January 2022

Trends in AI Research for the Visual Surveillance of Populations

CSET Data Brief

AUTHORS
Ashwin Acharya
Max Langenkamp
James Dunham

Executive Summary

Since 2014, computer vision models have dramatically improved their performance on benchmarks for image classification, image generation, facial recognition, and other tasks.[1] As these examples show, computer vision researchers aim to solve a wide variety of problems. Yet previous bibliometric studies have examined international output of computer vision research as a whole, without distinguishing among these many research tasks. In principle, analyses of academic research can inform us about the global growth of research and the interests and incentives of each nation's researchers. But only a small segment of computer vision research may relate to any particular area of interest. In this brief, we focus on "visual surveillance research," the development of algorithms such as facial recognition that could be used to surveil individuals or groups.[2]

These algorithms are often applied for benign, commercial uses, such as tagging individuals in social media photos. But progress in computer vision could also empower some governments to use surveillance technology for repressive purposes.

Using a dataset of English-language papers published between 2015 and 2019, we applied natural language processing methods to identify the computer vision papers in this corpus and the research tasks they described.[3] We found:

- **Papers relevant to visual surveillance accounted for less than 10 percent of all computer vision research.** This proportion was fairly consistent across geographic regions and time, varying from about 5 percent to 8 percent. In China, for example, we estimate that 6 percent of computer vision research papers were focused on visual surveillance tasks in both 2015 and 2019. The most common tasks in computer vision research are general-purpose ones that could improve performance in a wide variety of application areas.

- **Researchers with Chinese institutional affiliations were responsible for more than one third of publications in both computer vision and visual surveillance research.** This makes China by far the most prolific country in both areas**.** Chinese researchers' share of global visual surveillance research is growing at a similar rate to their share of computer vision research.

- **China produces a disproportionate share of research on three "emerging" surveillance tasks—those where worldwide publication counts grew at over 30 percent annually.** Chinese researchers are responsible for a large share of publications in person re-identification, crowd counting, and facial spoofing detection, and their share of publications in these areas is growing.

These conclusions derive from English-language research publications only. Further work could extend our task-extraction methodology to non-English publications, which would improve our estimates of international research output and would likely increase non-Anglophone countries' estimated share of global surveillance research. Future projects could also augment our bibliometric approach by including more direct indicators of nations' capabilities and interests in deploying surveillance technology, such as patent data, private research and development, camera deployment, and relevant government policies.

Table of Contents

Background and Methods

In recent years, progress in artificial intelligence has led to concern about the expanding capabilities of surveillance technology.[4] Stanford University's 2021 AI Index Report found that "surveillance technologies are increasingly fast, cheap, and ubiquitous."[5] Other research has cautioned that dual-use technology developed in the United States could empower population surveillance in non-democratic countries.[6] Indeed, such surveillance is now often augmented by AI. In the Xinjiang Uyghur Autonomous Region of China, the Chinese government is deploying its growing surveillance capabilities to track and repress a Muslim minority.[7] Security forces reportedly use the Integrated Joint Operation Platform (一体化联合作战平台) to monitor 13 million Uyghurs via facial recognition.[8] As the applications of AI in surveillance expand, policymakers express increasing interest in tracking and forecasting progress in AI research.[9]

Bibliometric analyses have the benefit of providing concrete metrics for important factors in technological development. Publications indicate the areas where a nation's researchers are focused, and cross-national comparisons of publication output can demonstrate international variation in research activity. At the same time, there are limitations to the insights available from bibliometric approaches. Research in the United States supports Chinese innovation and vice-versa; research in both countries can be useful to developers in smaller countries. The research literature also excludes classified research and unpublished industry research. Nonetheless, publication counts are useful as indicators of research fields' growth, as well as the interests and incentives of a nation's researchers.

In emerging fields like AI, attempts to analyze the trajectory of research are hindered by the technical nature of the field, the large and diverse quantity of research, and the speed of innovation. For instance, past research has identified China as the largest producer of AI papers, particularly in computer vision.[10] Yet computer vision is a broad field, comprising hundreds of thousands of papers and approximately one-third of all published AI research.[11]

To better understand trends in computer vision research, we analyzed publications in terms of research tasks: the particular problems that papers address. Papers With Code, a site that tracks AI benchmarks, lists more than a thousand research tasks in computer vision, from general ones like object detection to specific tasks such as handwritten digit recognition. Following prior work, we trained a SciREX model on data from Papers With Code. Given the text of papers' titles and abstracts, the model learned to identify references to tasks.[12]

We applied this model to the CSET merged corpus of scholarly literature from Dimensions, Web of Science, Microsoft Academic Graph, China National Knowledge Infrastructure, arXiv, and Papers With Code. This corpus is a set of over 100 million distinct publications from six academic datasets.[13] We identified computer vision papers using the SciBERT classifier, trained on arXiv preprints.[14] Since both the SciREX and SciBERT models are trained on English-language documents, our analysis is limited to English-language research. This means that in national comparisons it underestimates non-English research output, and in particular it likely underrepresents China's share of world research.

We attributed research contributions to countries by associating publications with the organizations at which their authors worked at the time of publication. When a publication's authors worked at institutions in multiple countries, we divided up credit for the paper evenly across all contributing countries. For example, a paper with authors from Tsinghua University in China and Carnegie Mellon University in the United States would add half of a paper to China's publication count and half of a paper to the United States' publication count.[15] When we calculated world totals, we counted all publications with at least one associated country.

Tasks in Computer Vision Research

The SciREX task model identified a large number of unique task references. For instance, across 68,400 computer vision papers published in 2019 with at least one extracted task, it identified over one million task terms. The vast majority of these task references were rare, appearing in fewer than 10 papers in 2019. These terms either did not refer to significant areas of research, or were specialized terms for more common tasks. We identified 1,400 task terms that appeared in 20 or more computer vision papers in 2019, of which only a hundred appeared in two hundred or more papers. To identify major trends in computer vision research, we focused on these hundred most frequent tasks in 2019.

We reviewed the hundred tasks that most commonly appeared in computer vision papers published in 2019 and found that they included many general-purpose tasks. For example, a reference to "object detection" might appear in a paper that improves on object detection in general, or to domain-specific progress in detecting a particular type of object. These common general-purpose tasks appeared in roughly half of the 68,400 computer vision papers published in 2019.

By contrast, tasks focused on a particular application were relatively rare; most common tasks fell into the general-purpose task area. Based on our manual grouping of tasks, the most common *applied* task area for common computer vision tasks was the visual surveillance of human populations.

Within the application domain of visual surveillance, face recognition was the most common computer vision research task, appearing in over a thousand papers in 2019. Other common visual surveillance tasks were person re-identification, action recognition, and emotion/expression recognition. When we looked at which computer vision tasks had the highest growth in number of publications from 2018 to 2019, we found two smaller but fast-growing visual surveillance tasks: crowd counting and face spoofing detection.

Table 1. The most frequent computer vision tasks in 2019 tended to be general-purpose. Surveillance was the most common applied task area.

