorities released a draft of a dedicated national data security standards for facial recognition data.33 Several cities have also introduced or are considering regulations to restrict the use of facial recognition and are fining companies for data privacy infringements.34

Another notable public backlash unfolded in response to the rise of deepfakes – false or altered images, videos and audio generated using AI. In August 2019, the release of the face-swapping app ZAO caused almost-immediate outrage among users over data privacy and copyright infringements.35 Regulators responded swiftly, demanding that the app’s parent company take corrective action. WeChat subsequently restricted some access to the app on its messaging platform, citing security risks.

Authorities soon thereafter took initial steps to regulate the use and spread of deepfakes. One policy document – released apparently in reaction to the ZAO incident – requires online information service providers to review and label any audiovisual content that is produced using new technologies such as deep learning.36 The rules also prohibit the use of deep learning to create, publish or transmit fake news.

While these examples show that civil society influence government regulations and corporate actions related to AI ethics to some degree, their ability to do so is ultimately constrained by China’s political system. China’s data protection regime, for example, leaves the government with unrestrained power to harvest and use citizens’ data for its intrusive public security and law enforcement activities.37

4. AI ETHICS AND GOVERNANCE IN SPECIFIC SECTORS

Government agencies, tech companies, scholars and to some extent the public actively explore AI ethics and governance issues in China. Some of the resulting principles and research bear strong similarities to international efforts. This section examines whether and how these are put into practice by delving into specific ethics and governance issues confronted in relation to AI applications in healthcare, autonomous driving and public security. This shows that the government and companies have a long way to go in turning their normative discourse into practical governance mechanisms.

4.1 Healthcare

China’s government has prioritized healthcare as an area for boosting AI applications to alleviate some of the pressure on its strained system. Chinese companies have made headway
in introducing AI in healthcare, with many applying AI technologies to disease control and prevention as well as medication and vaccine development during the Covid-19 outbreak.\(^{38}\)

Nevertheless, there are numerous legal and ethical concerns surrounding medical AI, such as the data security of health IT systems. According to the Committee of Health Information Security and New Technology Application, 60 percent of the reported data breaches that happened in China in 2017–2018 occurred in the healthcare industry.\(^{39}\) Scholars have pointed out that China’s approach to medical AI ethics puts more emphasis on public health than individual wellbeing, as exemplified by the mass collection of citizens’ health data at the cost of individual privacy.\(^{40}\) It also remains to be seen whether AI will improve access to healthcare and social equity or exacerbate inequalities by improving the quality of care only for the rich.

Chinese research investigating the legal and ethical challenges brought by medical AI has skyrocketed since 2019.\(^{41}\) Researchers have examined a variety of issues such as the legal liability of AI systems when errors occur, the transformation of health workers’ role, data protection and privacy and ethical decision-making in using autonomous systems.

One predominant approach to regulating AI in healthcare is to apply existing risk assessment frameworks for medical equipment. The authorities’ focus has been mainly on setting up industry standards and equipment-assessment procedures to ensure data security and safety in each application. For example, Chinese regulators have established evaluation and certification measures for medical products and services that incorporate big data and issued assessment criteria for medical devices that incorporate deep learning.\(^{42}\)

AI regulation appears to be more advanced in healthcare than in other sectors, with a dozen policies, standards and assessment criteria issued in recent years by government entities ranging from the State Council to healthcare regulators such as the National Health Commission and the National Medical Products Administration.

However, the government’s focus on data security and its approach to treating medical AI as medical equipment leave unaddressed many ethical issues, such as bias embedded in AI systems. Algorithmic bias refers to outcomes that are systematically less favorable to a particular social group due to unrepresentative, incomplete or flawed training data. Moreover, the ethical review process in the healthcare system, which allows the examination of individual AI use cases, also appears to be underdeveloped: since the developers of most AI systems used in the sector are not healthcare companies but tech firms, they are not under the purview of existing ethical review committees. Meanwhile, in several provinces and cities these committees lack technical AI expertise.\(^{43}\)

4.2 Autonomous driving

Autonomous driving is one field of AI application where China is widely believed to be on its way to taking the global lead. In 2020, pilots for Chinese self-driving vehicles advanced rapidly while the United States suspended testing on pandemic-related grounds. However, advances in autonomous driving – fueled by the government’s ambition for China to become a global leader in smart-car development by 2035 – have been accompanied and at times moderated by concerns about ensuring the safety of intelligent connected vehicles (ICV).

