

# Computer-vision research powers surveillance technology

<https://doi.org/10.1038/s41586-025-08972-6>

Received: 28 September 2023

Accepted: 3 April 2025

Published online: 25 June 2025

Open access

 Check for updates

Pratyusha Ria Kalluri<sup>1</sup>✉, William Agnew<sup>2</sup>✉, Myra Cheng<sup>1</sup>, Kentrell Owens<sup>3</sup>, Luca Soldaini<sup>4</sup> & Abeba Birhane<sup>5</sup>✉

An increasing number of scholars, policymakers and grassroots communities argue that artificial intelligence (AI) research—and computer-vision research in particular—has become the primary source for developing and powering mass surveillance<sup>1–7</sup>. Yet, the pathways from computer vision to surveillance continue to be contentious. Here we present an empirical account of the nature and extent of the surveillance AI pipeline, showing extensive evidence of the close relationship between the field of computer vision and surveillance. Through an analysis of computer-vision research papers and citing patents, we found that most of these documents enable the targeting of human bodies and body parts. Comparing the 1990s to the 2010s, we observed a fivefold increase in the number of these computer-vision papers linked to downstream surveillance-enabling patents. Additionally, our findings challenge the notion that only a few rogue entities enable surveillance. Rather, we found that the normalization of targeting humans permeates the field. This normalization is especially striking given patterns of obfuscation. We reveal obfuscating language that allows documents to avoid direct mention of targeting humans, for example, by normalizing the referring to of humans as ‘objects’ to be studied without special consideration. Our results indicate the extensive ties between computer-vision research and surveillance.

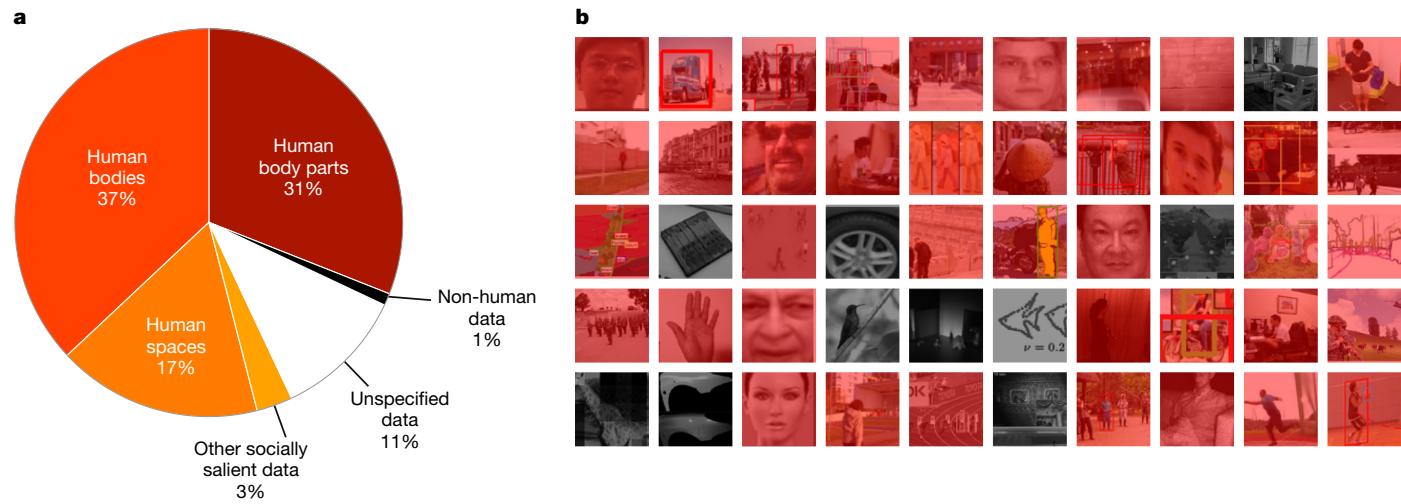
Various groups, ranging from grassroots communities to academics to legislators, have drawn attention to and organized against the rise of mass surveillance<sup>8,9</sup>, arguing that artificial intelligence (AI) research—particularly in computer vision—serves as a foundation for the design, development and implementation of modern surveillance<sup>1,2,4–7</sup>. If these claims are true, the rapidly growing field of computer vision is contributing to the legacy of surveillance technologies, technologies that have infringed on privacy, limited free expression, exacerbated disparities and created conditions facilitating abuse of power<sup>4,6,9–15</sup>. A precise account of the pathway from computer-vision research to surveillance is valuable because it empowers individuals and communities to make informed decisions regarding their role and could effectively influence the development of computer vision and surveillance. We characterize the nature and extent of the surveillance AI pipeline and illuminate the critical role of computer vision in facilitating surveillance.

Computer vision refers to AI that focuses on measuring, recording, representing and analysing the world from visual inputs such as image and video data. Computer vision has historical roots in military and carceral surveillance, where it was historically developed to identify targets and gather intelligence in war, law enforcement and immigration contexts<sup>16–18</sup>. Over time, the priorities of computer vision have continued to be shaped by a confluence of social influences beyond individual researchers’ interests, including the interests of academic institutions, funding agencies, governments, companies and larger systemic

pressures<sup>19</sup>. The field of computer vision now generally conceptualizes itself as a scientific, statistical, engineering and data-driven endeavour inspired by human vision<sup>20</sup>. An emphasis is often placed on mathematically well-founded approaches to training computers to interpret, classify, identify patterns, model and reproduce the visual world<sup>20,21</sup>. Stated topics of interest include general and application-agnostic computer-vision techniques as well as applications such as robotics and autonomous driving<sup>20,22</sup>. Prestigious computer-vision conferences have also highlighted other applications, often referred to as ‘computer vision for social good’, such as computer-vision tools to facilitate designing new proteins, creating art and modelling climate change<sup>23,24</sup>. Yet, prominent computer-vision tasks, such as facial recognition, continue to be tightly tied to military and carceral use, heavily shaping core aspects and uses of these subfields<sup>18,25</sup>. This motivates interrogation into the extent to which the field of computer vision as a whole has been shaped in a way that continues to power mass surveillance.

Drawing upon surveillance studies, surveillance is defined as an entity gathering, extracting or attending to data connectable to other persons, whether individuals or groups<sup>10</sup>. Frequently, bodies, behaviours, relationships, and social and physical environments are datafied, modelled and profiled. This formal conceptualization of surveillance includes a wide range of activities. In recent years, surveillance has become ‘extensive’: entities, who are often minimally visible, use big datasets and aggregation to extend their reach, accessing previously

<sup>1</sup>Computer Science Department, Stanford University, Stanford, CA, USA. <sup>2</sup>Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, USA. <sup>3</sup>Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA. <sup>4</sup>Allen Institute for Artificial Intelligence (AI2), Seattle, WA, USA. <sup>5</sup>School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland. ✉e-mail: pkalluri@stanford.edu; wagnew@andrew.cmu.edu; birhanea@tcd.ie



**Fig. 1 | Human data extraction in computer-vision papers and downstream patents.** **a**, Relative frequencies of data types extracted from computer-vision papers and patents. For each year from 2010 to 2019, we randomly sampled and analysed ten paper–patent pairs ( $n = 200$ ). Most of the annotated computer-vision papers and patents (88%, s.d. = 5.7%) refer to data about humans. Most of the papers and patents (68%, s.d. = 4.7%) specifically refer to data about human bodies and body parts. Only 1% (s.d. = 0.7%) of the papers and patents targeted exclusively non-socially salient data. **b**, Examples of images analysed in computer-vision papers. For a random sample of the computer-vision papers ( $n = 50$ ), we display one example of an image analysed

by the paper. For papers that analysed any images containing humans, we display an example of these images of humans (highlighted in red). For papers that did not analyse any images of humans, we display an example of these non-human-depicting images (highlighted in grey). Many papers analysed images of humans. Images in **b** are adapted from the following references and are, unless otherwise stated, from IEEE, under a Creative Commons licence CC BY ND. Top row, left to right: refs. 51–60. Second row, left to right: refs. 61–70. Third row, left to right: refs. 71–80. Fourth row, left to right (except second image): refs. 81–89. Fifth row, left to right: refs. 90–99. Fourth row (second image): ref. 100, arXiv, under a non-exclusive licence to distribute.

unseen persons, locations or information<sup>2,19,26</sup>. Prominent examples are practices where entities in positions of power observe, monitor, track, profile, sort or police individuals and populations in private and public spaces through devices such as CCTV, digital traces on social network sites or biometric monitoring of bodies<sup>14</sup>. Surveillance frequently occurs at loci of influence and control, such as targeted recommendation and personalization algorithms that have become widely used on the internet<sup>27</sup>. Through such ubiquitously connected networks, data are gathered, shared and aggregated<sup>28</sup>. Many scholars emphasize that surveillance is inextricable from purposes such as influence, management, coercion, repression, discipline and domination<sup>10,29</sup>.

