The Surveillance AI Pipeline

PRATYUSHA RIA KALLURI*, Computer Science Department, Stanford University, USA
WILLIAM AGNEW*, Paul G. Allen School of Computer Science and Engineering, University of Washington, USA
MYRA CHENG*, Computer Science Department, Stanford University, USA
KENTRELL OWENS*, Paul G. Allen School of Computer Science and Engineering, University of Washington, USA
LUCA SOLDAINI*, Allen Institute for Artificial Intelligence (AI2), USA
ABEBA BIRHANE*, Mozilla Foundation & School of Computer Science and Statistics, Trinity College Dublin, Ireland

ABSTRACT

A rapidly growing number of voices have argued that AI research, and computer vision in particular, is closely tied to mass surveillance. Yet the direct path from computer vision research to surveillance has remained obscured and difficult to assess. This study reveals the *Surveillance AI pipeline*. We obtain three decades of computer vision research papers and downstream patents (more than 20,000 documents) and present a rich qualitative and quantitative analysis. This analysis exposes the nature and extent of the Surveillance AI pipeline, its institutional roots and evolution, and ongoing patterns of obfuscation. We first perform an in-depth content analysis of computer vision papers and downstream patents, identifying and quantifying key features and the many, often subtly expressed, forms of surveillance that appear. On the basis of this analysis, we present a topology of Surveillance AI that characterizes the prevalent targeting of human data, practices of data transferal, and institutional data use. We find stark evidence of close ties between computer vision and surveillance. The majority (68%) of annotated computer vision papers and patents self-report their technology enables data extraction about human bodies and body parts and even more (90%) enable data extraction about humans in general.

Moreover, we unearth widespread patterns of documents using language that obfuscates the extent of surveillance: documents frequently claim to study “objects” or similarly generic terms, but brief definitions and figures reveal the field has come to conceptualize the term “object” as subsuming humans. This casts all papers and downstream patents referencing “objects”, which constitutes the majority of the field, in a new light, aligned with the production of surveillance. Through a large-scale computational analysis of three decades of computer vision papers with downstream patents, we find that the large majority of these papers (71%) are used in surveillance patents. Comparing the 1990s to the 2010s, the number of papers with downstream surveillance patents increased more than five-fold. The large-scale analysis further sheds light on who is producing computer vision research leading to surveillance. We find consistent evidence against the narrative that a few rogue entities are dedicated to and driving surveillance. Rather, we expose the overwhelming norm that when institutions, nations, or subfields author papers with downstream patents, the majority of these papers are used in surveillance patents; 74% of institutions, 83% of nations, and 70% of subfields authoring papers with downstream patents follow this norm.

As a result, institutions, namely elite universities and “big tech” corporations dominating the production of computer vision research, have authored substantial computer vision research used in surveillance patents. In total, we find computer vision research has been used in more than 11,000 surveillance patents. This analysis reveals the paths by which computer vision research has powered the ongoing expansion of surveillance.

[*] These authors contributed equally to the realization of this project.

Authors’ addresses: Pratyusha Ria Kalluri*, pkalluri@stanford.edu, Computer Science Department, Stanford University, 353 Jane Stanford Way, Palo Alto, USA; William Agnew*, wagnew3@cs.washington.edu, Paul G. Allen School of Computer Science and Engineering, University of Washington, 185 E Stevens Way NE, Seattle, USA; Myra Cheng*, myra@cs.stanford.edu, Computer Science Department, Stanford University, 353 Jane Stanford Way, Palo Alto, USA; Kentrell Owens*, kentrell@cs.washington.edu, Paul G. Allen School of Computer Science and Engineering, University of Washington, 185 E Stevens Way NE, Seattle, USA; Luca Soldaini*, lucas@allenai.org, Allen Institute for Artificial Intelligence (AI2), 2157 N Northlake Ave, Seattle, USA; Abeba Birhane*, birhanea@tcd.ie, Mozilla Foundation & School of Computer Science and Statistics, Trinity College Dublin, Dublin, Dublin, Ireland.

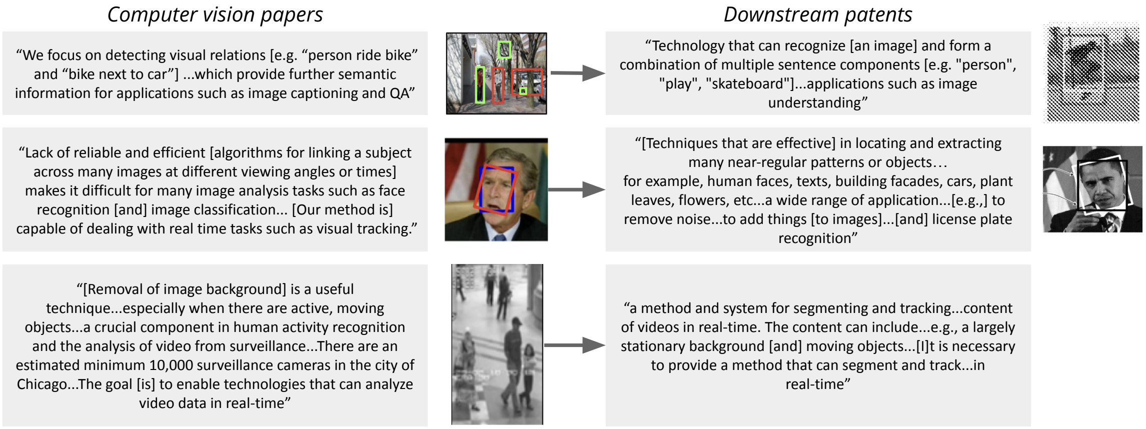
1 INTRODUCTION

Over the past few decades, many voices, from grassroots communities to policymakers, have drawn attention to and organized against the rise of mass surveillance [3, 4, 24, 27, 55]. Moreover, they have asserted that artificial intelligence (AI) research, and computer vision research in particular, is a primary source for designing, building, and powering modern mass surveillance [10, 52, 65, 67, 72]. These concerns are grounded in the historical and ongoing legacy of surveillance technologies that exacerbate disparities, limit free expression, and create conditions that facilitate discrimination and abuse of power [24, 64]. Yet there remains a sharp divide between those within the field of computer vision and those outside of it. Computer vision insiders are intensely familiar with emerging thrusts of computer vision research, and are publishing in record-breaking numbers [7], yet are frequently siloed from the downstream applications of their research. Meanwhile, public attitudes indicate nuanced distrust and fear regarding the normalisation of surveillance technologies [24]. Furthermore, there remain steep barriers from rigorously understanding, unmasking, and thus intervening on, these technologies. Ascertaining the details of these pathways is urgent and crucial given these pathways are societally consequential at a global scale: the threat of mass surveillance based in computer vision has been compared to plutonium for the reason that its harms far outweigh its benefits [67]. This study introduces the *Surveillance AI pipeline* to illuminate the role of computer vision in surveillance.

Computer vision is a subfield of AI that focuses on measuring, mapping, recording, and monitoring the world from visual inputs such as image and video data. Computer vision has historical roots in military and carceral surveillance [22, 62]. As a technology that emerged in military contexts, it was historically developed to identify targets and gather intelligence in war, law enforcement, and immigration contexts. While the field of computer vision generally emphasizes training computers to interpret and understand the visual world, in the prominent task of facial recognition, for example, the identification of suspects for law enforcement remains one of the primary motives [62]. Further surveying the genealogy of the history of facial recognition datasets, Raji and Fried [62] illustrate that their military history has heavily shaped every aspect of the developed technologies, including data collection practices, prioritized tasks, and evaluation metrics. In our study, we interrogate whether and how these histories of computer vision shape computer vision papers and the downstream patents.

Surveillance studies offers a vast body of literature on the dynamics, histories, and consequences of surveillance. *Surveillance at the most general level is defined as an entity gathering, extracting, or attending to data connectable to persons, whether individuals or groups* [50]. In the current computer age, surveillance is frequently "extensive": entities, who are often minimally visible, use big datasets and aggregation to extend their reach, accessing previously unseen persons, locations, or information. Prominent examples are practices where entities in position 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 [23, 52]. Through ubiquitously connected networks, data is aggressively gathered, shared, and aggregated. Behaviours, relationships and social and physical environments are datafied, modelled, and profiled. Many scholars emphasize that surveillance is inextricable from purposes such as influence, management, coercion, repression, discipline, and domination [14]. A foundational understanding in surveillance studies is that, once established, technologies enabling surveillance continue to operate as surveillance regardless of whether they are currently in use. Even when monitoring is not actively leveraged, the very possibility of such monitoring suffices to foster conditions of fear and self-censorship. A long legacy of surveillance studies scholarship studies this approach as a key means of social control [30, 38].

Fig. 1. **Computer vision papers and downstream patents.**
These examples of computer vision papers and downstream patents are randomly drawn from our corpus. For each paper and patent, an excerpt describing its goals and applications is shown, with an illuminating data sample if any were provided.



Computer vision papers

Downstream patents

"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"

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

"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."

"[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"

"[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"

"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"

Throughout this project, we ground our conceptualization and assessment of Surveillance AI in surveillance studies and critical AI literature. In doing so, we are able to connect our study to the broader understandings, features, and consequences of surveillance. We present a more extensive review of contextualizing literature in Appendix F.

The obfuscation of the Surveillance AI pipeline results from a confluence of forces. First, computer vision research is perceived by many as a neutral, purely intellectual endeavor, separate from downstream impacts and applications. In fact, AI research at large rarely discusses connection to societal needs or potential negative consequences [18]. Despite strong arguments emerging from surveillance studies, science and technology studies (STS), critical data studies, and more connecting AI and surveillance [10, 23, 48, 52, 65, 67, 72], the direct path from the computer vision field to surveillance remains obscured. Furthermore, surveillance often operates in the dark, and surveillance technology producers take extra measures to hide their existence [21, 44].

It is difficult to gather direct evidence and details regarding the connections between research and surveillance applications: computer vision research papers and documentation (i.e., what research is being done) are written in ways that are not accessible to many outside the field; those who can parse this work are not accustomed or incentivized to elucidate the details of surveillance emerging; and research appears to trickle down in a multi-stage process. As a result, many aspects of the connection between computer vision research and surveillance remain shrouded in mystery.