A small number of prominent fatal accidents caused by automated driving systems in China and abroad have caused widespread public discussions that highlighted to government...
and industry players the need to pay closer attention to safety. Though the government is still pushing the commercialization of ICVs, safety concerns seem to have taken on a greater importance.

For example, the National Development and Reform Commission, China’s industrial planner, downgraded previous ambitious market-share targets for autonomous vehicles in its finalized development strategy for intelligent vehicles released in February 2020. Regulators aim to complete a basic standards system for autonomous vehicles by 2025, including a standards framework for driver-assistance functions and low-level autonomous driving that was scheduled for completion by the end of 2020. Since 2017, China has also rolled out regional standards for road-testing autonomous vehicles, alongside national road-testing rules for ICVs.

Chinese auto and tech companies are actively involved in domestic and international efforts to improve the safety of autonomous vehicles. Baidu – the designated AI national champion for autonomous driving – plays a key part in leading industry discussions through its Apollo open autonomous driving platform and technical research into safety frameworks. Baidu was also one of 11 major international companies that published a white paper proposing a framework for implementing existing auto industry safety standards in autonomous driving.

Governance efforts have also tackled information and algorithmic security. These feature strong involvement from and cooperation between government-affiliated research institutes, industry players and academia. Industry players and research institutes jointly issued China’s first technical specification for ICVs, which proposes evaluations for different information-security dimensions including data, network and hardware. The effort was spearheaded by Baidu’s Apollo Cyber Security Lab, which has been a key promoter of the implementation of information-security standards for ICVs in China.

Another research center affiliated with the MIIT has proposed a data-security system that classifies different types of autonomous-driving data and their associated risks. Baidu has also tackled issues related to algorithmic security, with the company’s Security X-Lab conducting extensive research on adversarial attacks that pose threats to deep-learning models in autonomous driving.

4.3 AI-enabled surveillance and the targeting of ethnic minorities

AI plays a pivotal role in the CCP’s vision of data-driven governance and control. The AIDP calls for the “intelligentization of social governance” and the enhancement of public safety and security capabilities through “intelligent monitoring.” A 2018 White Paper by the China Academy of Information and Communications Technology (CAICT) recommends that AI be more vigorously integrated with public-security work to improve social governance by automating tasks like censorship, “smart” security and public-opinion monitoring. The use of AI in public security in China has grown explosively, amid an effort to apply it to mass surveillance.

In Xinjiang, where at least one million Muslims are being held in detention camps, technology complements conventional police methods in the repression of Uighurs and other ethnic minorities. While not all surveillance methods deployed there involve advanced technologies and decisions to designate people as terrorists are still made by humans, AI plays a big role. AI companies profit from government demand for high-tech surveillance applications and tailor products and solutions to the needs of public security.
Exhibit 3

Megvii’s contradictory actions on AI ethics
AI unicorn spearheads self-regulation efforts on ethical AI while developing minority recognition software

AI ethics champion  Human rights entanglements

JULY 2019
Megvii publishes AI principles, which put forward six tenets to guide its own responsible development of AI.

The company sets up an AI Ethics Committee, which is made up of Megvii executives and external experts.

Megvii introduces an ‘AI Ethics Code of Conduct’ for employees and a ‘Protocol for the Correct Usage of AI Products’ for customers.

JANUARY 2020
Megvii establishes a dedicated AI Governance Research Institute.

JULY 2020
Co-founder and CEO gives a keynote speech at the World AI Conference’s AI Governance Forum, highlighting how his company ensures data privacy protection and ethical usage of its AI products.