TASK AREA	2019 PAPERS	TOP 10 MOST FREQUENT TERMS IN THE AREA
General*	30,994	Classification, Segmentation, Detection, Computer vision, Image processing, Object detection, Recognition, Feature extraction, Image segmentation, Image classification
Visual surveillance**	3,116	Face recognition, Action recognition, Face detection, Facial expression recognition, Video surveillance, Security, Human action recognition, Re-ID, Surveillance
Medical	2,493	Diagnosis, Medical imaging, Breast cancer, Medical image analysis, Diabetic retinopathy, Fusion, Medical image segmentation, Computer-aided diagnosis, Treatment
Remote sensing	1,263	Remote sensing, SAR, Target detection, Remote sensing images
Autonomous vehicles	724	Autonomous driving, Autonomous vehicles
Other	1,988	SR, Pose estimation, Hyperspectral image classification, Super-resolution, Pedestrian detection, Agriculture, Image super-resolution, Depth estimation

Source: CSET merged corpus. Results generated July 22, 2021.
* Truncated to 10 most common terms. See the project GitHub for the full 100 top CV task terms.
** These terms provided a starting point for the list used in our analysis. As shown in Figure 2, we ultimately identified 3,713 visual surveillance papers published in 2019.

We confirmed the relevance of these tasks to visual surveillance by reviewing their descriptions on the Papers With Code site and examining review papers that mentioned them. For each of the six tasks, these sources mentioned surveillance applications.

Surveillance Tasks

Based on their Papers With Code descriptions and the text of review papers that mentioned these tasks, we identified six computer vision tasks relevant to the visual surveillance of populations. These tasks may offer commercial or public-safety benefits, but could also be used to support state repression.

- **Face recognition:** Identifying a face in a photo or video by comparing it to a database of face images. It is used to identify individuals in various spaces, public and private.[16]
- **Person re-identification:** Identifying a person across time and space, typically in a video. It may use the same techniques involved in face recognition and gait recognition. It can also identify individuals in public and private spaces.[17] This task is also referred to as person recognition.
- **Action recognition:** Identifying a person's actions (e.g., running), typically from a video. It can identify individuals with abnormal behavior in crowds.[18]
- **Emotion recognition**: Classifying a person's emotion (e.g., anger) from a photo or video. In addition to its non-security-oriented and commercial purposes, some researchers, firms, and government agencies propose applying emotion recognition to identify security threats in crowded public areas.[19]
- **Crowd counting:** Counting the number of people in an image. It can be used to monitor civilian protests.[20]
- **Facial spoofing detection:** Detecting cases where a surveilled person is attempting to fool a face recognition system. While sometimes used in biometric login systems or to prevent fraud, it may also prevent journalists and activists from hiding their identity.[21]

For the most part, these tasks did not appear in the same publications; we found that only 2 percent of visual surveillance papers referenced more than one of these tasks.[22]

A task is often referred to with multiple terms, so we expanded our search to identify synonyms and sub-tasks for our six surveillance tasks. For example, “action recognition,” “human activity recognition,” and “HAR” all refer to the same task. However, not all of these variants would appear in our initial list of common and fast-growing tasks. In each paper, the SciREX model grouped *coreferences*—terms that likely refer to the same task. To capture alternative phrasings of our six tasks of interest, we used SciREX to identify common coreferences of our initial set of terms. We looked at these coreferences’ Papers With Code pages and reviewed papers to determine whether they were relevant to the task in question. See Appendix A for the final list of terms for each task.

We excluded some tasks that are relevant to surveillance but focus on vehicles or objects rather than humans. For example, “vehicle re-identification” is the task of recognizing the same vehicle across multiple images. These tasks are relevant to surveillance in general, but this brief limits its analysis to the surveillance of people. We also decided not to include general surveillance tasks such as “video surveillance”: while clearly relevant, they encompass a broader swathe of surveillance applications than those considered in this brief.

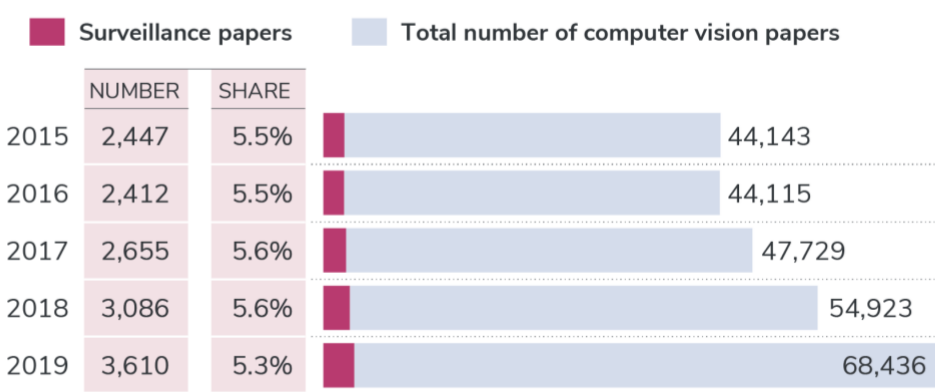
We also excluded many cross-domain tasks—those that are relevant to human surveillance but also to many other computer vision applications. For instance, “image denoising” is a common task used to enhance images applicable both to wedding photos and surveillance footage. Our analysis, therefore, may be considered a conservative summary of surveillance-related research, compared to an analysis of computer vision that considers all tasks with potential relevance to human surveillance or surveillance more broadly. For more detail on our methodology for selecting surveillance tasks, see Appendix A.

Findings

Visual surveillance papers constituted a small fraction of all computer vision papers.

Our analysis indicates that, despite the significant attention that visual surveillance receives, it made up a small fraction of all computer vision research.[23] Using our search process, we identified visual surveillance tasks in only 6 percent of computer vision papers. Computer vision research spans a wide range of tasks, and while surveillance tasks were among the most frequent, they made up only a small portion of the whole.

Figure 1. Visual surveillance papers accounted for less than 10 percent of all computer vision research.



Source: CSET merged corpus. Results generated July 22, 2021.

