MERICS
REPORT

Annex

AI ENTANGLEMENTS

Balancing risks and rewards of European-Chinese collaboration

Rebecca Arcesati, Wendy Chang, Antonia Hmaidi, Kai von Carnap

MERICS
Mercator Institute for China Studies

ANNEX

1. Example papers

Some research outputs present ethical and human rights risks

Example collaborations relating to public security and biometric surveillance applications

PAPER TITLE	CHINESE AFFILIATIONS OF CO-AUTHORS	EUROPEAN AFFILIATIONS OF CO-AUTHORS	YEAR
"NAS-FAS: Static-Dynamic Central Difference Network Search for Face Anti-Spoofing"	CAS Institute of Automation; Northwestern Polytechnical University; University of CAS	University of Oulu (FI)	2021
"ArvaNet: Deep Recurrent Architecture for PPG-Based Negative Mental-State Monitoring"	Nanjing Agricultural University; Northeastern University; Shenyang Ligong University; Hohai University; Ministry of Public Security of China	University of Lincoln (GB)	2021
"Cross-Ethnicity Face Anti-Spoofing Recognition Challenge: A Review"	CAS Institute of Automation; Westlake University; Baidu; Macau University of Science and Technology; Beihang University; Beijing Jiaotong University; National Engineering Laboratory of Deep Learning Technology and Application	University of Barcelona (ES)	2021
"Group Abnormal Behaviour Detection Algorithm Based on Global Optical Flow"	Xi'an University of Posts and Telecommunications	University of Huddersfield (GB)	2021
"Self-Attention Transfer Networks for Speech Emotion Recognition"	Tianjin Normal University; National Laboratory of Pattern Recognition, CAS Institute of Automation	King's College London (GB); Fairfield University (GB); Imperial College London (GB); University of Augsburg (DE)	2021
"High-Speed Multi-Person Pose Estimation with Deep Feature Transfer"	Hangzhou Normal University	Northumbria University (GB)	2020
"Effective Crowd Anomaly Detection Through Spatio-Temporal Texture Analysis"	Xi'an University of Posts and Telecommunications	University of Huddersfield (GB); Sheffield Hallam University (GB)	2019

Source: MERICS corpus of AI research outputs extracted from OpenAlex, filtered according to a combination of relevant keywords (e.g., "biometrics", "emotion recognition", etc.) and co-authoring Chinese institutions that are either public security and policing organs or civilian entities with close ties with them.

Co-authoring AI research with Chinese cybersecurity schools

Example collaborations relating to cybersecurity and cyber operations

PAPER TITLE	CHINESE AFFILIATIONS OF CO-AUTHORS	EUROPEAN AFFILIATIONS OF CO-AUTHORS	YEAR
"TSAEns: Ensemble Learning for KPI Anomaly Detection"	National University of Defense Technology	Technical University of Berlin (DE)	2022
"FuzzGAN: A Generation-Based Fuzzing Framework for Testing Deep Neural Networks"	Shandong University	Helmholtz Center for Information Security (DE)	2022
"A Novel Intrusion Detection Method Based on Lightweight Neural Network for Internet of Things"	Jiangsu Police Officer College; Nanjing University of Posts and Telecommunications; Shanghai Jiao Tong University	Manchester Metropolitan University (GB); RWTH Aachen University (DE)	2021
"An Automata Based Intrusion Detection Method for Internet of Things"	Xidian University	University of Pau and Pays de l'Adour (FR)	2021
"Zero-Shot Fine-Grained Entity Typing in Information Security Based on Ontology"	Zhengzhou University; PLA Information Engineering University	University of Ulster (GB)	2021
"CECoR-Net: A Character-Level Neural Network Model for Web Attack Detection"	Shanghai Jiao Tong University; Tencent	Télécom Paris (FR)	2019
"A Novel RNN-GBRBM Based Feature Decoder for Anomaly Detection Technology in Industrial Control Network"	State Grid Jiangsu Electric Power Company; Beijing University of Posts and Telecommunications; NARI Group Corporation	Saft (FR)	2017
"An Efficient Intrusion Detection Method Based on Dynamic Autoencoder"	Jiangsu Police Officer College; Nanjing University of Posts and Telecommunications; Shanghai Jiao Tong University	Manchester Metropolitan University (GB); RWTH Aachen University (DE)	2017
"Deep Learning-Based Real-Time VPN Encrypted Traffic Identification Methods"	Luoyang Institute of Science and Technology; PLA Information Engineering University	Tampere University of Applied Sciences (FI)	2020

Source: MERICS corpus of AI research outputs extracted from OpenAlex, filtered according to a combinaton of relevant keywords (e.g., „automated anomaly detection", „automated exploit," etc.) and co-authoring Chinese institutions that either are among the certified Chinese World-Class Cybersecurity Schools (流网络安全学院) and/or have documented links to Advanced Persistent Threat (APT) cyber actors.[1]