Our contributions. In this paper, an interdisciplinary team of researchers leveraged broad expertise including machine learning, AI, robotics, computer vision, privacy, science technology and society studies (STS), critical data studies, and critical AI studies to conduct an in-depth content analysis and large-scale computational analysis of three decades of computer vision papers and downstream patents. Notably, we study self-reported uses and claims in the papers and patents. Thus, our findings are robust to claims of unintentional, unanticipated dual use by "bad actors"; the extent and types of Surveillance AI we uncover are those that are intentionally indicated and anticipated. Our key contributions are fourfold:

(1) **We present a topology illuminating core dynamics of Surveillance AI**. We present an organizing system for understanding essential features of computer vision papers and patents and for identifying when they

constitute Surveillance AI. For the many subtle variants of Surveillance AI that emerge, we offer examples, textual evidence, and analysis, capturing their nature.

(2) **We assess the extent of Surveillance AI**, quantifying the prevalence with which computer vision papers and patents serve as Surveillance AI, and which variants of Surveillance AI dominate.

(3) **We present a large-scale analysis of more than 20,000 computer vision papers and downstream patents to capture the roots and evolution of Surveillance AI, across corporations, universities, nations, subfields, and years**. We reveal, for example, that it is not a few outlier institutions dedicated to surveillance. Rather it is the overwhelming norm that when institutions author papers used in patents, the majority of these papers will be used in surveillance patents.

(4) **We shed light on widespread obfuscating language in papers and patents that contributes to perpetuating the paradigm of Surveillance AI.** This underscores the additional, hidden layers of interconnectedness between computer vision research and surveillance.

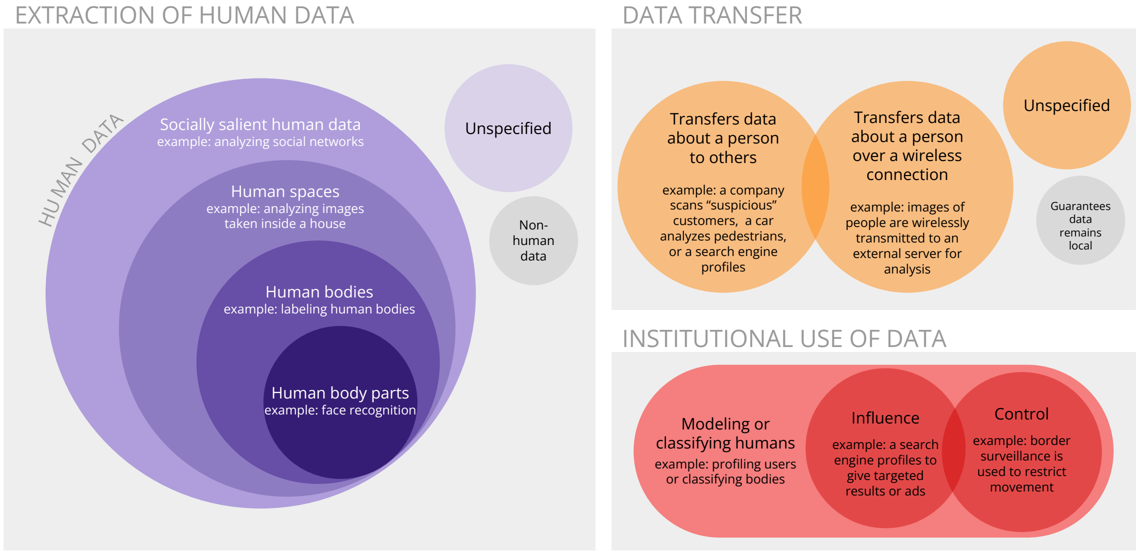
We make visible the pathways from computer vision research to surveillance applications, and we aim for this mapping to serve as a tool for communities to strategically organize around and against surveillance; policy-makers to identify regulatory targets to curb surveillance; researchers to contend with the consequences of the field and (re)shape the research agenda; and the public to exercise the right to knowledge and power over the apps, gadgets, and devices that mediate and infiltrate their daily lives with surveillance.

2 METHODOLOGY

Data. To study the pathway from computer vision research to applications, we analyzed computer vision research papers and their downstream patents. Research papers and patents have several unique advantages making them revealing artifacts for this analysis. First, they are primary sources written in researchers' and patenters' own words, with professional and institutional standards that they accurately describe their research and technologies and be able and willing to defend these documents' accuracy. Additionally, they must report their authors, primary affiliated institutions, and years of publication, enabling reliable analysis 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. The connections between research papers and citing patents serve as a rich datatrail of the path from research to applications [11, 41]. We choose to study papers published at the annual Conference on Computer Vision and Pattern Recognition (CVPR) because it is the longest standing and highest impact computer vision conference by a large margin: by standard h5-index it is among the top five highest impact publications *in any discipline*, alongside Nature and Science. As the premier computer vision conference, trends in CVPR are an "indicator of hot topics for the AI and machine learning community" [5]. Acceptance and publication at CVPR marks approval of research as work that exemplifies the core values of the computer vision community. As such, papers published at CVPR both represent state-of-the-art in current computer vision and effectively reveal the values held in high regard within the community.

We leverage the Microsoft Academic Graph [66] and the paper-patent linkage data by Marx and Fuegi [51] to construct our corpus. First, we obtain all CVPR papers published from 1990-2020. Then, for each paper, we obtain all patents in which the paper was cited, which we refer to as the paper's *downstream patents*. In total, our corpus links more than 19,000 papers to more than 23,000 downstream patents. In Figure 1, we present randomly sampled pairs consisting of a CVPR paper and a downstream patent, providing a snapshot of our data.

Fig. 2. **The topology of Surveillance AI.** We present a topology of key dimensions of Surveillance AI. In particular, the relationship between complex technologies and surveillance can be clarified by attending to the *extraction of human data*, the *practices of data transferal*, and the *institutional uses*. At each stage, we identify and describe prominent variants. In Section 3 we present textual examples and analysis and quantify the prevalence of these features of Surveillance AI.


EXTRACTION OF HUMAN DATA
HUMAN DATA
Socially salient human data
example: analyzing social networks
Human spaces
example: analyzing images taken inside a house
Human bodies
example: labeling human bodies
Human body parts
example: face recognition
Unspecified
Non-human data
DATA TRANSFER
Transfers data about a person to others
example: a company scans "suspicious" customers, a car analyzes pedestrians, or a search engine profiles
Transfers data about a person over a wireless connection
example: images of people are wirelessly transmitted to an external server for analysis
Unspecified
Guarantees data remains local
INSTITUTIONAL USE OF DATA
Modeling or classifying humans
example: profiling users or classifying bodies
Influence
example: a search engine profiles to give targeted results or ads
Control
example: border surveillance is used to restrict movement

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 each year from 2010 to 2020, we randomly sampled ten paper-patent pairs that consisted of a CVPR research paper published in this 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 close reading of documents and a rigorous qualitative methodology. Such an orientation is necessary when the key concepts that will emerge from a body of study are not known a priori, a deep characterization is valuable, and documents are complex or dense, expressing their key concepts with subtle language unique to the corpus [32, 68]. An interdisciplinary six-person team analyzed the documents using an integrated inductive-deductive methodology. In the inductive component, each document was read line by line including figures, inductively coding key emergent features of the technology's treatment of human data and iteratively accumulating a list of these key features and their relationships.

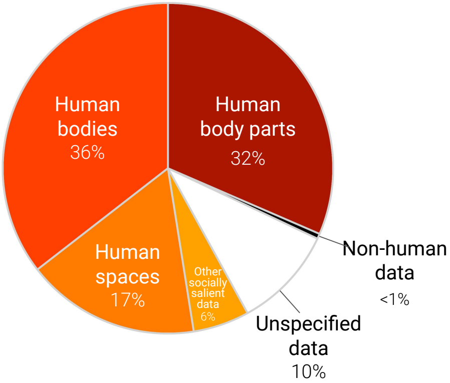
We complemented this with an additional deductive component in order to ensure that we actively looked for and captured instances of papers and patents with key features that inhibited usage for surveillance, even if rare. In this deductive component, we coded for two such features. The inductive and deductive codes are discussed in detail in Section 3 and Appendix B and C.

Our annotation team had several strengths: our team included both published experts in computer vision and field outsiders, allowing for expert insights and translation, as well as fresh perspectives that could illuminate computer vision disciplinary biases. We utilized the constant comparative method. The first fifty papers and patents were annotated by more than one annotator, which was particularly valuable for surfacing cases where a single sentence or figure influenced the appropriate code; the existence of such cases is discussed in Section 4. 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, along with the relationships amongst these dimensions and features, which we present in the form of a topology. At the end of all coding, we additionally quantized the annotations corresponding to the most salient emergent dimension, in order to present a quantitative analysis of this primary dimension.

Our guiding aim was to cast light on the nature of the dense bodies of computer vision research and applications and elucidate connections to surveillance. These papers and patents can each be dozens of pages, difficult to obtain and link, and written in a manner that assumes the reader has substantial expertise in computer vision, disciplinary jargon, academic research, and patent applications. On the basis of our in-depth, interdisciplinary content analysis, we present the Surveillance AI topology (Figure 2), bringing to the fore the dimensions, features, and dynamics of computer vision's treatment of human data and, on the basis of surveillance studies, highlighting those that constitute surveillance. Our analysis identified three key dimensions capturing these technologies' treatment of human data: *(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 does the data remain under the control of the datafied person or become transferred to others? *(3) Use of data* — For what purpose is the data used? These three dimensions are discussed in detail, with examples and analysis, in Section 3 and Appendix B and C. In Section 1, we introduced the conceptualization of surveillance, drawn from Surveillance Studies, as an entity gathering, extracting, or attending to data connectable to persons, whether individuals or groups. We found this conceptualization particularly aligned with one of the emergent dimensions we identified. Specifically, under this conceptualization, the first dimension identified in our analysis (the type of data targeted, and specifically the extent to which it is connectable to humans) reveals the connection between these technologies and surveillance. Given this, we discuss this dimension of the topology in detail in Section 3, provide examples, map the emergent features to varieties of surveillance, and we further quantize the annotations, presenting the relative frequencies of data types in Figures 3 and A1. This conceptualization casts the remaining dimensions of the topology as secondary with respect to categorizing whether a document contains surveillance, and these dimensions are 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 essential to understanding the roles, dynamics, and consequences of surveillance. Given the importance of these dimensions, in Appendix B and C we include a full discussion of these dimensions, demonstrative examples, the emergent features and patterns identified, and connections to nuanced dynamics of surveillance that have been discussed in surveillance studies literature.

