Research Report

JENNIFER BOUEY, LYNN HU, KELLER SCHOLL, WILLIAM MARCELLINO, JAMES GAZIS, AMMAR A. MALIK, KYRA SOLOMON, SHENG ZHANG, ANDY SHUFER

China's AI Exports

Developing a Tool to Track Chinese Development Finance in the Global South—Technical Documentation

For more information on this publication, visit **www.rand.org/t/RRA2696-1**.

Published by the RAND Corporation, Santa Monica, Calif.

About This Report

With average annual commitments reaching $85 billion in the Belt and Road Initiative era, the Chinese government is now the world's largest provider of development financing. Supported by large funding, many Chinese technology companies can deploy state-of-the-art artificial intelligence (AI) tools in development contracts in recipient countries. The AI exports facilitated by these arrangements are likely to bolster China's growing global AI technology-related supply chains, trade flows, technology standards, and regulatory systems.

This report is the technical documentation for the China's AI Exports Database (CAIED), which is a tool that tracks Chinese government-supported development finance projects that used or enabled AI technology in the Global South between 2000 and 2017. The goal of the report is to explain the motivation, data source, and methodology of building the tool. This report also provides a detailed codebook for the database and explanations on how to use the interactive world map and country selector.

RAND researchers collaborated with AidData to create a first-of-its-kind interactive tool, showing China's AI influence in the developing world. The publications in this series are the first published studies to use quantitative and qualitative metrics to examine China's AI technology exports. The analysis summarized in this series reflects data available to the research team from AidData's Global Chinese Development Finance (GCDF) Dataset version 2.0, which covers the years 2000–2017. An updated GCDF 3.0 was made available in November 2023, and the research team is working on analyzing the new dataset, which extends the time horizon covered through 2021.

Using advanced data mining tools, the team identified 155 projects enabling AI exports and described their characteristics in richer detail. Users can filter the world map by AI technology categories associated with the development finance projects and learn further details of the exported China AI project by clicking on the country of interest.

In addition, the study team produced a RAND report titled *China's AI Exports: Technology Distribution and Data Safety* (Bouey et al., 2023). The report included an analysis of the geographic distribution of China's AI exports and identified the top sectors and technologies in which AI components were focused. The report also used a country case study of Kenya and Pakistan, two of the countries that received many of China's official AI projects, to explain the characteristics of the Chinese AI projects and the recipient countries' feedback. This report showcases key insights from stakeholder interviews and provides policy recommendations for policymakers in both AI technology donor and recipient countries.

RAND National Security Research Division

The research was funded by RAND-initiated research and conducted within the Acquisition and Technology Policy Center of the RAND National Security Research Division (NSRD), which operates the National Defense Research Institute (NDRI), a federally funded research and development center sponsored by the Office of the Secretary of Defense, the Joint Staff, the Unified Combatant Commands, the Navy, the Marine Corps, the defense agencies, and the defense intelligence enterprise.

Funding

Funding for this research was made possible by the independent research and development provisions of RAND's contracts for the operation of its U.S. Department of Defense federally funded research and development centers.

Acknowledgments

We would like to thank several individuals within the RAND Corporation for providing the support and resources necessary to conduct this effort. RAND-Initiated Research Director Lisa Jaycox provided encouragement and thoughtful suggestions throughout the project. We appreciate the NSRD quality assurance team, Saci Haslam, Aaron Frank, and Jim Powers, for their coordination, reviews, comments, and approval. The report also benefited from hard work of the RAND publication team; Samantha Bennett was especially helpful for the final format of the report. All errors or omissions remain the sole responsibility of the authors.

Contents

CHAPTER 1

Background

The United States and China are competing across multiple arenas of technology, security, and global influence. Whether one is referring to the "age of artificial intelligence" (Trump White House, 2020), "one of the most powerful technologies of our time" (White House Briefing Room, 2023), or the widespread agreement among top CEOs and scientists that artificial intelligence (AI) development includes a risk of extinction (Center for AI Safety, undated), there is little disagreement that AI will be a defining technology of the 21st century. It is possible that the United States may not win the race to develop, implement, and normalize very powerful foundation models for AI applications or their use in business and government. Chinese AI implementations have been prioritized with funding and industrial policy support for decades, including investments in education, training, and subsidies.

AI development in China benefits from the central government's long-term industrial policy and local governments' heavy investment in AI-assisted surveillance tools that enhance governments' ability to identify and suppress grassroots antigovernment movements and voices of dissent. China's economic model of authoritarian capitalism included a decentralized finance mechanism for innovation that has fueled the development of China's AI industry. In addition to AI's use in surveillance, Chinese AI technology is rapidly finding applications in the military, transportation, health, education, finance, and other sectors.

China is also dominating digital infrastructure growth in developing countries, a process bolstered by the multi-billion-dollar Belt and Road Initiative, which pushed local governments and state-owned enterprises to invest more aggressively abroad. Nations—including U.S. allies—are reliant on China's technology infrastructure, such as 5G and renewable power equipment. AI technology, as a tool to enhance governance and service delivery, is integrated into the Digital Silk Road initiatives (数字丝绸之路) of Smart City (智慧城市) and Safe City (安全城市). China's AI exports, in short, are likely to continue to grow and influence the global AI technology–related supply chain, technology standards, and regulatory systems.

In this context, there are several overarching questions for the U.S. policymakers who are interested in competing with or simply understanding China's AI ecosystems. For example, how will China's AI exports affect great power competition, with implications for military, economic, and soft power? How will China's AI exports affect human rights in countries with few regulations pertaining to AI utilization? Finally, to what extent do China's AI exports further enhance its AI development advantages by providing additional data sources and funds for China's AI research? To date, no study has identified reliable data sources to systematically analyze China's AI exports.

Our work fills this void. AidData, an international development research lab based at William & Mary, has chronicled all Chinese government–supported overseas development financing activities to produce the world's most comprehensive public information source on this topic. The Global Chinese Development Finance (GCDF) Dataset version 2.0 provides granular details on 13,427 Chinese projects across 145 countries and worth $843 billion (Malik et al., 2021). For each project, this dataset details more than 70 fields of information featuring available project details, such as goals, year, amount, and the recipient; financing details, including interest rate, repayment period, and evidence of collateral; implementation features,

including precise georeferenced location and exact dates of project implementation; and more than 91,000 sources with stable hyperlinks to key documents.