A foundational understanding in surveillance studies is that technologies and processes may enable surveillance, even when they are not labelled as surveillance, not universally perceived as nefarious or not inflicting immediate, visible violence<sup>4,15</sup>. Technologies enabling the possibility of monitoring human data suffice to foster conditions of fear and self-censorship where this approach is a key means of social control<sup>12</sup>. In some cases, technologies that monitor humans to enable surveillance may be proposed and perceived by particular communities as connected to benevolent purposes. Importantly, these technologies continue to constitute tools enabling surveillance. Because value assessments are subjective and contested, the same technologies may be perceived and experienced by other communities as oppressive where such technologies are frequently ‘spun into’ mass surveillance by entities in positions of power<sup>25,30–32</sup>. In Supplementary Information sections 1.3, 1.4 and 1.6, we provide a more extensive review of the contextualizing literature and descriptions of the types of data transferal and institutional uses that contribute to surveillance.

To study the pathway from computer-vision research to surveillance, we collected and analysed a corpus linking more than 19,000 computer-vision research papers from the longest standing computer-vision conference, the *Conference on Computer Vision and Pattern Recognition* (CVPR), to more than 23,000 downstream citing patents. Using a mixed methods content analysis and large-scale lexicon-based analysis, we characterized the roots, extent, evolution and obfuscation

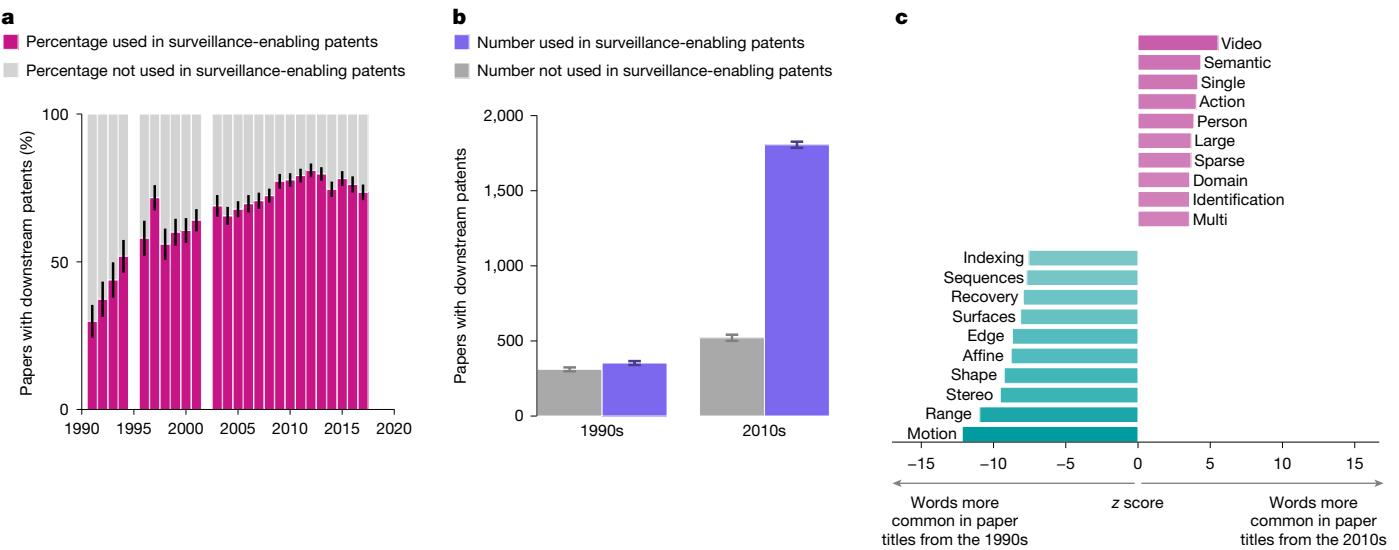
of computer-vision-based surveillance, collectively forming what we identify as the surveillance AI pipeline.

## The extent of human data extraction

Extensive evidence shows public distrust and fear concerning the capturing and monitoring of human data, including substantial concern about computer-vision technologies operating on data ranging from online personal data traces to biometric and body data<sup>9,25,33</sup>. To identify the potentially numerous and subtly expressed variants of human data extraction actively enabled by computer vision, we conducted a mixed methods content analysis of a randomly sampled subset of the corpus. We analysed 100 computer-vision papers and 100 downstream patents, annotating all declarations and demonstrations of human data extraction. (Herein we refer to these as ‘Paper X/Patent X’ and further details can be found in ref. 34). In the context of manual content analysis, this constitutes an in-depth, large-scale analysis.

## Quantitative analysis

We quantify the uncovered types of human data targeted in the computer-vision papers and downstream patents in Fig. 1a. We additionally present a stratification of this data to compare human data extraction in papers versus patents in Supplementary Information Fig. 2. We found that 90% of papers and 86% of downstream patents extracted data relating to humans. Most (71% of the papers and 65% of the patents) explicitly extracted data about human bodies and body parts. In particular, 35% of papers and 27% of patents targeted human body part data, and at least another third of both papers and patents (36% of papers and 38% of patents) claimed or demonstrated targeting human bodies at large. Another portion of papers and patents (18% of papers and 16% of patents) extracted data about human spaces. We found a small number of papers and patents (1% of papers and 5% of patents) presented their technology as useful only for analysing non-body-related socially salient human data. Finally, the remaining small portion of papers and patents (9% of papers and 13% of patents) claimed to capture and analyse ‘images’, ‘text’, ‘objects’ or similarly



**Fig. 2 | The rise of computer vision with downstream surveillance.** **a**, Across three decades of patented computer-vision papers ( $n = 11,917$ ), there has been a steady increase in the proportion used in surveillance-enabling patents. Whiskers represent the standard deviation. **b**, Number of computer-vision papers used in surveillance-enabling patents. Comparing the 1990s to the 2010s, the number of computer-vision papers used in only non-surveillant patents has remained relatively stable, whereas the number of computer-vision papers used in surveillance-enabling patents has risen more than fivefold.

Whiskers represent the standard deviation. **c**, Differences in word frequencies between paper titles from the 1990s versus those from the 2010s. We report highly polarized words (with z scores computed using weighted log odds ratios). All word associations shown are statistically significant ( $P < 0.01$ ). There is a clear qualitative shift from the more generic paper focus of the 1990s (turquoise bars) to an increased focus on analysing semantic categories and humans (for example, ‘semantic’, ‘action’ or ‘person’) in the 2010s (pink bars).

generic terms, leaving unstated whether they anticipated these categories would include humans or human data. Strikingly, only 1% of papers and 1% of patents were dedicated to extracting only non-human data, revealing that both computer-vision research and its applications are extensively involved with datafying humans, specifically human bodies.

### Qualitative analysis

Our findings challenge narratives that most kinds of computer vision and data extraction are largely benign or harmless and that only a small portion is harmful. Rather, we found that the computer-vision papers and patents prioritize intrusive forms of data extraction that are well established in surveillance studies scholarship. The four targets of human data extraction that emerged during the content analysis form a series of increasingly focused categories: socially salient human data, human spaces, human bodies and human body parts. We present the uncovered types of human data extraction alongside textual examples in Table 1.

The papers and patents broadly prioritize tasks targeting ‘human body part’ data, particularly facial analysis, and sometimes enable activity classification. This validates the substantial concerns that have been put forth regarding the collection, aggregation and sharing of biometric and related body part data<sup>35</sup>. Biometrics (for example, faces, fingerprints and gait), which constitute uniquely personal data that is often inseparable from identities, have proliferated as a form of surveillance in recent years<sup>36</sup>. Their pervasiveness has significantly infringed on fundamental human rights, including rights to privacy and freedom of expression and movement<sup>9,37</sup>.