Automated analysis. To study the breadth and variation of surveillance across years, institutions, nations, and subfields, we conducted a large-scale computational analysis of more than 23,000 downstream patents. Specifically, during the in-depth manual content analysis the team of annotators developed a list of surveillance indicator words that indicated surveillance (in particular, words that indicated the targeting of human body parts, human bodies, human spaces, or traces of socially salient human data; Section 3 provides detailed discussion of each of these types of targeting and discussion of how they enable surveillance). To validate each surveillance indicator word, we scanned the corpus for all patents containing this word, randomly sampled ten of these patents, and conducted manual inspection. We removed from the list all words that manual inspection identified as not reliable indicators (typically because they had frequent additional word senses; e.g. a "store" could be a human space but was frequently a technical term related to

Fig. 3. **The extraction of human data in computer vision papers and downstream patents.** The large majority (90%) of the annotated papers and patents extract data about humans. The majority of the papers and patents (68%) reported that they specifically enable extracting data about human bodies and body parts. Less than 1% of the papers and patents targeted only non-humans.



data or memory storage, so was removed from the list). The resulting list of surveillance indicator words was approved by consensus, and we list these words in Appendix G.2.

For each paper, we scanned its downstream patents to identify patents containing one or more of these surveillance indicator words, which we refer to as downstream surveillance patents. We present the distribution of surveillance patents across institutions, nations, subfields, and years, along with contextualizing discussion, in Section 5, with additional analysis in Appendix D and E. We present additional methodological details in Appendix G.

3 THE TARGETING OF HUMAN DATA

There is extensive evidence of public distrust and fear concerning the capturing and monitoring of human data, including substantial concern about computer vision technologies operating on online personal data traces and biometric and body data [24, 54]. Interrogation of the role of computer vision in originating these practices is well-justified, yet it is in reality made extremely difficult for non-experts to access the inner workings of computer vision's relationship to human data. We present an empirically grounded characterization, illuminating the types of data targeted in computer vision, along with to what extent parts or the whole of computer vision is explicitly dedicated to targeting human data, which types of human data, as well as the level of sensitivity of such data.

Quantitative summary

In Figure 3 we present a quantitative summary of the types of data targeted in computer vision papers and downstream patents. Additionally, in Figure A1 we stratify this data to compare papers versus patents. On the basis of our in-depth content analysis, we find that 90% of papers and patents extracted data relating to humans. Furthermore, the majority (68%) explicitly extracted data about human bodies and body parts. No papers and only 1% of patents were dedicated to targeting non-human data, showing that both computer vision research and applications are overwhelmingly concerned with datafying, analyzing, tracking, and monitoring humans and specifically human bodies.

The topology of targeted data

Our content analysis identified four main types of human data targeted in computer vision papers and patents. These targeted human data types form a series of nested categories as follows: *human body parts*, *human bodies*, *human spaces* and *socially salient human data* (Figure 2). As an example of this nested relationship, whereas the monitoring of human spaces surveys spaces human bodies may enter (or avoid entering), the tracking of human bodies hones in on the case when a body has definitely entered the space being surveyed, extracting an even deeper layer of intimate details. For each paper and patent, we identified the innermost of these data types, *i.e.* the most intimate, that the document targeted. We illustrate each type of data targeted with examples and textual evidence. For each type, we also connect the targeting of this human data type to specific insights from surveillance studies, surfacing prominent concerns regarding the consequences of computer vision targeting this type of data.

At least a quarter of both papers and patents (38% and 25% respectively) claimed targeting human body part data as a major strength of their technology. Papers and patents most frequently emphasized analysis of faces, including detection of eyes, eye movement, faces, "suspicious" facial expressions, and, extremely frequently, facial recognition. Other uses of human body part data include fingerprint detection and activity recognition technologies that emphasize tracking body parts, sometimes even using explicit "body part models". Papers and patents broadly assumed these tasks as valuable. Biometrics such as faces, fingerprints, and gait, which constitute uniquely persona data that is often inseparable from our identities, has proliferated as a form of surveillance in recent years, and its pervasiveness is believed to significantly infringe on people's privacy and threaten human rights [24].

Human bodies

"...people monitoring in public areas, smart homes, urban traffic control,
mobile application, and identity assessment for security and safety..." (Paper 53)

Papers and patents that claimed they were useful for analyzing human body parts contributed to a larger, overwhelming trend in the data: *the majority of papers and the majority of patents enabled extracting human body data.* In addition to body part and facial recognition, these technologies were frequently aimed at mass analyzing datasets of humans in the midst of everyday movement and activity (shopping, walking down the street, sports events) for purposes such as security monitoring, people counting, action recognition, pedestrian detection, for instance in the context of automated cars, or unnamed purposes. The dominance of analysis of human bodies in everyday settings aligns with the view of new surveillance by Browne [23] who characterizes the new practices of surveillance as often *undetected* – for example cameras hidden in everyday benign objects – or even *invisible.* In these forms, data is collected without consent of the target, and then shared, permanently stored and aggregated. Browne [23] characterizes surveillance as focused on monitoring and cataloguing that which was previously left unobserved, with the human body as a primary site of surveillance.

Human spaces

"...a scene could be decomposed into a set of semantic objects...
(accompanied by an example image taken inside an office)" (Paper 40)

The analysis of human spaces was the most common concretization of scene analysis, understanding, or recognition, which are often presented as a core contribution of the field in papers and patents alike. This type of targeted data was data generated from living spaces – personal and communal – such as people's homes, offices, roads, town squares, auditoriums, or borders. In 16% of papers and 18% of patents, the targeting of such spaces was established, even while the potential for humans to be in the scene was rendered unstated, a given, or overall inconsequential. Purported purposes for these can include product design (automated vacuum cleaners), traffic pattern prediction, identifying objects in a scene or assisting in monitoring of large uncontrolled border crossing areas. Importantly, surveillance works by first making previously unobserved phenomena, events, interactions and places amenable to observation [26]. The rendition of homes, streets, neighbourhoods, villages and towns to surveillance technology marks these spaces as no longer scenes where residents, live, meet and talk but another object of target for data collection, tracking, categorizing, and predicting [72]. The consequence of the gradual rendering of more and more of these spaces is extremely subtle yet has profound implications for the future of humanity. It accumulates to what Zuboff calls the condition of "no exit", where there are fewer and fewer spaces left to "disconnect", seek respite and be left to just be [72].

Socially salient human data

"Free-hand human sketches [e.g., of another person's item of clothing] are used as queries to perform instance-level retrieval of images" (Paper 81)

Relatively few papers and patents present their technology as useful for monitoring, tracking or predicting only non-body-related socially salient human data (less than 10% of both papers and patents). GDPR articulates that socially significant data includes data containing traces of the mental, economic, cultural, social status, identities, preferences, or location details of humans. Individuals themselves may not be under direct focus however, data about individuals, groups, societies, cultural identities, events, situations, which contain traces of personal details are collected and analysed. Similar to the above category, capturing socially salient human data contributes to the gradual cataloguing, documenting, mapping, and monitoring of human affairs in its rich complexities [52, 72].

Non-human data

"The invention discloses a method for classifying and identifying a plant image set" (Patent 74)

Unlike the other data types we have presented here, which were inductively found only through close reading of the papers and patents, the annotation team deductively included non-human data in the topology from the start. This was to ensure that we captured mentions of any non-surveillance technologies in papers and patents, even if rare. Non-human data refers to data collected and analysed containing no traces of personal data or human affairs. Of the papers and patents we examined, 0% of papers and 1% of patents limited themselves to non-human data.

Unspecified

"...[computing] the location and amplitude (visibility) of edges in natural images..." (Paper 56)

Finally, the remaining portion (10%) of papers and patents claimed to capture and analyze "images", "text", "objects", or similarly generic terms without disclosing whether they anticipated these categories including humans or human data. Through close inspection, we find this label does not imply that the technology described in the paper or patent cannot be used on human-related data or even that human data was not a desired use case by the authors. To the contrary, we find that dense patent language can hide the human data analysis in the upstream papers and, conversely, papers that do not speak to the potential for use with human data often lead to patents that explicitly monitor human data. To further characterize the actuality and consequences of these unspecified data types, we discuss the obfuscating and "object"-related language of Surveillance AI in Section 4.

We present this analysis, illuminating the specific ways computer vision is extracting human data and the extent of this, so individuals, communities, and organizations may leverage these insights and form informed perspectives and effective strategies to resist, challenge, influence, and/or regulate the targeting of human data.

4 THE OBFUSCATING LANGUAGE OF SURVEILLANCE AI

Across the computer vision papers and downstream patents analyzed, a striking trend emerged of obfuscating language that minimized or sidestepped mentions of potential surveillance and discussion of its harms. We highlight two salient themes that emerged:

1. Papers and patents cast humans as merely another entity under the umbrella term "objects".

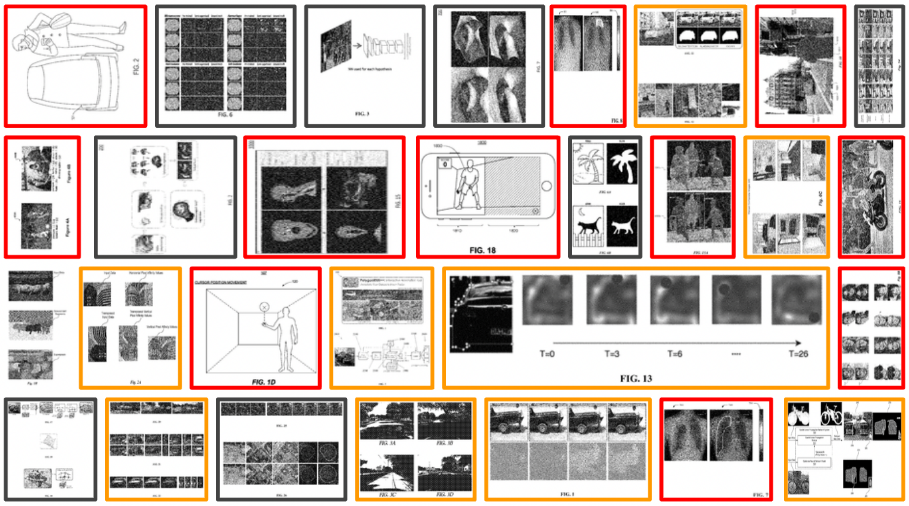
"We will simply use the term objects to denote both interactional objects and human body parts" (Paper 84)
"Using these methods, objects such as people and vehicles may be identified and quantified based on image data." (Patent 85)
"Since the surveillance system detects and can be interested on vehicles, animals in addition to people,
hereinafter we more generally refer to them with the term moving object." (Paper 53)