Some research outputs could aid Chinese information operations

Example collaborations with key Chinese defense research institutes

PAPER TITLE	CHINESE AFFILIATIONS OF CO-AUTHORS	EUROPEAN AFFILIATIONS OF CO-AUTHORS	YEAR
"Containment of Rumor Spread in Complex Social Networks"	Xidian University; Xi'an University of Science and Technology	University of Cagliari (IT); Aix-Marseille University (FR); Université de Toulon (FR)	2020
"Region-Aware Face Swapping"	Zhejiang University	CE Technologies (GB)	2022
"Unsupervised Image Generation with Infinite Generative Adversarial Networks"	Zhejiang University	Clemson University (GB); University of Leeds (GB)	2021
"Camouflage Generative Adversarial Network: Coverless Full-Image-to-Image Hiding"	Central South University	Loughborough University (GB); University of Portsmouth (GB)	2020
"Sentiment Classification for Chinese Text Based on Interactive Multitask Learning"	Zhengzhou University; PLA Information Engineering University	University of Ulster (GB)	2020
"SwapGAN: A Multistage Generative Approach for Person-to-Person Fashion Style Transfer"	National University of Defense Technology	Leiden University (NL)	2019
"Five Shades of Untruth: Finer-Grained Classification of Fake News"	Shandong University	Max Planck Institute for Informatics (DE)	2018
"Sentiment Analysis and Spam Detection in Short Informal Text Using Learning Classifier Systems	Beihang University	University of Pavia (IT)	2018
"Fusing and mining opinions for reputation generation"	Xidian University	Aalto University (FI); Polish Academy of Sciences (PL); Systems Research Institute (PL)	2017

Source: MERICS corpus of AI research outputs extracted from OpenAlex, filtered according to a combination of relevant thematic keywords (e.g., "generative adversarial network", "deepfake," "face generation," etc.) and co-authoring Chinese defense research institutions that are either directly controlled by or linked to the military.

Co-authoring AI research with the Chinese military

Example collaborations relating to various defense applications

PAPER TITLE	CHINESE AFFILIATIONS OF CO-AUTHORS	EUROPEAN AFFILIATIONS OF CO-AUTHORS	YEAR
"Real-Time Positioning and Tracking for Vision-Based Unmanned Underwater Vehicles"	State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing; Wuhan University	ETH Zurich (CH)	2022
"Transferring Transformer-Based Models for Cross-Area Building Extraction from Remote Sensing Images"	PLA Information Engineering University	Universität der Bundeswehr München (DE)	2022
"UAV Target Tracking Method Based on Deep Reinforcement Learning"	China State Shipbuilding; Northwestern Polytechnical University	London South Bank University (GB)	2022
"Neural Mixed Platoon Controller Design"	Beihang University; State Key Laboratory of Vehicle NVH and Safety Technology	University of Sussex (GB); University of Glasgow (GB)	2022
"Deep Learning for 3D Vision"	National University of Defense Technology; PLA Information Engineering University	University of Oxford (GB); University of Florence (IT)	2022
"Maximal Admissible Mode Decision Delay in Terminal Guidance"	National University of Defense Technology; University of Salamanca; Aviation Key Laboratory of Science and Technology on Airborne Guided Weapons, China Airborne Missile Academy	University of Salamanca (ES)	2019
"Generalized Impact Time and Angle Control Via Look-Angle Shaping"	Beijing Institute of Technology	Technical University of Munich (DE)	2019
"An Efficient Link Prediction Index for Complex Military Organization"	National University of Defense Technology	Karolinska Institutet (SE)	2017
"Integrated Missile Guidance and Control Using Optimization-Based Predictive Control"	Beijing Institute of Technology	Cranfield University (GB)	2017

Source: MERICS corpus of AI research outputs extracted from OpenAlex, filtered according to a combination of relevant thematic keywords (e.g., "target tracking," "scene analysis," "UAV," etc.) and co-authoring with Chinese defense research institutions that are either directly controlled by or linked to the military.

2. Additional documentation

Chinese participation in EU-funded, AI-relevant projects

Selected projects involving China-based partners

PROJECT NAME	STATUS	CHINESE CONSORTIUM PARTNERS
DORNA (Development of High Reliability Motor Drives for Next Generation Propulsion Applications)	Ongoing	Zhejiang University; Harbin Institute of Technology; Zhejiang University
CORESMA (COVID-19-Outbreak Response Combining E-Health, Serolomics, Modelling, Artificial Intelligence and Implementation Research)	Concluded	Beijing Institute of Chemical Technology
ULTRACEPT (Ultra-Layered Perception with Brain-Inspired Information Processing for Vehicle Collision Avoidance)	Ongoing	Tsinghua University; Xi'An Jiao Tong University; Huazhong University of Science and Technology; Northwestern Polytechnical University; Lingnan Normal University; Guizhou University; CAS Institute of Automation; Guangzhou University
IDENTITY (Computer Vision Enabled Multimedia Forensics and People Identification)	Concluded	CAS Institute of Automation; Hong Kong Baptist University; South China University of Technology
LARKC (Large Scale Semantic Computing\nSemantic Web technologies\ndistributed reasoning\nprobabilistic reasoning\nweb-scale inference\ninformation retrieval)	Concluded	Beijing University of Technology
OPTIMAL (Smart and CO2 Neutral Olefin Production by Artificial Intelligence and Machine Learning)	Ongoing	East China University of Science and Technology; Southeast University; Hunan University; Xi'An Jiao Tong University; Nanjing Improve Automation Technology Co Ltd

Source: Database of all 49 projects that touch on AI or AI-tangential fields of science, pulled from CORDIS, the official database of EU-supported R&D projects. Since 2015, China-based partners in EU-funded projects have been funded by the Chinese government (the same applies to Hong Kong-based participants since 2016).

