

Speech in the Panaudikon: On the Relationship between Speech Processing Research and Surveillance

Vasilisa Rudykh, Peter Bell, Zeerak Talat

University of Edinburgh
Edinburgh, United Kingdom
v.rudykh@sms.ed.ac.uk, peter.bell@ed.ac.uk, z@zeerak.org

Abstract

The potential of using speech technologies for surveillance purposes are tacitly acknowledged within the field of speech processing, however there has been little research on how theories from surveillance studies interact with speech processing research. In this paper, we analyse 20% of papers published in 2025 from Interspeech, one of the main venues for speech research, through Steve Mann’s compass of veillances. We find that engagement with considerations of surveillance is almost entirely absent in the literature, despite risks being readily apparent in a number of research tasks, with some tasks being more prone to risks of surveillance. Encouragingly, we find tasks and findings that provide avenues for counter-surveillance research, yet these introduce other risks. We argue that identifying mechanisms for addressing the risk of surveillance requires discussion, and therefore provide a framework for researchers to identify how their work engages with surveillance, and how they can limit the risks of it. To this end, we present a list of questions for researchers to engage in how their work might contribute to surveillance.

1 Introduction

Surveillance is a factor that touches upon many of our daily lives, from being recorded by closed-circuit television (CCTV) in public spaces (Mann 2016) with and without facial recognition technologies, to the content we post on online having consequences for our ability to obtain visas (Brown 2014), to the general tracking of our online footprint (Puglisi, Rebollo-Monedero, and Forné 2016). Yet, statements that we have voiced have been less subject to widespread surveillance, in part due to the difficulties in processing speech in the wild (Zhao et al. 2025). In fact, when Amazon made headlines in 2021 for its Rekognition system being deployed for transcribing phone calls from prisons in the United States of America, the company’s spokespeople acknowledged that “heavy accents or poor audio quality can lead to variations” (Sherfinski and Asher-Schapiro 2021). Recently, speech processing technologies have seen rapid improvement and deployment. From speech-based assistant systems such as Amazon Alexa and Apple Siri, to speech transcription services included in video-calling software and speech synthesis systems—such as those provided

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

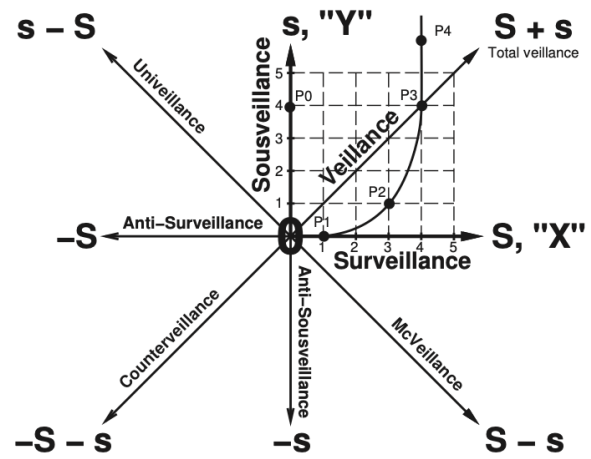


Figure 1: Steve Mann’s 8-point Veillance Compass (Mann 2013). Reproduced with license.

by large language model providers—speech technologies are increasingly being embedded into our society.

Historically, surveillance has required painstaking manual data collection and processing labour, improvements in machine learning systems thus present an avenue for more efficient data labour, and thereby more efficient surveillance mechanisms. While the prospect of improved surveillance capabilities may raise concerns for many, for some, it is a “vision” to be fulfilled. In January 2026, the United Kingdom’s Home Secretary, Shabana Mahmood, stated:

“[M]y ultimate vision for that part of the criminal justice system was to achieve, by means of AI and technology, what Jeremy Bentham tried to do with his Panopticon. That is that the eyes of the state can be on you at all times [...] I think there’s big space here for being able to harness the power of AI and tech to get ahead of the criminals” – *Shabana Mahmood* (as quoted in Morrison 2026).

While risks of dual use of speech technologies have been explicitly discussed around synthesis (Hutiri, Papakyrakopoulos, and Xiang 2024; Burgess, Carlon, and Doyuran 2025), research around surveillance has centred questions of

privacy-preservation (Tomashenko et al. 2026) and little attention has been given to how research areas in speech processing engage with surveillance—where privacy focuses on maintaining privacy of personal information and surveillance attends to the processing of personal information. This is concerning because (i) even where text can more easily be processed, speech signals carry more information about the speaker, their emotional state, sex, and other characteristics (Al-Ali et al. 2025); (ii) it is unclear which research areas of speech processing elicit risks of surveillance, which risks are surfaced, and how to address them; and (iii) the lack of a holistic overview of the risks of the field erects barriers to public participation in democratic processes—including resistance campaigns such as the campaigns against facial recognition which have resulted in restrictions in their use in 16 states and outright bans in 15 cities in the United States of America (State of Surveillance 2025).

In this paper, we seek to close this gap by analysing speech processing papers for how they engage with questions of dual use and surveillance. Drawing on a theory of *veillance*, we annotate 229 (20%) of papers published at Interspeech 2025. Discouragingly, we find a dearth engagement with dual use and risks of surveillance. Specifically, only three papers mention dual use and only two papers mention surveillance in our sample. Still, highly encouragingly, we find several areas of speech research that affords counters to surveillance, including text-to-speech, voice conversion, and privacy-preserving methods such as speaker anonymisation. Further, we provide an in-depth discussion around the nuances of surveillance. Although questions of surveillance are often presented as clear-cut, the potential for speech synthesis technologies complicates this view: on one hand, synthesis may provide additional mechanisms for privacy in the context of surveillance, yet the very same tools may also further perpetuate harm to individuals, e.g. by providing mechanisms for scammers to remain undetected. We argue that rather than seeking to limit, prohibit, or encourage certain areas of research, a more appropriate method would be to encourage discussion around how work might engage with surveillance, and provide guidance for mechanisms towards this. Finally, the paper provides concrete recommendations for the field and a set of questions for researchers to interrogate how their work engages with surveillance and how risks might be mitigated.

Paper Structure In the remainder of this paper, we provide an overview of work on surveillance (§ 2). Then we present our data collection and annotation methodology (§ 3). We then consider the outcomes of our analyses (§ 4), and discuss on the nuances of surveillance in speech processing research (§ 5). We close the paper with recommendations for the field and questions to help researchers reflect on the surveillance capacity of their work (§ 6).