Beyond its borders, and with annual commitments of more than $85 billion, China is now the world's largest provider of development financing. In recent years, China has outspent the United States on a two-to-one basis (Malik et al., 2021). This project enhances the policymaker's ability to understand and respond to this spending with original analysis that mixes financial tracking with RAND Corporation's custom-made natural language processing (NLP) tools for understanding recipient attitudes toward Chinese spending. This is necessary because Beijing does not participate in any foreign aid transparency initiatives, such as the Organisation for Economic Co-operation and Development's (OECD's) Common Reporting Standard or the International Aid Transparency Initiative. The details of how, where, and what Chinese official-sector financiers have achieved around the world are shrouded in mystery. Beijing has responded to other economic analyses by closing off previously public data sources (Wei, 2023). By focusing on information coming from recipient countries, we have a more resilient tracking method for this attempt at informal censorship.

CHAPTER 2

AidData's GCDF 2.0 Dataset

For this project, we use the 2.0 version of AidData's GCDF Dataset, which covers every major world region, every low-income and middle-income country, all sectors, and all types of financial and in-kind transfers from government and state-owned institutions in China (Malik et al., 2021).

The AidData dataset captures the known universe of projects with development, commercial, and representational intent supported by official financial and in-kind commitments (and pledges) from the Chinese government over an 18-year period (2000–2017),[1] with details on the timing of project implementation over a 22-year period (2000–2021). The dataset also assigns every project to one of three status categories (Pipeline: Commitment, Implementation, or Completed) according to sources that were available as late as August 2021.

AidData captures the full range of projects that align with the OECD's definition of official development assistance (ODA) and other official flows (OOF). Therefore, any project that benefits from financial or in-kind support from any official-sector institution in China is included, regardless of its purpose, the extent of financial concessions, funding source, and overseas destination. The only type of official financing we do not seek to capture is official investment, although we do capture debt financing that facilitates investment.

AidData captures grants, technical assistance, loans, buyer's credits, seller's credits, debt forgiveness, debt rescheduling, debt refinancing, scholarships, and training activities. By monetary value, most of the ODA and OOF transfers (*flows*) that are captured in the dataset come from official-sector loans, seller's credits, and buyer's credits. Yet the majority of the projects in the dataset represent other types of financial or in-kind support. For many flow types other than loans, seller's credits, and buyer's credits, AidData is not able to identify monetary commitment values. However, these project records still provide valuable information for users who are interested in capturing the full set of Chinese ODA- and OOF-financed activities in a world region, country, or subnational area.

The AidData dataset records all officially financed projects and activities related to ODA and OOF, regardless of the sector that they support. AidData coders follow the OECD's classification guidelines to identify the three-digit sector that a given project or activity is meant to support (OECD, undated).

AidData used the 2.0 version of the Tracking Underreported Financial Flows (TUFF) methodology to systematically search for projects supported by official financial and in-kind transfers from China across 165 countries and territories (Custer et al., 2021). The resulting dataset covers official financial and in-kind transfers from China to every low-income, lower-middle-income, and upper-middle-income country and territory across every major world region, including Africa, Asia, Oceania, the Middle East, Latin America and the Caribbean, and Central and Eastern Europe. The dataset also covers 11 high-income countries that

[1] RAND researchers collaborated with AidData to create a first-of-its-kind interactive tool, showing China's AI influence in the developing world. The publications in this series are the first published studies to use quantitative and qualitative metrics to examine China's AI technology exports. The analysis summarized in this series reflects data available to the research team from AidData's Global Chinese Development Finance Dataset version 2.0, which covers the years 2000–2017. An updated GCDF 3.0 was made available in November 2023, and the research team is working on analyzing the new dataset, which extends the time horizon covered through 2021.

were included to help ensure comprehensive coverage in each world region to the extent possible. In total, the dataset identifies Chinese government–financed projects in 145 countries and territories, meaning that no projects were identified in 20 countries and territories despite systematic searches.

For this project, we merged AidData's Global Huawei Finance Dataset and AidData's Global Military Finance Dataset with AidData's GCDF 2.0 Dataset to include Huawei Technologies Co. Ltd. ("Huawei")-financed projects and projects with military intent that are backed by official financial flows. The GCDF 2.0 dataset does not include projects backed by official financial transfers with military intent or projects financed with aid or debt from Huawei or any of its subsidiaries because of a lack of agreement about whether Huawei should be treated as an official source of financing. However, given that some analysts and decision-makers have expressed interest in learning more about Huawei-financed projects and projects with military intent that are backed by official financial flows, we have included these *ODA-adjacent* and *OOF-adjacent* projects in the new dataset.

Within this merged dataset, we also used AidData's "Recommended for Aggregates" filter to include only formally approved, active, and completed Chinese government–financed projects and exclude all canceled projects, suspended projects, and pledged projects that never reached the Formal Approval (official commitment) stage. This filter also removes all overarching *umbrella agreement* projects in which funds were not allocated for a specific project or purpose (or set of projects or purposes) until a subsequent date and, thus, have subsidiary projects or transactions that are likely captured elsewhere in the dataset. These types of umbrella agreements include economic and technical cooperation agreements issued by China's Ministry of Commerce, master facility agreements issued by China Eximbank, lines of credit issued by China Development Bank, and framework agreements issued by a variety of official-sector institutions in China. This filter is useful to conduct analysis that requires the aggregation of projects supported by official financial (or in-kind) commitments from China—including analysis of monetary amounts and project counts—and to avoid double-counting.

CHAPTER 3

Methodology for Building the China's AI Exports Database

We built the China's AI Exports Database (CAIED) based on the AidData GCDF 2.0 Dataset. In the following sections, we explain the steps of constructing the CAIED: searching AI projects from the GCDF dataset, generating new AI application indicators, selecting project-level variables, and adding additional country-level information.

Phase I: Developing String Lists for Searches

We conducted an initial review of the AidData GCDF 2.0 dataset, extracting the relevant strings identified during data collection activities. The initial string used for this preliminary scan was as follows: remote medicine, machine learning, algorithms, closed-circuit television (CCTV), Smart City, Safe City, surveillance. The preliminary scan returned a list of projects ($N = 70$), which the study team used to develop the test string list for the project search.

By reviewing and screening the project descriptions of the 70 projects, the study team extracted a list of strong keywords and a list of weak keywords (see Table 3.1). The study team was also able to extract keywords specifically related to Chinese AI technology from the *China AI Development Report* (China Institute for Science and Technology Policy at Tsinghua University, 2018). Using these identified keywords, the study team conducted a further literature search on Google Scholar and extracted keywords relating to Chinese AI from multiple recently published studies (Amazon Web Services, undated; Gaikwad, Gawali, and Yannawar, 2010; Leung, Braun, and Cuzzocrea, 2019; Nadkarni, Ohno-Machado, and Chapman, 2011; Tao and Tan, 2005; AI Index, 2021; Wang et al., 2012; Zhao et al., 2003). Finally, the study team consulted more than ten researchers and professionals working in the technology industry to suggest additional keywords that might help us identify AI projects from the AidData GCDF 2.0 Dataset.