Universities and firms have set up labs to conduct joint AI R&D, some of which warrant scrutiny

Curated profiles of select labs, (Part 1)

LAB NAME	TYPE	KEY ACTORS AND PARTNERS	DESCRIPTION
China-Austria Belt and Road Joint Laboratory on Artificial Intelligence and Advanced Manufacturing (中国—奥地利人工智能与先进制造”一带一路”联合实验室)	Academic	Hangzhou Dianzi University (CN); Vienna University of Technology (AU)	An official Belt and Road Joint Lab approved by MOST in 2021, focused on intelligent manufacturing and environmental protection. Behind it is a foreign talent recruitment base under the Chinese government's Project 111 program. Hangzhou Dianzi conducts defense AI research.
Sino-European Laboratory in Computer Science, Automation and Applied Mathematics (LIAMA) (中法信息自动化应用数学联合实验室)	Academic	INRIA (FR); CAS Institute of Automation (CN); other European and Chinese universities	Partnership dating back to the 1990s, with projects spanning human-machine interaction, neuromorphic chips, brain-inspired AI, computer vision, machine-aided design, and embedded systems. Tao Jianhua (陶建华), deputy director of the National Laboratory of Pattern Recognition, was the China director.
Belt and Road Joint Laboratory of Measurement and Control Technology (测控技术“一带一路”联合实验室)	Academic	Huazhong University of Science and Technology (CN); multiple Polish universities (PL)	An official Belt and Road Joint Lab approved by MOST in 2021. Applications of measurement and control technology in micro- nano-magnetic physics, decarbonization and medical imaging are priorities, but there is also work on brain-like computing and brain-machine intelligence.
Phillips China Innovation Hub	Corporate	Phillips (NL)	In 2018, established an AI lab to develop solutions for the local market in the areas of medical imaging, patient monitoring, health informatics and remote care.
Sino-German Joint Laboratory of Neuroinformatics (西北工业大学中德神经信息联合实验)	Academic	TU Berlin (DE); Northwestern Polytechnical University (CN); University of Kent (GB)	Builds on 20 years of collaboration in brain-computer interfaces and control theory and appears to focus heavily on UAV applications. Almost all resulting publications and patents have been produced in China. NWPU is the PLA's prime drone supplier.

Source: MERICS. Sources available upon request.

Universities and firms have set up labs to conduct joint AI R&D, some of which warrant scrutiny

Curated profiles of select labs, (Part 2)

LAB NAME	TYPE	KEY ACTORS AND PARTNERS	DESCRIPTION
Tsinghua-Bosch Center for Machine Learning (清华大学－博世机器学习联合研究中心)	Corporate-academic	Bosch (DE); Tsinghua University (CN)	Opened in 2020 as part of a five-year research collaboration agreement, focuses on basic ML theories and key technologies, combined with an applied focus on intelligent manufacturing and autonomous driving. Jun Zhu (朱军), director of the Basic Theory Research Center at Tsinghua Institute for AI, is director.
Sino-Swiss Laboratory for Data-Intensive Neuroscience (中瑞数据密集型神经科学联合实验室)	Academic	EPFL (CH); CAS Institute of Automation (CN)	Set up in 2015 to work on neuroinformatic and brain-inspired intelligence, bringing together CAS's Brain-Intensive Cognitive Intelligence Lab, (BRAICOG) and the research team led by Sean Hill, co-director of the Swiss national brain initiative Blue Brain.
Siemens AI Lab	Corporate	Siemens (DE)	Lab pursuing mainly applied research on industrial AI applications, also in partnership with Tsinghua University. Tsinghua and Siemens also run a Joint Research Center for Industrial Intelligence and the Internet of Things (JCIIOT).
Lagrange Mathematics and Computing Research Center (LMCRC)	Corporate	Huawei (CN)	Focuses on fundamental research in mathematics and computer science and building on Huawei's Paris Research Center's prior R&D work in areas like mathematical and algorithmic sciences. Collaborates with local AI research institutions such as the Ecole Normale Supérieure.
Huawei Future Computing Lab	Corporate	Huawei (CN)	Located within Huawei's Zurich Research Center, conducts fundamental research into future computing architecture, software, and algorithms. The lab has strong ties with Huawei's chip design arm, HiSilicon, and cooperates with ETH Zurich on research and talent training.

Source: MERICS. Sources available upon request.