Establishing the conceptualization of human as merely a kind of object explicitly, as many papers and patents do, enables the rest of those documents and, crucially, *all other papers and patents* to merely discuss problems related to *objects* or *scenes*, as they can rely on the understanding of human as object that has been established by peers. Because humans are considered objects and scenes often contain people, such abstractions indicate that the field understands any paper or patent that discusses objects and scenes – which constitutes the majority of the field – as potentially enabling surveillance of humans. Many papers conflate humans with objects, making no note of how performing tasks like detection or segmentation on people has extremely specific, and socially consequential impacts. For instance, a paper about panoptic segmentation, in giving context about the body of literature that it draws from, makes no distinction between non-human detection and face detection: "Early work on face detection...helped popularize bounding-box object detection. Later, pedestrian detection datasets helped drive progress in the field" (Paper 96). The lede of a paper about parsing object interactions does the same: "major task of fine-grained interaction action analysis is to detect the interacting objects or human body parts for each video frame (in the rest of the paper, we will simply use the term objects to denote both interactional objects and human body parts)" (Paper 84). Considering humans as objects implies that any knowledge produced related to object-focused tasks can be directly applied to human data. This assumption neatly abstracts away the ways that such methods can be applied to surveillance. This phenomenon also ties to literature about traditional science's sharp divide between subject and object, which positions scientists as the studiers of "objects" out there. This "splitting of subject and object" faciliates "denial of responsibility and critical inquiry" [42]. This contextualizes the field's homogenization of all possible data, including human data, into objects to be studied, often without consent and without consideration of their sources or impacts.

2. *What is not said*: Even when the text of papers and patents makes no mention of human data, figures or datasets may contain many, sometimes exclusively, 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 find has already played out: we find that other documents lean on these norms, claim to target "objects", in actuality 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 paper's first figure 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 in its text or figures; yet inspection of the paper's datasets reveals they demonstrate their technology by detecting regions of interest such as humans walking on a sidewalk (Paper 1). Figure 4 illuminates this

Fig. 4. **Figures in downstream patents, illustrating the prominence of targeting human bodies and spaces.** Images that illustrate the targeting of humans are widespread in patents, as is shown in this random sample of patent images. We highlight those containing human bodies (red) and those containing human spaces (orange).



pattern via a random sample of annotated patents' images, many of which we find contain human bodies and spaces despite lacking explicit mention of these entities in the text. This further entrenches a field norm that, on one hand, humans and objects of all kinds may be targeted in parallel, casting the vastly different implications as inconsequential, and, on the other hand, that humans can be central targets of technologies without needing to leave a textual trace, let alone discuss, surveillance. This norm obscures the extent of Surveillance AI from both outsiders attempting to understand the field and insiders not cognizant of these practices. In this way, the modeling and categorization of humans has become so pervasive it can only be understood as a task that has become widely acceptable across the field of computer vision, rendering it a potential application of virtually all computer vision papers and patents.

5 THE ROOTS AND EVOLUTION OF SURVEILLANCE AI

Surveillance AI does not emerge in a vacuum. Over time, researchers from many nations, institutions, and subfields have conducted this research and developed applications, whether cognizant of and attentive to its downstream applications or not. This work has been actively funded and commercialized by external parties, and it continues to evolve. Through a large-scale computational analysis of three decades of computer vision's downstream patents (analyzing more than 23,000 patents), we uncover rapid growth of Surveillance AI, surface fieldwide norms, and reveal the patterns of institutions, nations, and subfields that have contributed to the rise of Surveillance AI.

Quantitative results

First, we present the evolution across decades of the computer vision papers used in patents (Figure 5). We find a substantial increase in the proportion of these papers used in surveillance patents. Comparing decades, we find that the 1990s produced relatively fewer computer vision papers with downstream patents, and half of these were used in surveillance patents. Two decades later, the 2010s produced more than three times as many computer vision papers

Fig. 5. **Between 1990 and 2020, computer vision research used in surveillance patents has increased significantly.** *Left:* Out of computer vision papers with downstream patents, the proportion used in surveillance patents has gradually increased. *Right:* Comparing the 1990s to the 2010s, the amount of computer vision research used in surveillance patents has increased more than five-fold.

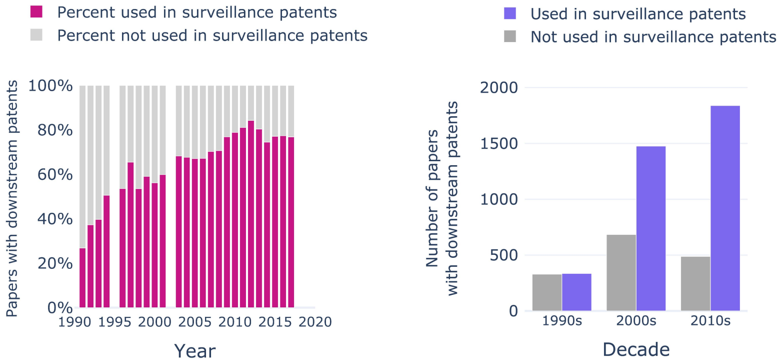
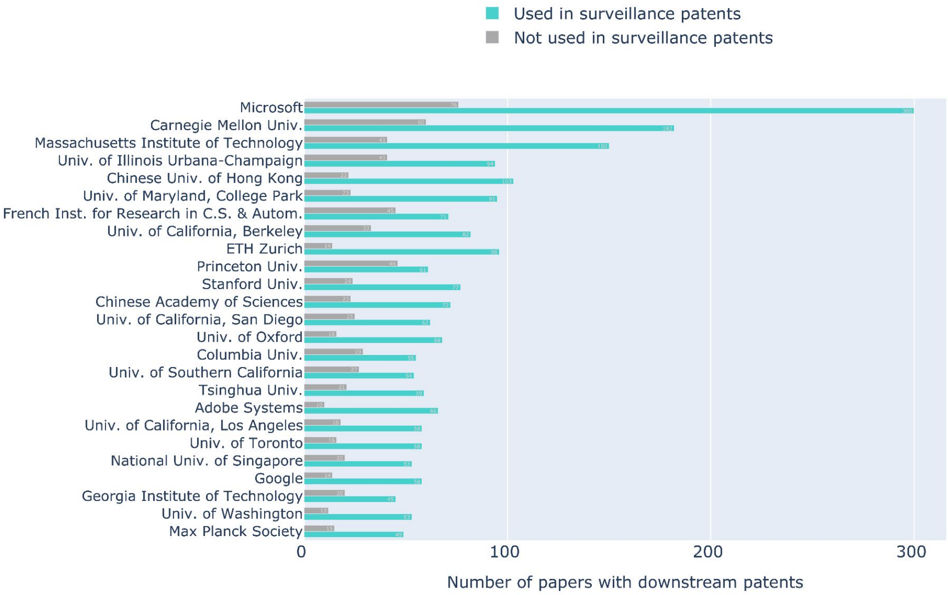
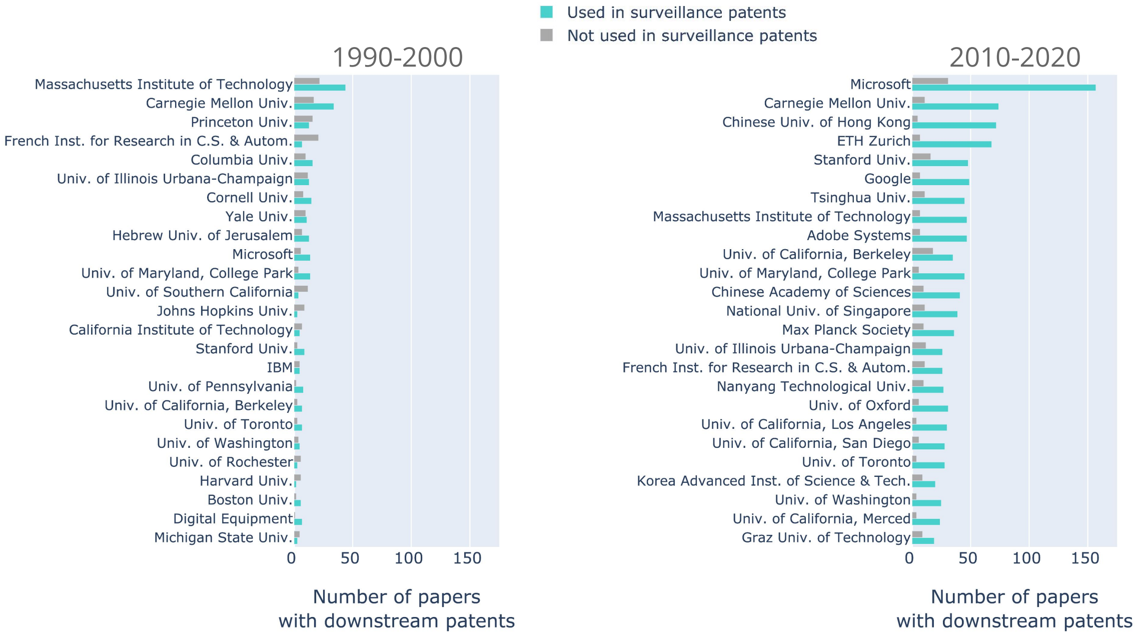


Fig. 6. **Top institutions producing computer vision research with downstream patents.** For all of the top 25 institutions, the majority of these papers are used in surveillance patents. (For all of these institutions, teal bars are larger than grey bars.) Many of these are prestigious institutions, including elite universities and "big tech" corporations.



with downstream patents and 79% were used in surveillance patents. The twin forces of the increase in computer vision papers with downstream patents and the increase in the proportion of these used in surveillance patents combined to large effect: whereas the 1990s produced 335 computer vision papers used in surveillance patents, the 2010s produced more than 1,800 computer vision papers used in surveillance patents. In total, the 1990s to the 2010s constituted a more than five-fold increase in the number of computer vision papers used in surveillance patents.