2 Background

Compass of Veillances

Surveillance practices have long existed as a way to exercise control over marginalised groups, from the Roman Empire's provincial information collectors, the *Frumentarii* (Sinnigen

1962), to the enslaved people during the 18th and 19th century United States of America (Browne 2015) to modern day surveillance enabled through CCTV (Mann 2016) and the internet (Fuchs et al. 2013). In *Dark Matters: On the Surveillance of Blackness*, Browne (2015) argues that not everyone is equally subject to surveillance. Drawing a history from the transatlantic slave trade to modern biometrics, she argues that marginalised people, and Black people in the United States in particular, have been subject to an othering through technological means—first through the distribution of news and wanted ads in which enslaved people who had escaped were described as property bearing distinct markers—such as scars—which identified them; then through lantern laws which sought to make hyper-visible the Black people—limited to three people per lantern—who were out after dark; and now through modern biometrics. Rather than function purely as mechanism for being watched, surveillance of Black people and marginalised communities have served as mechanisms of control, to limit their movement and their ability to organise (Browne 2015). But, have other forms of looking and exercising of control occurred? During the French Revolution, citizens would routinely attend discussions of the national convention in an effort to exert power over the legislature (Maslan 1995). Similarly, in modern times, mobile phone recordings of law enforcement officers have functioned as mechanisms to mitigate unlawful conduct by the officers (Andrejevic 2011). What we can observe is that (i) the act of watching applies a force to an individual and (ii) watching can be done from a position of power—e.g., through CCTV—or as a means of holding power to account (Mann 2016)—e.g., through citizen recordings.

In his consideration of computer vision and its potentials for misuse, Steve Mann developed his compass of veillances (Mann 2013). In this theory he outlines eight axes of the act of seeing, arguing that each form of axes gives rise to different politics and different relationships between actors (see Figure 1): *Surveillance* refers to a third-party watching “from above” (Mann 2016), typically referring to the state or an actor in a position of power. *Surveillance* functions as a way for the state to enact power, and to limit the ability of the surveilled to act freely. Notably, in Bentham's (Bentham 2010) writings on the Panopticon—a prison concept in which imprisoned people can be under constant surveillance—he noted that the purpose of the design of the prison was to instil the experience of being watched as a mechanism for ensuring that prisoners did not break the rules of the prison at any point. In contrast to surveillance, *Sousveillance* refers to a public watching “from below”, referring to practices such as citizen's recordings. *Sousveillance* describes a mechanism through which the public is empowered against the state by using recordings as a mechanism to hold the state, and its actors accountable. However, as Jansen (Andrejevic 2011) argues, for *sousveillance* to be effective, there must be some expectation of accountability and a realisation thereof. Without either, any recording of the agents in power is rendered meaningless in ensuring they act in concordance with the laws or rules they are subject to. *Unveillance* refers to the ability of a first-party in a communicative context to engage in a record-

ing of another first-party. For example, in an interaction between parties that are equal amongst each other, either party can engage in univeillance by recording the other part (regardless of the other's knowledge or consent). *Veillance*, in Mann's theory, refers to the general act of monitoring and encompasses sur-, uni- and sousveillance. In opposition to monitoring, the compass contains *Anti-surveillance*, which refers to methods by which one limits the ability of surveillance to function, either by disabling the surveillance mechanisms themselves or by rendering oneself invisible to them. *Anti-sousveillance* on the other hand refers to any methods which prevent sousveillance from occurring—e.g., police covering up identification information (badge numbers) or covering their faces with masks to prevent identification. *Anti-sousveillance* becomes *McVeillance*—being recorded while one is prohibited from recording—when it is combined with surveillance. Often such prohibitions are instantiated in policy—such as in airport passport security lines where individuals are not allowed to record images or photos while being under the surveillance of airport cameras.

Counterveillance refers simply to the ability to prevent being watched or recorded, and covers anti-surveillance, anti-sousveillance, and *McVeillance*.

Affordance Theory

In order to assess what particular form of veillance a given paper gives rise systematically, we turn towards the concept of *affordances* (Gibson 1983). The concept of affordances seeks to describe how humans are able to sense how to engage with and interact with our environment such that it is conducive to our goals and purposes. It purports that objects afford certain outcomes more than others in specific environments. For example, a small pebble near a lake has the affordance of being thrown across the surface of the lake. A large rock, however, does not have the affordance of being possible to skip across the surface of a lake. Similarly, the same small pebble may function as a poor tool breaking other stones, whereas the larger rock, with its larger surface area, is a better candidate for breaking stones. That is, where the small rock affords skipping, in the context of a human by a still lake, the larger rock affords being used by a tool-making animal as a tool for breaking other rocks, as a function of its size. More precisely, the affordance of an object refers to the action potential of the object, i.e., what an individual with a particular purpose can do with the object.

In terms of speech technologies and their capacity for surveillance, the affordance of a research object can be understood as what a person can do with the object without significant technological reconfiguration (Majchrzak 2013). For example, an ASR system developed for language **A** may be possible to deploy for language **B**, however the transcripts that would arise from it are likely to be riddled with errors. We can therefore say that our ASR model trained for language **A** affords transcription for language **A** more so than it affords transcription for language **B**.

In terms of Mann's axes of veillance, we consider how a given paper affords different axes of veillance. For example, a lightweight ASR system such as the model proposed by Kong, Huang, and Ou (2025) affords sousveil-

lance and univeillance more than it affords surveillance, because its compute requirements are low enough that it can be developed and used with minimal resources, thereby allowing a technically-minded person the ability to set up the system, develop it, and run it locally to process whatever speech they may wish to. *Sousveillance* and *univeillance* may often have overlapping affordances, but are distinct along axes of power. For example, while an ASR system for children's speech affords surveillance—by the hospital recording—and *univeillance*—through recording from another figure of authority—yet does not afford *sousveillance* as recording does not place children in a position of power.

Speech (and Surveillance)

Prior work in speech has addressed different aspects of privacy and surveillance. Privacy-focused research has addressed concerns such as voice biometric misuse and data protection (Bäckström 2025), where surveillance is considered as a risk associated with privacy breaches. Existing work has also explored how audio processing can enhance surveillance tasks, including audio event detection, monitoring, and recognition (Crocco et al. 2014). Critical studies have explored how voice data can become a source of identification, profiling, and both commercial and state surveillance (Palumbo and Prey 2024; Couldry and Turow 2022). However, these discussions primarily consider the applications and risks of speech technologies rather than examining how advances in speech processing research enable different veillances, including but not limited on surveillance.

Machine Learning and Surveillance

Machine learning methods are particularly well-suited for surveillance purposes as they are developed to organise and sort information—across and within different modalities of data—at scale. As improved machine learning methods—e.g., in its capability in identifying patterns, processing data at greater scales, or act on new data modalities—are introduced, often for the purposes of developing better representation of data and afford better performance on downstream tasks, they also become better tools for surveillance: the ability to process, organise, and index large amounts of data affords the scaling of surveillance. Indeed, the rise of mass video surveillance is tightly connected to the scientific progress in computer vision (Kalluri et al. 2025). Notably, given that performance of machine learning models, particularly deep neural networks, is heavily dependent on large amounts of data and substantial computational power, there is a noticeable imbalance in who can practice and strengthen mass surveillance and who cannot. Turning to the question of how machine learning distributes power (Kalluri 2020), the computational and the data needs of machine learning function as processes which centralise power (Mumford 1964). For example, institutions who practise surveillance are able to collect and store large amounts of data for model development and have the means to access a large number of powerful GPUs for computing; larger datasets improve the performance of models used for surveillance, which in turn affords more precise surveillance. The collection of these datasets is not always ethically justified. For instance, data may be

collected without the data subject’s knowledge or consent. For a discussion on the ethical implications of large-scale data collection, we refer the reader to Ungless et al. (2024) for language model-related considerations, Leschanowsky et al. (2025) for speech processing (speaker recognition), and Prabhu and Birhane (2020) for computer vision.