We ran an initial search within AidData's GCDF 2.0 Dataset using all of the keywords and extracted strings with positive matches. These strings were combined into a test string list for the test run: CCTV, smart city, safe city, surveillance, facial recognition, machine learning, image, e-governance, data, remote, algorithm, video, consultation, AI (Artificial Intelligence), recognition, ICT (information and communication technology), information system, network, monitoring, software, camera, OCR (optical character recognition), scanning, computing, digital, drone, 3D, natural language, chip, sensor, semiconductor, robot, laser, radar, network security, urban security, information technology, IoT (Internet of Things), blockchain, object detection, automation, unmanned, UAV (unmanned aerial vehicle), augmented reality.

To address case sensitivity in string detection, we employed a function to convert all project description text to lowercase during the string detection process. The initial results from the test run reported 2,412 projects, with only 30 strings returning valid output: CCTV, smart city, safe city, surveillance, image, data, remote, algorithm, video, consultation, recognition, ICT, information system, monitoring, software, camera,

TABLE 3.1

Summary of Keywords from Literature Review

Source	Strong Keywords	Weak Keywords
Keywords extracted from preliminary scan	intelligent applications	surveillance, CCTV
		video surveillance
	intelligent video surveillance	information system
		security surveillance network
		video monitoring system, software
		video surveillance
		Safe City, big data
		smart camera
	Intelligent alert system	remote monitoring
Keywords extracted from the *China AI Development Repor*t (China Institute for Science and Technology Policy at Tsinghua University, 2018)	speech recognition, speech synthesis, voiceprint recognition, human-machine dialogue	speech recognition, speech synthesis, speech interaction, speech evaluation, human-machine dialogue, voiceprint recognition
	facial recognition, visual recognition, affective computing, expression recognition, behavior recognition, gesture recognition, body recognition, mobile vision, OCR, handwriting recognition, text recognition, image processing, image recognition, pattern recognition, spatial recognition	biometrics (face recognition, iris recognition, fingerprint recognition, vein recognition, etc.), affective computing, emotion recognition, expression recognition, behavior recognition, gesture recognition, body recognition, video content recognition, object and scene recognition, mobile vision, optical character recognition (OCR), handwriting recognition, text recognition, image processing, image recognition, pattern recognition, SLAM, spatial recognition, 3D scanning, 3D reconstruction, etc.
	natural language processing	natural language interaction, natural language understanding, semantic understanding, machine translation, text mining (semantic analysis, semantic computing, classification, clustering), information extraction, human-machine interaction
	basic algorithm and platform	machine learning, deep learning, open source framework, open platform
	basic hardware (semiconductor, GPU)	chips, lidars, sensors, etc.
	basic enabling technology	cloud computing, big data
	intelligent robotics (including solutions)	industrial robotics, service robotics, personal/home robotics, home security robot, in-vehicle robot
	smart driving (including solutions)	intelligent driving, driverless driving, autonomous driving, assisted driving, advanced driver assistance system (ADAS), laser radar, ultrasonic radar, millimeter wave radar, GPS positioning, high-precision map, vehicle chip, human-car interaction
	drone (including solutions)	consumer drones, professional drones
	AI+	finance, insurance, judiciary administration, entertainment, tourism, healthcare, education, logistics and warehousing, smart home, smart city, network security, video surveillance, commerce, human resources, corporate services

Table 3.1—Continued

Source	Strong Keywords	Weak Keywords
Keywords extracted from literature review	speech recognition	speech processing, signal processing, pattern recognition
	face recognition	face recognition, face perception, facial recognition techniques
	affective computing	smart surveillance, perceptual interface, human-computer interaction, emotional speech processing, prosody
	pattern recognition	
	text recognition	unsupervised, feature learning, text detection, character recognition, text recognition, pretraining
	natural language	natural language, NLP, lexer
	deep learning	convolutional nets, recurrent nets, deep learning, image processing, video processing, linear classifier, multilayer architectures, backpropogation
	sensors	sensor information fusion, supervised learning, data mining
		expert systems, fuzzy systems, natural language processing, speech recognition, pattern recognition, computer vision, decision-support systems, knowledge bases, neural networks
	intelligent robotics	smart technology, artificial intelligence, robotics, and algorithms (STARA)
		general artificial intelligence, narrow artificial intelligence
		AI robotics
	smart driving artificial intelligence	internet of things (IoT), physical internet (PI), industry 4.0
		blockchain-AI, intelligent transportation systems, intelligence transport system
	drone artificial intelligence	drones, object detection, feature extraction, detectors, classifier, deep learning, detection, image processing
		drone network, drone communication, drone communication security
	AI+ artificial intelligence	radiology, computed tomography, image analysis
	China AI development mechanisms	representation learning, deep learning, supervised learning, unsupervised learning healthcare AI applications, AI-based screening and referral system Arterys, IDx-DR, Guardian Connect (Medtronic)
		Chinese language information processing
		new concept weapons, cyber warfare automation in manufacturing, digital environmental protection, social credit system
		internet plus
		computer vision, natural language processing, trans-media analysis and reasoning, intelligent adaptive learning, collective intelligence, automated unmanned systems, intelligent chips, brain-computer interfaces
		unmanned aerial vehicles (UAVs), voice and image recognition

Table 3.1—Continued

Source	Strong Keywords	Weak Keywords
	Additional Keywords	
NLU, natural language generation (NLG), autonomous vehicles, self-driving cars, chatbot(s), large language models (LLMs), training data, test data, validation data, data science, modelling, augmented reality, long short-term memory (LSTM), support vector machine (SVM), t-distributed stochastic neighbor embedding (TSNE), automated reasoning, ensemble methods, feature learning, feature selection, generative adversarial network, machine perception, Monte Carlo, swarm AI/swarm intelligence, entity recognition/named-entity recognition (NER), dimensionality reduction, principal component analysis (PCA), python, reinforcement learning (RL), dataframe(s), Turing, big data, recommendation system, data mining, page rank, search engine, anomaly detection, sentiment analysis, graph optimization		

NOTE: AI+ = AI + human intelligence; GPU = graphics processing unit; NLU = natural language understanding; SLAM = simultaneous localization and mapping.

OCR, scanning, computing, digital, 3D, chip, sensor, robot, laser, network security, information technology, IoT, automation, UAV.

After summarizing the statistics from the frequency table of the strings, we identified two issues related to false positive results. The first issue is nonexact matches triggering false positives for many strings. During the analysis of the strings used in the test run, it was observed that nonexact matches were causing false positives for several strings. Examples of these false positives include the following: The string "iot" returned matches such as "antibiotics," "riots," and "physiotherapy" instead of the intended meaning; the string "ict" returned matches such as "victims," "district," and "conflict" instead of the intended meaning; the string "data" returned matches such as "AidData" and "Chinese Loans to Africa Database" instead of the intended meaning.