The papers and patents targeting ‘human bodies’ at large frequently targeted humans in the midst of everyday activities (for example, walking, shopping and attending group events), and the named purposes included body detection, tracking and counting, as well as security monitoring and human activity recognition. The dominance of analysis of human bodies in everyday settings aligns with the view of new surveillance by Browne<sup>4</sup>, who characterized the new practices of surveillance as often undetected—for example, cameras hidden in everyday benign objects—or even invisible. In these forms, data are frequently collected

without the consent of the target and then shared, permanently stored and aggregated. Browne<sup>4</sup> characterized surveillance as focused monitoring and cataloguing of that which was previously left unobserved, with the human body as a primary site of surveillance.

Beyond human bodies, the analysis of ‘human spaces’, such as homes, offices and streets, is widespread. Scene analysis, understanding or recognition are presented as a core contribution of the field in this portion of papers and patents. Additionally, analyses of ‘other socially salient human data’ appeared in a small portion of papers and patents. The consequence of gradually rendering these public and social spaces and making them amenable to observation is a fundamental mechanism of surveillance<sup>38</sup>. These forms of extraction contribute to the gradual cataloguing, documenting, mapping and monitoring of human affairs in its rich complexities<sup>1,7</sup>. It accumulates to what Zuboff calls the condition of ‘no exit’, where there are diminishing spaces in which to opt out or ‘disconnect’<sup>7</sup>. Taken together, our analysis indicates that the forms of data extraction presented in these papers and patents align with and facilitate established, intrusive forms of surveillance.

### The rise of surveillance AI

To study the roots and evolution of computer-vision based surveillance, we conducted a large-scale lexicon-based analysis. We identified surveillance-enabling patents by scanning the corpus of patents for those containing words in the verified surveillance indicator lexicon. The full method, including descriptions of the surveillance indicator keywords and the extensive lexicon verification process, is described in ‘Lexicon-based analysis’ in Methods. Figure 2a,b presents the evolution of CVPR papers and downstream patents. We found a substantial increase in papers used in surveillance-enabling patents. Comparing decades, we found that the 1990s produced significantly fewer computer-vision papers with downstream patents than the 2010s, and only half of these were used in surveillance-enabling patents (53%, s.d. = 1%,  $n = 664$ ). Two decades later in the 2010s, there had been a tripling of computer-vision papers with downstream patents,

**Table 1 | Targets of human data extraction in computer-vision papers and patents**

Human body parts	Example: “The acquisition system may include a biometric sensor (for example, an electronic fingerprint sensor, or an optical eye scanner, or a camera arranged to acquire a portrait image of an authorized person’s face)” (Patent 71)
	These technologies most frequently target faces, including the detection of eyes, eye movement, faces, ‘suspicious’ facial expressions and facial recognition. Some targeted other body parts, typically to enable human activity recognition.
Human bodies	Example: “People monitoring in public areas, smart homes, urban traffic control, mobile application, and identity assessment for security and safety” (Paper 53)
	These technologies most frequently target humans in the midst of everyday activities (for example, walking, shopping and attending group events) for purposes including body detection, tracking and counting as well as security monitoring and human activity recognition.
Human spaces	Example: “A scene could be decomposed into a set of semantic objects” (accompanied by an example image taken inside an office) (Paper 40)
	These technologies most frequently target living spaces, personal and communal, such as people’s homes, offices, roads, town squares or borders. The purpose is often to identify unspecified objects in these spaces. Other specific purposes include modelling traffic and monitoring large border crossing areas.
Other socially salient human data	Example: “Free-hand human sketches [for example, of another person’s item of clothing] are used as queries to perform instance-level retrieval of images” (Paper 81)
	These technologies target data containing traces of the mental, economic, cultural, social status, identities, preferences or location details of humans, most frequently to narrow users’ search results.

We identified four targets of human data extraction in the computer-vision papers and patents. These targets of extraction form a series of increasingly focused categories: socially salient human data, human spaces, human bodies and human body parts. This table describes each category, with textual examples and a qualitative description. See ref. 34 for the dataset of the papers and patents.

78% of which were used in surveillance-enabling patents (78%, s.d. = 1%,  $n = 2,327$ ). The twin forces of the increase in computer-vision papers with downstream patents and the increase in the proportion of these used in surveillance-enabling patents combined to large effect: from the 1990s to the 2010s, there was a more than fivefold increase in the number of computer-vision papers used in surveillance-enabling patents.

We gained further insight into the evolution of computer vision by inductively identifying patterns of linguistic evolution. To study the linguistic evolution that has occurred over the past several decades, we compared the log odds ratios of word frequencies in paper titles from the 1990s versus those from the 2010s. We used an informative Dirichlet prior to obtain measures of statistical significance and control for variance in word frequencies<sup>39</sup>. Figure 2c shows highly polarized word associations in both directions with computed z-scores. More methodological details are presented in ‘Longitudinal analysis’ in Methods. We see a clear, qualitative shift from more generic application-ambiguous language in the 1990s (for example, ‘shape’, ‘edge’ and ‘surfaces’; turquoise bars) to an increased focus in the 2010s on analysing semantic categories and humans (for example, ‘semantic’, ‘action’ and ‘person’; pink bars). As this was an inductive analysis that identified the dominant patterns of linguistic evolution, this finding indicates that not only has there been a major change towards enabling surveillance but it is one of the most salient changes that have occurred in the field over the past several decades. Taking our results together, we infer that the language and patenting practices in computer vision have evolved in ways that increasingly focus on analysing humans and enabling surveillance.

## The normalization of surveillance AI

Surveillance technology does not emerge in a vacuum. Research and subsequent applications are actively conducted, incentivized, funded and commercialized by numerous stakeholders. In the previous section, we applied a large-scale lexicon-based analysis to identify surveillance-enabling patents. In this section, we consider the institutional affiliations, national affiliations and subfields of the computer-vision papers, and we study the links from these entities to the identified surveillance-enabling patents. The full method, including descriptions of the surveillance indicator keywords and the extensive lexicon verification process, is described in ‘Lexicon-based analysis’ in Methods.

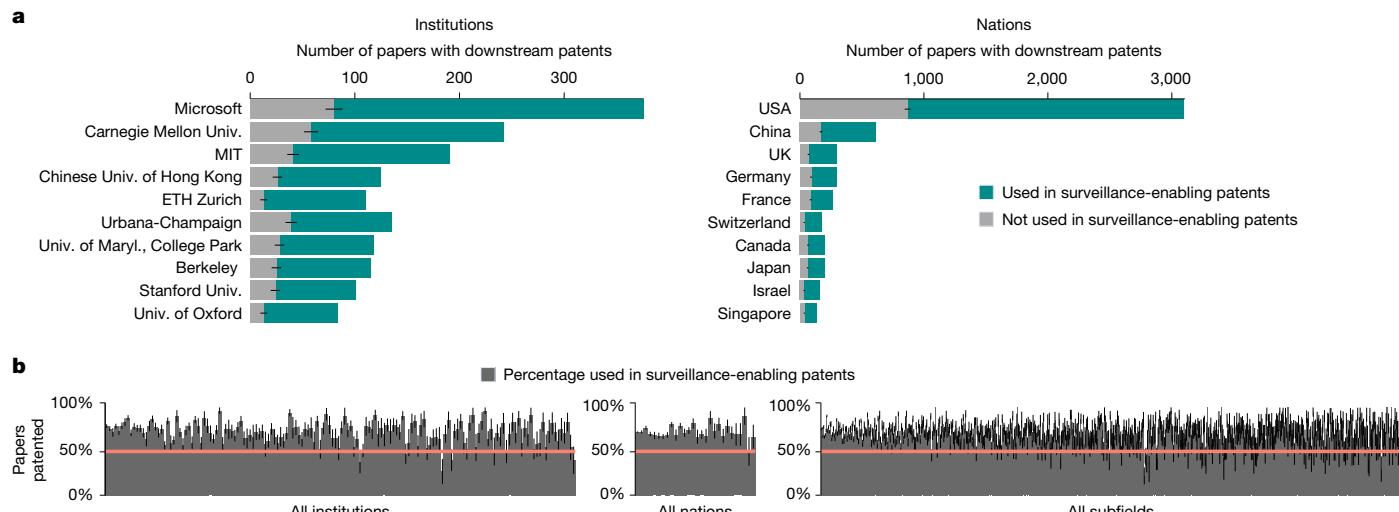
Figure 3a presents the top ten institutions and nations authoring the most CVPR papers with downstream surveillance-enabling patents, as found in the corpus. As shown, for each of these top institutions and nations, most of the patented papers have been used in surveillance-enabling patents. Many of these are top institutions including ‘big tech’ corporations and elite universities, including many that have been identified as top producers of computer science papers and computer-vision papers generally<sup>40</sup>. Additionally, many of the institutions we identified as authoring a substantial number of papers with downstream surveillance-enabling patents have been identified by previous research as aligning with well-established historical legacies of the military–industrial–academic complex<sup>41</sup>.