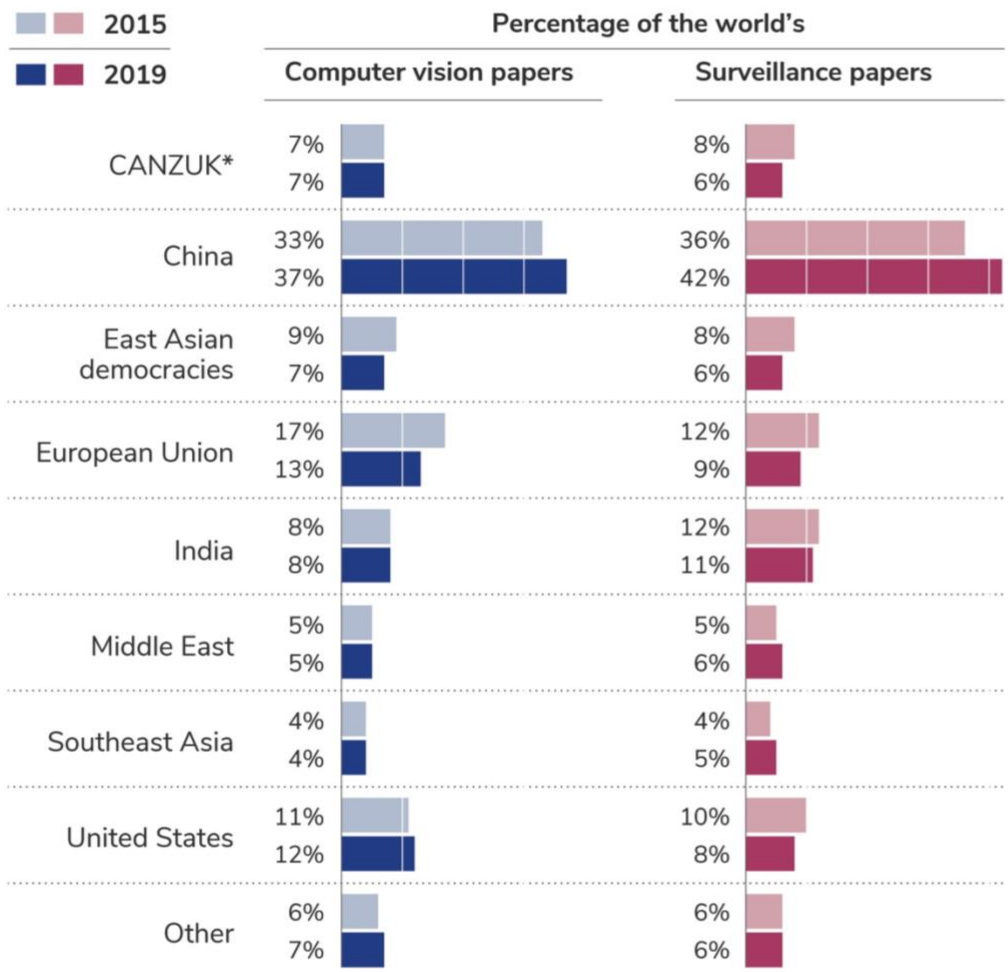
As shown in Figure 1, researchers around the world published roughly 260,000 computer vision research papers between 2015 and 2019.[24] Over the same period, we estimate they published approximately 14,000 visual surveillance papers, or 5.5 percent of all computer vision publications. The share of computer vision research focused on surveillance tasks remained fairly stable over time.

China was by far the most prolific publisher of both computer vision and visual surveillance research.

International comparisons of AI capabilities often start with the United States and China; some focus entirely on them. While they are not the only contributors to AI progress, their large economies and talent pools mean that these two countries have an outsized influence in AI research and development. As a result, some prior analyses have compared the United States and China to the European Union, rather than any individual European state.[25] Here, we aimed to map global contributions to visual surveillance research, not just publications from a few leading countries. We could not easily represent or analyze all individual countries, and most of these countries contribute only a small share of global research. We therefore grouped countries into a tractable number of regions.

We considered the top sources of computer vision papers—China, the United States, and India—as their own regions, and aggregated other countries such as those in the European Union or the Association of Southeast Asian Nations into regional groups.[26] Some of these regions, such as the European Union, contain countries with close economic, academic, and political ties. Other regions, such as the East Asian democracies of Japan, South Korea, and Taiwan, group countries with less formal ties but that have similar economic, cultural, or geopolitical characteristics (i.e., technologically advanced Asian democracies with strong ties to the United States). The countries in some regions, such as the Middle East and Southeast Asia, may have a wider distribution of technological capability and research output.

Figure 2. China was a large and growing contributor to computer vision and surveillance research.[27]



Source: CSET merged corpus. Results generated July 22, 2021.
* "CANZUK" refers to Canada, Australia, the United Kingdom, and New Zealand.

Globally, research output in computer vision and visual surveillance has grown over time, but China's publication rate rose especially rapidly between 2015 and 2019. As a result, China grew as a contributor to visual surveillance research, from 36 percent of global surveillance research in 2015 to 42 percent in 2019. By contrast, other regions held steady or lost research share.

China's leadership in surveillance research reflected trends in computer vision publications.

We found that the proportion of computer-vision research that addresses surveillance tasks varied moderately across countries and over time, and that Chinese computer vision research was not unusually focused on surveillance.

Figure 3. For most regions, between 4 percent and 8 percent of computer vision papers included visual surveillance tasks.

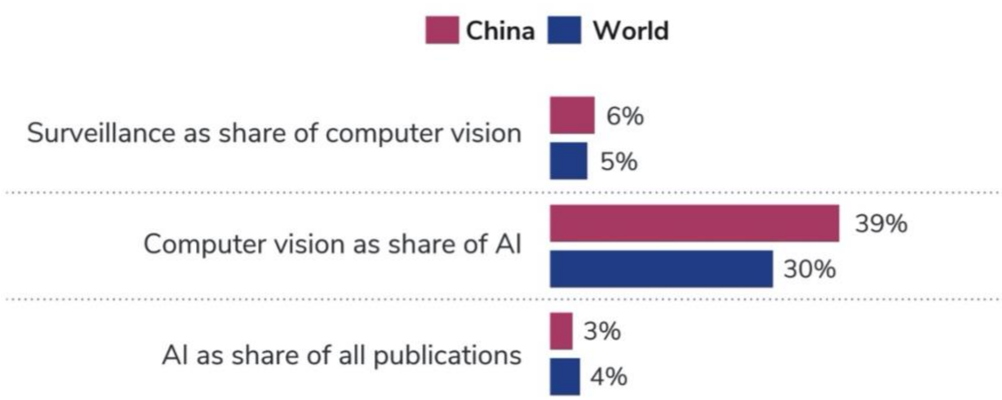
	2015	2019	Change, 2015-2019
CANZUK	6.4%	5.1%	-1.3
China	6.2%	6.1%	-0.1
East Asian democracies	5.2%	4.2%	-1.0
European Union	4.0%	3.9%	-0.1
India	7.8%	7.2%	-0.6
Middle East	5.4%	6.8%	1.4
Southeast Asia	6.3%	7.0%	0.7
United States	5.2%	3.7%	-1.5
Other	5.3%	5.0%	-0.3
World	**5.7%**	**5.4%**	**-0.3**

Note: Columns show the share of each region's computer vision publications that referenced surveillance tasks.
Source: CSET merged corpus. Results generated July 22, 2021.

As Figure 3 shows, researchers in most regions published between 4 percent and 8 percent of their computer vision papers on visual surveillance tasks, with China at 6 percent. The regions with the smallest surveillance proportion in 2019 were the United States and the European Union, both with 4 percent of computer vision research dedicated to visual surveillance. The regions with the highest surveillance proportion were India, Southeast Asia, and the Middle East, in which between 7 percent and 8 percent of computer vision research publications addressed visual surveillance tasks. Thus, while Chinese researchers generally published a larger proportion of their computer vision research on surveillance than those in the United States, this proportion was not unusually high.

Overall, China's focus on visual surveillance, as a share of its computer vision research, was not much higher than the world average and has remained stable over time. However, China's AI research was particularly focused on computer vision: as shown in Figure 4, 39 percent of Chinese AI research output was in computer vision, compared to the world average (over all countries) of 30 percent. China's focus on computer vision also stayed constant over time.