Universities and firms have set up labs to conduct joint AI R&D, some of which warrant scrutiny

Curated profiles of select labs, (Part 3)

LAB NAME	TYPE	KEY ACTORS AND PARTNERS	DESCRIPTION
Huawei Noah Ark Lab (华为诺亚方舟实验室)	Corporate	Huawei (CN)	Huawei's AI research arm, with European branches in Paris and London. Focuses on frontier R&D in the areas of computer vision, NLP, AI theory, decision making and reasoning, and human-computer interaction and collaborates with several universities.
Video Competence Lab (VCL)	Corporate	Huawei (CN)	Housed within Huawei's Dublin Research Center, focuses on biometric surveillance applications of computer vision, like action recognition and person re-identification.
DReaMS Lab (Dialogues, REAsoning and Multi-linguality for Search)	Corporate-academic	Huawei (CN); Vrjie University of Amsterdam (NL); University of Amsterdam (NL)	Funded by Huawei, conducts research on knowledge representation and reasoning, language processing and computational lexicology to optimize multilingual web search systems in different European languages.
International Research Centre for Advanced Robotics & Mechanism (IRCARM) (先进机器人学与机构学国际联合研究中心)	Academic	Shenyang Institute of Automation, CAS (CN); Centre for Robotics Research (CORE), King's College London (GB); Edinburgh Centre for Robotics (ECR) (GB); Heriot-Watt University (GB)	Founded in 2016, focuses on robotics and automation, machine vision and human-robot interaction, with applications in the power industry, manufacturing, and marine exploration. SIA is a key player in China's state-backed autonomous undersea vehicle program.

Source: MERICS. Sources available upon request.

3. Methodology and scope: AI-related publications

3.1 Data sources and approach

Publication data was extracted from OpenAlex,[2] an open-source database of academic journal articles based on Microsoft Academic Graph. Inspired by a "bag-of-words" methodology developed by CSET, we compiled a list of relevant keywords to find AI-related publications. We performed full-text searches using OpenAlex's API for each of the keywords and extracted document data including DOI, authors and their institutional affiliations, title, publication date, any citation and funding information (where available) and OpenAlex-defined concepts. We compared the results to the set of papers to which OpenAlex assigned the "Artificial Intelligence" concept. We found the two sets of be of similar size but preferred the more granular keyword-based approach in identifying our target papers. This also enabled us to look for subsets of AI papers in specific subfields.

OpenAlex is a good but young resource for scientific publications. OpenAlex itself performed some comparisons with other leading scientific literature databases, like Scopus, Web of Science and Dimensions, and found its own database to be at least on par. We compared OpenAlex with Scopus to analyze Sweden-China collaborations and found very similar results, usually within 5 percent, with either Scopus or OpenAlex having more papers. The key difference was papers with more than 100 co-authors, which OpenAlex does not systematically use entity resolution for. Since we were interested in the strength of collaboration links between different institutions, we decided to ignore papers with more than 20 co-authors.

There is some debate in the literature on the merits of using fractional scaling: If there are five co-authors, a publication would count as 1/5 for each of these co-authors. If there is only one author, the publication would count as one. This approach is sensible if the goal is to assess country contributions to the overall scientific knowledge base but was not appropriate for this project, where links between different institutions were the unit of analysis. Instead, we chose to count each paper only once for each link: If a paper has two authors claiming affiliation with the same Chinese institution, and one author from a European institution, we only counted one link.

In the interest of completeness, we explored the use of data extracted from the China National Knowledge Infrastructure (CNKI), China's largest repository of scientific literature, to capture Chinese-language research that also fit our criteria. We found that while many AI papers are published on CNKI, very few represent collaborative efforts with European and American institutions. This echoes previous findings by CSET.[3] Therefore, we chose not to include CNKI data in our analysis.

3.2 Entity resolution

OpenAlex has in-built entity resolution which uses ORCID ids for researchers and ROR for institutions. While we found this resolution to be generally reliable for academic institutions, for companies OpenAlex does not attribute all subsidiaries and labs to the same mother company. Moreover, OpenAlex does not go by the headquarters of the parent company, but instead by location of the specific lab. Since we were primarily interested in the collaboration links between entities in European countries and China, we attributed all labs and subsidiaries to a company's headquarters. As such, Siemens China will be attributed to Germany, while Huawei Germany will be attributed to China.

We also combined all company labs and subsidiaries into one entity. OpenAlex often attributes to a university all research divisions that are subordinate to it. This makes sense for practical reasons, as labs sometimes only exist for a short time and then are re-named, while university and research institutes tend to exist for longer. It does, however, make risk assessment more difficult, as many Chinese universities and research institutions host both civilian and defense- or security-focused divisions.[4]

3.3 Dataset

We built a set of 36,951 academic papers that represents collaborative AI research having at least one author claiming affiliation with an institution based in China and one claiming affiliation with an institution based in Europe (EU member states, Switzerland, Norway, or the United Kingdom). To identify AI papers, we performed a full search in OpenAlex using the AI keywords list mentioned earlier. This query was combined with a filter representing the OpenAlex "Artificial Intelligence" concept to avoid false positives. The target timeframe was papers published between January 1, 2017, and December 31, 2022. Since OpenAlex is a living platform, with papers from 2022 still being added, the figures we provide for 2022 are current as of May 2023.