3 Methodology

Data Collection

We collected our initial sample of all 1,147 paper retrieved using the SemanticScholar Search API, published at Interspeech 2025.¹ This sample includes papers and four keynotes. Removing the keynotes, we are left with 1,143 papers as our initial sample. We select papers published at Interspeech for two reasons. First, Interspeech is a premier venue for speech technology and processing research, and the work published at Interspeech covers a breadth from linguistically focused research (e.g., Oh, Kim, and Chung 2025), to more signal processing oriented work (e.g., Kim and Chung 2025), and to system demonstrations (e.g., Dewhurst et al. 2025). Second, the themes of Interspeech over the past few years have centred inclusion and ethics, thus encouraging work to address questions of speech technologies’ social impact. From our initial sample of 1,143 papers, we select 20 for validating our annotation schema, and 229 (20%) for annotation.² After the first round of validating our schema, we select 15 papers from the 229 and conduct a second validation round.

Data Annotation

Two of the authors engaged in the annotation process, each annotating an independent sample from one another. To ensure consistency in annotation, we compute annotator agreement on 35 samples that were jointly annotated by both annotators, and conducted regular check-ins between the annotators to ensure that any questions and differences in understanding were identified and resolved. We annotated papers along the 12 dimensions:

Author Affiliation: We first recorded the author affiliations, and conduct a second pass over papers to group affiliations into academia, industry, government, and research institutes.

Languages Covered: Where research deals with speech signals, rather than room noise, echo, and other environmental factors, we record the languages that the research covers.

Tasks: We conduct a fine-grained annotation of the task that papers are conducting (e.g., automatic speech recognition (ASR) for dysarthric speech) on the basis of the description in the paper. Once we had concluded the annotation, we conducted a second pass, to group the tasks into one or more coarse grained categories through an inductive coding.³

¹We conducted the search on the 5th of January, 2026 using https://github.com/zestlaboratory/Search_Interspeech.

²We ensure that there is no overlap our schema validation sample and our annotation sample.

³We identify the following task categories: *ASR, Speech Translation, Speaker Identification, Speech Synthesis, Speech Enhancement, Anti-Spoofing, Emotions or Traits, Clinical Settings, Multimodal Speech Decoding, Speech Representation, Dialogue Sys-*

Datasets Used: We record the datasets that are included for training, evaluation, or analysis.

Use of Biomarkers: We note if biomarkers are used in the study, with the following options: “models biomarkers”, “predicts biomarkers”, “other”, and “N/A”.

Artefact Release: We label information about how research artefacts are distributed into the following categories: “open-source”, “closed source”, “restricted access”, “not mentioned”, and “N/A”. We choose the label on the basis of what is mentioned in the papers, or if a link is provided to a website or repository for distributing the research artefacts.⁴

Dual Use Considerations: We select one or more categorical labels for dual use considerations from: “Discusses Dual Use”, i.e., instances where concrete dual use concerns are raised and discussed; “Mentions Dual Use” which covers instances where concrete or abstract mentions of dual use without further discussion, e.g. “However, these techniques also pose significant risks, particularly in the spread of highly realistic misinformation” (Kim et al. 2025); “No Mention” where there is no mention of dual use; or “N/A” where dual use considerations are not relevant, e.g., in papers that perform evaluation of pre-existing methods.

Surveillance Considerations: We also categorically label how surveillance is treated in papers, allowing selection between: “Discusses Surveillance”; “Mentions Surveillance”; “No Mention”; or “N/A”.

Mann’s Veillance Compass Axes: We label each paper according to Veillance Compass Axes from Mann (2013), with the exception of Veillance, because in practice it exclusively co-occurs with *sur-* and *sousveillance*; *Counterveillance* because it similarly exclusively occurs along instances of *anti-sousveillance* and *anti-surveillance*; and *McVeillance*, as technologies by themselves do not set policies that allow some to record while limiting others. The final set of axes that we label for are: *surveillance*, *sousveillance*, *univeillance*, *anti-surveillance*, *anti-sousveillance*, and *N/A*. We label each paper according to the compass areas by drawing on affordance theory (Gibson 1983). This approach means that two papers for ASR may be labelled differently. For example, a paper which seeks to develop high performing ASR may be labelled as *surveillance* and not *sousveillance* if it requires a large number of GPUs to train and/or run. A paper may be labelled for multiple axes of veillance, for instance where a paper that affords *anti-surveillance* also affords *anti-sousveillance*.

In our consideration of veillances, we treat the processing, rather than the recording, of speech as the seeing act. We therefore disregard the question of temporality, allowing for the processing of speech to occur at a different point in time than the recording. Furthermore, we assume that results

tems, Environmental Monitoring, Research Analysis, and Other.

⁴We take the authors words at face value and do not check if the contents of the webpages or repositories contain the research artefacts.

Annotation category	Type	Options (if categorical)
Author Affiliation	Free Text	
Author Affiliation Type	Categorical	Academia, Governmental, Industry, Research Institute
Languages Included	Free Text	
Tasks	Free Text	
Task Category	Multi-label	ASR, Speech translation, Speaker identification, Speech synthesis, Speech enhancement, Anti-spoofing, Emotions or Traits, Clinical settings, Multimodal speech decoding, Speech representation, Dialogue systems, Environmental monitoring, Research analysis, Other
Datasets Used	Free text	
Use of Biomarkers	Multi-label	Models biomarkers, Predicts biomarkers, Other, N/A
Open-Sourcing	Categorical	Open-source, Closed source, Restricted access, Not mentioned, N/A
Dual Use	Categorical	Discusses dual use, Mentions dual use, Not mentioned, N/A
Surveillance	Categorical	Discusses surveillance, Mentions surveillance, Not mentioned, N/A
Veillance Compass Areas	Multi-label	Surveillance, Anti-surveillance, Sousveillance, Anti-sousveillance, Univeillance, N/A
Other Risks	Free text	
Limitations	Categorical	Discusses limitations, Mentions limitations, Not mentioned, N/A
Other Funding Source	Free text	

Table 1: Overview of annotation schema.

Course-grained Category	Fine-grained Category	% Agreement
Veillance Axes	anti-sousveillance	91.43%
	anti-surveillance	85.71%
	sousveillance	68.57%
	surveillance	71.43%
	univeillance	68.57%
	N/A	85.71%
Biomarkers	Models biomarkers	85.71%
	Predicts biomarkers	91.43%
	Biomarkers: Other	97.14%
	N/A	80.00%
	Open Source	85.71%
	Dual Use	100%
	Discusses Surveillance	100%
	Risks	91.43%
	Limitations	77.14%

Table 2: Annotator agreement.

reported in papers are reliable as the aims of our paper is to examine engagement with questions around surveillance.