Figure 4. In 2019, Chinese AI research was more focused on computer vision than the world average.



Note: "World" statistics include China.
Source: CSET merged corpus. Results generated July 22, 2021.

As shown in Figure 5 below, Chinese research output in surveillance, computer vision, and AI all grew at a similar rate of about 14 percent annually, far faster than Chinese research overall. Chinese researchers' high output of visual surveillance research was because a particularly large share of Chinese computer vision research targeted surveillance tasks. This was commensurate with the large and growing amount of Chinese AI research, which was disproportionately concentrated in computer vision.

Figure 5. Chinese visual surveillance, computer vision, and AI research growth rates were comparable, and slightly outpaced the world average.

CHINA	Number of publications		Annualized growth rate,
	2015	2019	2015-2019
Surveillance	897	1,519	14%
Computer vision	14,711	25,325	15%
AI	38,923	64,136	13%
All publications	2,097,589	2,232,802	2%

WORLD	Number of publications		Annualized growth rate,
	2015	2019	2015-2019
Surveillance	2,447	3,610	10%
Computer vision	44,143	68,436	12%
AI	152,205	231,478	11%
All publications	5,435,039	5,940,103	2%

Note: “World” statistics include China.
Source: CSET merged corpus. Results generated July 22, 2021.

Surveillance task areas were diverse in terms of size and growth rate.

As seen in Figure 5, both visual surveillance research and computer vision broadly grew at a rapid pace in recent years, indicating a high level of interest from researchers and funders. While visual surveillance as a whole grew at a similar pace to overall computer vision research, the sub-areas of person re-identification, face spoofing detection, and crowd analysis saw exceptionally high growth rates (more than 30 percent annualized).

Growth was most dramatic for the subfield of person re-identification, which quadrupled from 150 papers per year in 2015 to almost 600 in 2019. Much of this growth came from Chinese research, which was the source of 56 percent of all person re-identification research papers from 2015 to 2019. For illustration, the most cited person re-identification publication in 2019 was "Bag of Tricks and a Strong Baseline for Deep Person Re-Identification." In this conference paper, researchers from several top Chinese institutes reported training a neural network that achieved high performance in tracking individuals across six different cameras in a supermarket.[28] Performance on this specific benchmark (the Market-1501 dataset) improved significantly from 2014 to 2019, but has since plateaued at 95 percent accuracy.[29] The "Bag of Tricks" paper suggested that applying general-purpose methods to improve neural net training was the biggest driver of this progress, rather than the development of methods for person re-identification in particular. This result demonstrates that general-purpose AI research has impacts across many application areas, suggesting that future analyses of surveillance-relevant research should also consider the progress being made on AI algorithms as a whole.

While face spoofing detection grew similarly quickly in percentage terms, the scale of research on this task remained small, with just 83 papers in 2019. The most cited spoofing detection publication in 2019 was "Multi-adversarial Discriminative Deep Domain Generalization for Face Presentation Attack Detection." In this conference paper, authors from a Hong Kong university noted that most spoofing detection methods fail to generalize across datasets, since they rely on detecting cues for "real" faces that might be specific to their training data. The authors reported building a general-purpose neural network that outperformed state-of-the-art systems on multiple face spoofing tasks.[30] The rapid progress demonstrated in this paper, combined with the small size of the field, suggests that face spoofing detection research could continue to grow rapidly in percentage terms. However, there may also be low-hanging fruit for research that improves the ability of individuals to circumvent face recognition algorithms.[31]

Finally, the field of crowd analysis was also fast-growing and relatively small, at 189 papers in 2019. Similar to surveillance tasks, crowd analysis has both relatively benign and malicious applications, everything from businesses tracking retail traffic to authoritarian regimes surveilling public gatherings.[32] The most highly-cited crowd counting publication in 2019 was "Learning from Synthetic Data for Crowd Counting in the Wild." In this conference paper, authors from a Chinese "Seven Sons" university noted that crowd counting had become a "hot topic" due to its applications in areas including "video surveillance [and] public security."[33] However, they found that the task was difficult to apply in practice due to variations in real-world circumstances.[34]

We consider these three tasks to be *emerging* tasks; the number of papers addressing them grew rapidly relative to computer vision, which in turn grew significantly faster than academic research as a whole. By contrast, our other three surveillance tasks are *established* tasks. Established tasks have hundreds of publications per year and a relatively stable growth rate—face recognition, for example, was particularly large and had a particularly slow growth rate. Two of the three emerging task areas, crowd counting and facial spoofing detection, had relatively few papers, but all three grew at a rapid pace.

Figure 6. Globally, computer vision research grew quickly relative to academic research as a whole, which grew at 2 percent annually.[35] Most surveillance tasks also grew rapidly, especially the three emerging tasks.

		Total papers in 2019	Annualized growth rate, 2015-2019
	Computer vision	68,436	12%
ESTABLISHED TASKS	Face recognition	1,402	1%
	Action recognition	857	11%
	Emotion recognition	573	13%
EMERGING TASKS	Person re-identification	593	39%
	Crowd counting	189	30%
	Facial spoofing detection	83	49%

Source: CSET merged corpus. Results generated July 22, 2021.

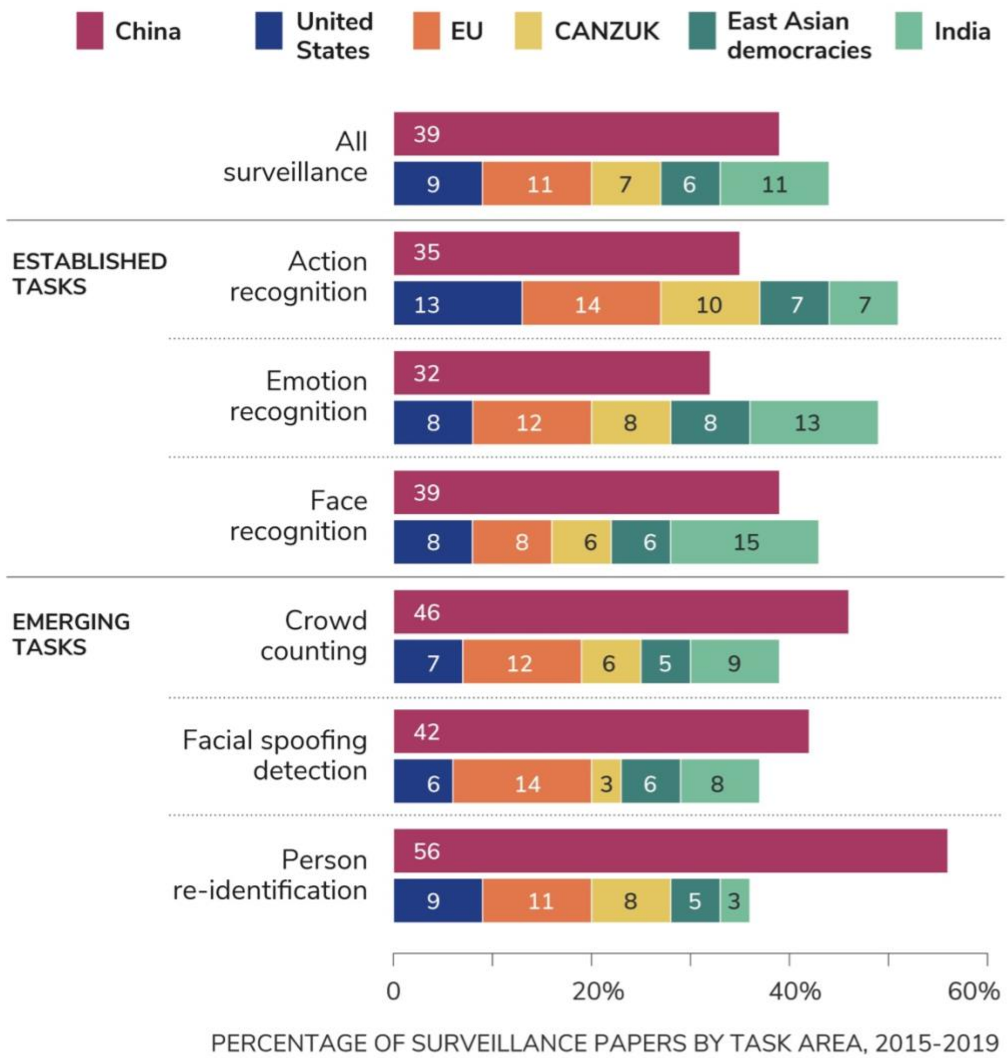
The three emerging visual surveillance task areas—person re-identification, crowd analysis, and face spoofing detection—grew rapidly each year. All three task areas demonstrated annualized growth rates of over 30 percent. Person re-identification is noteworthy for being a relatively large task area that underwent rapid growth in this period. A large portion of this growth can be attributed to Chinese research output, which accounted for 56 percent of person re-identification research from 2015 to 2019.[36]