To address this issue, we implemented a solution by creating a new section in the string processing script. This new section was designed to distinguish between strings for vague matches and strings for strict matches. For the strict matches, the team added "\\b" as boundaries for the strings, enabling strict matching. The vague match strings list includes strings such as "surveillance," "automation," "robot," "recognition," "e-governance," and "facial recognition." On the other hand, all other strings were set to strict match and constructed with boundaries in the script.

The second issue identified is that certain strings in the initial list were too broad in their scope. To address this issue, the study team reviewed and refined the strings and narrowed their focus. The refined string list that emerged from this collaboration is as follows: CCTV, smart city, safe city, surveillance, facial recognition, machine learning, image, e-governance, data center, government data, remote, algorithm, video, consultation, AI, artificial intelligence, recognition, ICT, information system, neural network, monitoring, software, camera, OCR, scanning, computing, digital environmental protection, drone, 3D, natural language, chip, sensor, semiconductor, robot, laser, radar, network security, urban security, information technology, IoT, blockchain, object detection, automation, unmanned, UAV, augmented reality, CT (Computed Tomography), PET (Positron Emission Tomography), MRI (Magnetic Resonance Imaging), ultrasound.

These refined strings provide a more targeted and specific representation of the concepts and technologies related to the analysis of Chinese development finance projects.

Phase II: Adjusting String Lists and Conducting Manual Reviews

Test runs were performed using the new string detection script and modified string list. We further enhanced the analysis by merging AidData's military projects and Huawei projects with the Chinese development finance projects. The combined dataset was used for a second round of string detection. In this round, 697

projects were analyzed, and 36 strings returned valid output. To ensure the accuracy of the results, the study team established a manual review workflow for thorough screening of the project outcomes.

Following a meticulous manual review of the 697 projects and excluding those that were clearly not related to AI (false positives), the final result revealed 81 projects categorized as "yes" (confirmed AI application) and 101 projects classified as "maybe" (potential AI application). The 81 "yes" projects were projects that clearly used AI technologies, such as speech recognition, facial recognition, NLP, deep learning, intelligent robotics, and drones. We also consulted experts with specialty in AI to confirm that the "yes" projects used AI technology (see the next section for details). The 101 "maybe" projects were projects that can possibly include AI technology, but the specific technology was not explicitly mentioned in the project description.

Phase III: Refining Search Keywords and Conducting Recovery Search

To make sure that our string list is comprehensive, we consulted researchers who have expertise in AI technology applications in four sectors to which most of the "yes" (confirmed AI application) projects belong: communication, government and civil society, health, and education. We specifically asked about the experts' suggestions on the comprehensiveness of the string list. We also sought their opinions on how to code the "maybe" projects by providing them with the project descriptions and the potential AI application technologies.

From experts' responses, we were able to collect the additional search keywords:

- general AI keywords: cat scanner, broadband, fiber optic, biometric, smart
- communication-sector AI keywords: 4G, 5G, 6G, radio, transceiver, training (chips), Global System for Mobile communication, Code Division Multiple Access (CDMA), crypto, Coalition for Content and Provenance and Authenticity (C2PA), computational thinking, digitization
- education-sector AI keywords: intelligent tutoring, language models, education technology, automated essay scoring, latent semantic analysis
- health-sector AI keywords: diagnosis, quality improvement, health information technology (HIT), electronic health record (EHR), electronic medical record (EMR), clinical decision support (CDS, or sometimes CDSS for clinical decision support systems), digital health, smart devices, medical device, digital medicine, eHealth and variants (e-Health, etc.)
- government-sector AI keywords: disaster, disaster risk reduction, community resilience, Sendai framework, preparedness, mitigation, response, recovery, UN Disaster Risk Reduction (UNDRR), humanitarian action, triple nexus, Grand Bargain.

Our study team screened the additional keywords and created the following string list to conduct a recovery project search: 5G, 6G, CDMA, broadband, digitization, computational thinking, C2PA, biometric, crypto, e-health, eHealth, digital medicine, smart devices, health information technology, electronic health record, digital health, generative, automated, intelligent tutoring, clinical decision support, CAT scan.

We then screened all the projects from this recovery search and filtered out 51 additional "maybe" (potential AI application) projects.

Phase IV: Generating the AI Application Indicator

Refining Coding Criteria and Rules

Since there is no universal rule in defining AI technology, we generated an indicator to differentiate between two AI project categories. One category contains *AI application* projects that directly use AI technologies, such as facial recognition, speech recognition, and algorithms that facilitate medical diagnosis. The other contains *AI infrastructure* projects, such as building data centers, connecting 5G network technology, laying fiber optical cables, training in AI technology, and installing AI hardware (such as CCTV cameras) that provide the necessary platform for AI deployment.

Using the suggestions and feedback from the experts we consulted, we were able to develop a set of rules for coding the "maybe" projects into AI application projects or AI infrastructure projects. The study team also paid special attention to the implementing agencies or companies and the implementation year to determine whether a project might have any AI components.

AI Rating Scales (Levels 1–5)

1. something for which AI is impossible or irrelevant
2. something that could have AI applied to it (e.g., broadband)
3. something that is AI-empowered or -powered, but most of the value is not AI (e.g., X-ray scanners)
4. something explicitly for AI use (e.g., CCTV for smart cities)
5. data center for AI.

Coding Rules on the AI Application Indicator

- We excluded **level 1** projects that belong to the following categories:
  - e-government projects that implement only documentational management systems
  - projects that involve only training of human capital (not specifically mentioning AI content in training)
  - medical devices and technologies that have no AI components (all the health projects except the ones identified above in the "yes" and AI infrastructure categories)
  - disease surveillance projects in the database that were implemented quite early and were designed for specific countries
  - generalized equipment, such as equipment for a mining project or a hospital, without additional detail
  - projects that are mostly non-AI, such as an electricity grid, oil rigs, or a hospital building (even if the project has a small AI software component) and are not implemented by a Chinese AI company[1]
  - CDMA or radio station projects that do not mention any AI component.
- We coded **level 2** and **level 3** projects that belong to the following categories as AI infrastructure projects:
  - projects that involve the building of data centers
  - CDMA or radio station projects that were implemented after 2012 and specifically mention any AI components

[1] We gathered a list of Chinese AI companies from multiple sources, including various research institutions' reports and articles, AI company rankings on private company websites, and AI expert consultations.