Other Risks While surveillance and dual use are a particular set of risks, we here capture other risks of harms that are raised by research papers. We categorise papers into “Discusses Risks”; “Mentions Risks”; “No Mention”; or “N/A”.

Limitations All research methods have their limitations, some of which may relate to surveillance or the potential for the work to engage in surveillance. We therefore extract limitations according to the categories “Discusses Limitations”; “Mentions Limitations”; or “No Mention”.

Funding Source: We extract mentions of different funding sources that are listed in the acknowledgements.

Where dual use, surveillance, limitations, and other risks are at least mentioned, we also extract the relevant segments of

the papers (for an overview of our categories, see Table 1).

Annotator Agreement We randomly sample 20 papers from the initially collected data sample—which are not included in the 229 papers randomly sampled for annotation—and another 15 from our sample of 229 papers for annotation, resulting in 35 shared papers for the annotators. We compute percentage-wise agreement on all categorical and multi-label dimensions (see Table 2)—with the exception author affiliation type, and task category—on all 35 samples used for validation. We exclude author affiliation type and task category, as these only arose as a necessity for analysis and therefore have not been subject to multiple annotations. For the *Veillance Compass Axes*, we compute agreement for each axis individually.

We find substantial agreement between the two annotators, with perfect agreement around discussions of dual use and surveillance. The lowest agreement scores are for sousveillance and univeillance; however in discussion of annotation differences, we found that the difference was caused because one annotator did not label the least afforded veillance, if they perceived a difference in their affordances, whereas the other annotated all types of veillances that were afforded, regardless of differences in how much they were afforded. For example, in the disagreement regarding whether a paper producing a model for *Speech enhancement* in noisy conditions should be tagged as univeillance, one annotator’s position was that it affords all kinds of veillances equally, while the other annotator argued that it is less likely for one-to-one setting to be noisy, thus it affords surveillance and sousveillance more than univeillance. We retain the disagreements in our data to allow further studies of how speech processing can afford surveillance.

Annotator Information The annotators involved in this work are both authors of this paper, and were not compensated for their work. They are both in their 30s, are speakers of low resource languages, and have received formal training

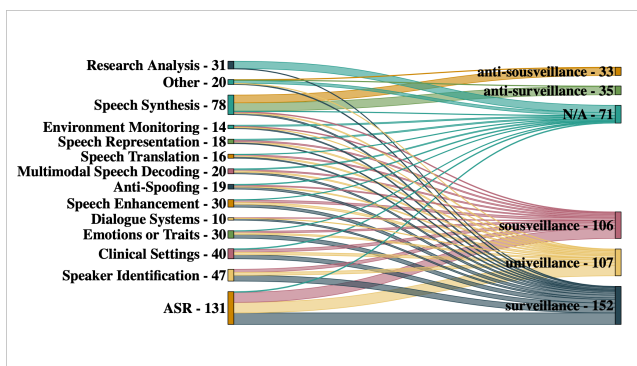


Figure 2: Sankey diagram showing the relationship between task categories and axes (including N/A) from Veillance Compass. Papers may belong to multiple task categories and axes; link widths indicate the number of papers per category-axis pair.

in computer science, and are currently working in research positions. They identify as a cis woman and non-binary person, respectively. The former is currently undertaking a research degree, while the latter is junior faculty.

4 Results

Tasks and Veillances

In our analysis of the relationship between tasks and veillances (see Figure 2), we find that 56% of all papers in our dataset are labelled as *surveillance*, yet only 2 papers (< 1%) in our annotated data mention surveillance at all, both of which only mention surveillance so in passing, with reference to how their work affords surveillance. For example, Liang et al. (2025) note that:

[lipreading] also plays an important role in silent secret communication scenarios and long-distance surveillance.

Calling to attention the negative potential of their work while Tao, Jia, and Hu (2025) call attention to

Accurate knowledge of source positions contributes to a range of practical audio-only applications, such as robot navigation, cochlear implants and intelligent surveillance systems

highlighting the positive potential for surveillance systems.

The discrepancy of proportion of papers that mention surveillance and the proportion tagged with surveillance is even more stark when we consider that we map 170 papers (74%) of the our annotated data to one or more axes of the veillance compass. Considering Figure 2, we can see how tasks distribute to the different the veillance axes: *Environment Monitoring*, *Speech Representation*, *Speech Translation*, *Multimodal Speech Decoding*, *Anti-Spoofing*, *Speech Enhancement*, *Dialogue Systems*, *Emotions or Traits*, *Clinical Settings*, *Speaker Identification*, and *ASR* all exclusively contribute to *sur*-, *sous*-, and *univeillance* of the compass axes. Of these, *ASR* is the largest contributor, contributing with 50 out of

53 papers on *ASR* to *sur*-, *sous*-, and *univeillance*. Within *ASR*, three papers are labelled as N/A for surveillance, two of which are focused analyses of *ASR* (Serditova, Tang, and Steffens 2025; Pratap Singh, Sahidullah, and Kinnunen 2025) while the third focuses on voice conversion-resistant watermarking (Haiyun et al. 2025). Therefore, these papers do not produce any models or data, which does not allow for them to contribute towards any type of veillance. Of the coarse-grained tasks, we find that only the *Speech synthesis* and *Other* categories contribute towards anti-surveillance and anti-sousveillance.

Within the categories of anti-surveillance and anti-sousveillance, the majority of papers describe efforts towards various forms of speech synthesis, tasks such as voice conversion, text-to-speech (TTS), and spoofing. These tasks all share the feature that they afford concealing identifiable markers of ones own voice, while still producing speech. Famously, the activist-hacker collective “Anonymous” have used speech synthesis systems to avoid detection and arrest. Speech synthesis tasks have recently come under greater scrutiny for the risks of social harms associated with them, highlighting the importance of taking a nuanced approach to discussing the benefits and harms of different areas of work in speech technology research.

Task-based Analysis As previously noted, not all tasks give rise to the same set of risks. Here, we provide a fine-grained analysis of what each task category gives rise to.

ASR: An important component of surveillance is not just recording, but also the ease with which monitoring can occur. *ASR* affords making speech more easily searchable and thereby afford easier monitoring and indexing of the contents of speech and thus affords surveillance, sousveillance and univeillance.

Speech Translation: Where *ASR* provides easier searchability, speech translation affords cross-lingual monitoring of the contents of speech.

Speaker Identification: Speaker identification tasks affords tracking individuals, and disambiguating their speech from others.

Speech Enhancement: Where speech is distorted by noise or otherwise poor recording quality, speech enhancement can allow for recovering speech and thereby can afford surveillance, sousveillance, and univeillance when combined with tools such as *ASR* and *Speech Translation*.