3.4 Social Network Analysis

For the social network analysis, we looked at the links between institutions. We counted each paper once per link, if at least one author from institution A and at least one author from institution B were mentioned as authors. In addition to looking at the strength of the ties between institutions, we also looked at larger patterns and network characteristics.

3.5 Curated list of AI-related keywords

To identify AI-relevant literature, we ran queries based on relevant keywords, adapting lists previously developed by CSET researchers which include both terms associated with AI more generally and terms specific to certain subfields and application domains.[5] We also developed our own curated keyword lists (not included below) to find papers with relevance to specific application domains we were interested in, including cybersecurity, information operations, biometric surveillance, and other military applications of AI.

abstract event representation	active learning
adaptive learning	additive learning
adversarial network	affective computing
AGI	AI alignment
algorithm	ANI
anomaly detection	architecture search
artificial emotion	artificial general intelligence
artificial intelligence	artificial narrow intelligence
artificial pedagogy	artificial psychology
artificial scientist test	artificial superintelligence
ASI	associative learning
attention modeling	autoencoder
Automated navigation system	automated target recognition
autonomous knowledge acquisition	autonomous learning
autonomous navigation	autonomous reasoning

autonomous system	autonomous vehicle
autonomous weapon	average link clustering
backpropagation	binary classification
blue brain project	boltzmann machine
brain in a vat	brain simulation
brain-computer chip	brain-computer interfaces
brain-inspired	cascade
category theory	causal inference
causal reasoning	character recognition
classification algorithm	classification label
clustering method	cognitive architecture
cognitive computing	cognitive enhancement
complete AI	complete link clustering
computational creativity	computer vision
computing overhang	concept formation
conceptual binding	conscious perception
consciousness and memory model	consciousness uploading
control problem	cross-domain generalization
Crowd detection	cumulative learning
cyborg	declarative knowledge
declarative memory	deep learning
discontinuous leap	effective BCI
efficient concept formation	efficient learning
elastic memory	embodiment
emotion recognition	ensemble learning
episode retrieval	episodic memory
evolutionary algorithm	evolutionary learning
executive control	facial expression recognition
facial identification	facial recognition
feature binding	feature extraction
feature learning	feature matching
feature selection	feature vector
feedforward network	fingerprint recognition
flexible memory	formal neural networks
fuzzy clustering	gait recognition
general-purpose system	generative adversarial network
generative model	global workspace
gradient algorithm	gradual neuron replacement
graph matching	graphical model
handwriting recognition	hard takeoff
hierarchical clustering	hierarchical decomposition
hierarchical model	higher order cognition
high-level machine intelligence	HLAI
HLMI	human-level AI
human-robot interaction	hypothetical reasoning
image annotation	image classification
image matching	image processing

image registration	image representation
image retrieval	implicit learning
incremental clustering	incremental learning
inductive inference	inductive reasoning
information extraction	information fusion
information retrieval	integrated AI
integrative designs	intelligence explosion
intelligence science	intentional memory
intentionality	intuitive understanding
iris recognition	k-nearest neighbor
knowledge discovery	knowledge representation
knowledge-based system	language identification
language model	long-term potentiation ML
LTP machine learning	machine ethics
machine learning	machine learning neurofidelity
machine perception	machine translation
mental representation	metalearning
MicroPSI	mind modeling
mind uploading	mirroring
modeling of reasoning	motive selection
multi-class classification	multi-label classification
multi-objective reinforcement learning	multitask learning
natural language generation	natural language processing
natural language understanding	neural network
neuroinformatics	neuron replacement
neurosymbolic representation	Non-Axiomatic Reasoning System
non-linear dynamical systems	object detection
object permanence	object recognition
object tracking	object vision
OCR	one-shot learning
OpenCog	OpenPSI
Optical character recognition	parameter
pattern matching	pattern recognition
pattern reconstruction	perceptual consistency
Person detection	Person re-identification
perverse instantiation	position invariance
post-biological	predictive coding
predictive modeling	procedural memory
procedure learning	progressive learning
psyche + AI	random forest
recommender system	recurrent network
recursive self-improvement	reinforcement learning
reinterpretation	relative strength index
reward replacement	scene classification
scene understanding	scene vision
seed AI	semantic memory
semantic vision	semi-supervised learning

sense making	sensory memory
sentiment classification	shape sorting
simulated environment	single-link clustering
singularity	sparse representation
spatial learning	speech processing
speech recognition	speech synthesis
spike timing dependent plasticity	spiking neural network
statistical learning	strong AI
superintelligence	supervised learning
support vector machine	surface autonomous vehicle
target tracking	technological singularity
text mining	text processing
thinking machine	transfer learning
transformer-based	transhuman
transhumanism	UAV
uncertainty estimation	underwater autonomous system
unified knowledge	unmanned aerial vehicle
unsupervised learning	variable binding
vehicle detection	video classification
video processing	whole brain emulation
zero shot learning	

4. Methodology and scope: Investment in AI companies

We relied on a dataset provided by financial data provider Crunchbase[6] to study AI-related investment flows between European countries and China and Hong Kong. While its coverage of the investment markets is not perfect, Crunchbase data allows us to derive meaningful insights. When checking for its coverage of Chinese AI firms specifically, CSET experts found that Crunchbase tends to capture most transactions in which non-China based investors participated.[7]

We collected information on investment into AI companies located in select European countries (EU member states, Switzerland, Norway, and the United Kingdom) as well as in China and Hong Kong between January 1, 2017, and December 31, 2022, to mirror the timeframe of our analysis of AI-related publications. The information we collected includes names and business descriptions of the target companies, names of the investors, types of investment, and value raised. We acquired the data on June 30, 2023, so our analysis does not account for transactions completed during our target timeframe that Crunchbase may have added after this date.