Chinese researchers were especially prolific in emerging surveillance research areas.

When we look at the constituent tasks within visual surveillance, we see that China led the United States and other regions combined in most task areas over the 2015–2019 period. While China published less surveillance research overall than the combination of the United States with its allies and partners, China still led these regions in publishing research on emerging surveillance tasks.

Moreover, China's share of emerging surveillance areas grew considerably over time, especially in facial spoofing detection and person re-identification. As of 2019, Chinese researchers published nearly half of all research on crowd analysis and facial spoofing detection and almost two-thirds of research on person re-identification.

Figure 7. Over the 2015–2019 period, China outpaced the United States and its allies and partners in emerging surveillance tasks (crowd counting, facial spoofing detection, and person re-identification).



Source: CSET merged corpus. Results generated July 22, 2021.

Conclusion

In examining the global landscape of English-language research on computer vision tasks relevant to visual surveillance, we uncovered three broad findings. First, surveillance tasks constituted a relatively small fraction of computer vision research output as a whole. Research on the six visual surveillance tasks we focus on in this brief made up only about 6 percent of computer vision research output. Even a more expansive definition of surveillance, including all papers with tasks like "video surveillance" and "public safety," would have accounted for fewer than 7 percent of computer vision papers. However, progress on common general-purpose computer vision tasks such as image segmentation could improve performance in many domains, including surveillance.

Second, Chinese researchers were by far the largest contributors of both computer vision and surveillance research output, publishing more than one-third of research in both areas. China was particularly prominent in three visual surveillance tasks: person re-identification, facial spoofing detection, and crowd counting. All of these areas have applications related to population surveillance.[37]

Third, the share of both computer vision and visual surveillance from China increased over time. The United States, together with its allies and partners, published a similar amount of research in these areas as China published alone. However, these other regions' share of global surveillance research was stable or declined while China's grew.

Publication counts are a useful indicator of activity in a given research field, reflecting the interests and incentives of researchers. There are, however, some limitations to this indicator: some research activity is not reported in publications, and paper counts do not distinguish between high- and low-impact papers. Trends in classified research or in proprietary software development could differ from those in published research—they may focus on different tasks or on applying research to real-world circumstances. Our task extraction method also restricted our analysis to English-language literature, although our findings about the large scale of Chinese computer vision and surveillance research would only be strengthened by the inclusion of Chinese-language publications. Chinese work on emerging surveillance tasks was clearly significant compared to the rest of the world, but our analysis may have underestimated the growth rate of Chinese research in other surveillance tasks. Our analysis may have also failed to include surveillance research tasks that are primarily discussed in the Chinese-language literature.

Even the best data about research publications would not allow us to determine national capabilities in surveillance or other technologies. Innovations developed in major research-producing countries such as the United States and China could be adapted for the private sector in smaller countries with a technologically sophisticated workforce and funding base, such as Singapore or Israel. Innovations also spread between major producers of research, a trend that has given rise to concerns that openly published U.S. research may be enabling Chinese surveillance systems.[38] For practical applications of surveillance, other important factors include the deployment of cameras, the government's ability to collect and interpret data from a variety of sources, and the norms and institutions that empower governments to monitor and control citizens or empower citizens to resist them. Other analyses attempt to address these aspects of surveillance, providing a richer understanding of how these technologies are used today.[39]

Categorizing AI publications based on the research tasks they address adds a valuable level of detail to bibliometric analyses. Our examination of computer vision tasks offers a new perspective on the international distribution of surveillance research and highlights the degree to which computer vision research is focused on general-purpose tasks that could improve performance both in surveillance and in other application domains. Future work could make use of these task classification methods to provide a more detailed map of trends in AI research.

Authors

Ashwin Acharya is a research analyst at CSET, where Max Langenkamp was a semester research analyst and James Dunham is a data scientist.

Acknowledgments

For their feedback, we thank Catherine Aiken, Andrew Lohn, Igor Mikolic-Torreira, Dewey Murdick, and Helen Toner. We particularly thank José Hernández-Orallo for his comments as an external reviewer. For editing, we thank Melissa Deng, Alex Friedland, and Corey Cooper.

Document Identifier: doi: 10.51593/20200097

Appendix A: Selection of Surveillance Task Terms

Identifying Terms Related to Surveillance Tasks

As described in the Methodology section, we identified six common or fast-growing surveillance tasks within computer vision. However, we needed to ensure that we captured variations on these terms, as AI practitioners often refer to the same task with different terms. In our initial results, for example, we encountered both “person re-identification” and “re-ID”; these terms refer to the same task. For each paper, the SciREX pipeline predicts which terms are “coreferences”, meaning they have a common referent task. To identify all terms related to our surveillance tasks, we searched for terms that are often predicted to be coreferences of terms that we know to be relevant. For example, we found that the term “HAR” was often predicted as a coreference of “activity recognition.” In order to filter out large quantities of relatively rare or unrelated terms, we filtered for terms that were 1) common coreferences of our terms of interest and 2) appear in a substantial number of papers.[40]

We then investigated these related terms by searching for them in the Papers With Code data—for example, a search of the site shows that “HAR” is short for “human activity recognition.” If a term did not appear on the Papers With Code website, we looked at papers containing the term to determine whether it relates to our tasks of interest. (Specifically, we looked at the top 10 most highly cited papers with the term published between 2015 and 2019.) For example, we found that many top papers on “spoofing attacks” focused not on avoiding surveillance via facial spoofing but on forms of spoofing more relevant to bypassing biometric logins, such as iris, finger, or palmprint spoofing. As a result, we did not include this term as a related term for facial spoofing detection.