2017
Megvii becomes a technical support unit of the Public Security Video Laboratory in Xinjiang.

APRIL 2019
The New York Times reports on Megvii’s involvement in the development of ethnic minority recognition software.

MAY 2019
An investigation by Human Rights Watch (HRW) reports that Megvii’s facial recognition software Face++ was found in an app used by the police for mass surveillance, predictive policing and the internment of Uighurs. (HRW later clarifies that the company appeared not to have actively collaborated in developing the app).

JULY 2019
Megvii argues it cannot prevent misuses of its AI technologies by third parties.

OCTOBER 2019
The US Department of Commerce adds Megvii to its Entity List.

DECEMBER 2020–JANUARY 2021
IPVM publishes new evidence of Megvii’s development of technology that can detect Uighur faces.

Sources: MERICS; Human Rights Watch; IPVM; The New York Times
Algorithms designed to perform racial profiling are particularly troubling. The aim is to further the government's goal of monitoring, tracking and controlling minorities, and to support predictive policing and extra-judicial confinement in detention camps. Yitu, Megvii, SenseTime and CloudWalk, among others, have been associated with the development of minority-tracking machine-learning software.\textsuperscript{55} iFlytek has provided speech recognition technology to help the police monitor communications in minority languages.\textsuperscript{56}

The ethical issue goes much beyond algorithmic bias: the government weaponizes AI against specific ethnic groups. Across China, several video-surveillance projects have mandated Uighur-detection capabilities. Even Alibaba has developed facial recognition software that can specifically detect Uighurs and other ethnic minorities in images and videos.\textsuperscript{57}

The government’s approach to AI ethics will continue to reflect priorities that, though conflicting from a liberal democratic perspective, are aligned in China’s political system. These are making AI beneficial for society, addressing citizens’ privacy concerns and positioning China as a constructive player in global AI governance while harnessing AI to safeguard political security. Companies are bound by these considerations, no matter whether they align with their own visions for how technology should be developed and used.

More importantly, the growing importance of AI in the surveillance state creates conflicting incentives for private companies. Not only is their ability to meaningfully push back against the party-state’s interests limited, but they also have political, financial and reputational incentives to develop and supply AI products tailored to the system’s needs (see Exhibit 3).

Many companies do not even seem to view the use of AI in the party-state’s public security work as troubling.\textsuperscript{58} Ultimately, Beijing’s security-centric approach risks leading to the prioritization of political goals over ethical considerations in how AI research, development and adoption will take shape in China in the years to come.

5. IMPLICATIONS FOR EUROPE: AI RESEARCH COLLABORATIONS WITH CHINESE ACTORS SHOULD BE GUIDED BY ROBUST ETHICAL STANDARDS

China wants to play an driving role in global AI ethics and governance activities. MOST’s Governance Principles emphasize “open collaboration” across borders and call for “a broad consensus on an international AI governance framework, standards and norms.” The AIDP expresses China’s ambition to drive such a consensus and strengthen international research into common AI challenges.

China’s participation in intergovernmental AI governance efforts is still limited. Some of the highest-profile multilateral efforts lack Chinese participation, which is unsurprising given their emphasis on democratic values and human rights. These include the Global Partnership on AI and the OECD Principles on AI.\textsuperscript{59}

Nevertheless, Chinese actors’ participation in international initiatives is growing. As a member of the G20, China signed on to the group’s non-binding AI principles, drawn from the OECD principles.\textsuperscript{60} At UNESCO, Chinese experts spearheaded a consensus on AI and education and contributed to an expert group drafting AI ethics recommendations.\textsuperscript{61}

Chinese actors are also driving global AI standard-setting. In 2019, Beijing hosted the first meeting of an influential AI standardization committee and Chinese companies are...
involved in ethics-related projects at leading international standards bodies. In the academic sphere, Chinese scholars conduct research with international counterparts on issues like the use of facial recognition during the Covid-19 pandemic. Researchers at tech companies also engage in international research projects on technical challenges.