Speech Synthesis: Through mechanisms such as voice conversion, tasks within speech synthesis affords mechanisms for individuals to mask their own voice, and the biological markers that it communicates, and thereby allows for people to protect their identity and limit identity-based tracking. Yet synthesis does not, by design, prevent risks of identification or tracking of the contents of speech.

Anti-Spoofing: In the goal of its task, anti-spoofing allows for identifying (segments of) audio which has been synthesised, thereby mitigating the benefits of speech synthesis when it is used defensively, and can—in this way—afford reconstructing surveillance, sousveillance, and univeillance of content that a speaker may have synthesised for protecting their identities.

Emotions or Traits (paralinguistics): Tasks within paralinguistics seek to identify information about the speaker, such as their emotional states, accents, warmth, and so forth. Through such efforts, paralinguistics offers a mechanism through which one can go beyond identifying the speaker or the content of speech, to also infer personal characteristics and sensitive attributes of a speaker.

Clinical Settings: Speech technologies developed for clinical settings afford the identification of sensitive medical information such as health disorders, disabilities, and neurological conditions.

Multimodal Decoding: In our analysis, multimodal decoding covers a variety of tasks such as sentence prediction using EEG and producing speech from lipreading. Collectively, the risks of surveillance, sousveillance, and univeillance that arise from this set of tasks is producing speech from non-audio signals, thereby allowing for reconstruction of communication that a given speaker may have been explicitly trying to keep private.

Speech Representations: While developing better speech representations constitutes basic research within modelling speech signals, it also affords improvements on downstream tasks, including ASR, and can therefore aid across tasks in terms of anti-surveillance and anti-sousveillance as well as their counterparts.

Dialogue Systems: Unlike many other types of systems, dialogue systems allow for many similar mechanisms of surveillance; however, dialogue systems are designed to be opt-in systems, thereby allowing the speaker control over whether they will engage with them. However, in many cases, such as customer service, individuals may only be able to access services through dialogues systems, thereby effectively coercing people into using them. Yet, even within those instances, they are inherently more limited in that they require the active participation of the speaker.

Environment Monitoring: Environment monitoring does not directly engage with speech. However, such systems afford collecting information about the physical surroundings and contexts within which a speaker is situated at the moment of recording and thereby affords surveillance and univeillance of the speaker and their immediate context, rather than the speech itself.

Research Analyses & Other: Research analyses and the other category cover many different areas. The shared factor for them, however is that they are not readily used for either monitoring or resisting being monitored. The exception to this rule is the development of analytical datasets and methods for assessing fairness (e.g., Sakuma et al. 2025), which provides ethnicity, nationality, and race, which can allow the developers of systems to identify the populations for which their tools fail, and thus aid in enrolling users into surveillance infrastructures that they would otherwise have evaded.

Dual Use, Risks, and Limitations

In our findings, we note that the vast majority of papers do not include discussions around dual use, risks that arise from the work, or even the limitations of the work itself. Indeed, only 1.3%, 5%, and 25% of papers discuss or mention these, respectively. While dual use has not been common to report

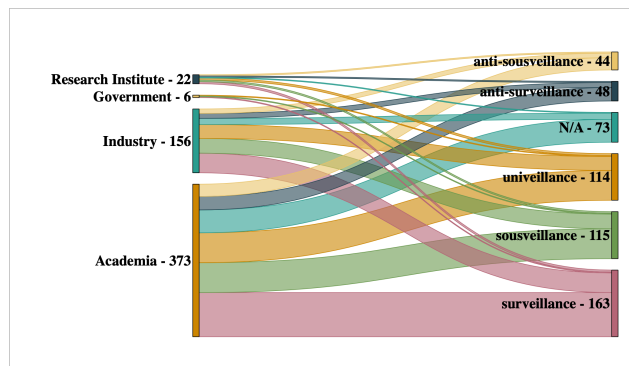


Figure 3: Relationship between author affiliations and axes (including N/A) from Veillance Compass.

in many computational fields, we follow the recommendations set out by Kaffee et al. (2023) for natural language processing. The questions presented there around the processes, development, and description of research artefacts will also be useful for the speech technology research field.

One example of engaging with the question of risk is provided by Kondo et al. (2025), where they remind readers that

Users must be mindful that the idols are actively involved in the entertainment industry, and care must be taken to avoid harming the reputation of the recorded groups, their management companies, fans, or any associated parties

when using their speech corpus of Japanese idol speakers.

Affiliations, Languages, and Artefact Availability

While governmental agencies and research institutes occur too infrequently in our sample to draw any conclusions, we find an interesting pattern among academia and industry (see Figure 3). Although academia has primarily focused on sur-, sous-, and univeillance (~ 56%), approximately 17% of all academic work focuses on anti-surveillance and anti-sousveillance. While industry papers have a similar proportion (~ 16%), this is a shadow number of work that occurs in industry. Kalluri et al. (2025) examine patents related to computer vision technologies and find that there is a research to surveillance pipeline in industry. That is, while there is no reason that academic—and public industry—work would not align along the identified patterns, only a portion of private industry work is made public and propriety work tends toward veillance products, often for purposes such as profiling and targeted marketing (Iqbal et al. 2022).

Unsurprisingly, the most prevalent language of study is English, covered by approximately half of the papers in our annotation sample. Following English, Chinese languages (Mandarin, Cantonese) and dialects account for ~12% of papers, and multilingual speech processing—covering a wide array of languages—account for ~10% of papers. Although one should expect that languages that are covered in a paper to be clearly stated, we find that this information is often omitted and it therefore often necessary to consult the papers describing datasets to identify which languages are included.

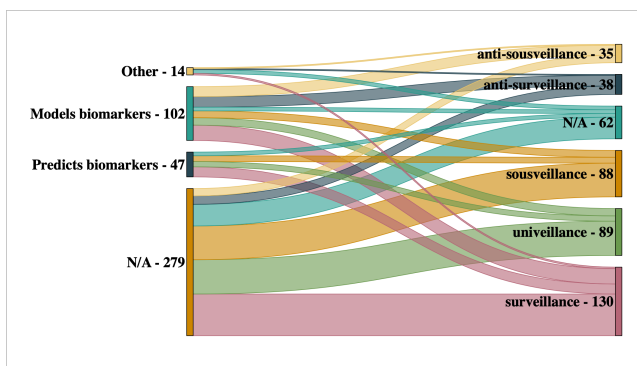


Figure 4: Relationship between biomarker usage types and axes (including N/A) from Veillance Compass.

In terms of resource availability, we find that 74 papers report that their data, code, or methods are available, while only one paper limits access. The remaining papers either are papers for which there are no research artefacts of note to share (20 papers), or the authors have not explicated how to access their research artefacts (133 papers).

Biomarkers

We find that 33.2% of papers include modelling or predicting biomarkers spanning across all veillance axes. Papers which engage in the prediction of biomarkers exclusively engage in research into sur-, sous-, and univeillance, whereas 37% of work that models biomarkers contributes to anti-surveillance and anti-sousveillance (see Figure 4). This distinction is expected: the project of predicting biomarkers relates to the goal of extracting information about the speaker, whereas the modelling of biomarkers is important for realistic voice synthesis.