To understand the influence of nations on surveillance-enabling patents, we additionally present in Fig. 3a the distribution of ties to surveillance across nations. The nations were obtained from the location of paper authors’ institutional affiliations. The top two nations producing papers with downstream surveillance-enabling patents are the USA and China by a large margin, with the USA producing more of these papers than the next several nations combined. Our findings correspond to previous reports about AI-driven surveillance across nations, which state that on a global scale, China and the USA are the main drivers in supplying both AI and advanced surveillance technologies<sup>42</sup>.

These findings provide the basis for a salient question: are only a few key entities and authors contributing to surveillance or are ties from research to surveillance found across the field? We found substantial evidence showing a pervasive fieldwide norm: when an institution or nation authors computer-vision papers with downstream patents, most have been used in surveillance-enabling patents (Fig. 3b; the vertical grey bars for institutions and nations are frequently above the orange 50% threshold). This norm describes the behaviour of 71% of institutions (575 out of 805) and 78% of nations (45 out of 58), which may provide evidence for the wide-spanning normalization of computer vision used in surveillance. Similarly, we found substantial evidence against the notion that there are merely a few implicated applications of computer-vision research within a broader non-surveillance-oriented field. Rather, we found an extension of the above norm: when a subfield produces computer-vision papers with downstream patents, most have been used in surveillance-enabling patents (Fig. 3b; the vertical grey bars for most subfields are above the orange 50% threshold). It may be expected that the stated norm describes frequently implicated subfields such as facial recognition, but in fact, we found that the norm describes most subfields (69%; 2,922 out of 4,247). We present the details of which entities are implicated in this norm in Supplementary Information section 1.5. Our findings indicate that, across institutions, nations and subfields, the practice of producing computer vision that enables surveillance is a pervasive norm.

## The obfuscating language of surveillance AI

Finally, in addition to our analysis of ties to surveillance, we found trends for using obfuscating language that minimized or sidestepped mentions of surveillance. Drawing upon our manual inspection of



**Fig. 3 | Field-wide dominance of downstream surveillance. a,** Institutions and nations that have produced the most CVPR papers with downstream surveillance-enabling patents. For each institution or nation, most patented papers have been used in surveillance-enabling patents. **b,** Percentage of patented papers used in surveillance-enabling patents, stratified by institution, nation and subfield. For each institution, nation or subfield that has published at least ten papers with downstream patents, we show the

percentage of these papers that have been used in surveillance-enabling patents (vertical grey bars) ( $n = 13,804$ ,  $n = 18,272$  and  $n = 19,413$ , respectively). We found a pervasive norm: if an institution, nation or subfield authors papers with downstream patents, most are used in surveillance-enabling patents. (Vertical grey bars are frequently above the 50% threshold; orange line). Whiskers represent the standard deviation.

100 computer-vision papers and 100 downstream patents, we highlight and describe two salient qualitative themes that emerged:

#### Theme 1. What is said: humans are subsumed under the term ‘objects’.

“Since the surveillance system detects and can be interested on vehicles, animals in addition to people, hereinafter we more generally refer to them with the term moving object”. (Paper 53)

Establishing the explicit conceptualization of humans as merely a kind of object, as many papers and patents do, enables the rest of those documents and may, crucially, enable all other papers and patents to merely discuss problems related to objects or scenes, as they can rely on the understanding that humans are objects. Because humans are considered objects and because scenes contain humans, documents can rely on the covert assumption that any paper or patent that discusses objects or scenes—most of the field—may enable human data extraction and surveillance. Theme 2 describes one form of reliance on this covert assumption. Reflecting the continuing tight relationship between computer vision and surveillance, a paper about panoptic segmentation makes no distinction when summarizing the field: ‘Early work on face detection ... helped popularize bounding-box object detection. Later, pedestrian detection datasets helped drive progress in the field’ (Paper 96). During our qualitative analysis we frequently encountered papers and patents stating that they use the term ‘objects’ as shorthand for human entities, including humans, human body parts, vehicles, students and pedestrians.

#### Theme 2. What is not said: even when documents do not mention humans, the figures or datasets may contain images of humans.

The pattern of papers and patents claiming to target objects, while briefly defining these terms as subsuming humans, sets a clear precedent that we found has already played out: we found that other documents lean on these norms. Although claiming to target objects, in actuality, they target humans and, thus, leave no textual trace of the human data extraction they are engaged in. For example, one paper describes itself as improving object classification and makes no

mention of humans. Yet, close inspection of the first figure in the paper reveals (in 3-point font) that it classifies so-called objects into classes including ‘person’, ‘people’ and ‘person sitting’ (Paper 5). A second paper describes itself as identifying salient regions of images and does not mention humans. Yet, inspection of the datasets reveals that the authors demonstrate their technology by detecting regions of interest, such as humans walking on a sidewalk (Paper 1). Figure 1b presents examples of images targeted in a random sample of the computer-vision papers ( $n = 50$ ). Our annotators observed that papers frequently analysed images depicting human bodies, including these in datasets and often featuring these in figures, despite many papers lacking explicit mention of these entities.

The nature of these themes is such that, first, humans and objects of all kinds may be targeted in parallel, despite the vastly different implications, and, second, that humans can be primary targets of technologies without leaving a textual trace of surveillance.

## Discussion

The field of computer vision has frequently emphasized its place as a scientific and statistical endeavour inspired by human vision, referencing a range of commercial and industrial applications and highlighting the use of computer vision for good<sup>20</sup>. Yet, based on the studies presented in this paper, we contend that such characterizations of the field underrecognize or misrecognize the extent to which the field, taken as a whole, simultaneously engages in the mass extraction of human data<sup>7,43,44</sup>. Cutting across research motivations and subfields, we found a fieldwide norm in which the analysed computer-vision papers and patents extensively and increasingly extract human body data and other socially salient data. The normalization of such extraction is particularly striking when considered alongside evidence that the field frequently fails to address concerns regarding the use of human data. Exemplifying this are the themes in which the analysed papers and patents obfuscate their extraction of human data and conceptualize humans as objects to be studied without special consideration. These patterns align with existing literature that has established that AI research frequently fails to mention or mitigate concerns regarding human agency, consent or privacy and fails to engage with these

# Article

ethical considerations in many of the ways expected of other fields that analyse human data<sup>17,18,45</sup>. Our findings do not comment on the intentions of computer-vision researchers. Rather, they bring into focus the systematic pattern of extracting human data and enabling surveillance.

Although many computer-vision researchers conceptualize the overall rise and proliferation of computer-vision technologies as field success, this rapid proliferation might alternatively be understood as the perpetual practice of rendering visible what was previously shielded and unseen, a practice that surveillance studies scholars such as Browne<sup>4</sup> view as the core of surveillance. Technologies that enable the monitoring of human data, which may be perceived as differentially malevolent or benevolent by different communities, nonetheless have historically established consequences: these technologies engender fear and self-censorship; it is lucrative and standard practice for entities in positions of relative power to use these technologies to access, monetize, coerce, control or police individuals or communities with lesser power; and these technologies are frequently deputized by state surveillance organizations<sup>2,4,25</sup>. Crucially, in addition to individualized consequences, the rapid generation and proliferation of technologies monitoring humans accumulates to what Zuboff<sup>7</sup> calls the condition of ‘no exit’, where there are fewer and fewer spaces left to opt out, ‘disconnect’ and seek respite.

The uncovered features of computer vision tie into a broader literature about the historical narrative of neutrality in science. Scientific findings are frequently presented as facts that emerge from an objective ‘view from nowhere’ in a historical, cultural and contextual vacuum. Such views of science as ‘value-free’ and ‘neutral’ have been deconstructed by a variety of scholarships, including the philosophy of science, science and technology studies, and feminist and decolonial studies. A purported view from nowhere is always a view from somewhere and usually a view from those with the greatest power<sup>46</sup>. Social and cultural histories and norms, funding priorities, academic trends, researcher objectives and research incentives, for example, all inevitably constrain and shape the production of scientific knowledge<sup>47,48</sup>. An assemblage of social forces have shaped computer vision, resulting in a field that now mass-produces highly specific technologies. Viewing computer vision in this light, it becomes clear that shifting away from surveillance requires, not a small shift in applications, but rather a reckoning and challenging of the foundations of the discipline.