4.1 Investor location

We first identified China-based investors and Europe-based investors by filtering for investment entities and companies whose headquarter location, as logged by Crunchbase, was listed as China, Hong Kong, EU member states, Switzerland, Norway, or the United Kingdom.

We recognize that investor domicile may not always capture the country of origin of the capital, i.e., that of the ultimate controlling entity behind a given investor. Relying on investor

domicile is therefore likely to lead to over- or undercounting relevant transactions. One authoritative study of VC investment between the United States and China over the period 2000-2019 found that providers that rely on investor domicile undercounted China-origin investments into US companies by 25 percent.[8] As we did not attempt to determine the ultimate source of capital of the observed investment, our figures should not be read as a complete picture of Europe-China investment relations in the AI sector.

4.2 AI companies

We further filtered our two datasets for transactions into AI companies only. This includes all transactions into companies that Crunchbase labels as "Artificial Intelligence," "Intelligent Systems," "Machine Learning," "Natural Language Processing," or "Predictive Analytics". We further applied a regular expression filter developed by CSET,[9] where a Boolean logic command filters the company descriptions for character strings that are indicative of AI companies but might have not been labeled as such by Crunchbase. The filter was built in and applied by Excel VBA.

SIMPLIFIED REPRESENTATION OF REGULAR EXPRESSION QUERY FOR AI COMPANIES
Either meet the following expression condition
(machine|artificial)(\W*\w*){0,2}intelligence|\bAI\b|A\.I(\W|\b)
AND
analy|predict|robot|cluster|adapt|diagnos|automat|detect|personaliz|label| augment|autonom|sensor|sensing|recommend|optimiz|chatbot|\bbot(s|\b)| (digital|virtual) assistant|semiconductor|chipset|\bGPU\b|\bASIC\b|\bFPGA\b| high(|-)performance computing|knowledge graph
OR either of the following expression
(reinforcement|transfer|one-shot|one shot|zero-shot|zero shot|supervised|unsupervised) ?-?(machine)?learning (self(-|)driving|driverless|autonomous)(\W*\w*\W*){0,4} (vehicle|truck|car|vehicle|automobile|technolog|navig|transport|robot|machine) (driverless|autonomous|automat)\w*(\W*\w*\W*){0,2}(driv|navig) (machine|deep)(|-)learning cognitive computing|synthetic data|neural net|predictive analytic|generative adversarial|network hyperparameter (computer|machine) vision \b(R|D)NN\b|\bGAN\b (natural language|speech) (processing|understanding)|\bNLP\b feature (extraction|learning|matching|selection) autoencod|tensorflow|\bkeras\b|\btheano\b q(-|)(learning|value|network) (support vector|Boltzmann) machine machine (translation|perception) (facial|speech|face|voice|music|image|character|text|emotion|video|gesture) (recogn i|classif)|(recogni|classif)\w*(facial|speech|face|voice|music|image|character|text|e motion|video|gesture)

We identified 98 transactions into 76 European AI companies where a China-based investor was involved and 81 transactions into 64 Chinese AI companies which saw a contribution by investors based in the Europe. Combined, the dataset comprises 179 transactions into 140 companies.

4.3 Sectors

To conduct a more granular investigation of AI-related investments, we categorized 140 AI companies based on the main field in which they apply AI technology. We applied the classification taxonomy for AI application, or "TINA" (Taxonomy of INtelligence Applications), developed by CSET and used most recently in their investigation of US investment into Chinese AI companies.[10] The taxonomy suggests 17 primary application areas of AI technologies.

Two researchers classified the 140 companies within both datasets, according to TINA categories and based on information that was either publicly available or provided by Crunchbase. Matching classifications occurred in two thirds of all cases. For the remaining third where differing classification decisions occurred, a third researcher made the final decision.

4.4 Investment sums and missing values

When the value of an investment transaction was not disclosed in Crunchbase, we ran our own calculations to estimate the missing values. Our estimates are based on the average amount of funding raised by companies in the same target geography, at the same funding stage and in the same year. Based on these three characteristics, we were able to triangulate a close approximate transaction value. For example, the missing transaction value for a 2019 series-C funding round of a German AI company with a China-based participation was calculated based on the average of all series-C rounds in 2019 into a German AI company. Where no annual average could be calculated because no similar transaction was recorded, we took the average from the timeframe 2017 to 2022.