5 Discussion

Thus far, we have outlined Mann’s veillance compass, developed and analysed a dataset of Interspeech papers, and presented our findings. Now, we shift to questions and insights that have arisen from the papers in our annotation sample, our discussions, and the wider literature in the field.

Does a Tool Afford the Same for Everyone?

Both annotators hold degrees in computer science, and would be able to implement and run the tools, even where code is not given is for the papers presented, given the necessary amount of compute and appropriate data. Indeed, much of the work that is analysed requires significant background knowledge of programming and computing, and would therefore not be as easily affordable to a lay person.

In a similar way, where a technology, e.g., multilingual ASR, requires large-scale compute to be able to run, it may, in theory, generally afford the its use, but the system will practically only be available to a small handful of entities that have a sufficient budget for computing equipment. Such technologies may then be provided as a service, which may or may not collect and retain data provided by users of the

service. In cases where there are high infrastructural requirements, the technologies themselves afford surveillance to a greater degree than sousveillance. In contrast, when a technology requires little expertise or computing infrastructure to deploy and use, it affords being available to a much wider group of people, and can thus more easily afford sousveillance, even as it also offers surveillance capabilities for entities that would seek to centralise.

What Are the Implications of a Tool Affording Surveillance?

There are several implications of a tool that affords surveillance depending on the context. For example, work on language technologies for low-resource languages has the potential of aiding people in meeting their needs in new ways that are easier to access (Yan et al. 2025), however, the creation of language resources and technologies can also subject a population that had not been under much surveillance to greater degrees of surveillance, and to other forms of harms (Nigatu and Raji 2024).

Yet, simply because a tool or a task holds the potential for surveillance does not mean that we should abandon research on it. Instead, it is important that we develop research infrastructures that afford identifying the potential risks that are introduced by research along with mechanisms for mitigating and addressing such risks. That is, rather than seeking to sanctions particular lines of work, we should seek to develop infrastructures that can allow for research to be safely conducted and research artefacts produced and shared without negative ramifications to speaker communities. Barring such research infrastructures, we argue that it is incumbent upon researchers that they carefully engage with questions about how to release and make available the research artefacts that are produced, when pursuing research directions that afford surveillance.

Developing Countermeasures

Developing countermeasures against surveillance can, by design, aid people in avoiding being held accountable, e.g., for illegal activities, but it can also help a wider community engage in modern life with greater degrees of privacy. It is therefore important the resources are dedicated towards the development of accessible tools and mechanisms that can help mitigate the risks of being under surveillance, while at the same time engaging with care when considering how such tools are released.

While research in NLP has extensively addressed questions of social harms of technologies (e.g., Talat and Blodgett 2025; Bender and Friedman 2018), such methods are not necessarily readily portable to speech research due to the additional information conveyed through human speech. For example, unlike text, speech communicates biometric and paralinguistic information in addition to linguistic information, which makes it possible to track speakers, identify them, and evaluate specific biomarkers (Kröger, Lutz, and Raschke 2019) in ways that are either not possible or are made much more difficult through text alone. We therefore argue that it is necessary for the speech research community

to develop approaches, best practices, and research infrastructures that take into account the wide variety of surveillance risks that speech technologies pose.

Contemporary Research Practices

We find it of great concern that the majority of papers in our dataset entirely omit discussing the social impacts, ethical concerns, or even limitations of their work. While it has negative implications for the research community—as research is not properly hedged and appropriate and inappropriate applications of the research are made opaque—it also has negative implications for identifying potential mechanisms for mitigating the risks of surveillance that are posed by the research and its artefacts, and can have an inhibiting effect on the development of methods for preserving privacy.

Relatedly, we take at face value what is described in papers, and do not systematically validate whether artefacts indeed are available at their described level. It is therefore possible that research that claims to be open-source has not actually been released or can otherwise be accessed.⁵

A surprising and hope-inducing finding from C M et al. (2025)—who experiment with ASR performance—is that ASR performance decreases when microphones are capturing speech through a speaker. We find this work particularly desirable direction for thinking about easily accessible interventions that may help individuals circumvent surveillance mechanisms, particularly if they are participating in pro-democratic activities such as protests.

6 Recommendations

Drawing on the results of our annotation and the subsequent discussion, we propose a set of recommendations for conducting research in speech processing with consideration of potential risks tied to surveillance. Additionally, we provide a list of questions that might be used for evaluating such risks. Firstly, languages on which research is conducted should be explicitly mentioned (Bender 2019). Secondly, due to the risks of malicious use of research outputs, research artefacts should clearly specify permissible and impermissible uses, e.g., through licensing. Thirdly, we recommend that conferences develop centralised repositories for research outputs which balance protection and transparency by allowing for securely sharing research artefacts and clearly communicating risks while simultaneously ensuring that code is available for reproduction. The depositing into such repositories could be included into the processes of preparing camera-ready versions of papers.

Incentives for Developing Tools

In the development of tools that afford surveillance or mitigate it, we should critically engage with the incentives of different stakeholders and how they might use such tools. To examine how a tool might be used, researchers can ask:

- What does our work afford in terms of surveillance or mitigation thereof?

⁵Indeed, in the small set of links provided in papers that we followed—to validate whether merely samples were made public or code and/or data—some were empty webpages or broken links.

- Which stakeholders might wish to make use of our tool and to what ends?
- What are mechanisms we can use to ensure that our intended use of the tools aligns with their actualised use?

Different stakeholders may also have different motivations for working on tasks. For example, industry researchers may have less agency to work on measures that mitigate risks of surveillance due to the potential risks to the public image of the company, should their method be used for negative or harmful outcomes.

Field-level Recommendations

We recommend that venues in the speech research community allow for sections in papers that are not counted towards the page limit, specifically for discussing surveillance risks, dual use, and limitations. While it is disappointing that concerns like dual use, limitations, surveillance, and wider ethical implications are not further discussed in papers, we believe that this issue is systemic. Given limited space, researchers are likely to dedicate as much space as possible towards describing their research in detail at the cost of discussing limitations or the social implications of their work. By allowing for sections that do not count towards the page count, the field can encourage researchers to critically engage with their work, which we believe will have outsized benefits for the field as a whole.

Paper-level Recommendations

For researchers writing papers, we recommend that they attempt to answer the following questions, which we believe would be a valuable starting point for identifying how their work engages with questions of surveillance.

- How might our work be used to improve surveillance capabilities of speech technologies?
- What are (social or technological) mechanisms to mitigate the surveillance impacts of your work?
- Whose speech does the research outcome target?
 - Have they agreed that tools for their language should be made publicly available?
 - * If **No**: How are you distributing the artefacts arising from your research?
 - * If **Yes**: Under what conditions, if any?
 - If the group is vulnerable (e.g., children, elderly people), what protections have been implemented to protect the people who may be targeted using this work?
- What outcomes does your research afford?
- Is your research easy to implement and use (in case of model)?
- How much computational power does your model need for training and inference?
- How much data was used to develop your model?
- What are the limitations of your research?
- What open questions remain or have arisen from your work?