  - communication projects that installed any fiber optics, signal towers, satellites, or other supporting infrastructure for AI
  - e-government projects that built electrical management systems with AI technology
  - projects that installed CCTV cameras for detection and recognition purposes
  - projects on human capital training programs that especially involve AI technology or were implemented by a China AI company
  - projects that involve container scanners manufactured by a Chinese AI company
  - seismic monitoring projects that collect data that have a very high likelihood of being processed by machine learning models
  - medical imaging projects that could theoretically involve AI technologies.
- Coded **level 4** and **level 5** projects that belong to the following categories as AI application projects:
  - projects that substantially involve biometric identity detection, anti–electricity theft, or anomaly detection technologies
  - health projects that substantially involve automated radiology
  - safe and/or smart city projects that either substantially involve surveillance cameras or were implemented by a Chinese AI company.

Coding the "Maybe" Projects from the Manual Review

According to these rules that we developed and refined, three analysts of the team individually coded the 101 "maybe" projects using the AI application indicator. The codes from the three analysts are the same for 81 "maybe" projects. For the remaining 20 "maybe" projects that involve discrepancies in the codes, the analysts compared and discussed the code and used the majority result as the final code (e.g., if two analysts code a project as an AI application project and one analyst codes it as an AI infrastructure project, the final AI application indicator for this project will be AI project).

Coding the "Maybe" Projects from the Recovery Search

In addition, the study team followed the same steps to code the AI applications indicator for the 51 "maybe" projects screened out from the recovery search. All of the 51 projects fall under the exclusion criteria.

This manual review process allowed us to exclude a total of 549 projects from the 697 projects that returned from the adjusted search. We selected 94 AI application projects and 54 AI infrastructure projects to include in the CAIED.

Salvaging Projects That Were Originally Excluded

To ensure that the original manual coding that we conducted also followed the same rule, the study team conducted another manual review of all the 549 projects that were excluded from the original manual review. We found seven projects that were recorded as AI infrastructure projects according to the refined coding criteria and rules we created. Four of these AI infrastructure projects are surveillance projects that involved installations of cameras. The other three projects belong to e-government, medical imaging, and security scanners. These seven AI infrastructure projects were incorporated into the CAIED.

Finalizing the Selection of AI Application and AI Infrastructure Projects

Lastly, the study team conducted a manual review of any potential duplicate projects to confirm that projects with similar purposes and sectors were, indeed, separate financial transactions. In the final CAIED, we

include 94 AI application projects and 61 AI infrastructure projects (54 from manually coding the "maybe" projects and seven from salvaging the excluded projects).

Phase V: Selecting Variables from the AidData GCDF 2.0 Dataset

The AidData team and RAND jointly selected a subset of 15 key variables from the broader AidData GCDF 2.0 dataset. This includes **Recipient Country**, **Title**, **AidData Project ID**, **Sector**, **Amount**, **Implementation Status**, **Commitment Year**, **Completion Year**, **Description**, **Flow Type**, **Concessionality**, **Flow Class**, **Funding Agencies**, **Receiving Agencies**, and **Implementation Agencies**. We shortened the AidData project titles for concision. These key variables offer a comprehensive overview of the individual AI projects included in the dataset while also providing granular details, such as the implementation timeline and agencies involved in the project. The variables showcase financial and in-kind transfers from a wide range of official donors and lenders; capture the terms and conditions of these financial flows; track the implementation of projects over time and geographic space; and include detailed narrative descriptions that explain how China's AI projects are being designed, implemented, monitored, and evaluated in practice. Reporting specifics on the financial flow amount, type, class, and sector, consistent with OECD-Development Assistance Committee (DAC) reporting directives, allows users the opportunity to compare China's financing with other sources of development finance, including members of the OECD-DAC and multilateral institutions.

Phase VI: Adding in Technology Categories and Country-Level Information

To help users better understand the content of the projects at a glance, we created the technology category variable using the information from the detailed project description provided in the AidData GCDF 2.0 dataset. The technology category variable currently includes the eight most commonly used general AI technology types of Chinese AI exports in the dataset, including Advanced Computing and Data Storage, e-Government, Medical Imaging, Remote Sensing and Seismic Monitor, Safe and Smart City, Security Scanners, Unmanned Vehicles, and Other. The Other category includes educational AI-training projects.

In addition, we created an indicator showing the purpose of AI technology usage using the information from the detailed project description and flow class provided in the AidData GCDF 2.0 dataset. The usage variable currently has two categories, Military and Civilian.

To provide users with context, we added several columns from outside AidData. We included three metrics of freedom and democracy and two metrics of data protection.

Because of the perceptions of China and the impact of AI-powered surveillance, we included three metrics of the political state of the country. The first was the **political regime** (of the year 2022), drawing from Our World in Data, which, in turn, relied on results from the Varieties of Democracy (V-Dem) Institute democracy indices in the *Democracy Report 2022* (Herre, Ortiz-Ospina, and Roser, undated; V-Dem Institute, 2022). The well-respected V-Dem indices are developed by political scientists and funded by the World Bank and several governments: V-Dem is a credible and independent source. The second, the **Electoral Democracy Index** (of the year 2021), draws on the same source and represents an aggregate score based on freedom of association, clean elections, freedom of expression, elected officials, and suffrage. The third was the **Freedom House rating**, "the most widely read and cited report of its kind": It is less precise but more popular and so was also included (Freedom House, undated).

The United Nations Conference on Trade and Development keeps track of which countries have introduced or passed data protection laws—Data Protection Metrics. These were incorporated in our **Data Pro-**

tection Law Status column (United Nations Conference on Trade and Development, undated). Additionally, the law firm DLA Piper grades every country's data protection level on a more detailed scale. We replicate that scoring here. A second metric from an independent source, DLA Piper's mapping of the levels of global data protection laws, increases confidence in results, and the greater precision of the DLA Piper data may be useful for those who are less concerned about the risks of a potentially biased source, such as an independent law firm with business in the countries in question (DLA Piper, undated).

All data acquired from the external data sources listed above were last accessed by the authors as of April 28, 2023.

CHAPTER 4

Codebook

Variables Inherited from the AidData GCDF 2.0 Dataset

Recipient. This field captures the country in which the entity receiving the official financial or in-kind transfer is located. If multiple entities from multiple recipient countries are involved, this field records the geographical region to which the recipient countries belong.

AidData ID. This field provides the unique identification number that AidData has assigned.

Sector. This field provides a sector name indicating the primary sectoral focus of the project. It is founded on the OECD's sector categorization scheme. There are 12 three-digit OECD sector codes: agriculture, forestry, and fishing (310); communications (220); disaster prevention and preparedness (740); education (110); emergency response (720); government and civil society (150); health (120); industry, mining, and construction (320); other multisector (430); other social infrastructure and services (160); trade policies and regulation (330); and transport and storage (210).