Rapidly evolving AI research agendas, narratives, norms and policies afford opportunities to intervene. For individuals and communities concerned about surveillance, there are historical precedents and frequent examples in which key figures have made informed decisions regarding the role they wish to play, for example, by adopting critical technical practice, exercising the right to conscientious research including the right to conscientious objection, collectively protesting against and cancelling surveillance projects, and changing their focus to study the ethical dimensions of a field, educate the public or put forward informed advocacy<sup>49,50</sup>. In this context, this paper serves to illuminate the roots, extent, evolution and obfuscation of the surveillance AI pipeline and, in doing so, aims to provide access to information with which individuals and communities may understand, influence or disrupt these pathways to surveillance.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-025-08972-6>.

- Monahan, T. & Wood, D. M. *Surveillance Studies: A Reader* (Oxford Univ. Press, 2018).
- Lyon, D. in *Emerging Digital Spaces in Contemporary Society: Properties of Technology* (eds Kalantzis-Cope, P. & Gherab-Martin, K.) 107–120 (Springer, 2010).

- Scheuerman, M. K., Hanna, A. & Denton, E. Do datasets have politics? Disciplinary values in computer vision dataset development. *Proc. ACM Hum.–Comput. Interact.* **5**, 317 (2021).
- Browne, S. *Dark Matters: On the Surveillance of Blackness* (Duke Univ. Press, 2015).
- Agre, P. E. Surveillance and capture: Two models of privacy. *Inf. Soc.* **10**, 101–127 (1994).
- Stark, L. Facial recognition is the plutonium of AI. *XRDs: Crossroads* **25**, 50–55 (2019).
- Zuboff, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Profile Books, 2019).
- Building community power to abolish the police state. Stop LAPD Spying Coalition <https://stoplapdspying.org/> (accessed 1 September 2023).
- Chang, M. et al. *Countermeasures: The Need for New Legislation to Govern Biometric Technologies in the UK* (2022).
- Marx, G. T. in *International Encyclopedia of the Social & Behavioral Sciences* 2nd edn (ed. Wright, J. D.) 733–741 (Elsevier, 2015).
- Monahan, T. & Murakami Wood, D. *Introduction: Surveillance Studies as a Transdisciplinary Endeavor* (2018).
- Foucault, M. *Discipline and Punish: The Birth of the Prison* (Pantheon Books, 1977).
- Deleuze, G. *Postscript on the Societies of Control* (MIT Press, 1992).
- Allmer, T. Critical surveillance studies in the information society. *tripleC: Commun. Capitalism Crit.* **9**, 566–592 (2011).
- Richards, N. M. The dangers of surveillance. *Harv. Law Rev.* **126**, 1934 (2013).
- Dobson, J. E. *The Birth of Computer Vision* (Univ. Minnesota Press, 2023).
- Raji, I. D. & Fried, G. About face: a survey of facial recognition evaluation. Preprint at <https://arxiv.org/abs/2102.00813> (2021).
- Broussard, M. *Artificial Unintelligence: How Computers Misunderstand the World* (MIT Press, 2018).
- Königs, P. Government surveillance, privacy, and legitimacy. *Philos. Technol.* **35**, 8 (2022).
- Szeliski, R. *Computer Vision: Algorithms and Applications* (Springer, 2020).
- Forsyth, D. & Ponce, J. *Computer Vision: A Modern Approach* (Pearson, 2011).
- Call for papers. *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR 2024)*; <https://cvpr.thecvf.com/Conferences/2024/CallForPapers>.
- Keynotes and panels. *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR 2024)*; <https://cvpr.thecvf.com/Conferences/2024/KeynotesAndPanels>.
- Zitnick, L. et al. Spherical channels for modeling atomic interactions. In *Proc. 35th Conference on Neural Information Processing Systems* (eds Koyejo, S. et al.) 8054–8067 (Curran Associates, 2022).
- Hill, K. *Your Face Belongs to Us: The Secretive Startup Dismantling Your Privacy* (Simon and Schuster, 2023).
- Watt, E. The right to privacy and the future of mass surveillance. *Int. J. Hum. Rights* **21**, 773–799 (2017).
- Pridmore, J. & Zwick, D. Marketing and the rise of commercial consumer surveillance. *Surveill. Soc.* **8**, 269–277 (2011).
- Andrejevic, M. in *Routledge Handbook of Surveillance Studies* (eds Ball, K. et al.) 91–98 (Routledge, 2012).
- Zuboff, S. in *Social Theory Re-wired* (eds Longhofer, W. & Winchester, D.) 203–213 (Routledge, 2023).
- Csernatoni, R. & Lavallée, C. in *Emerging Security Technologies and EU Governance* (eds Calcaro, A. et al.) 206–223 (Routledge, 2020).
- Almansoori, M., Gallardo, A., Poveda, J., Ahmed, A. & Chatterjee, R. A global survey of Android dual-use applications used in intimate partner surveillance. In *Proc. on Privacy Enhancing Technologies* (eds Kerschbaum, F. & Mazurek, M. L.) 120–139 (Privacy Enhancing Technologies Board, 2022).
- Selinger, E. & Durant, D. Amazon’s ring: surveillance as a slippery slope service. *Sci. Cult.* **31**, 92–106 (2022).
- Nesterova, I. Questioning the EU proposal for an artificial intelligence act: the need for prohibitions and a stricter approach to biometric surveillance. *Inf. Polity* **27**, 147–162 (2022).
- Kalluri, P. R. & Agnew, W. Code and data for ‘Computer vision research powers surveillance technology’. *Github* <https://github.com/wagnew3/Computer-Vision-Research-Powers-Surveillance-Technology> (2025).
- The fight to stop face recognition technology. *American Civil Liberties Union* [www.aclu.org/news/topic/stopping-face-recognition-surveillance](http://www.aclu.org/news/topic/stopping-face-recognition-surveillance) (accessed 1 September 2023).
- Awumey, E., Das, S. & Forlizzi, J. A systematic review of biometric monitoring in the workplace: analyzing socio-technical harms in development, deployment and use. In *Proc. 2024 ACM Conference on Fairness, Accountability, and Transparency* 920–932 (2024).
- Murray, D. et al. The chilling effects of surveillance and human rights: insights from qualitative research in Uganda and Zimbabwe. *J. Hum. Rights Pract.* **16**, 397–412 (2023).
- Cohen, J. E. in *Cambridge Handbook of Surveillance Law* (eds Gray, D. & Henderson, S. E.) 455–469 (Cambridge Univ. Press, 2017).
- Monroe, B. L., Colaresi, M. P. & Quinn, K. M. Fightin’ words: Lexical feature selection and evaluation for identifying the content of political conflict. *Polit. Anal.* **16**, 372–403 (2008).
- Ahmed, N. & Wahed, M. The de-democratization of AI: deep learning and the compute divide in artificial intelligence research. Preprint at <https://arxiv.org/abs/2010.15581> (2020).
- Leslie, S. W. et al. *The Cold War and American Science: The Military-Industrial-Academic Complex at MIT and Stanford* (Columbia Univ. Press, 1993).
- Feldstein, S. *The Global Expansion of AI Surveillance* (Carnegie Endowment for International Peace, 2019).
- Carrier, J. G. Misrecognition and knowledge. *Inquiry* **22**, 321–342 (1979).
- Véliz, C. *Privacy Is Power* (Melville House, 2021).
- Waelen, R. A. The ethics of computer vision: an overview in terms of power. *AI Ethics* **4**, 353–362 (2024).
- Haraway, D. in *Feminist Theory Reader* (eds McCann, C. et al.) 303–310 (Routledge, 2020).
- Ensmenger, N. L. *The Computer Boys Take Over: Computers, Programmers, and the Politics of Technical Expertise* (MIT Press, 2012).
- Birhane, A. et al. The values encoded in machine learning research. In *Proc. 2022 ACM Conference on Fairness, Accountability, and Transparency* 173–184 (ACM, 2022).