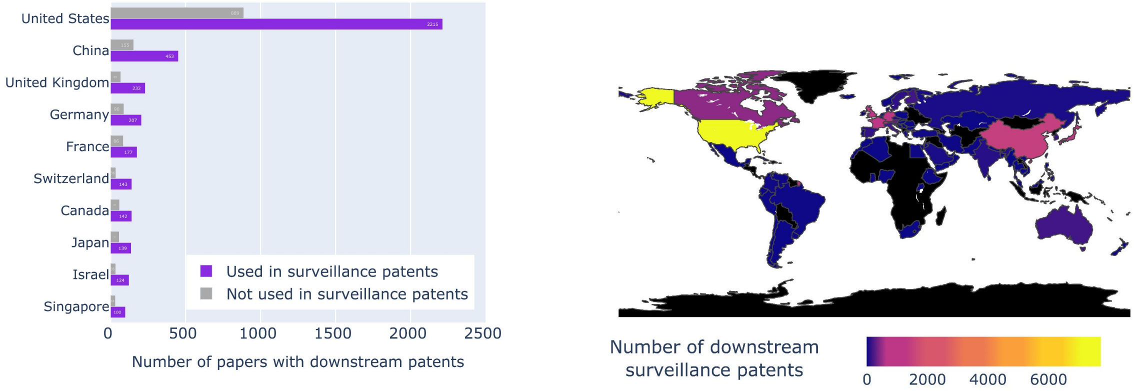
Fig. 7. **Top institutions producing computer vision research with downstream patents, in the 1990s and the 2010s.** In the 1990s, the number of papers with downstream patents is relatively small, and there is a mix of institutions tending and not tending toward surveillance patents. In the 2010s, the number of papers with downstream patents has increased substantially, and, for all of the top 25 institutions, the majority of these papers are used in surveillance patents. (For all of these institutions, teal bars are larger than grey bars.)



We next analyze which entities have driven this line of research. Are a few rogue entities producing research for surveillance or are ties from research to surveillance a fieldwide norm? Our findings do not support the narrative of only a few rogue entities producing research for surveillance. We find a striking norm: when an institution, nation, or subfield successfully authors computer vision papers with downstream patents, the majority are used in surveillance patents. This norm is widespread; 74% of institutions, 83% of nations, and 70% of subfields authoring papers with downstream patents follow this norm. We include additional details in Appendix D and E. Due to the pervasive norm of computer vision research being used for surveillance, it is the reality that institutions driving computer vision research are often driving research used in surveillance. In Figure 6, we show the top institutions producing computer vision papers used in patents. We see that for all of these top institutions, the majority of these papers are used in surveillance patents. We further find that many of these papers are used in multiple surveillance patents. These institutions are often prestigious institutions, including "big tech" corporations and elite universities, many of which are also the top producers of computer science papers generally [12, 16], reflecting tight ties between those driving research and those driving surveillance. Taken together, these 25 institutions alone have authored papers used in 7,021 surveillance patents.

Finally, to understand the influence of nations and their research output on surveillance patents, we present in Figure 8 the distribution of surveillance across the authoring countries. The authoring countries are obtained from the location of paper authors' institutional affiliations. The top two nations producing papers with downstream surveillance

Fig. 8. **The pathway from nations to computer vision research and surveillance patents.** *Left:* The top nations producing computer vision research with downstream surveillance patents. *Right:* A map indicating the resulting number of downstream surveillance patents across the globe. The United States dominates the production of these computer vision papers, and the resulting papers have been used in more than 7,700 surveillance patents.



patents are the US and China, with the US producing more of these papers than the next several nations combined. Our findings correspond to previous reports about AI-driven surveillance across countries, which state that on a global scale, China and the United States are the major drivers in supplying advanced surveillance technologies, while the major users include both liberal democracies and other countries with less democratic governments [36].

Discussion

The prevalence of both elite universities and major tech corporations in Figures 6 and 7 reflects the research ties between universities and corporations that have shaped the field of computer science from its nascence [31]. Our findings align with historical precedent: the Cold War facilitated the rise of government-funded, military-oriented science and engineering projects in both universities and corporations. For instance, computer science departments at Stanford and MIT got their start from, and rose to their dominance with, wartime government funding [46]. Similar initiatives gave rise to the Stanford Industrial Park, which later became Silicon Valley. These projects were often motivated by technocratic ideologies of military surveillance, a paradigm that has shaped the directions of research ever since [45, 71]. The early conception of the Internet is also a tale of military and academic dominance resulting from Cold War-era think tanks [8]. As Silicon Valley rapidly expanded, intelligence agencies turned their attention and funding to the tech sector, incentivized by the dual utility of AI in both civilian and military applications; CIA's relationship with Google, for example, helped facilitate Google's dominance [13]. In recent years, tech companies have not only produced papers leading to surveillance patents but also bidded for and accepted various defense contracts [28]. Now, we find that these legacies continue as these institutions author the most papers with downstream surveillance patents. We also connect the nation statistics to the narrative of ongoing tensions between the United States and China to establish themselves as the global superpower in AI [57]. From this perspective, the institutional race to develop surveillance technologies is cast as a mission to defend against an enemy [25]. In both nations, the actions and policies of leading tech corporations are inextricable from state agendas [28].

6 A PARADIGM OF SURVEILLANCE

The studies presented in this paper ultimately reveal that the field of computer vision is not merely a neutral pursuit of knowledge; it is a foundational layer for *a paradigm of surveillance.* Our findings include these striking points: 90% of papers and patents emphasize it as a strength that their technologies can target human data. Not only is human data broadly targeted, but the *majority* (68%) of papers and patents explicitly focus on surveillance of human body parts (e.g., faces) and human bodies. Between the 1990s and 2010s, we have seen the rise of Surveillance AI, and it has become an overwhelming norm that computer vision papers used in patents are most likely to be used in surveillance patents. Moreover, even when a paper does not *explicitly* state surveillance as an application, it provides the methods to do so and is grounded in a historical context that makes it possible to target human surveillance while minimizing the acknowledgement of these intentions. In other words, the default stance of the field ties progress closely with surveillance. This is evident in the types of research questions that are valued and prioritized by the field, as well as the way that the papers are written – particularly the use of obfuscating language – for example the use of "object" in the analysis of humans, sometimes only exposing the anticipated human data in images and figures.

The uncovered features of computer vision tie into a broader literature about the veneer of neutrality in science. Scientific findings are frequently falsely 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 debunked by a variety of scholarships, from philosophy of science, STS 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 [42, 43, 47, 61]. Social and cultural histories and norms, funding priorities, academic trends, researcher objectives, and research incentives, for example, all inevitably constrain the direction and production of scientific knowledge. [9, 18, 33, 34]. An assemblage of social forces have shaped computer vision, resulting in a field that fuels the mass production of Surveillance AI.

Peering past the veneer of scientific neutrality, we find that the ongoing expansion of the field of computer vision is centrally and inextricably tied to the expansion of Surveillance AI. At its core, surveillance is the perpetual practice of rendering visible what was previously shielded and unseen [23]. This is precisely the goal of the discipline of computer vision. The continued progress of the field amounts to increasing the capabilities for recording, monitoring, tracking, and profiling of humans as well as the wider social and physical environment. These tasks, which may seem benign to those swimming in the waters of computer vision, in fact exemplify the ways that progress in the field of computer vision is inextricable from increasing surveillance capabilities.

Ultimately, whether a work in the field of computer vision predicates surveillance applications or not, it can and frequently will be used for these purposes. Given the ways that research throughout the field can be implicated and engaged in surveillance, even when the precise details are missing or obfuscated, our findings constitute only a lower bound on the extent of computer-vision based surveillance: there are likely many more works that have quietly contributed to Surveillance AI. Viewing computer vision in this light, it becomes clear that shifting away from the violence of surveillance requires, not a small shift in applications, but rather a reckoning and challenging of the foundations of the discipline.

ACKNOWLEDGEMENTS

We owe gratitude and accountability to the long history of work exposing the nature of surveillance and how technology shifts power, work primarily done by communities at the margins. Myra Cheng is supported by an NSF Graduate

Research Fellowship (Grant DGE-2146755) and Stanford Knight-Hennessy Scholars graduate fellowship. Pratyusha Kalluri is supported in part by an Open Phil AI Fellowship.

REFERENCES

[1] European Commission [n. d.]. *2018 reform of EU data protection rules.* European Commission. https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf

[2] ICO [n. d.]. *Enforcement Powers of the Information Commissioner Enforcement Notice.* ICO. https://ico.org.uk/media/action-weve-taken/enforcement-notices/4020437/clearview-ai-inc-en-20220518.pdf

[3] [n. d.]. Mijente. mijente.net.

[4] [n. d.]. Stop LAPD Spying Coallition. https://stoplapdspying.org/.

[5] 2021. CVPR 2021 Report Identifies 5 Trend Areas. https://www.computer.org/publications/tech-news/events/cvpr-2021-recap.

[6] ICO 2022. *ICO could impose multi-million pound fine on TikTok for failing to protect children's privacy.* ICO. https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2022/09/ico-could-impose-multi-million-pound-fine-on-tiktok-for-failing-to-protect-children-s-privacy/

[7] 2022. Record-Breaking Registrants and Technical Papers for 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). *Markets Insider* (2022). https://markets.businessinsider.com/news/stocks/record-breaking-registrants-and-technical-papers-for-2022-ieee-cvf-conference-on-computer-vision-and-pattern-recognition-cvpr-1031564264

[8] Janet Abbate. 2000. *Inventing the internet.* MIT press.

[9] Janet Abbate. 2012. *Recoding gender: Women's changing participation in computing.* Mit Press.

[10] Philip E Agre. 1994. Surveillance and capture: Two models of privacy. *The information society* 10, 2 (1994), 101–127.

[11] Mohammad Ahmadpoor and Benjamin F. Jones. 2017. The dual frontier: Patented inventions and prior scientific advance. *Science* 357, 6351 (2017), 583–587. https://doi.org/10.1126/science.aam9527 arXiv:https://www.science.org/doi/pdf/10.1126/science.aam9527

[12] Nur Ahmed and Muntasir Wahed. 2020. The De-democratization of AI: Deep Learning and the Compute Divide in Artificiatl Intelligence Research. *arXiv* (2020).

[13] Nafeez Mossadeq Ahmed. 2015. How the CIA Made Google. Inside the Secret Network behind Mass Surveillance, Endless War, and Skynet. *Insurge Intelligence, January* 22 (2015).

[14] Thomas Allmer. 2011. Critical surveillance studies in the information society. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 9, 2 (2011), 566–592. https://www.triple-c.at/index.php/tripleC/article/view/266

[15] Ruha Benjamin. 2019. Race after technology: Abolitionist tools for the new jim code. *Social forces* (2019).