China’s eagerness to engage in global AI ethics and governance initiatives could provide opportunities for the EU, where ethics is a cornerstone of the EU’s “human-centric” approach to AI. In April 2021, the European Commission proposed an AI regulation that introduces a regulatory structure centered on a risk-based classification of AI systems. Working with other countries on standards and regulations will be crucial to ensuring a beneficial AI future. And given China’s rapid advances in AI applications and the growing global reach of its companies, it will be critical for the EU to engage with Chinese actors.

Broad similarities between European and Chinese ethical interests and standards could pave the way for constructive collaboration. Some key tenets of MOST’s Governance Principles closely resemble EU guidelines, indicating that Chinese researchers are already looking to European insights and recommendations on AI. Stakeholders on both sides have shared interests in many areas, such as managing AI’s impact on sustainability or developing aligned safety frameworks for self-driving cars. In pursuing cooperation, European actors should be receptive to the multitude of ethical concerns and solutions identified by Chinese lawmakers, academics and citizens.

However, clear differences exist at the governance level that create substantive challenges and barriers to cooperation. The Chinese party-state’s prioritization of social stability and political security in its AI development is incompatible with Europe’s approach, which is rooted in the values of democracy, rule of law and respect for universal human rights. Chinese concepts like harmony and the construction of a “community of common destiny” (cited in the Beijing AI principles) are part of a political project aimed at optimizing social governance to ensure regime stability while promoting the party-state’s interests in international relations.

European policymakers must take the Chinese government’s international outreach on AI ethics with great caution. In China, as in other countries, there are notable gaps between the lofty ethical principles adopted by the government, industry and academia, and their implementation. Europe should closely watch how those differences and contradictions develop as China becomes a key player in global AI governance.

Any engagement with Chinese government actors must be informed by an awareness that political interests outweigh ethical considerations. China’s government and businesses are contributing to a global discourse that focuses on ethics without addressing crucial human rights questions. China is certainly not alone in using (and exporting) AI for surveillance, including in ways that raise serious ethical concerns. However, what sets China apart is the scale of the party-state’s ambitions to harness AI to strengthen its authoritarian governance system.

Against this backdrop, European countries must work with like-minded democracies to advance standards rooted in liberal-democratic values. They must develop robust ethical standards and guidelines for AI research collaborations with Chinese actors and use the recently reformed EU export controls to prevent European hardware, software and research from enabling unethical applications in China and elsewhere. Lastly, the EU should continue pushing Beijing to respect non-binding principles it endorsed, like the G20 AI principles.


19 | Baidu, ByteDance, Megvii, Meituan, and Xiaomi. Members of the Artificial Intelligence Industry Alliance (AIIA) and the Artificial Intelligence Industry Technology Innovation Strategic Alliance (AITISA) include Baidu, Alibaba, Tencent, JD.com, Huawei, Qihoo 360 and Yitu.


Backlash was particularly intense among citizens in Hangzhou, where officials contemplated expanding the use of those apps for long-term health assessments. See for example: Weibo (2020). Weibo user post. 


ACKNOWLEDGEMENTS

The author would like to thank the many MERICS colleagues who engaged with this report through substantive input, ideas, suggestions and feedback, in addition to reviewing previous drafts of the text.

CONTACT
Rebecca Arcesati
Analyst, MERICS
rebecca.arcesati@merics.de

EDITORIAL TEAM
Claudia Wessling
Director Communications
and Publications, MERICS
claudia.wessling@merics.de

Nick Bouchet
Freelance Editor

GRAPHICS
STOCKMAR+WALTER Kommunikationsdesign

LAYOUT
Alexandra Hinrichs
Graphic Designer, MERICS

PUBLISHER
MERICS | Mercator Institute for China Studies
Klosterstraße 64 | 10179 Berlin
Tel.: +49 30 3440 999 0
Mail: info@merics.de
www.merics.org