After we added a first set of related terms to our set of surveillance terms, we iterated our search to see if there are new relevant terms that overlap with this augmented set. That is, we looked for terms that 1) met our quantitative filters for relevance and prevalence to our existing list of terms of interest, including our newly added terms and 2) after investigation (of their Papers With Code page or relevant papers) appear to be closely related to our chosen surveillance tasks. After three rounds of iteration, we found that there were no new relevant terms to add.

This left us with a final set of 39 surveillance-related terms, reported in the table below. The additions we made to this table over several rounds of iteration are documented in the project GitHub as build and iterate all-task dictionary.sql. The search we used to find relevant tasks is documented in related tasks via task clusters.sql.

Final List of Surveillance Task Terms

Based on the process outlined above, we associated the following terms with each surveillance task. For example, if the SciREX task classifier identified "action classification" as one of a paper's task terms, we would count that paper as an action recognition paper.

Table A-1. Our list of terms associated with each surveillance task.

Task	Terms
Action recognition	action classification, action recognition, activity recognition, har, human action recognition, human activity recognition
Crowd counting	counting people, crowd analysis, crowd behavior analysis, crowd behavior recognition, crowd count, crowd counting, crowd density, crowd density estimation, crowd scene understanding, crowded scenes, people counting, people tracking
Facial spoofing detection	face anti-spoofing, face antispoofing, face liveness detection, face presentation attack detection, face presentation attacks, face spoofing, face spoofing attacks, face spoofing detection
Face recognition	face detection, face recognition, facial recognition
Emotion recognition	emotion recognition, expression recognition, facial emotion recognition, facial expression analysis, facial expression recognition, fer
Person re-identification	person detection, person re-identification, person reidentification, re-id, reid

Source: CSET analysis.

Surveillance-Related Terms We Chose Not to Include

In our search for surveillance task terms, we found a number of terms and tasks that relate to surveillance broadly construed. While these terms would be appropriate for a broader analysis of surveillance overall, we chose not to include them because they are not strictly relevant to our visual-human-surveillance tasks of interest.

In all cases, our search process would still identify papers that focus on the visual surveillance applications of these tasks, as long as the SciREX classifier identified them as targeting a surveillance task. For example, a paper that focused on the facial recognition applications of face alignment may be identified as both a face alignment and a face recognition paper; if so, it would be included in our search results.

We did not include in our search terms:

- More basic research that enables surveillance tasks, such as face alignment (a prerequisite for face recognition). Progress on these tasks indirectly can lead to better surveillance algorithms, but can also improve performance on other tasks. For example, face alignment can help improve digital photo filters.

- Broad tasks that sometimes relate to surveillance applications, such as event detection and anomaly detection. Unlike our chosen tasks, these are fairly broad, applying across domain areas.

- General terms such as “public safety” and “video surveillance.” These terms clearly relate to surveillance, but do not refer to a particular research task, making them harder to interpret. They may also include papers that focus on non-human surveillance, such as remote sensing of vehicles.
    - 26 percent of CV papers containing these general terms are captured by our specific surveillance tasks.

        - The relevant SQL query is *surveillance-general tasks - overlap with domain-specific surveillance.sql*
    - When we looked for other common terms that occur alongside these general surveillance terms, we did not identify any new domain-specific surveillance tasks.
        - The relevant SQL query is *surveillance-general tasks - find highly overlapping tasks.sql*
- Other application areas that make use of progress in surveillance-relevance domains, including biometric login systems (which make use of face recognition), human-computer interaction (which may include emotion recognition), and gesture recognition for video gaming (which relates to activity recognition).
- General task terms that include our surveillance tasks but also incorporate unrelated tasks that relate to non-surveillance applications such as biometric access systems. For example, the term "spoofing attacks" refers to both face spoofing attacks (which are relevant to anti- and pro-surveillance efforts) and fingerprint spoofing attacks (which are more specifically relevant to biometric login systems).

Further work could make use of these terms for a more detailed look at areas of surveillance-relevant computer vision research.

Appendix B: International Collaboration on Surveillance Research

We examined international collaboration rates and relationships among the top 10 publishers of visual surveillance research. We found that China was by far the most prolific publisher in this area, and that China and the United States (the top two countries) were the most common collaboration partners for third parties.

We also found that international collaborations made up a large share of the research output in this area; for most Western countries, including the United States, United Kingdom, Australia, France, and Italy, more than half of their visual surveillance publications were international collaborations. As a result, counting international collaborations as fractions of a paper gave relatively more weight to countries such as India and China, which published fewer international collaborations. Nonetheless, as the first two columns of Table B-1 show, the distribution of papers across countries was fairly similar whether we count international collaborations as a full paper toward each contributing country or divide them among their contributors.

Table B-1. The top ten publishers of visual surveillance research, 2015–2019.

Country	All visual surveillance papers with country affiliation	Fractional count of surveillance papers*	Number of papers that are international collaborations	Share of papers that are international collaborations	Top ten collaborating countries (number of collaborations)
China	6,293	5,600	1,313	21%	United States (525), Australia (219), United Kingdom (207), Singapore (127), Japan (72), Canada (60), Finland (34), France (33), Taiwan (28), Korea (28)
United States	1,894	1,326	1,052	56%	China (525), United Kingdom (106), India (67), Australia (58), Italy (50), Canada (45), Singapore (43), Germany (35), France (30), Korea (29)
India	1,681	1,596	161	10%	United States (67), Korea (18), China (11), Italy (9), United Kingdom (8), Singapore (7), Germany (7), Norway (6), Australia (5), France (5)
United Kingdom	751	483	481	64%	China (207), United States (106), Australia (31), Spain (28), Germany (26), Pakistan (24), Italy (23), France (21), Netherlands (18), Greece (13)
Australia	496	319	322	65%	China (219), United States (58), United Kingdom (31), Singapore (23), Netherlands (8), South Korea (7), Iraq (7), Saudi Arabia (7), Malaysia (6), India (5)
South Korea	457	390	122	27%	United States (29), China (28), India (18), Pakistan (16), Australia (7), United Kingdom (5), Canada (5), Vietnam (5), Saudi Arabia (4), Brazil (4)

France	395	272	221	56%	China (33), United States (30), Tunisia (29), Algeria (21), United Kingdom (21), Italy (21), Switzerland (16), Spain (15), Germany (12), Belgium (10)
Japan	387	300	163	42%	China (72), United States (22), Vietnam (17), United Kingdom (11), Singapore (7), Bangladesh (6), Italy (6), Taiwan (5), Malaysia (5), Romania (4)
Italy	307	218	158	51%	United States (50), United Kingdom (23), France (21), China (20), Spain (15), India (9), Romania (7), Germany (6), Japan (6), Canada (6)
Germany	304	236	120	39%	United States (35), United Kingdom (26), China (16), France (12), Netherlands (10), Spain (7), India (7), Italy (6), Canada (6), Switzerland (5)

Source: CSET analysis. Generated October 1, 2021.
* i.e., the counting methodology used throughout this brief.