These questions become all the more critical when the speaker communities are marginalised along social or linguistic lines. By addressing these questions, our hope is that researchers will be better positioned to address the question of surveillance and social harms of speech technologies, and be better suited to identify pathways towards more socially just and desirable speech technologies.

Potential Research Directions

Our final set of recommendations pertains to open research directions. While there is a great deal of research on how to build better and more impressive models for a variety of tasks, we found very limited evidence of research that seeks to engage with **developing privacy-preserving methods** or with **developing methods that can disturb the performance of surveillance technologies**. Moreover, there is space in the speech field for thinking about how to distribute and share data and models that maximise their social benefit, i.e., allow for the tool to exist without its surveillance affordances being realised. We believe that attending to these research areas can provide safer and more inclusive outcomes for the research field as a whole.

7 Limitations

This study has several limitations. First, the work has no longitudinal validity, as it examines one specific year in one specific venue. Second, we discovered that some papers labelled as open-source may not actually be publicly available online. Third, while our sample is 20% of all Interspeech 2025 papers, certain affiliation types, particularly governmental agencies and research institutes, appear very infrequently in our sample, and we therefore cannot draw strong conclusions on the results of such affiliations, or the tasks that they tend to work on. Additionally, although we are confident that our annotation of languages covered is correct, there may be instances where newer versions of the datasets that include more languages have been used.

8 Conclusion

In this paper, we highlight the importance of the relationship between speech technologies and surveillance by providing the historical background of surveillance and explaining how speech processing contributes to the rise of mass surveillance. We then analyse randomly selected 20% sample of papers from Interspeech 2025. We find that on one hand, research on processing speech does afford different veillances, and is particularly prone to surveillance. We also find that papers rarely engages with questions of the harms that arise from the research presented, thereby limiting the capability of the field to figure out methods to mitigate such issues. We therefore call for the speech research field and researchers to carefully consider how their work may build towards futures with greater degrees of surveillance, and ways to mitigate such futures.

Generative AI Disclosure

We have not used generative AI, or any other form of AI, for any step in this research project.

References

- Al-Ali, H.; Ghavamipour, A. R.; Caselli, T.; Turkmen, F.; Talat, Z.; and Aldarmaki, H. 2025. Personal Attribute Leakage in Federated Speech Models. ArXiv:2510.13357 [cs].
- Andrejevic, M. 2011. Watching Back. In Jansen, S. C.; Pooley, J.; and Taub-Pervizpour, L., eds., *Media and Social Justice*, 177–191. New York: Palgrave Macmillan US. ISBN 978-1-349-29139-7 978-0-230-11979-6.
- Bender, E. 2019. The BenderRule: On Naming the Languages We Study and Why It Matters. *The Gradient*.
- Bender, E. M.; and Friedman, B. 2018. Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. *Transactions of the Association for Computational Linguistics*, 6: 587–604.
- Bentham, J. 2010. *The Panopticon writings*. Radical thinkers. London [u.a]: Verso. ISBN 978-1-84467-666-8.
- Brown, I. 2014. Social Media Surveillance. In Ang, P. H.; and Mansell, R., eds., *The International Encyclopedia of Digital Communication and Society*, 1–7. Wiley, 1 edition. ISBN 978-1-118-29074-3 978-1-118-76777-1.
- Browne, S. 2015. *Dark Matters: On the Surveillance of Blackness*. Duke University Press. ISBN 9780822359197.
- Burgess, J.; Carlon, D.; and Doyuran, E. B. 2025. Voice AI and authenticity: current issues and emerging challenges. Working paper, ARC Centre of Excellence for Automated Decision-Making and Society.
- Bäckström, T. 2025. Privacy in Speech Technology. *Proceedings of the IEEE*, 113(7): 668–692.
- C M, V.; Pal, S.; Mantri, N.; and Agrawal, G. K. 2025. Effect of Loudspeaker Emitted Speech on ASR performance. In *Interspeech 2025*, 3170–3173. ISCA.
- Couldry, N.; and Turow, J. 2022. Market-Driven Voice Profiling: A Framework for Understanding. *Advertising & Society Quarterly*, 23.
- Crocco, M.; Cristani, M.; Trucco, A.; and Murino, V. 2014. Audio Surveillance. *ACM Computing Surveys (CSUR)*, 48: 1 – 46.
- Dewhurst, M.; Collins, J.; Lo, J. J. H.; Alderton, R.; and Kirkham, S. 2025. Nosey: Open-Source Hardware for Acoustic Nasalance. In *Interspeech 2025*, 2265–2269. ISCA.
- Fuchs, C.; Boersma, K.; Albrechtshlund, A.; and Sandoval, M. 2013. Introduction: Internet and Surveillance. In Fuchs, C.; Boersma, K.; Albrechtshlund, A.; and Sandoval, M., eds., *Internet and Surveillance*, 22–53. Routledge, 0 edition. ISBN 978-1-136-65527-2.
- Gibson, J. J. 1983. *The senses considered as perceptual systems*. Westport, Conn: Greenwood Press. ISBN 978-0-313-23961-8.
- Haiyun, L.; Wu, Z.; Xie, X.; Xie, J.; Xu, Y.; and Peng, H. 2025. VoiceMark: Zero-Shot Voice Cloning-Resistant Watermarking Approach Leveraging Speaker-Specific Latents. 5108–5112.
- Hutiri, W.; Papakyriakopoulos, O.; and Xiang, A. 2024. Not My Voice! A Taxonomy of Ethical and Safety Harms of