[16] Emery D. Berger. 2017. *CSrankings.* https://csrankings.org

[17] Federico Bianchi, Pratyusha Kalluri, Esin Durmus, Faisal Ladhak, Myra Cheng, Debora Nozza, Tatsunori Hashimoto, Dan Jurafsky, James Zou, and Aylin Caliskan. 2022. Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. *arXiv preprint arXiv:2211.03759* (2022).

[18] Abeba Birhane, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. 2022. The values encoded in machine learning research. In *2022 ACM Conference on Fairness, Accountability, and Transparency.* 173–184.

[19] Abeba Birhane and Vinay Uday Prabhu. 2021. Large image datasets: A pyrrhic win for computer vision?. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV).* IEEE, 1536–1546.

[20] Sarah R Blenner, Melanie Köllmer, Adam J Rouse, Nadia Daneshvar, Curry Williams, and Lori B Andrews. 2016. Privacy policies of android diabetes apps and sharing of health information. *Jama* 315, 10 (2016), 1051–1052.

[21] Thomas Brewster. 2022. Meet The Secretive Surveillance Wizards Helping The FBI And ICE Wiretap Facebook And Google Users. https://www.forbes.com/sites/thomasbrewster/2022/02/23/meet-the-secretive-surveillance-wizards-helping-the-fbi-and-ice-wiretap-facebook-and-google-users/

[22] Meredith Broussard. 2018. *Artificial unintelligence: How computers misunderstand the world.* mit Press.

[23] Simone Browne. 2015. *Dark matters: On the surveillance of blackness.* Duke University Press.

[24] Madeleine Chang. 2022. Countermeasures: The need for new legislation to govern biometric technologies in the UK.

[25] Wendy Hui Kyong Chun. 2006. Control and freedom. *Power and Paranoia in the Age of Fiber* (2006).

[26] Julie E Cohen. 2017. Surveillance vs. privacy: effects and implications. *Cambridge Handbook of Surveillance Law, eds. David Gray & Stephen E. Henderson (New York: Cambridge University Press, 2017)* (2017), 455–69.

[27] Kate Conger, Richard Fausset, and Serge F Kovaleski. 2019. San Francisco bans facial recognition technology. *The New York Times* 14 (2019), 1.

[28] Kate Crawford. 2021. *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence.* Yale University Press.

[29] Lydia de la Torre. 2018. A guide to the california consumer privacy act of 2018. *Available at SSRN 3275571* (2018).

[30] Gilles Deleuze. 1992. *Postscript on the Societies of Control.* The MIT Press.

[31] Paul N Edwards. 1996. *The closed world: Computers and the politics of discourse in Cold War America.* MIT press.

[32] Satu Elo and Helvi Kyngäs. 2008. The qualitative content analysis process. *Journal of advanced nursing* 62, 1 (2008), 107–115.

[33] Nathan Ensmenger. 2015. "Beards, sandals, and other signs of rugged individualism": masculine culture within the computing professions. *Osiris* 30, 1 (2015), 38–65.

[34] Nathan L Ensmenger. 2012. *The computer boys take over: Computers, programmers, and the politics of technical expertise.* Mit Press.

[35] Virginia Eubanks. 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor.* St. Martin's Press.

[36] Steven Feldstein. 2019. *The global expansion of AI surveillance*. Vol. 17. Carnegie Endowment for International Peace Washington, DC.

[37] U. Finardi. 2011. Time relations between scientific production and patenting of knowledge: the case of nanotechnologies. *Scientometrics* (2011).

[38] Michel Foucault. 1977. *Discipline and Punish : the Birth of the Prison.* Pantheon Books.

[39] Thomas Germain. 2023. Innocent Black Man Jailed After Facial Recognition Got It Wrong, His Lawyer Says. https://news.yahoo.com/innocent-black-man-jailed-facial-200800345.html?guccounter=1

[40] Chris Gilliard. 2020. Caught in the Spotlight. *Urban Omnibus* 9 (2020).

[41] Björn Hammarfelt. 2021. Linking science to technology: the "patent paper citation. *Journal of Documentation* (2021).

[42] Donna Haraway. 2020. Situated knowledges: The science question in feminism and the privilege of partial perspective. In *Feminist theory reader.* Routledge, 303–310.

[43] Sandra Harding. 2013. Rethinking standpoint epistemology: What is "strong objectivity"? In *Feminist epistemologies.* Routledge, 49–82.

[44] Kashmir Hill. 2020. The secretive company that might end privacy as we know it. In *Ethics of Data and Analytics.* Auerbach Publications, 170–177.

[45] Melvyn P Leffler and Odd Arne Westad. 2010. *The Cambridge history of the cold war.* Vol. 1. Cambridge University Press.

[46] Stuart W Leslie et al. 1993. *The Cold War and American science: The military-industrial-academic complex at MIT and Stanford.* Columbia University Press.

[47] Helen E Longino. 2020. Science as social knowledge. In *Science as Social Knowledge.* Princeton university press.

[48] David Lyon. 2010. Surveillance, power and everyday life. *Emerging digital spaces in contemporary society: Properties of technology* (2010), 107–120.

[49] Gianclaudio Malgieri. 2020. The concept of fairness in the GDPR: a linguistic and contextual interpretation. In *Proceedings of the 2020 Conference on fairness, accountability, and transparency.* 154–166.

[50] Gary T. Marx. 2015. Surveillance Studies. In *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)* (second edition ed.), James D. Wright (Ed.). Elsevier, Oxford, 733–741. https://doi.org/10.1016/B978-0-08-097086-8.64025-4

[51] Matt Marx and Aaron Fuegi. 2022. Reliance on science by inventors: Hybrid extraction of in-text patent-to-article citations. *Journal of Economics & Management Strategy* 31, 2 (2022), 369–392.

[52] Torin Monahan and David Murakami Wood. 2018. *Surveillance Studies: A Reader.* Oxford University Press.

[53] Mozilla. 2022. Privacy Not Included. https://foundation.mozilla.org/en/privacynotincluded/

[54] Irena Nesterova. 2022. Questioning the EU proposal for an Artificial Intelligence Act: The need for prohibitions and a stricter approach to biometric surveillance. *Information Polity* Preprint (2022), 1–16.

[55] Access Now. 2021. Ban biometric surveillance. *Brooklyn, Access Now* (2021).

[56] Ciara O. 2023. Data watchdogs issued nearly €3bn in fines in 2022. https://www.irishtimes.com/business/2023/01/17/data-watchdogs-issued-nearly-3bn-in-fines-in-2022/

[57] State Council of the People's Republic of China. 2017. Next generation artificial intelligence development plan.

[58] Cathy O'neil. 2017. *Weapons of math destruction: How big data increases inequality and threatens democracy.* Crown.

[59] Amandalynne Paullada, Inioluwa Deborah Raji, Emily M Bender, Emily Denton, and Alex Hanna. 2021. Data and its (dis) contents: A survey of dataset development and use in machine learning research. *Patterns* 2, 11 (2021), 100336.

[60] Kenny Peng, Arunesh Mathur, and Arvind Narayanan. 2021. Mitigating dataset harms requires stewardship: Lessons from 1000 papers. *arXiv preprint arXiv:2108.02922* (2021).

[61] Robert N Proctor, Robert Proctor, et al. 1991. *Value-free science?: Purity and power in modern knowledge.* Harvard University Press.

[62] Inioluwa Deborah Raji and Genevieve Fried. 2021. About face: A survey of facial recognition evaluation. *arXiv preprint arXiv:2102.00813* (2021).

[63] Urbano Reviglio and Rogers Alunge. 2020. "I am datafied because we are datafied": An Ubuntu perspective on (relational) privacy. *Philosophy & Technology* 33, 4 (2020), 595–612.

[64] Neil M. Richards. 2013. The Dangers of Surveillance. *Harvard Law Review* (2013).

[65] Morgan Klaus Scheuerman, Alex Hanna, and Emily Denton. 2021. Do datasets have politics? Disciplinary values in computer vision dataset development. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–37.

[66] Arnab Sinha, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-June Hsu, and Kuansan Wang. 2015. An overview of microsoft academic service (mas) and applications. In *Proceedings of the 24th international conference on world wide web.* 243–246.

[67] Luke Stark. 2019. Facial recognition is the plutonium of AI. *XRDS: Crossroads, The ACM Magazine for Students* 25, 3 (2019), 50–55.

[68] Mojtaba Vaismoradi, Hannele Turunen, and Terese Bondas. 2013. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & health sciences* 15, 3 (2013), 398–405.

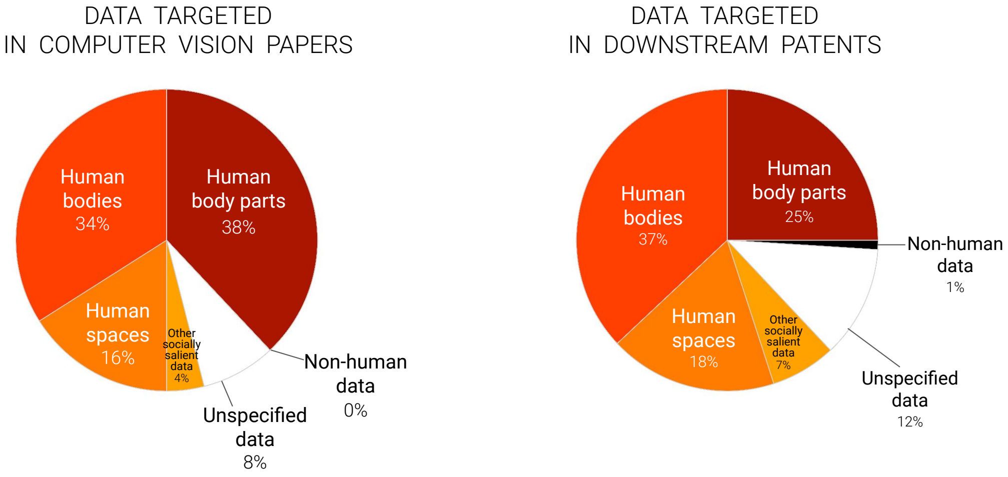
[69] Carissa Véliz. 2021. *Privacy is power.* Melville House New York.

[70] Sarah Myers West. 2019. Data capitalism: Redefining the logics of surveillance and privacy. *Business & society* 58, 1 (2019), 20–41.