49. Agre, P. E. in *Social Science, Technical Systems and Cooperative Work: Beyond the Great Divide* (eds Bowker, G. et al.) Ch. 6 (Psychology Press, 1997).
50. Butcher, S. I. *Origins of the Russell-Einstein manifesto*. Technical report (Pugwash Conferences on Science and World Affairs, 2005).
51. Cevikalp, H. & Triggs, B. Face recognition based on image sets. In Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2567–2573 (IEEE, 2010).
52. Salakhutdinov, R., Torralba, A. & Tenenbaum, J. Learning to share visual appearance for multiclass object detection. In Proc. CVPR 2011 1481–1488 (IEEE, 2011).
53. Khamis, S., Morariu, V. I. & Davis, L. S. A flow model for joint action recognition and identity maintenance. In Proc. 2012 IEEE Conference on Computer Vision and Pattern Recognition 1218–1225 (IEEE, 2012).
54. Chen, C.-Y. & Grauman, K. Watching unlabeled video helps learn new human actions from very few labeled snapshots. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 572–579 (IEEE, 2013).
55. Lin, G., Shen, C., Shi, Q., Van den Hengel, A. & Suter, D. Fast supervised hashing with decision trees for high-dimensional data. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 1963–1970 (IEEE, 2014).
56. Yim, J. et al. Rotating your face using multi-task deep neural network. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 676–684 (IEEE, 2015).
57. Song, S. et al. Multimodal multi-stream deep learning for egocentric activity recognition. In Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops 24–31 (IEEE, 2016).
58. Arvanitopoulos, N., Achanta, R. & Susstrunk, S. Single image reflection suppression. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 4498–4506 (IEEE, 2017).
59. Bloesch, M., Czarnowski, J., Clark, R., Leutenegger, S. & Davison, A. J. Codeslam—learning a compact, optimisable representation for dense visual slam. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 2560–2568 (IEEE, 2018).
60. Shi, L., Zhang, Y., Cheng, J. & Lu, H. Skeleton-based action recognition with directed graph neural networks. In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition 7912–7921 (IEEE, 2019).
61. Wang, W., Wang, Y., Huang, Q. & Gao, W. Measuring visual saliency by site entropy rate. In Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2368–2375 (IEEE, 2010).
62. Zhang, Y., Jia, Z. & Chen, T. Image retrieval with geometry-preserving visual phrases. In Proc. CVPR 2011 809–816 (IEEE, 2011).
63. Ranjbar, M., Vahdat, A. & Mori, G. Complex loss optimization via dual decomposition. In Proc. 2012 IEEE Conference on Computer Vision and Pattern Recognition 2304–2311 (IEEE, 2012).
64. Fidler, S., Sharma, A. & Urtasun, R. A sentence is worth a thousand pixels. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 1995–2002 (IEEE, 2013).
65. Bae, S.-H. & Yoon, K.-J. Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 1218–1225 (IEEE, 2014).
66. Zhao, R., Ouyang, W., Li, H. & Wang, X. Saliency detection by multi-context deep learning. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 1265–1274 (IEEE, 2015).
67. Wang, L., Qiao, Y., Tang, X. & Van Gool, L. Actionness estimation using hybrid fully convolutional networks. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 2708–2717 (IEEE, 2016).
68. Liu, W. et al. Sphereface: deep hypersphere embedding for face recognition. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 212–220 (IEEE, 2017).
69. Wan, F., Wei, P., Jiao, J., Han, Z. & Ye, Q. Min-entropy latent model for weakly supervised object detection. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 1297–1306 (IEEE, 2018).
70. Ranjan, A. et al. Competitive collaboration: Joint unsupervised learning of depth, camera motion, optical flow and motion segmentation. In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition 12240–12249 (IEEE, 2019).
71. Socher, R. & Fei-Fei, L. Connecting modalities: semi-supervised segmentation and annotation of images using unaligned text corpora. In Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 966–973 (IEEE, 2010).
72. Harker, M. & O’Leary, P. Least squares surface reconstruction from gradients: direct algebraic methods with spectral, Tikhonov, and constrained regularization. In Proc. CVPR 2011 2529–2536 (IEEE, 2011).
73. He, J., Balzano, L. & Szlam, A. Incremental gradient on the Grassmannian for online foreground and background separation in subsampled video. In Proc. 2012 IEEE Conference on Computer Vision and Pattern Recognition 1568–1575 (IEEE, 2012).
74. Khosla, A., Hamid, R., Lin, C.-I. & Sundaresan, N. Large-scale video summarization using web-image priors. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 2698–2705 (IEEE, 2013).
75. Tang, K., Yang, J. & Wang, J. Investigating haze-relevant features in a learning framework for image dehazing. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 2995–3000 (IEEE, 2014).
76. Chen, X., Ma, H., Wang, X. & Zhao, Z. Improving object proposals with multi-thresholding straddling expansion. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 2587–2595 (IEEE, 2015).
77. Hu, S. et al. A polarimetric thermal database for face recognition research. In Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops 119–126 (IEEE, 2016).
78. Dansereau, D. G., Eriksson, A. & Leitner, J. Richardson-Lucy deblurring for moving light field cameras. In Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops 70–81 (IEEE, 2017).
79. Hu, R., Dollár, P., He, K., Darrell, T. & Girshick, R. Learning to segment everything. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 4233–4241 (IEEE, 2018).
80. Acuna, D., Kar, A. & Fidler, S. Devil is in the edges: learning semantic boundaries from noisy annotations. In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition 11075–11083 (IEEE, 2019).
81. Wu, Y., Shen, B. & Ling, H. Online robust image alignment via iterative convex optimization. In Proc. 2012 IEEE Conference on Computer Vision and Pattern Recognition 1808–1814 (IEEE, 2012).
82. Gengti, B., Yang, L., Xu, C. & Hua, X.-S. Content-aware ranking for visual search. In Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 3400–3407 (IEEE, 2010).
83. Angelova, A. & Zhu, S. Efficient object detection and segmentation for fine-grained recognition. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 811–818 (IEEE, 2013).
84. Lin, J., Liu, Y., Hullin, M. B. & Dai, Q. Fourier analysis on transient imaging with a multifrequency time-of-flight camera. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 3230–3237 (IEEE, 2014).
85. Bernard, F. et al. A solution for multi-alignment by transformation synchronisation. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 2161–2169 (IEEE, 2015).
86. Yu, H., Wang, J., Huang, Z., Yang, Y. & Xu, W. Video paragraph captioning using hierarchical recurrent neural networks. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 4584–4593 (IEEE, 2016).
87. Zhang, H., Kyaw, Z., Chang, S.-F. & Chua, T.-S. Visual translation embedding network for visual relation detection. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 5532–5540 (IEEE, 2017).
88. Volpi, R., Morerio, P., Savarese, S. & Murino, V. Adversarial feature augmentation for unsupervised domain adaptation. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 5495–5504 (IEEE, 2018).
89. Wu, D., Dai, Q., Liu, J., Li, B. & Wang, W. Deep incremental hashing network for efficient image retrieval. In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition 9069–9077 (IEEE, 2019).
90. Vijayanarasimhan, S. & Kapoor, A. Visual recognition and detection under bounded computational resources. In Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1006–1013 (IEEE, 2010).
91. Balzer, J., Hofer, S. & Beyerer, J. Multiview specular stereo reconstruction of large mirror surfaces. In Proc. CVPR 2011 2537–2544 (IEEE, 2011).
92. Saberian, M. J. & Vasconcelos, N. Boosting algorithms for simultaneous feature extraction and selection. In Proc. 2012 IEEE Conference on Computer Vision and Pattern Recognition 2448–2455 (IEEE, 2012).
93. Zhou, Z., Jin, H. & Ma, Y. Plane-based content preserving warps for video stabilization. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 2299–2306 (IEEE, 2013).
94. Danelljan, M., Khan, F. S., Felsberg, M. & Van de Weijer, J. Adaptive color attributes for real-time visual tracking. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 1090–1097 (IEEE, 2014).
95. Moreno, D., Son, K. & Taubin, G. Embedded phase shifting: robust phase shifting with embedded signals. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 2301–2309 (IEEE, 2015).
96. Hu, R. et al. Natural language object retrieval. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 4555–4564 (IEEE, 2016).
97. Fanello, S. R. et al. Ultrastereo: efficient learning-based matching for active stereo systems. In Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 6535–6544 (IEEE, 2017).
98. Nguyen, P., Liu, T., Prasad, G. & Han, B. Weakly supervised action localization by sparse temporal pooling network. In Proc. IEEE Conference on Computer Vision and Pattern Recognition 6752–6761 (IEEE, 2018).
99. Zhang, J. & Peng, Y. Object-aware aggregation with bidirectional temporal graph for video captioning. In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition 8327–8336 (IEEE, 2019).
100. Raheja, J. L., Das, K. & Chaudhary, A. An efficient real time method of fingertip detection. In Proc. 7th International Conference on Trends in Industrial Measurements and Automation 447–450 (TIMA, 2011).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