Amount (2017 U.S. dollars, millions). This field captures the monetary value of the official commitment (or pledge) issued by the funding agency in constant 2017 U.S. dollars. To calculate this value, AidData first converts the financial commitment amount in its original currency of denomination to nominal U.S. dollars at the average exchange rate in effect during the commitment year and then converts this amount to constant 2017 U.S. dollars using the OECD's deflation methodology to adjust for inflation and ensure comparability over time and space.

Implementation Status. This field identifies the latest status of a project. In this database, projects are assigned to one of three categories: Pipeline: Commitment, Implementation, or Completion. A project assigned to Pipeline: Commitment status is one that is backed by an official commitment (i.e., a binding, written agreement that governs the provision of financial or in-kind support for a specific project or purpose was signed by an official-sector donor or lender in China and an entity in a recipient country) but has not yet entered implementation. A project assigned to *Implementation* status is one that is backed by an official commitment and has begun implementation with financial or in-kind support from the source of the commitment. A project assigned to *Completion* status is one that is backed by an official commitment and that reached completion with financial or in-kind support from the sources of the commitment. For the purpose of this study, we used the Recommended_for_Aggregates filter in AidData's GCDF 2.0 Dataset to exclude all pledged, suspended, and canceled projects. A project assigned to the *Pipeline: Pledge* category in AidData's GCDF 2.0 is one that an official-sector institution in China indicated that it was interested in supporting (or willing to consider supporting), but that interest did not result in an official commitment. Projects assigned to this category include those that are identified in letters of intent, term sheets, memoranda of understanding, and nonbinding announcements. Projects assigned to the *Suspended* and *Canceled* categories are those that were backed by an official commitment but subsequently suspended or canceled. The coding of the "Status" field in the 2.0 dataset is drawn from sources that were available as late as September 2021.

Commitment Year. This field captures the year in which an official financial commitment (or official commitment to provide in-kind support) was codified through the signing of a formal agreement by an offi-

cial donor or lender in China and one or more entities in a recipient country or a set of recipient countries. In the event that an official commitment was made for a project that entered implementation but the official commitment year is not identifiable, AidData records the first year of project implementation as a proxy for the official commitment year. In the event that an official commitment was made for a project that has not yet reached implementation and the official commitment year is not identifiable, AidData records the year in which the underlying commercial contract (supported by the official commitment) was issued. If this information is unavailable, AidData records the first year in which an informal pledge was made as a proxy for the official commitment year.

Completion Year. This field captures the year in which a project supported by an official financial (or in-kind) commitment from China was completed. Whenever possible, this field reflects the precise calendar day when a project was completed, which is captured in the "Actual Completion Date" field. For projects that involve the construction of buildings or infrastructure, the "Completion Year" field is intended to capture the last year of construction. In cases in which the last year of construction is unavailable but a proxy for the last year of construction (e.g., a road or railway is opened for use, a power plant reaches its commercial operation date and begins selling electricity to customers) can be identified, AidData records the proxy for the last year of construction. For projects that do not involve construction but involve the provision of personnel, training, analytical or advisory support, equipment, supplies, or commodities, the "Completion Year" field captures the last year in which some type of support was delivered to an entity (or set of entities) in the recipient country. For projects that involve only financial transactions (cash donations, loans issued to shore up foreign exchange reserves, forgiveness or rescheduling of outstanding debts), the "Completion Year" field captures the year in which the last disbursement was made (or the year in which new terms and conditions went into effect for a previously signed loan or export credit agreement).

Description. This field provides a detailed summary of the main purposes and activities of the project; the funding, receiving, and implementing agencies involved in the project; the terms and conditions of the financial transaction(s) supporting the project; the timing of project implementation and completion; the challenges that arose during project implementation and how funding, receiving, and implementing agencies responded to these challenges; and main achievements and shortcomings of the project. For loan- and export credit–financed projects, AidData also records the monetary value and timing of underlying commercial contracts, disbursements, and repayments in this field.

Flow Type. This field captures the type of financial or in-kind transfer supporting the project. Each project is assigned to one of nine categories: loan, export buyer's credit, supplier's credit/export seller's credit, debt forgiveness, debt rescheduling, grant, scholarships/training in donor country, freestanding technical assistance, and vague (to be determined [TBD]) if there is insufficient information to specify a flow type.

Concessional. This marker designates whether the flow is concessional in nature or not. For loan- and export credit–financed projects, this marker identifies whether the loan terms were provided on concessional terms or not. The concessionality determination is based on OECD's grant element calculator and 25-percent grant element threshold. For grants, scholarships, technical assistance, training activities, and debt forgiveness, the concessional marker is set to "yes." If the concessionality is unknown, this marker is set to "vague."

Flow Class. According to OECD-DAC guidelines for ODA and OOF, this field assigns projects to one of three categories: *ODA-like*, *OOF-like*, or *vague* (official finance)(OECD, undated-a). Projects are assigned to the ODA-like category if they meet three criteria. First, the primary purpose of the project must be the promotion of economic development and welfare in the recipient country. Second, the official commitment supporting the project must be concessional in nature (i.e., grant, technical assistance, scholarship, debt forgiveness, or loan with a grant element of at least 25 percent). Third, the project must take place in a country that qualifies for ODA based on its income level. Projects that are supported by an official financial or in-kind transfer but do not meet all three of these criteria are assigned to the OOF-like category. Projects that

are backed by an official commitment but cannot be reliably categorized as ODA-like or OOF-like because of insufficiently detailed information are assigned to the vague (official finance) category. Projects in this residual category primarily consist of (a) those with an unspecified flow type (i.e., values of vague TBD) and (b) those financed with development-intent loans for which AidData lacks the borrowing terms (interest rates, grace periods, or maturity dates) needed for concessionality determinations.

For Huawei-financed projects, the flow class is assigned as *corporate-aid (CA)* for any donations, grants, and freestanding technical assistance given by Huawei and *vague (commercial)* for any loans and export credits from Huawei.

For military-financed projects, the flow class is assigned as *military*, consistent with OECD-DAC reporting directives, if they seek to promote the security interests of the country from which the financial transfers originate or strengthen the lethal force capabilities of military institutions in the recipient country. When military institutions are involved in a project with a humanitarian or development purpose, these are not coded as military flow class. Other activities that involve the military but do not qualify as military flow class include peacebuilding and peacekeeping operations (when peacekeepers are executing specific development-related activities, such as medical personnel providing medical training and medical care, engineering personnel deployed to build transportation infrastructure); security system management and reform efforts that reinforce civilian control; removal of landmines and explosive remnants of war; reintegration of demobilized military personnel into the economy; conversion of production facilities from military to civilian outputs; repatriation and demobilization of armed factions and disposal of their weapons; technical cooperation to control, prevent, and/or reduce the proliferation of small arms and light weapons; and efforts to demobilize, disarm, release, reintegrate, repatriate, and resettle child soldiers. The direct provision of equipment or funds to military institutions (e.g., defense ministries, different branches of the armed forces) do not meet the OECD-DAC reporting directives for development-related purposes and are identified as military projects in AidData's dataset.