[71] Charles Albert Ziegler and David Jacobson. 1995. *Spying without spies: origins of America's secret nuclear surveillance system.* Greenwood Publishing Group.

[72] Shoshana Zuboff. 2019. *The age of surveillance capitalism: The fight for a human future at the new frontier of power: Barack Obama's books of 2019.* Profile books.

Fig. A1. **The extraction of human data in computer vision papers versus downstream patents.**



APPENDIX

A THE TARGETING OF HUMAN DATA IN PAPERS VERSUS PATENTS

In Figure A1 we present a quantitative summary of the types of data targeted in computer vision papers compared to downstream patents. We find similar trends. On the basis of our in-depth content analysis, we find that 92% of papers and 87% of patents extracted data relating to humans. Furthermore, the majority (72% of papers and 62% of patents) explicitly extracted data about human bodies and body parts. In some cases, papers that do not speak to the potential for use with human data led to patents that explicitly report monitoring human data. In general it was more common that papers did speak to targeting human data, and in particular targeting human body parts, and then patents leveraged these papers for overall targeting of human bodies, or even targeting of now unnamed data types. No papers and only 1% of patents were dedicated to targeting non-human data, showing that both computer vision research and applications are similarly concerned with analyzing, tracking, and monitoring humans and specifically human bodies.

B THE TRANSFER OF HUMAN DATA

An additional central, organizing feature of surveillance is the mass collection, permanent storage, aggregation, and sharing of data, frequently without consent or awareness by the target individual, group or community [23]. Regulatory bodies such as Europe's General Data Protection Regulation (GDPR) [1] and the California Consumer's Privacy Act of 2018 (CCPA) [29] have aimed to put mechanisms and regulations in place to ensure and enforce individual and collective privacy rights. GDPR outlines *fair*, *lawful* and *transparent* data collection practices [49], deeming much of the current ubiquitous and aggressive nonconsensual mass data collection, transferal and sharing by surveillance companies/technologies unlawful. Subsequently, surveillance companies such as Clearview AI [2] as well as TikTok and Meta [6] are often found in breach of these data protection rights and face fines. European data regulation authorities for example, issued nearly €3bn in fines in 2022 alone [56]. Still, problematic and unlawful data collection, sharing, and transferal practices have become the norm. From targeted online ads to wide ranging services (including, insurance, retail and finance) to "smart" home devices, future prediction is a core objective of surveillance technology [72], which heavily relies on the vigorous collection, aggregation and transferal of data. Many studies of public attitudes reveal intense concern alongside a need for knowledge regarding the practices of data transferal.

We identified four categories capturing technologies' transferal of human data: *the paper or patent anticipates transferring the data on a wireless connection; the data is transferred to another person or institution; the data is kept entirely locally; and whether and where data is stored or transferred is left ambiguous.* We found that stating data transferal, storage or management information is rarely mentioned in papers but relatively more common and conveyed in patents.

Data is transferred to others

"We developed methods for face recognition from sets of images...of the same unknown individual" (Patent 0)

This category captured scenarios in which data about a person is not guaranteed to remain solely with that person and may instead be transferred to one or more other persons or institutions. An example of this is a home video surveillance system that gives the system administrator access to videos of other persons, and may also share those videos with the manufacturer or other entities, such as law enforcement. In a world of 'data economy' [58] where AI systems are hungry for data, data collected from our digital devices, fitness tracking technology and cameras provide insights about ourselves as well as our surroundings [40]. Rarely, if at all, such data remains under the control of the data subject and is shared with third parties; institutes, data brokers, or other persons. Even when privacy polices are outlined, data is not guaranteed to remain under the control of the person. Examining 211 diabetes apps, Blenner et al. [20], for example, found that of apps with privacy policies, 79 percent shared data while only about half of them admitted doing so. Similarly, a recent review of the privacy and data sharing policies of IoT devices and apps, found that despite restrictions in privacy policies, personal data is aggressively collected, shared and sold to third parties [53].

Data transfer over a wireless connection

"image data...may not be saved in intermediate form, but may simply be "piped"
to a next stage over a bus, cable, wireless signal or other information channel" (Patent 5)

Some patents indicate that image or video analysis will be done in the cloud and illustrate this in diagrams outlining their system. Others do not explicitly mention that their artifact will be used to transfer data to an institution, but described the wireless capabilities of their artifact. In both of the described scenarios we understand these as having the fully and intentionally anticipated capability for wireless data transfer. The collection, aggregation and categorization of data is one of the key characteristics of surveillance and an increasingly lucrative business [23, 72]. Even while appearing everyday and seemingly benign to many, ubiquitously connected technologies are instrumental for documenting, mapping, monitoring and facilitating widespread, networked surveillance. The under-regulated data broker industry and analytics companies, who infer individual features from consumer data in order to predict behaviour are an essential component of the surveillance ecosystem [63, 69, 70]. And, despite diverse understandings of the ideal that ought to be possible with internet and connectivity, in reality all connectivity serves an, at times shockingly productive, venue for data collection, aggregation, analytics, prediction, and ultimately surveillance [72]. According to Zuboff, "Every avenue of connectivity serves to bolster private power's need to seize behavior for profit."

Data remains exclusively local

Surveillance is not mere designing, building and deploying technologies, but is also marked by the struggle for power and control. A tracking technology such as a health monitor, for example, that exclusively remains under the control of a particular person, might serve only that particular user. This can potentially include papers and patents where all data collected is guaranteed to be kept and processed entirely at the control of the data subject, for example, on a personal server. Because this is entirely possible, we included this deductive code: the inclusion of this category served to actively search for and document any possible technology aimed placing total agency in the hands of the end user; however, *none of the papers or patents fell into this category.*

Unspecified

Data transferal or storage information is sometimes undisclosed in patents and is rarely stated in papers. Note that this label does not prevent or limit any data from being transferred to others. Instead, it means that the work does not specify where or how the data is stored, shared, or transferred. Given that surveillance technologies tend to operate in the dark where technology vendors take extra measurements to hide their existence [21, 44], opacity in these category of papers and patents can signify purposeful obfuscation.

C THE INSTITUTIONAL USE OF DATA

Surveillance is not mere passive observation but also extends on some capacity to control, regulate, or modulate behaviour [52]. These can be seemingly invisible influences, for example limiting choices or opportunities or directing people towards certain decisions (and away from others) through, for example, recommendation or personalization tools developed by big corporations. This form of influence and behaviour modulation is subtle and at times not recognized. Other times, papers and patents present a relatively direct surveillance application of their technology, where data is transferred and controlled by institutions, such as state and military bodies for the purpose of exercising power. Typically, surveillance technologies and norms are implemented and practised as convenience and a solution to "efficiency, productivity, participation, welfare, health or safety" whereby social control is framed as an unintended consequence [23]. We identify three key data uses described in papers and patents.

Modeling or categorizing humans

The methods proposed in these works attempt to make humans amenable to modeling and categorization. This form of surveillance might be to used collected data to generate models of humans without specifying the intended use case for these models. An example of this could be pedestrian detection with no explicit purpose. Alternatively this form of surveillance might explicitly model and categorize to facilitate soft influence or hard control of humans.

Soft influence

"Applications including...real-time language translation,
online search optimizations, and personalized user recommendations" (Patent 35)

Soft influence includes online targeted ads, recommendations, and other forms of personalization. While surveillance used to exert soft influence does not require people to carry out certain behaviors, it can *coerce* them towards behaviors that the surveillor believes are desirable and constraining the other options available to them. An example of this is gaze tracking in mixed reality environments for targeted advertising. Note that while we use the term "soft" relative to the category of "hard control," such influences can have significant, life-changing harms. For instance, personalized recommendations and advertisements can and have targeted vulnerable populations with misleading products and excluded marginalized communities from opportunities for employment, credit, and housing [58].

Hard control

"Applications include...assisting in automated patrol of large uncontrolled border crossing areas,
such as the border between Canada and the US and/or the border between Mexico and the US." (Patent 5)

Patents (more so than papers) often state exemplar applications of their technologies where data is transferred to institutions of power. Examples of applications mentioned include tracking and identifying people for the purpose of border surveillance (for example, restricting movement), recognizing and tracking vehicles in cluttered urban scene using autonomous drones for the purpose of law enforcement, and detecting “anomalous” and suspicious activities.

This topology, which is developed and confirmed through extensive manual analysis of papers and downstream patents, contributes a picture of the *how* and *why* of Surveillance AI by laying bare the work creating and sustaining surveillance technologies.

D ADDITIONAL ANALYSIS OF THE INSTITUTIONAL ROOTS AND EVOLUTION OF SURVEILLANCE AI

In Figure A2, we analyze institutions, and we list the proportion and quantity of papers that are used in surveillance patents versus used in only non-surveillance patents. We find an overwhelming pattern that when an institution’s paper leads to downstream patents, a majority of those papers are used in surveillance patents. We also observe a long tail effect, where a handful of institutions produce the bulk of papers used in patents (and thus also the bulk of papers used in surveillance patents). Analyzing the entireity of the corpus, for 81% of institutions and 90% of nations that author papers with downstream patents, at least half of these papers are used in surveillance patents