In total, we calculated 39 missing transaction values. Their contribution to the total investment volumes was less than 10 percent and amounted to EUR 139 million for European AI companies (totaling EUR 1.43 billion) and EUR 730 million for Chinese AI companies (totaling EUR 11.16 billion). We recognize that averaging introduces a potential over- or underestimation of the real investment volumes. However, by basing averages on country, year, and funding rounds of the same investment stage, we were able to fill gaps and derive more reliable figures.

5. Select list of Chinese entities presenting security or ethical risks

To sample papers co-authored with entities presenting a heightened degree of risk, we ran a combination of actor-specific and thematic queries. Drawing on the information available in the ASPI China Defence Universities Tracker,[11] as well as on additional literature reviews and desk research into AI-focused research centers and labs hosted within Chinese universities, we compiled a list of entities and grouped them into families. Our list is not exhaustive, nor is it an attempt to comprehensively cover the collaborators that present risks.

The following is a list of the institutions that we were able to id in OpenAlex. When an entity belonged to multiple categories, it is displayed under one category for easier consultation.

Since most Chinese labs and research institute lack a corresponding id in OpenAlex, except some like the CAS Shenyang Institute of Automation, in most cases we could not determine the specific division with which the Chinese co-authors were affiliated. Further research and id resolution work should seek to illuminate those affiliations.

Not included below are Chinese tech firms that have been found to contribute to human rights abuses through the provision of surveillance products, like Sensetime or iFlytek. Any links with those firms warrant further investigation. For Huawei, we ran a separate query to explore the company's overall collaborative outputs with European partners.

ENTITIES DIRECTLY CONTROLLED BY THE CENTRAL MILITARY COMMISSION (MILITARY END-USE RISK)	
PLA Information Engineering University	中国人民解放军信息工程大学
National University of Defense Technology	中国人民解放军国防科技大学
Naval University of Engineering	中国人民解放军海军工程大学
Chinese Academy of Engineering Physics	中国工程物理研究院
Rocket Force University of Engineering	中国人民解放军火箭军工程大学
China Aerodynamics Research and Development Center	中国空气动力研究与发展中心
PLA Electronic Engineering Institute	电子对抗学院 (formerly known as 中国人民解放军电子工程学院 and now under NUDT)
Air Force Engineering University	空军工程大学
International College of Defense Studies	中国人民解放军国防大学国际防务学院 (under National Defense University, 中国人民解放军国防大学)
Space Engineering University	中国人民解放军战略支援部队航天工程大学
Northwest Institute of Nuclear Technology	西北核技术研究所
Academy of Military Science	中国人民解放军军事科学院
Navy Aviation University	中国人民解放军海军航空大学
Army Engineering University of PLA	中国人民解放军陆军工程大学
Army Military Transportation Academy	中国人民解放军陆军军事交通学院
PLA Airforce Aviation University	中国人民解放军空军航空大学
PLA Army Service Academy	中国人民解放军陆军勤务学院

SEVEN SONS OF NATIONAL DEFENSE AND SEVEN SONS OF THE ARMS INDUSTRY (MILITARY END-USE RISK)	
Beijing Institute of Technology	北京理工大学
Harbin Engineering University	哈尔滨工程大学
Northwestern Polytechnical University	西北工业大学
Harbin Institute of Technology	哈尔滨工业大学
Nanjing University of Science and Technology	南京理工大学
Beihang University	北京航空航天大学
Nanjing University of Aeronautics and Astronautics	南京航空航天大学
Changchun University of Science and Technology	长春理工大学
Shenyang Ligong University	沈阳理工大学
North University of China	中北大学
Xi'An Technological University	西安工业大学
Chongqing University of Technology	重庆理工大学

STATE-OWNED DEFENSE-INDUSTRIAL CONGLOMERATES (MILITARY END-USE RISK)	
Aviation Industry Corporation of China	中国航空工业集团公司
China Aerospace Science and Industry Corporation	中国航天科工集团有限公司
China National Nuclear Corporation	中国核工业集团公司
China Academy of Space Technology	中国空间技术研究院
China Electronics Technology Group Corporation	中国电子科技集团公司
China Aerospace Science and Technology Corporation	中国航天科技集团公司
China North Industries Group Corporation	中国兵器工业集团有限公司
China South Industries Group	中国南方工业集团公司
China Electronics Corporation	中国电子信息产业集团有限公司
Aero Engine Corporation of China	中国航空发动机集团
China State Shipbuilding Corporation	中国船舶集团有限公司

ADDITIONAL ENTITIES SUPERVISED BY THE STATE ADMINISTRATION OF SCIENCE, TECHNOLOGY, AND INDUSTRY FOR NATIONAL DEFENCE (SASTIND) (MILITARY END-USE RISK)	
University of Electronic Science and Technology of China	电子科技大学
Hunan University of Science and Technology	湖南科技大学
Shenyang Aerospace University	沈阳航空航天大学
Southwest University of Science and Technology	西南科技大学
Central South University	中南大学
Sun Yat-Sen University	中山大学
Tianjin University	天津大学
Jilin University	吉林大学
Peking University	北京大学
Wuhan University	武汉大学
Yanshan University	燕山大学
Jiangsu University of Science and Technology	江苏科技大学
East China University of Technology	东华理工大学
Xiangtan University	湘潭大学
Hangzhou Dianzi University	杭州电子科技大学
Zhejiang University	浙江大学
Xi'An Jiao Tong University	西安交通大学
University of South China	南华大学
University of Science and Technology Beijing	北京科技大学
Nanchang Hangkong University	南昌航空大学