Funding Agencies. This field captures the name of the agency that issued the official financial or in-kind commitment. The agency identified in this field must be based in the country (e.g., China) from which the official financial or in-kind commitment originated. For projects assigned to the *Pipeline: Pledge* category, this field captures the name of the official-sector agency that issued the pledge. The same "origin rule" applies to funding agencies that issued pledges rather than commitments. If multiple Chinese funding agencies are involved, the entries are pipe-delimited Unicode string delimited by a pipe "|" character.

Receiving Agencies. This field provides the name of the agency designated to receive and manage the financial or in-kind transfer. For projects that are financed with loans or export credits, the receiving agency is the entity responsible for debt repayment. If a receiving agency (borrower) on-lends the proceeds of a loan or export credit to an additional entity or entities, all entities with formal repayment responsibilities are identified in the "Receiving Agency" field (as pipe-delimited entries). Likewise, if more than one entity is responsible for receiving and managing incoming grant funds or an in-kind transfer, all of these entities are identified in the "Receiving Agency" field (as pipe-delimited entries).

Implementing Agencies. This field provides the name of the agency responsible for implementing the project activities. If more than one agency is responsible for implementing the project activities, all such agencies are identified in the "Implementing Agencies" field (as pipe-delimited entries).

Variables Constructed by the Study Team

Short Title. This field briefly describes the name or nature of the project by extracting the most-important information from the project descriptions. The identification numbers of other transactions that are linked to the project are also recorded in this field.

AI Application. This field can be *AI application* or *AI infrastructure*. The latter refers to projects, such as satellite or earthquake monitoring support, that do not directly use AI but are deeply integrated with its use and will lose some functionality without AI (please see the Phase IV section of Chapter 3 for details).

Technology Category. This field can take the values "advanced computing and data storage," "e-government," "medical imaging," "remote sensing and seismic monitor," "safe and smart city," "security scanners," "unmanned vehicles," and "other." They serve as descriptions of some of the most common types of Chinese AI exports in the dataset.

Variables Acquired from External Sources

Political Regime (2022). This field can be "closed autocracy," "electoral autocracy," "electoral democracy," or "liberal democracy."

Electoral Democracy Score (2021). This field has another ranking for the degree to which the country is an electoral democracy.

Freedom House Rating. This field is an alternative scoring for the freedom level of a country: free, partly free, or not free.

Data Protection Law Status. This field logs whether data protection laws are in effect, under discussion, or do not exist.

Data Protection Law Level. This field can be "limited," "moderate," "robust," or "heavy," corresponding to an assessment of how strong data privacy laws are in the country.

Please see Phase VI in Chapter 3 for details.

CHAPTER 5

Features of the Map and Link to the Database

In the interactive world map, users can see all 155 AI application or AI infrastructure projects exported from China between 2000 and 2017. The height of each spike reflects the count of identified projects per country, and the intensity of color for each spike shows the total transaction amount for all projects in the selected country. Users can choose to view both AI application and AI infrastructure projects, just AI application projects, or just AI infrastructure projects by applying the filter on the top left of the world map. Users can also access detailed project-level and country-level information by applying the "Browse by Country" filter below the world map.

When the "Browse by Country" filter is applied, a country profile that showcases selected information from the CAIED will appear below the world map. In the individual country's profile, users will be able to see the country's democracy, freedom, and data protection status at a glance. Users will also see the distribution of projects by technology category in the bar chart. In addition, users can view the short title, ID, AI application indicator, technology category, sector, aid amount, implementation status, commitment year, completion year, and full description for each of the projects in the country.

Users can access the full CAIED by navigating to the URL provided (www.rand.org/t/TLA2696-1) to view the full set of variables in the database (see Chapter 4 for details).

Users can also acquire more information on any individual project of interest by accessing the original CAIED through the file download link provided in the map description. Using the AidData ID provided in the CAIED, users can look up projects by searching the project ID in the downloaded AidData GCDF 2.0 dataset (AidData, 2021) under the variable AidData TUFF Project ID or by searching the project ID on the AidData project view page (AidData, undated).

CHAPTER 6

Next Steps

AidData's GCDF 2.0 dataset, used for this project, tracks financing from official entities in China and covers financial commitments up to 2017, with implementation details up to late 2021. However, the 2.0 version of AidData's dataset is unique in that it captures the full range of projects that align with the OECD definitions of ODA and OOF. Any project that benefits from financial or in-kind support from any official-sector institution in China is included. The projects in the 2.0 dataset are supported by 334 unique official-sector institutions in China, including central government agencies (e.g., the Ministry of Commerce, the Ministry of Foreign Affairs, the Ministry of Agriculture), regional and local government agencies (e.g., Chongqing Municipal Health Commission, Tianjin Municipal Government), state-owned enterprises (e.g., China National Petroleum Corporation, China National Aero-Technology Import and Export Corporation, China Machinery Engineering Corporation), state-owned policy banks (e.g., China Development Bank, China Eximbank), state-owned commercial banks (e.g., Bank of China, China Construction Bank, the Industrial and Commercial Bank of China), state-owned funds (e.g., the Silk Road Fund), and nonprofit government organizations (e.g., Hanban, the China Foundation for Poverty Alleviation). Other existing datasets capture only official financial transfers from China to a single sector (e.g., energy) or region (e.g., Latin America) or track only certain types of financial flows (e.g., loans) and funding sources (e.g., China's policy banks).

Moreover, we included AidData's Global Huawei Finance Dataset and AidData's Global Military Finance Dataset for analysis. In future iterations of the project, we will seek to expand the scope of China's development finance projects that are scanned for potential AI applications. For this project, we used the Recommended_for_Aggregates filter in AidData's GCDF 2.0 dataset to exclude all pledged, suspended, and canceled projects. It is important to systematically track these projects and subject them to analysis, so an expansion of this project could help provide an even more complete picture of China's overseas development finance portfolio as it relates to AI.

In November 2023, AidData will release data on China's official finance for the years 2000 to 2021, with implementation details up to the calendar date in 2023. Therefore, the next version of this project will benefit from incorporating the newest data on China's official financial commitments.