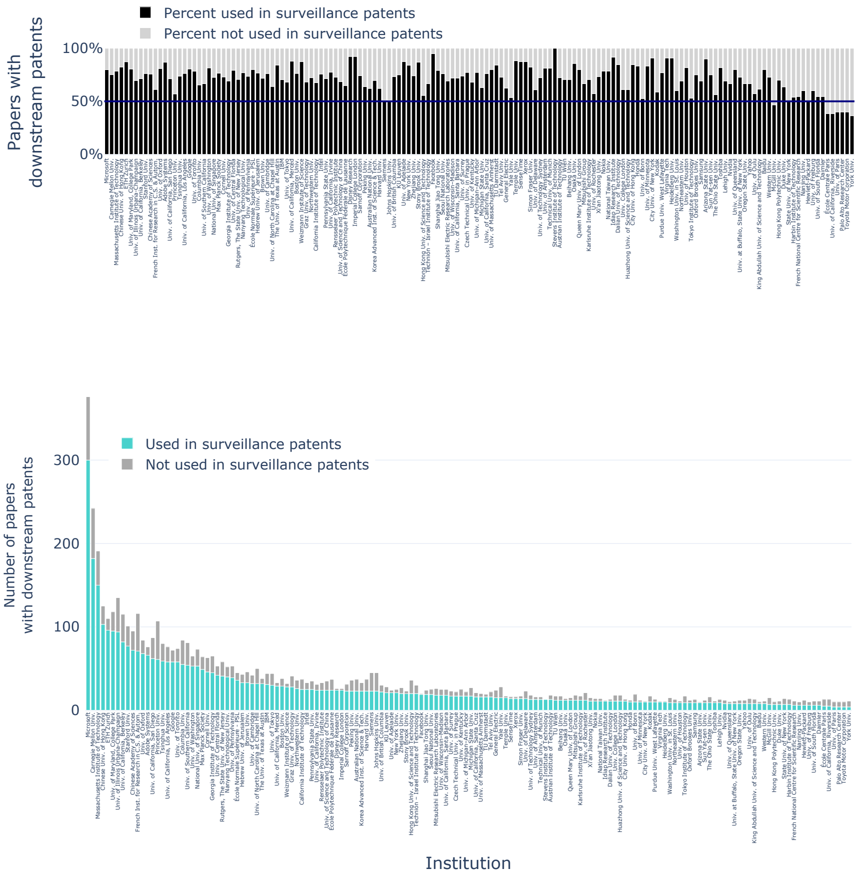
E SURVEILLANCE ACROSS SUBFIELDS

Computer vision is a broad field comprised of several subfields, such as object detection, medical imaging, and 3D reconstruction. Although some research in these subfields is more recognizable as surveillance (e.g., facial recognition) than others, we found that research across many subfields has contributed to the creation of Surveillance AI. For 78% of subfields that author papers with downstream patents, at least half of these papers are used in surveillance patents. Figure A3 displays the distribution of surveillance-related research papers across different subfields. As can be seen in Figure A3, there are several topics explicitly related to the tracking of human bodies, such as “face detection” and “motion detection.” Yet interestingly, many of the top subfields have no explicit relation to the modeling of human data and instead are simply common topics, like “background subtraction” and “computer graphics.” The wide distribution of topics across papers cited in surveillance patents reveal that the work of many subfields, even ones not explicitly connected to human data, have contributed to Surveillance AI.

F ADDITIONAL BACKGROUND ON SURVEILLANCE AND COMPUTER VISION

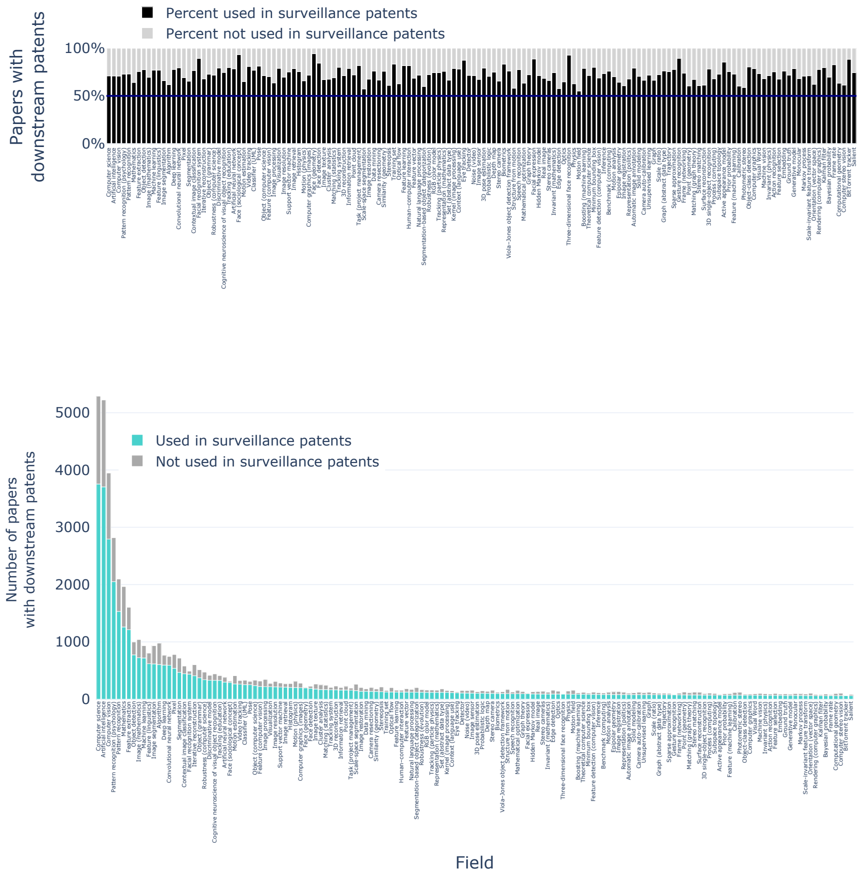
Surveillance. Surveillance is a technology of social control intrinsically tied to the production of power relations. The observing, tracing and monitoring is practised by those in a relative position of power to those being observed. The enactments of surveillance frequently reify boundaries, borders, and bodies along racial lines, the consequence of which is often discriminatory treatment of individuals and communities that are negatively portrayed, which Browne terms “Racializing surveillance” [23]. Surveillance perpetually influences its subjects in making them more “amenable to observations, prediction, and suggestion” [26]. Surveillance practices produce social norms and standards and exercise the “power to define what is in or out of place” [23]. State powers and military institutions track, monitor and profile citizens, immigrants, “offenders” or “suspects”; companies watch and monitor their employees; tech corporations track, sort and profile users; education institutes track and monitor their pupils, often with the justification of enhancing security, productivity, safety, or efficiency. As most surveillance technologies are designed, developed, and deployed by and for institutions of power as the paying customers and primary stakeholders, the safety, welfare, and interest of individuals and communities where these technologies are deployed are an afterthought, if considered at all. As a result, while institutions of power benefit the most from the production and deployment of

Fig. A2. **Institutions producing computer vision research with downstream patents.**
All institutions that have produced at least 10 computer vision papers with downstream patents are listed. *Top:* The percent of these papers used in surveillance patents. *Bottom:* The number of these papers used in surveillance patents. Institutions are listed in order, beginning with the institutions authoring the most research used in surveillance patents. There is a clear norm that, of an institution's papers with downstream patents, most will be used in surveillance patents.



surveillance technologies, communities subjected to surveillance (often communities at the margins of society) are disproportionately negatively impacted [15, 23, 35, 48]. Current facial recognition technologies used in law enforcement, for example, disproportionately negatively impact racial minorities. In the context of US law enforcement, for example, facial recognition surveillance has so far led to at least four wrongful arrest, all of whom are Black men [39], facilitating and expanding racialized carceral systems. Relational and collective conceptions of surveillance are therefore critical for a comprehensive understanding of surveillance that can account for power asymmetries that permeate the surveillance ecology.

Fig. A3. **Computer vision subfields producing papers with downstream patents.**
Computer vision subfields producing at least 10 papers with downstream patents. *Top:* The percent of these papers used in surveillance patents. *Bottom:* The number of these papers used in surveillance patents. Subfields are in order, beginning with the subfields authoring the most research used in surveillance patents and listing the first 150 subfields. There is a clear norm that, of a subfield's papers with downstream patents, most will be used in surveillance patents.



Computer Vision. The emergence of the World Wide Web and with it, the 'availability' of vast amount of image and video data, has been a central contributing factor for the rapid rise of the field in the past decade. Critiques have been formulated that draw attention not only to the histories, but the encoded values and ongoing practices within the field. Dataset collection, curation and management practices, in most cases remain devoid of careful considerations of issues such as informed consent, privacy, or dataset audits (for example, to mitigate negative social stereotypes often encoded in data) [19, 60]. Dataset collection and curation practices in computer vision are compared to the ethical equivalent of data theft [59] and erode privacy with most data collected without informed consent or procedures to opt-out [19, 59]. Dataset collection, documentation, and development in computer vision are driven by the underlying values of efficiency, universality, impartiality, and model work. Scheuerman et al. [65] et al. further note that "Efficiency is valued over care, a slow and more thoughtful approach to dataset curation. Universality is valued over contextuality, a focus on more specific tasks, locations, or audiences. Impartiality is valued over positionality" based on extensive analysis of canonical image datasets. Furthermore, the rapid rise and accessibility of generative models such as Stable Diffusion, not only exacerbates the erosion of privacy and the reproduction of social stereotypes, toxic and discriminatory predictions, the proliferation of these generated images at a massive scale also pollutes the digital ecology [17]. Against the backdrop of the rapid rise of computer vision and the growing array of critiques of the field, it is crucial we understand the extent and nature of the flow from the field's histories, values, and practices to major downstream applications such as surveillance.

G ADDITIONAL METHODOLOGICAL DETAILS

G.1 Data

Throughout our studies, we analyze the corpus of CVPR papers from 1990-2020 and their downstream patents. In 1990, 1995, and 2002, CVPR did not occur, so there are no papers from these years. For the analysis across years (displayed in Figure 5), we filter the corpus years, for reasons described here. In emerging and developing fields, the estimated time from a paper being published to a downstream patent being published is three to four years; this is the time from the paper being published to the downstream patent being filed as well as the time of the patenting process [37]. This appears 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 onward remained above 200 every year, until 2018 (exactly four 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 the years 2018 and 2019 since these were less than four years before our analysis began so many papers had not yet had the chance to be significantly patented leading to less reliable analysis. This had the added benefit that, in our analysis comparing the 1990s to the 2010s, both decades consisted of 8 years, putting these decades on a fair playing field for totaling when comparing the number of downstream patents of various types.

We used the paper-patent linkages inferred by Marx and Fuegi [51] to identify connections between papers and patents; manual verification found these linkages to have over 99% precision and 78% recall.

G.2 Automated analysis

We searched the abstract and body of each patent for the following surveillance indicator words: "ad", "advertisement", "airport", "apartment", "army", "baggage", "caste", "citizen", "combat", "convict", "crime", "criminal", "defense", "disability", "enemy", "ethnicity", "face", "facial", "facial recognition", "felon", "female", "foot traffic", "fraud", "friend", "gender", "geolocation", "hand", "iris", "irises", "jail", "kid", "license plate", "limb", "male", "man", "military", "nonbinary", "office", "pedestrian", "penitentiary", "prison", "prisoner", "purchase", "recommend", "reidentification", "security", "sex", "sexuality", "social network", "street", "surveil", "surveillance", "torso", "transgender", "underage", "woman", and "youth".

While we in general searched the patent abstracts and patent bodies, limitations of our parser resulted in, for a small number of patents, being able to obtain and search only the abstract. Our keyword counts thus constitute lower bounds, and the prevalence of surveillance is likely to be even greater than that which we document.