# Article

## Methods

### Corpus of computer-vision papers and downstream patents

To study the pathway from computer-vision research to surveillance, we collected and analysed a corpus linking more than 19,000 computer-vision research papers to more than 23,000 downstream patents. Research papers and patents have unique advantages making them revealing artefacts. First, they are primary sources written in the researchers' and patentees' own words and there exist professional and institutional expectations that they accurately describe the research and technologies. The connections between research papers and citing patents serve as a rich data trail of the path from research to applications<sup>101,102</sup>. These documents also include comprehensive metadata, as papers necessarily list their authors and their primary affiliated institutions and the publication year, thus enabling analyses of how these factors influence the pathway to applications. They are available online, and they have a consistent overall structure facilitating consistency of annotation and reliable comparisons. These papers and their collected downstream patents served as the basis for the mixed methods content analysis and large-scale lexicon-based analysis presented in this paper. We studied papers published in the proceedings of the longest standing computer-vision conference, CVPR, as metrics indicate it has the highest impact of all computer-vision conferences by an extremely large margin. By h5-index, CVPR proceedings are among the top five highest impact publications in any discipline, alongside *Nature* and *Science*. This research is widely seen to be an 'indicator of hot topics for the AI and machine learning community'<sup>103</sup>. Acceptance and publication are marks of approval of the research as work that exemplifies the core values of the computer-vision community. As such, these papers both represent the state-of-the-art in current computer vision and effectively reveal the values held in high regard within the community. We obtained all the proceedings published from 1990 to 2021, and, for each paper that has been cited in one or more patents, we obtained all citing patents. We refer to these as a patented paper and its downstream patents. Extended Data Fig. 1 presents randomly sampled pairs consisting of a paper and a downstream patent and is, thus, a snapshot of our corpus.

**Implementation.** We analysed the corpus of CVPR papers from 1990 to 2021. CVPR was not held in 1990, 1995 or 2002, so there are no papers from those years. In constructing our corpus, we leveraged and linked the papers in the Microsoft Academic Graph<sup>104</sup>, the paper–patent citation linkages inferred by Marx and Fuegi<sup>105</sup> and the patents in the Google Patents database. Manual verification found the paper–patent citation linkages to have over 99% precision and 78% recall. All papers presented at CVPR were published in English. For patents that were published in other languages, the English translations in Google Patents were used.

### Content analysis

Following best practices in content analysis, we conducted an in-depth analysis of a purposive sample of papers and patents distinctively informative of the development of computer-vision research and applications for enabling surveillance. For each year from 2010 to 2019, we randomly sampled ten paper–patent pairs that consisted of a CVPR research paper published in the year and a downstream patent. This formed a total of 100 papers and 100 downstream patents. In the context of content analysis, this constitutes a large-scale annotation.

We conducted the content analysis using a close reading of the documents and a rigorous qualitative methodology. An interdisciplinary six-person team analysed the documents using an integrated inductive–deductive methodology. In the inductive component, each document was read line by line, including figures. We inductively coded key emergent features in the treatment of human data by the technology and iteratively accumulated a list of these key features and their relationships. We complemented this with a deductive component to

ensure that we actively looked for and captured instances of papers and patents with key features that inhibited usage for surveillance, even if rare. The inductive and deductive codes were ultimately clustered into data type, data transfer and data use. The codes are discussed in this section as well as in 'The extent of human data extraction' section and Supplementary Information sections 1.3 and 1.4. During this process, our annotation team had several strengths: our team included both published experts in computer vision and field outsiders to allow expert insights and translation as well as fresh perspectives that could illuminate computer-vision disciplinary biases. We used the constant comparative method. Throughout the coding process, the team held frequent, extensive discussions to develop the precise meanings of codes and their relationships and to revise and refine the code list. At the end of all coding, the team unanimously agreed upon the key emergent dimensions and features of the treatment of human data by the technologies, along with the relationships among these dimensions and features. Additionally, as we coded papers and downstream patents, we encountered and discussed salient examples of obfuscating language being used to describe or avoid describing surveillance, and we present these findings in 'The obfuscating language of surveillance AI' section.

Based on our in-depth, interdisciplinary content analysis, we present the surveillance AI topology in Supplementary Fig. 1, which brings to the fore the dimensions, features and dynamics of the treatment of human data in computer vision and connects these to concepts in surveillance studies that elucidate the complexity and consequences of these particular findings. Our analysis identified three key dimensions capturing the treatment of human data by these technologies: (1) Data type—what type of data does the technology extract, attend to, capture, monitor, track, profile, compute or sort and to what extent is it human and personal? (2) Data transferal—to what extent do the data remain under the control of the datafied person or become transferred to others? (3) Use of data—for what purpose are the data used? These three dimensions are discussed in detail, with examples and analysis, in 'The extent of human data extraction' section and Supplementary Information sections 1.3 and 1.4.

We discuss the primary dimension of the topology in detail in 'The extent of human data extraction' section. In this primary dimension, the inductively identified types of human data extracted form a series of nested, increasingly focused categories: socially salient human data, human spaces, human bodies and human body parts. A fifth inductively identified target of data extraction was general and unspecified data, which tended to target generic tasks such as 'identifying objects', did not specify targeting human data but also did not commit to targeting only non-human data. In addition to these data types, which were inductively found only through a close reading of the papers and patents, the annotation team deductively included non-human data in the annotation scheme from the start. This was to ensure that we captured mentions of any non-surveillance technologies in papers and patents, even if rare. To enable a quantitative analysis of this primary dimension, we identified for each paper and patent the innermost (most focused) type of human data extracted. Half of the documents were annotated by more than one annotator, which was particularly valuable for allowing the annotators to become accustomed to types of cases in which a single sentence or figure influenced the appropriate code. The existence of such cases is discussed in 'The obfuscating language of surveillance AI' section. In these cases with several annotators, the final code of each document was determined through discussion until consensus. We then quantified the annotations for all documents and present the relative frequencies of the data types in Fig. 1a and Supplementary Fig. 2. Figure 1b additionally presents, for each of the first 50 annotated papers, one example of an image analysed by that paper. We found that the second and third dimensions of the topology were less consistently discussed in papers and patents. Nonetheless, key areas of surveillance studies scholarship are dedicated to how these dimensions

(data transfer and data use) are important to understanding the roles, dynamics and consequences of surveillance. Given the importance of these dimensions, Supplementary Information sections 1.3 and 1.4 include a full discussion of these dimensions, the inductive and deductive codes, demonstrative examples and findings, and connections to nuanced dynamics of surveillance that have been discussed in the surveillance studies literature.

### Lexicon-based analysis

**Surveillance indicator lexicon.** As introduced at the start of this article, drawing upon surveillance studies, surveillance is defined as an entity gathering, extracting or attending to data connectable to other persons, whether individuals or groups<sup>10</sup>.

During our manual content analysis of computer-vision papers and downstream patents, careful attention was paid to sentences in patents that revealed that a patent enabled the above conceptualization of surveillance, that is, sentences that revealed the patent was self-reported as enabling the gathering, extracting or attending to data connectable to other persons. During the content analysis, the team accumulated a list of candidate surveillance indicator keywords (one- or two-word phrases) that featured centrally in these sentences and that in one or more encountered documents played a key role in revealing that a patent enabled this conceptualization of surveillance.