ENTITIES DIRECTLY CONTROLLED BY THE MPS OR THE PEOPLE'S ARMED POLICE (SURVEILLANCE END-USE RISK)	
Logistics University of the Chinese People's Armed Police Force	中国人民武装警察部队后勤学院
Engineering University of the Chinese People's Armed Police Force	中国人民武装警察部队工程大学
People's Public Security University of China	中国人民公安大学
Criminal Investigation Police University of China	中国刑事警察学院

OTHER ENTITIES WITH NOTEWORTHY POLICE TIES (SURVEILLANCE END-USE RISK)	
CAS Institute of Automation	中国科学院自动化研究所
Hangzhou Normal University	杭州师范大学
Xi'An University of Posts and Telecommunications	西安邮电大学
Xinjiang University	新疆大学
Dalian Minzu University	大连民族大学

CYBERSECURITY SCHOOLS	
Xidian University	西安电子科技大学
Beijing University of Posts and Telecommunications	北京邮电大学
Shandong University	山东大学
Shanghai Jiao Tong University	上海交通大学
Wuhan University of Technology	武汉理工大学
Sichuan University	四川大学
Huazhong University of Science and Technology	华中科技大学
Southeast University	东南大学
University of Science and Technology of China	电子科技大学

ENDNOTES

1 | On Chinese cybersecurity schools and their links to state-approved hacking campaigns, see: Cary, Dakota. Center for Security and Emerging Technology (2021). "China's CyberAI Talent Pipeline." July. https://cset.georgetown.edu/publication/chinas-cyberai-talent-pipeline/. Accessed: August 29, 2023; Cary, Dakota. Center for Security and Emerging Technology (2021). "Academics, AI, and APTs: How Six Advanced Persistent Threat-Connected Chinese Universities are Advancing AI Research." March. https://cset.georgetown.edu/publication/academics-ai-and-apts/. Accessed: August 29, 2023.
2 | OpenAlex. https://openalex.org/.
3 | Chou, Daniel. Center for Security and Emerging Technology (2022). "Counting AI Research: Exploring AI Research Output in English- and Chinese-Language Sources." July. https://cset.georgetown.edu/publication/counting-ai-research/. Accessed: August 29, 2023.
4 | For example, 581 research outputs in our dataset involved Shenzhen University. The university hosts a National Laboratory of Automated Target Recognition, but we do not know whether any Chinese co-authors in our sample work for this lab. Shenzhen University's ATR National Defense Technology Key Laboratory (深圳大学ATR国防科技重点实验室). Shenzhen University. https://archive.is/la5iR. Accessed: August 29, 2023.
5 | Chou, Daniel. Center for Security and Emerging Technology (2022). "Counting AI Research: Exploring AI Research Output in English- and Chinese-Language Sources." July. https://cset.georgetown.edu/publication/counting-ai-research/. Accessed: August 29, 2023; Hannas, William, Chang, Huey-Meei, Chou, Daniel and Fleeger, Brian. Center for Security an Emerging Technology (2022). "China's Advanced AI Research: Monitoring China's Paths to "General" Artificial Intelligence." https://cset.georgetown.edu/publication/chinas-advanced-ai-research/. Accessed: August 29, 2023.
6 | Crunchbase. https://www.crunchbase.com/.
7 | Weinstein, Emily S. and Luong, Ngor. Center for Security and Emerging Technology (2023). "U.S. Outbound Investment into Chinese AI Companies." February. https://cset.georgetown.edu/publication/u-s-outbound-investment-into-chinese-ai-companies/. Accessed: June 27, 2023.
8 | Lysenko, Adam, Hanemann, Thilo, and Rosen, Daniel H. Rhodium Group (2020). "Disruption: US-China Venture Capital in a New Era of Strategic Competition." January 13. https://rhg.com/research/disruption-us-china-venture-capital-in-a-new-era-of-strategic-distrust/. Accessed: August 29, 2023.
9 | See Appendix 2 in: Arnold, Zachary, Rahkovsky, Ilya, and Huang, Tina. Center for Security and Emerging Technology (2020). "Tracking AI Investment: Initial Findings from the Private Markets." September. https://cset.georgetown.edu/wp-content/uploads/CSET-Tracking-AI-Investment.pdf. Accessed: August 29, 2023.
10 | Ibid.
11 | Australian Strategic Policy Institute. China Defence University Tracker. https://unitracker.aspi.org.au. Accessed: August 29, 2023.

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

Antonia Hmaidi
Analyst, MERICS
antonia.hmaidi@merics.de

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

Nicolas Bouchet
Freelance Editor

GRAPHICS AND 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

ISSN (Print): 2941-5799
ISSN (Online): 2941-5608