Abbreviations

AI	Artificial intelligence
CA	Corporate-aid
CAIED	China's AI Exports Database
CCTV	Closed-circuit television
DAC	Development Assistance Committee
GCDF	Global Chinese Development Finance (Dataset)
ICT	Information and communication technology
IoT	Internet of Things
NLP	Natural language processing
OCR	Optical character recognition
ODA	Official development assistance
OECD	Organisation for Economic Co-operation and Development
OOF	Other official flows
TUFF	Tracking Underreported Financial Flows
UAV	Unmanned aerial vehicle
V-Dem	Varieties of Democracy (Institute)

References

AidData, "Global Chinese Development Finance," webpage, undated. As of November 1, 2023: https://china.aiddata.org/

AidData, AidData's Global Chinese Development Finance Dataset, version 2.0, database, September 29, 2021. As of November 1, 2023: https://www.aiddata.org/data/aiddatas-global-chinese-development-finance-dataset-version-2-0

Artificial Intelligence Index, *The AI Index Report: Measuring Trends in Artificial Intelligence*, Stanford Institute for Human-Centered Artificial Intelligence, 2021. As of October 30, 2023: https://aiindex.stanford.edu/ai-index-report-2021/

Amazon Web Services, "What Is Deep Learning?" webpage, undated. As of October 27, 2023: https://aws.amazon.com/what-is/deep-learning/

Bouey, Jennifer, Lynn Hu, Keller Scholl, William Marcellino, Rafiq Dossani, Ammar Malik, Kyra Solomon, and Sheng Zhang, *China's AI Exports: Technology Distribution and Data Safety*, RAND Corporation, RR-A2696-2, 2023.

Center for AI Safety, "Statement on AI Risk: AI Experts and Public Figures Express Their Concern About AI Risk," webpage, undated. As of October 27, 2023: https://www.safe.ai/statement-on-ai-risk

China Institute for Science and Technology Policy at Tsinghua University, *China AI Development Report*, 2018.

Custer, Samantha, Axel Dreher, Thai-Binh Elston, Andreas Fuchs, Siddharta Ghose, Joyce Jiahui Lin, Ammar A. Malik, Bradley C. Parks, Brooke Russell, Kyra Solomon, Austin Strange, Michael J. Tierney, Katherine Walsh, Lincoln Zaleski, and Sheng Zhang, *Tracking Chinese Development Finance: An Application of AidData's TUFF 2.0 Methodology*, AidData, 2021. As of October 27, 2023: https://www.aiddata.org/publications/aiddata-tuff-methodology-version-2-0

DLA Piper, "Compare Data Protection Laws Around the World," online tool, undated. As of October 27, 2023: https://www.dlapiperdataprotection.com/

Freedom House, homepage, undated. As of October 27, 2023: https://freedomhouse.org/

Gaikwad, S. K., B. W. Gawali, and P. Yannawar, "A Review on Speech Recognition Technique," *International Journal of Computer Applications*, Vol. 10, No. 3, 2010.

Herre, B., E. Ortiz-Ospina, and M. Roser, Democracy, database, Our World in Data, undated. As of October 27, 2023: https://ourworldindata.org/democracy

Leung, C. K., P. Braun, and A. Cuzzocrea, "AI-Based Sensor Information Fusion for Supporting Deep Supervised Learning," *Sensors*, Vol. 19, No. 6, 2019.

Malik, Ammar A., Bradley Parks, Brooke Russell, Joyce Jiahui Lin, Katherine Walsh, Kyra Solomon, Sheng Zhang, Thai-Binh Elston, and Seth Goodman, *Banking on the Belt and Road: Insights from a New Global Dataset of 13,427 Chinese Development Projects*, AidData, September 29, 2021. As of October 27, 2023: https://www.aiddata.org/publications/banking-on-the-belt-and-road

Nadkarni, P. M., L. Ohno-Machado, and W. W. Chapman, "Natural Language Processing: An Introduction," *Journal of the American Medical Informatics Association*, Vol. 18, No. 5, 2011.

OECD—*See* Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development, "Development Finance Standards," webpage, undated-a. As of November 20, 2023: https://www.oecd.org/dac/financing-sustainable-development/development-finance-standards/

Organisation for Economic Co-operation and Development, "Official Development Assistance—Definition and Coverage," webpage, undated-b. As of November 20, 2023:
https://www.oecd.org/dac/financing-sustainable-development/development-finance-standards/officialdevelopmentassistancedefinitionandcoverage.htm

Tao, J., and T. Tan, "Affective Computing: A Review," *Proceedings*, Vol. 1, Affective Computing and Intelligent Interaction: First International Conference, October 2005.

Trump White House, "Artificial Intelligence for the American People," archival webpage, 2020. As of October 30, 2023:
https://trumpwhitehouse.archives.gov/ai/

United Nations Conference on Trade and Development, Data Protection and Privacy Legislation Worldwide, database, undated. As of October 30, 2023:
https://unctad.org/page/data-protection-and-privacy-legislation-worldwide

V-Dem Institute, D*emocracy Report 2022: Autocratization Changing Nature?* March 2022. As of November 20, 2023:
https://v-dem.net/media/publications/dr_2022.pdf

Wang, T., D. J. Wu, A. Coates, and A. Y. Ng, "End-to-End Text Recognition with Convolutional Neural Networks," *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 2012.

Wei, Lingling, "U.S. Think Tank Reports Prompted Beijing to Put a Lid on Chinese Data," *Wall Street Journal*, May 7, 2023.

White House Briefing Room, "FACT SHEET: Biden-Harris Administration Announces New Actions to Promote Responsible AI Innovation That Protects Americans' Rights and Safety," May 4, 2023. As of October 27, 2023:
https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/04/fact-sheet-biden-harris-administration-announces-new-actions-to-promote-responsible-ai-innovation-that-protects-americans-rights-and-safety

Zhao, W., R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Surveys (CSUR)*, Vol. 35, No. 4, 2003.

NATIONAL SECURITY RESEARCH DIVISION

With average annual commitments reaching $85 billion, the Chinese government is now the world's largest provider of development financing. Supported by large funding, many Chinese technology companies can deploy state-of-the-art artificial intelligence (AI) tools in development contracts in recipient countries. The AI exports facilitated by these arrangements are likely to bolster China's growing global AI technology-related supply chains, trade flows, technology standards, and regulatory systems.

This report is the technical documentation for the China's AI Exports Database (CAIED), which is a tool that tracks Chinese government–supported development finance projects that used or enabled AI technology in the Global South between 2000 and 2017. The goal of the report is to explain the motivation, data source, and methodology of building the tool. This report also presents a detailed codebook for the database and explanations on how to use the interactive world map and country selector.

CAIED was built on AidData's Global Chinese Development Finance (GCDF) Dataset version 2.0. Using advanced data mining tools, the authors identified 155 projects enabling AI exports and described their characteristics in rich detail. Users can filter the world map by AI technology categories associated with the development finance projects and learn further details of the exported China AI project by clicking on the country of interest.

www.rand.org

RR-A2696-1