After constructing this candidate keyword list, we began an extensive pruning process, as the aim was to create a reliable lexicon of keywords indicating technologies enabling surveillance. Our preference was to err on the side of more pruning, so that we ultimately undercounted rather than overcounted surveillance-enabling patents. Accordingly, we applied two phases of pruning. In the first phase, for each candidate keyword, we scanned the corpus for all patents containing this keyword and produced a random sample of ten such patents. Three team members conducted an independent manual inspection of these patents. After the manual inspection, the three team members came together to identify and remove, by consensus, candidate keywords that were not reliable indicators (typically because we found the keyword had other word senses or usages; for example, a 'store' could be a human space but was frequently a technical term related to data or memory storage, so 'store' was removed from the list). To strengthen the reliability of the lexicon, we undertook a second pruning phase two months later. During this second phase, for each candidate keyword, we obtained a random sample of nine patents containing the keyword to serve as a verification sample. Each of these patents was assigned to a team member. This team member was provided the full text of the patent as well as the paragraph and sentence containing the keyword. The team members independently annotated for each assigned patent whether it enabled the above conceptualization of surveillance. If a team member encountered any instances of a keyword being used as a word sense clearly different from the word sense expected and theoretically connected to the above conceptualization of surveillance, the keyword was removed from the lexicon. If a team member encountered any instances of a patent that did not enable the above conceptualization of surveillance, the respective keyword was removed from the lexicon. During this second phase, keywords were required to meet a strict threshold of, for the verification sample, obtaining 100% precision in predicting that a patent enables the above conceptualization of surveillance.

Following this extensive lexicon verification procedure, the final surveillance indicator list contains 30 keywords, which, for the verification sample, each met the strict threshold of 100% precision in predicting that a patent enables the above conceptualization of surveillance. The keywords are listed in Supplementary Information section 2.1.

**Downstream patent identification and analysis.** To study the breadth and variation of surveillance across years, institutions, nations and sub-fields, we conducted a large-scale lexicon-based analysis of more than

43,022 papers and patents. For each paper, we scanned its downstream patents to identify patents containing one or more of these surveillance indicator keywords. We refer to these as surveillance-enabling patents. Given that many patents do not explicitly state that they are intended for surveillance (for example, see 'The obfuscating language of surveillance AI' section), the identification and subsequent scanning for surveillance indicator keywords is a rigorous method for capturing the patents that enable surveillance. We present the distribution of surveillance-enabling patents across institutions, nations, subfields and years, along with a contextualizing discussion, in 'The rise of surveillance AI' and 'The normalization of surveillance AI' sections. We present further methodological details in Supplementary Information sections 2.1 and 2.3.

**Longitudinal analysis.** To conduct an analysis across years (Fig. 2a,b), we filtered the corpus years. In emerging and developing fields, the estimated time from a paper being published to a downstream patent being published is 3 to 4 years. This also incorporates the time spent during the patenting process<sup>106</sup>. This seems to be in line with our corpus, as the number of computer-vision papers with downstream patents stabilized in the early 2000s and, from the early 2000s onwards, remained above 200 every year until 2018 (exactly 4 years before our analysis began), at which point it suddenly dropped by nearly a half. Accordingly, for the analysis across years, we removed papers from 2018 and 2019, as these were published less than 4 years before our analysis began and the patenting process may not yet have played out for many of them, so that using them would have made the analysis less reliable. This filter had the added benefit that, in our analyses comparing the 1990s to the 2010s, both decades consisted of 8 years, putting these decades on a fair playing field for totalling when comparing the number of downstream patents of various types.

To study the linguistic evolution that has occurred, we computed the log odds ratio with a Dirichlet prior of words appearing in paper titles from the 1990s versus paper titles from the 2010s<sup>39</sup>. We removed stop words (as listed in the Natural Language Toolkit, as well as 'using' and 'via' because these are common stop words in computer-vision titles). We present ten highly polarized word associations in both directions with computed z scores in Fig. 2c. These are the strongest word associations by z score with the exception that, because we are interested in changes in the focus of papers and patents and not in the well-known evolution of specific types and names of models being used, we skipped the words 'machine learning model(s)' and 'neural network(s)'.

### Defining big tech and elite universities

Following Ahmed and Wahed<sup>40</sup>, we relied on the QS World University Rankings for our definition of an elite university. To determine what is considered big tech, we relied on the criteria established by Abdalla and Abdalla<sup>107</sup> and Birhane et al.,<sup>48</sup> namely, Alibaba, Amazon, Apple, DeepMind, Element AI, Facebook, Google, Huawei, IBM, Intel, Microsoft, Nvidia, Open AI, Samsung and Uber.

### Data availability

Instructions for downloading and creating the datasets used are available at <https://github.com/wagnew3/Computer-Vision-Research-Powers-Surveillance-Technology> (ref. 34).

### Code availability

The code used for this project is available at <https://github.com/wagnew3/Computer-Vision-Research-Powers-Surveillance-Technology> (ref. 34).

101. Hammarfelt, B. Linking science to technology: the 'patent paper citation' and the rise of patentometrics in the 1980s. *J. Doc.* **77**, 1413–1429 (2021).

# Article

102. Ahmadpoor, M. & Jones, B. F. The dual frontier: patented inventions and prior scientific advance. *Science* **357**, 583–587 (2017).
103. IEEE Computer Society Team. *CVPR 2021 Report Identifies 5 Trend Areas* (IEEE, 2021).
104. Sinha, A. et al. An overview of Microsoft Academic Service (MAS) and applications. In Proc. 24th International Conference on World Wide Web 243–246 (2015).
105. Marx, M. & Fuegi, A. Reliance on science by inventors: hybrid extraction of in-text patent-to-article citations. *J. Econ. Manag. Strategy* **31**, 369–392 (2022).
106. Finardi, U. Time relations between scientific production and patenting of knowledge: the case of nanotechnologies. *Scientometrics* **89**, 37–50 (2011).
107. Abdalla, M. & Abdalla, M. The grey hoodie project: big tobacco, big tech, and the threat on academic integrity. In Proc. 2021 AAAI/ACM Conference on AI, Ethics, and Society 287–297 (2021).

**Acknowledgements** We owe gratitude and accountability to the long history of work exposing the centralization of power in general, and how technology shifts power specifically — work primarily done by communities at the margins. A.B. is supported by Science Foundation Ireland through the ADAPT Centre of Digital Content Technology funded under the European Regional Development Fund (grant no. 13/RC/2106\_P2). We thank T. Gebru and D. Jurafsky for feedback on an earlier version of the article.

**Author contributions** All authors contributed to data analysis and annotation. All authors contributed to group discussions about the project and writing and revising the paper. P.R.K. contributed to project ideation, conceiving and developing the methodology, data collection, processing and analysis, software, visualization, and writing. W.A. contributed to project ideation, data collection, processing and analysis, software, and writing. M.C. contributed to data analysis, theoretical analysis, literature review and writing. K.O. contributed to data analysis and writing. L.S. contributed to data analysis, software and writing. A.B. contributed to project ideation, methodology development, theoretical analysis, literature review and writing.

**Competing interests** The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-025-08972-6>.

**Correspondence and requests for materials** should be addressed to Pratyusha Ria Kalluri, William Agnew or Abeba Birhane.

**Peer review information** *Nature* thanks Jathan Sadowski and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

## Computer vision papers

"[Removal of image background] is a useful technique...especially when there are active, moving objects...a crucial component in human activity recognition and the analysis of video from surveillance...There are an estimated minimum 10,000 surveillance cameras in the city of Chicago...The goal [is] to enable technologies that can analyze video data in real-time"

"Lack of reliable and efficient [algorithms for linking a subject across many images at different viewing angles or times] makes it difficult for many image analysis tasks such as face recognition [and] image classification... [Our method is] capable of dealing with real time tasks such as visual tracking."

"We focus on detecting visual relations [e.g. "person ride bike" and "bike next to car"] ...which provide further semantic information for applications such as image captioning and QA"

## Downstream patents

"a method and system for segmenting and tracking...content of videos in real-time. The content can include...e.g., a largely stationary background [and] moving objects...[I]t is necessary to provide a method that can segment and track...in real-time"

"[Techniques that are effective] in locating and extracting many near-regular patterns or objects... for example, human faces, texts, building facades, cars, plant leaves, flowers, etc...a wide range of application...[e.g.,] to remove noise...to add things [to images]...[and] license plate recognition"

"Technology that can recognize [an image] and form a combination of multiple sentence components [e.g. "person", "play", "skateboard"]...applications such as image understanding"

**Extended Data Fig. 1 | Computer vision papers and downstream patents.** Example pairs have been randomly drawn from our collected corpus of 43,022 such documents. For each paper and patent, an excerpt of its stated applications is presented. This provides a snapshot of our corpus.