- Speech Generators. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 359–376. Rio de Janeiro Brazil: ACM. ISBN 979-8-4007-0450-5.
- Iqbal, U.; Bahrami, P. N.; Trimananda, R.; Cui, H.; Gamero-Garrido, A.; Dubois, D. J.; Choffnes, D. R.; Markopoulou, A.; Roesner, F.; and Shafiq, Z. 2022. Tracking, Profiling, and Ad Targeting in the Alexa Echo Smart Speaker Ecosystem. *Proceedings of the 2023 ACM on Internet Measurement Conference*.
- Kaffee, L.-A.; Arora, A.; Talat, Z.; and Augenstein, I. 2023. Thorny Roses: Investigating the Dual Use Dilemma in Natural Language Processing. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, 13977–13998. Singapore: Association for Computational Linguistics.
- Kalluri, P. 2020. Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature*, 583(7815): 169–169.
- Kalluri, P.; Agnew, W.; Cheng, M.; Owens, K.; Soldaini, L.; and Birhane, A. 2025. Computer-vision research powers surveillance technology. *Nature*, 643: 73 – 79.
- Kim, M.; Park, S.; Cha, S.; and Seo, P. H. 2025. Cross-Modal Watermarking for Authentic Audio Recovery and Tamper Localization in Synthesized Audiovisual Forgeries. In *Interspeech 2025*, 5103–5107. ISCA.
- Kim, Y.; and Chung, Y. 2025. Modality-Specific Speech Enhancement and Noise-Adaptive Fusion for Acoustic and Body-Conduction Microphone Framework. In *Interspeech 2025*, 3833–3837. ISCA.
- Kondo, Y.; Kameoka, H.; Tanaka, K.; and Kaneko, T. 2025. JIS: A Speech Corpus of Japanese Idol Speakers with Various Speaking Styles. In *Interspeech 2025*, 4783–4787.
- Kong, X.; Huang, H.; and Ou, Z. 2025. Lightweight and Robust Multi-Channel End-to-End Speech Recognition with Spherical Harmonic Transform. In *Interspeech 2025*, 3439–3443. ISCA.
- Kröger, J. L.; Lutz, O. H.-M.; and Raschke, P. 2019. Privacy Implications of Voice and Speech Analysis - Information Disclosure by Inference. In *Privacy and Identity Management*.
- Leschanowsky, A.; Rusti, C.; Quinlan, C.; Pnacek, M.; Gorce, L.; and Hutiri, W. 2025. A Data Perspective on Ethical Challenges in Voice Biometrics Research. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 7(1): 118–131.
- Liang, Y.; Yang, K.; Liu, F.; Li, A.; Li, X.; and Zheng, C. 2025. LightL2S: Ultra-Low Complexity Lip-to-Speech Synthesis for Multi-Speaker Scenarios. In *Interspeech 2025*, 3783–3787. ISCA.
- Majchrzak, A. 2013. Technology Affordances and Constraints Theory (of MIS). In *Encyclopedia of Management Theory*. 2455 Teller Road, Thousand Oaks, California 91320: SAGE Publications, Ltd. ISBN 978-1-4129-9782-9 978-1-4522-7609-0.
- Mann, S. 2013. Veilance and reciprocal transparency: Surveillance versus sousveillance, AR glass, lifelogging, and wearable computing. In *2013 IEEE International Symposium on Technology and Society (ISTAS): Social Implications of Wearable Computing and Augmented Reality in Everyday Life*, 1–12. Toronto, ON, Canada: IEEE. ISBN 978-1-4799-0929-2 978-1-4799-1242-1.
- Mann, S. 2016. Surveillance (Oversight), Sousveillance (Undersight), and Metaveillance (Seeing Sight Itself). In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.
- Maslan, S. 1995. Resisting Representation: Theater and Democracy in Revolutionary France. *Representations*, 52: 27–51.
- Morrison, H. 2026. Labour want 'Panopticon' to have 'eyes of the state on you at all times'.
- Mumford, L. 1964. Authoritarian and Democratic Technics. *Technology and Culture*, 5(1): 1.
- Nigatu, H. H.; and Raji, I. D. 2024. "I Searched for a Religious Song in Amharic and Got Sexual Content Instead": Investigating Online Harm in Low-Resourced Languages on YouTube. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 141–160. Rio de Janeiro Brazil: ACM. ISBN 979-8-4007-0450-5.
- Oh, S.; Kim, S.; and Chung, M. 2025. Multimodal and Multitask Learning for Predicting Multiple Scores in L2 English Speech. In *Interspeech 2025*, 2425–2429. ISCA.
- Palumbo, D.; and Prey, R. 2024. Sounding out voice biometrics: Comparing and contrasting how the state and the private sector determine identity through voice. *Big Data & Society*, 11.
- Prabhu, V. U.; and Birhane, A. 2020. Large image datasets: A pyrrhic win for computer vision? *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1536–1546.
- Pratap Singh, V.; Sahidullah, M.; and Kinnunen, T. 2025. Causal Structure Discovery for Error Diagnostics of Children's ASR. In *Interspeech 2025*, 2900–2904.
- Puglisi, S.; Rebollo-Monedero, D.; and Forné, J. 2016. You Never Surf Alone. Ubiquitous Tracking of Users' Browsing Habits. In Garcia-Alfaro, J.; Navarro-Arribas, G.; Aldini, A.; Martinelli, F.; and Suri, N., eds., *Data Privacy Management, and Security Assurance*, volume 9481, 273–280. Cham: Springer International Publishing. ISBN 978-3-319-29882-5 978-3-319-29883-2. Series Title: Lecture Notes in Computer Science.
- Sakuma, A.; Sato, H.; Sugano, R.; Kumano, T.; Kawai, Y.; and Ogawa, T. 2025. Speaker-Distinguishable CTC: Learning Speaker Distinction Using CTC for Multi-Talker Speech Recognition. In *Interspeech 2025*, 5503–5507. ISCA.
- Serditova, D.; Tang, K.; and Steffens, J. 2025. Automatic Speech Recognition Biases in Newcastle English: an Error Analysis. In *Interspeech 2025*, 3204–3208.
- Sherfinski, D.; and Asher-Schapiro, A. 2021. U.S. prisons mull AI to analyze inmate phone calls. *Reuters*.
- Sinnigen, W. G. 1962. The Origins of the "Frumentarii". *Memoirs of the American Academy in Rome*, 27: 211.
- State of Surveillance. 2025. Facial Recognition Bans in America: State by State.

Talat, Z.; and Blodgett, S. L. 2025. Ethics and Bias in Natural Language Processing. In *Reference Module in Social Sciences*, B9780323955041012266. Elsevier. ISBN 978-0-443-15785-1.

Tao, L.; Jia, M.; and Hu, Y. 2025. Direct-path Relative Harmonic Coefficients Detection for Multi-source Direction-of-Arrival Estimation in Reverberant Environments. In *Interspeech 2025*, 2505–2509. ISCA.

Tomashenko, N.; Miao, X.; Champion, P.; Meyer, S.; Panariello, M.; Wang, X.; Evans, N.; Vincent, E.; Yamagishi, J.; and Todisco, M. 2026. The Third VoicePrivacy Challenge: Preserving Emotional Expressiveness and Linguistic Content in Voice Anonymization. ArXiv:2601.11846 [cs].

Ungless, E. L.; Vitsakis, N.; Talat, Z.; Garforth, J.; Ross, B.; Onken, A.; Kasirzadeh, A.; and Birch, A. 2024. Ethics Whitepaper: Whitepaper on Ethical Research into Large Language Models. ArXiv:2410.19812 [cs].

Yan, B.; Hamed, I.; Shimizu, S.; Lodagala, V. S.; Chen, W.; Iakovenko, O.; Talafha, B.; Hussein, A.; Polok, A.; Chang, K.; Klement, D.; Althubaiti, S.; Peng, P.; Wiesner, M.; Solorio, T.; Ali, A.; Khudanpur, S.; and Watanabe, S. 2025. CS-FLEURS: A Massively Multilingual and Code-Switched Speech Dataset. In *Interspeech 2025*, 743–747. ISCA.

Zhao, Y.; Shang, Z.; Wang, M.; Liu, X.; and Zhang, P. 2025. Restoring Harmonics: Enhancing Speech Quality with Deep Mask and Harmonic Restoration Network. In *Interspeech 2025*, 5168–5172. ISCA.