A full table of international collaboration counts in visual surveillance research is available in the project GitHub as *International collaborations in surveillance research.tsv*.

Endnotes

[1] Daniel Zhang et al., “Artificial Intelligence Index Report 2021” (Stanford University Center on Human-Centered Artificial Intelligence, 2021), https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report_Master.pdf.

[2] As a result of this focus, we chose not to include some search terms that are related to surveillance more broadly, and not just the visual surveillance of human populations, such as “video surveillance.” See Appendix A for more discussion on the selection of search terms.

[3] See James Dunham, Jennifer Melot, and Dewey Murdick, “Identifying the Development and Application of Artificial Intelligence in Scientific Text,” arXiv preprint, arXiv:2002.07143 (2020), https://arxiv.org/abs/2002.07143; for information on SciREX, see Sarthak Jain, Madeleine van Zuylen, Hannaneh Hajishirzi, and Iz Beltagy, “SciREX: A Challenge Dataset for Document-Level Information Extraction,” arXiv preprint, arXiv:2005.00512 (2020), https://arxiv.org/abs/2005.00512.

[4] See, e.g., Yuval Noah Harari, “The World After Coronavirus,” *The Financial Times*, March 20, 2021, https://www.ft.com/content/19d90308-6858-11ea-a3c9-1fe6fedcca75.

[5] Zhang et al., 2021.

[6] “Due to the dual-use nature of facial recognition and other biometrics-detection technology, U.S. organizations are at risk of indirectly contributing to these human rights violations through research collaborations, technology exports, and investments.” Martijn Rasser et al., “The American AI Century: A Blueprint for Action” (Center for a New American Security, December 17, 2019), https://www.cnas.org/publications/reports/the-american-ai-century-a-blueprint-for-action.

[7] See, e.g., Lindsay Maizland, “China’s Repression of Uyghurs in Xinjiang,” Council on Foreign Relations, March 1, 2021, https://www.cfr.org/backgrounder/chinas-repression-uyghurs-xinjiang.

[8] “How Mass Surveillance Works in Xinjiang, China,” Human Rights Watch, May 2, 2019, https://www.hrw.org/node/329492; See also Dahlia Peterson, “Designing Alternatives to China's Repressive Surveillance State” (Center for Security and Emerging Technology, October 2020), https://cset.georgetown.edu/publication/designing-alternatives-to-chinas-repressive-surveillance-state/.

[9] See, e.g., U.S. Department of Defense, "DOD Official Briefs Reporters on Artificial Intelligence Developments," July 8, 2020, https://www.defense.gov/Newsroom/Transcripts/Transcript/Article/2270329/dod-official-briefs-reporters-on-artificial-intelligence-developments/. See also the OECD AI Policy Observatory.

[10] Dewey Murdick, James Dunham, and Jennifer Melot, "AI Definitions Affect Policymaking" (Center for Security and Emerging Technology, June 2020), https://cset.georgetown.edu/publication/ai-definitions-affect-policymaking/.

[11] Murdick, Dunham, and Melot, "AI Definitions Affect Policymaking."

[12] Jain et al., "SciREX: A Challenge Dataset for Document-Level Information Extraction."

[13] CSET's merged corpus of scholarly literature includes Digital Science's Dimensions, Clarivate's Web of Science, Microsoft Academic Graph, China National Knowledge Infrastructure, arXiv, and Papers With Code. China National Knowledge Infrastructure is furnished for use in the United States by East View Information Services, Minneapolis, MN, USA.

[14] Dunham, Melot, and Murdick, "Identifying the Development and Application of Artificial Intelligence in Scientific Text."

[15] We attribute papers equally to each country where an author's affiliated institution is located, regardless of how many authors or institutions are in any country. For example, a paper with three authors from UK institutions and one author from a U.S. institution counts as half of a paper for each country. We use this method because our objective is to identify how many papers are associated with each country as a measure of country output across research tasks. An alternative approach like weighting papers by the number of authors or affiliations in each country would introduce more complex assumptions about authors' relative contributions and would require robust organizational entity resolution (e.g., recognizing that "Oxford University" and "University of Oxford" refer to just one distinct UK affiliation).

[16] For various examples, see Antoaneta Roussi, "Resisting the Rise of Facial Recognition," *Nature*, November 18, 2020, https://doi.org/10.1038/d41586-020-03188-2.

[17] See, e.g., Dat Tien Nguyen et al., "Person Recognition System Based on a Combination of Body Images from Visible Light and Thermal Cameras," *Sensors* 17, no. 3 (2017), https://doi.org/10.3390/s17030605. And Thi Thanh Thuy Pham et al., "Fully-Automated Person Re-Identification in Multi-Camera Surveillance System with a Robust Kernel Descriptor and Effective Shadow

Removal Method," *Image and Vision Computing* 59 (March 2017): 44–62, https://doi.org/10.1016/j.imavis.2016.10.010.

[18] See, e.g., Cheng-Bin Jin, Shengzhe Li, and Hakil Kim, "Real-Time Action Detection in Video Surveillance Using Sub-Action Descriptor with Multi-CNN," arXiv preprint, arXiv:1710.03383, 29, https://arxiv.org/abs/1710.03383.

[19] The UK firm WeSee has partnered with a law enforcement agency to analyze videos of interviewees using emotion recognition tools, and its CEO has proposed applying WeSee technologies to identify potential threats in train stations. See David Fulton, "The cameras that know if you're happy - or a threat," *BBC News*, July 17, 2018, https://www.bbc.com/news/business-44799239. In a recent paper, researchers at the Airport Research Institute in Beijing suggested using an emotion recognition system to identify potential flight risks. See Weishi Chen et al., "A passenger risk assessment method based on 5G-IoT," *EURASIP Journal on Wireless Communications and Networking* 2021, no. 1 (2021): 1-20. The Transportation Security Administration trialed a facial recognition system as early as 2007, prior to the modern era of deep learning methods. See Oscar Schwartz, "Don't Look Now: Why You Should Be Worried about Machines Reading Your Emotions," *The Guardian*, March 6, 2019, sec. Technology, https://www.theguardian.com/technology/2019/mar/06/facial-recognition-software-emotional-science.

[20] For example: "There are many potential real-world applications in crowd counting, e.g. surveillance in public for safety and security by detecting [*sic*] abnormally large crowd...", from Zhaoxiang Zhang, Mo Wang, and Xin Geng, "Crowd Counting in Public Video Surveillance by Label Distribution Learning," *Neurocomputing* 166 (October 2015): 151–63, https://doi.org/10.1016/j.neucom.2015.03.083.

[21] Some relevant examples include preventing the use of anti-surveillance clothing (see Alex Hern, "Anti-Surveillance Clothing Aims to Hide Wearers from Facial Recognition," *The Guardian*, January 4, 2017, https://www.theguardian.com/technology/2017/jan/04/anti-surveillance-clothing-facial-recognition-hyperface). However, there have also been cases of spoof detection methods used to prevent face spoofs that target activists (see Raphael Satter, "Deepfake Used to Attack Activist Couple Shows New Disinformation Frontier," Reuters, July 15, 2020, https://www.reuters.com/article/us-cyber-deepfake-activist/deepfake-used-to-attack-activist-couple-shows-new-disinformation-frontier-idUSKCN24G15E). We have chosen to include the task for its potential relevance, but acknowledge that the task is small and that its link to surveillance may be limited.

[22] The rate of overlap was particularly high for facial emotion recognition and facial spoofing detection; 8 percent of the former and 24 percent of the latter papers are also tagged with facial recognition. Since these tasks are closely related to face recognition, but are newer and more specific, these overlap

papers are likely in fact relevant to these tasks. While the overlap papers might be less relevant to face recognition in general, they make up a small portion of face recognition papers (about 4 percent), so they do not substantially alter our face recognition results. In the results we report below, we count overlapping papers towards all of their surveillance tasks. When we refer to the number of surveillance papers overall, we count unique papers and do not double-count papers with multiple tasks.

[23] See, e.g., Mahesh Saptharishi, "The New Eyes of Surveillance: Artificial Intelligence and Humanizing Technology," *WIRED*, August 2014, https://www.wired.com/insights/2014/08/the-new-eyes-of-surveillance-artificial-intelligence-and-humanizing-technology/; Tom Simonite, "Thanks to AI, These Cameras Will Know What They're Seeing," *WIRED*, April 17, 2010, https://www.wired.com/story/thanks-to-ai-these-cameras-will-know-what-theyre-seeing/; Niraj Chokshi, "How Surveillance Cameras Could Be Weaponized With AI," *The New York Times*, June 13, 2019, https://www.nytimes.com/2019/06/13/us/aclu-surveillance-artificial-intelligence.html.

[24] See the project ReadMe on GitHub for information on our methodology for identifying computer vision papers.

[25] See, e.g., Daniel Castro and Michael McLaughlin, "Who Is Winning the AI Race: China, the EU, or the United States? — 2021 Update" (Information Technology and Innovation Foundation, January 2021), https://itif.org/publications/2021/01/25/who-winning-ai-race-china-eu-or-united-states-2021-update.

[26] We include the United Kingdom in a group of four Anglophone countries—Canada, Australia, New Zealand, and the United Kingdom, or CANZUK—rather than in the European Union.

[27] East Asian democracies" refers to Japan, South Korea, and Taiwan. "European Union" or "EU" refers to the 27 EU countries: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, and Sweden. "Middle East" refers to Egypt, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Pakistan, Qatar, Saudi Arabia, Syria, Turkey, United Arab Emirates, and Yemen. "Southeast Asia" refers to the countries in the Association of Southeast Asian Nations (ASEAN): Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, and Vietnam. "Other" refers to all countries not otherwise listed.

[28] Hao Luo et al., "Bag of Tricks and a Strong Baseline for Deep Person Re-Identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019. For a summary of the Market-1501

dataset, see “Market-1501,” AI Tribune, https://www.aitribune.com/dataset/2018051063.

[29] See “Person Re-Identification on Market-1501,” Papers With Code, https://paperswithcode.com/sota/person-re-identification-on-market-1501.

[30] Rui Shao et al., "Multi-Adversarial Discriminative Deep Domain Generalization for Face Presentation Attack Detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, 10023–10031.

[31] This is one of the policy prongs recommended in Tim Hwang, "Shaping the Terrain of AI Competition" (Center for Security and Emerging Technology, June 2020), cset.georgetown.edu/research/shaping-the-terrain-of-ai-competition/.

[32] For discussions of crowd counting’s business applications, see e.g., Bryan Walsh, “The Coronavirus Must-Have: Crowd-Counting Apps,” *Axios*, July 25, 2020, https://www.axios.com/coronavirus-crowd-counting-apps-ce552923-922f-4445-a023-e0d96c2f3142.html.

[33] The “Seven Sons” are seven Chinese universities with close ties to the Chinese national security apparatus and a focus on dual-use science and technology. See e.g., Ryan Fedasiuk and Emily Weinstein, “Universities and the Chinese Defense Technology Workforce” (Center for Security and Emerging Technology, December 2020), https://cset.georgetown.edu/publication/universities-and-the-chinese-defense-technology-workforce/.

[34] Qi Wang et al., "Learning from synthetic data for crowd counting in the wild," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, 8198–8207.

[35] Richard W. Klavans, Kevin Boyack, and Dewey Murdick, “A novel approach to predicting exceptional growth in research,” *PLOS ONE* (September 2020).

[36] See the ReadMe on the project GitHub for a description of how we attribute papers to countries.

[37] “There are many potential real-world applications in crowd counting, e.g. surveillance in public for safety and security by detecting abnormally large crowds…” from Zhaoxiang Zhang, Mo Wang, and Xin Geng, “Crowd Counting in Public Video Surveillance by Label Distribution Learning,” *Neurocomputing* 166 (October 2015): 151–63, https://doi.org/10.1016/j.neucom.2015.03.083. See also Nguyen et al., “Person Recognition System Based on a Combination of Body Images from Visible Light and Thermal Cameras” and Pham et al., “Fully-Automated Person Re-Identification in Multi-Camera Surveillance System with a Robust Kernel Descriptor and Effective Shadow Removal Method.”

[38] Rassjer et al., "The American AI Century."

[39] For an analysis of Chinese security forces' research efforts, see Dewey Murdick et al., "The Public AI Research Portfolio of China's Security Forces: A High-Level Analysis" (Center for Security and Emerging Technology, March 2021), https://cset.georgetown.edu/publication/the-public-ai-research-portfolio-of-chinas-security-forces/. This CSET analysis examined the AI publications authored by China's security forces. It found that the vast majority of these publications were in Chinese, but only a small share of them were about surveillance of populations. The analysis found 470 computer vision publications published between 2010 and 2019 that were produced by the Ministry of Public Security, the branch of the Chinese government that oversees domestic policing and produces a significant share of its publications on surveillance-related research. By contrast, most of the security forces' 50,000 AI publications in that period were supported by Chinese military branches, whose computer vision research appeared to focus on topics such as remote sensing and target recognition. Further work extending our task-tagging methodology to Chinese publications could provide a more detailed understanding of the progress and interests suggested by the Chinese security forces' publications. For one analysis of security camera placements, see Paul Bischoff, "Surveillance camera statistics: which cities have the most CCTV cameras?," *Comparitech,* May 17, 2021, https://www.comparitech.com/vpn-privacy/the-worlds-most-surveilled-cities/. For a discussion of data fusion capabilities, see Dahlia Peterson, "How China harnesses data fusion to make sense of surveillance data," *Brookings TechStream*, September 23, 2021, https://www.brookings.edu/techstream/how-china-harnesses-data-fusion-to-make-sense-of-surveillance-data/.

[40] Specifically, if task A is one of our six surveillance tasks, we filter for terms B where 1) at least 5 percent of papers with B have A as a coreference, and 2) where including B would increase the number of papers in A by at least 5 percent. The latter filter excludes the large number of tasks that appear in very few papers, such as the hundreds of thousands of tasks that only appear in a single paper.