

Privacy in Speech and Language Technology

Simone Fischer-Hübner^{*1}, Dietrich Klakow^{*2}, Peggy Valcke^{*3}, and Emmanuel Vincent^{*4}

1 Karlstad University, SE. simone.fischer-huebner@kau.se

2 Saarland University – Saarbrücken, DE. dietrich.klakow@lsv.uni-saarland.de

3 KU Leuven, BE. peggy.valcke@kuleuven.be

4 Inria – Nancy, FR. emmanuel.vincent@inria.fr

Abstract

This report documents the outcomes of Dagstuhl Seminar 22342 “Privacy in Speech and Language Technology”. The seminar brought together 27 attendees from 9 countries (Australia, Belgium, France, Germany, the Netherlands, Norway, Portugal, Sweden, and the USA) and 6 distinct disciplines (Speech Processing, Natural Language Processing, Privacy Enhancing Technologies, Machine Learning, Human Factors, and Law) in order to achieve a common understanding of the privacy threats raised by speech and language technology, as well as the existing solutions and the remaining issues in each discipline, and to draft an interdisciplinary roadmap towards solving those issues in the short or medium term.

To achieve these goals, the first day and the morning of the second day were devoted to 3-minute self-introductions by all participants intertwined with 6 tutorials to introduce the terminology, the problems faced, and the solutions brought in each of the 6 disciplines. We also made a list of use cases and identified 6 cross-disciplinary topics to be discussed. The remaining days involved working groups to discuss these 6 topics, collaborative writing sessions to report on the findings of the working groups, and wrap-up sessions to discuss these findings with each other. A hike was organized in the afternoon of the third day.

The seminar was a success: all participants actively participated in the working groups and the discussions, and went home with new ideas and new collaborators. This report gathers the abstracts of the 6 tutorials and the reports of the working groups, which we consider as valuable contributions towards a full-fledged roadmap.

Seminar August 21–26, 2022 – <https://www.dagstuhl.de/22342>

2012 ACM Subject Classification Artificial Intelligence → Natural Language Processing; Security and Privacy → Human and Societal Aspects of Security and Privacy; Security and Privacy → Software and Application Security; Security and Privacy → Database and storage security

Keywords and phrases Privacy, Speech and Language Technology, Privacy Enhancing Technologies, Dagstuhl Seminar

Digital Object Identifier 10.4230/DagRep.12.8.60

* Editor / Organizer



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Privacy in Speech and Language Technology, *Dagstuhl Reports*, Vol. 12, Issue 8, pp. 60–102

Editors: Simone Fischer-Hübner, Dietrich Klakow, Peggy Valcke, Emmanuel Vincent



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

4.3 Vulnerable groups and legal considerations

Lydia Belkadi (KU Leuven, BE) lydia.belkadi@kuleuven.be

Meiko Jensen (Karlstad University, SE) meiko.jensen@kau.se


Dietrich Klakow (Saarland University – Saarbrücken, DE) dietrich.klakow@lsv.uni-saarland.de

Katherine Lee (Google Brain & Cornell University – Ithaca, US) katherinelee@google.com

Olga Ohrimenko (University of Melbourne, AU) oohrimenko@unimelb.edu.au

Jo Pierson (Free University of Brussels, BE) jo.pierson@vub.be

Emmanuel Vincent (Inria – Nancy, FR) emmanuel.vincent@inria.fr

License  Creative Commons BY 4.0 International license

© Lydia Belkadi, Meiko Jensen, Dietrich Klakow, Katherine Lee, Olga Ohrimenko, Jo Pierson, Emmanuel Vincent

4.3.1 Biometric systems

From a legal perspective, biometric verification (that is verifying the identity of a speaker) systems are often deemed to be not as risky as biometric identification systems (who out of a larger set of known speakers is speaking). Under data protection laws, legal scholars have discussed the definition and legal nature of biometric data. Indeed, Articles 4(14) of the GDPR and 3(13) of the Law Enforcement Directive define biometric data as “personal data resulting from specific technical processing [...] which allow or confirm the unique identification” of an individual. In particular, the definition seems to directly refer to biometric identification (i.e., “allow” the unique identification) and verification (i.e., “confirm” the unique identification) [1]. Article 9 of the GDPR further specifies that only biometric data “processed for the purpose of uniquely identifying” an individual are considered sensitive. In other words, the GDPR does not consider all processing of biometric data as sensitive and excludes verification purposes [2].

This distinction between identification and verification further permeates the risk assessment performed by the European Commission under the AI Act. This regulation aims to set out rules for the development, marketing, and use of AI systems. It further aims to steer AI uptake to reach a high level of protection of public interests (e.g., health, safety, fundamental rights). The AI Act relies on a risk-based framework spanning from unacceptable to minimal risks to support this approach. Accordingly, AI practices entailing severe risks to public interests are prohibited or more strictly regulated.

The current draft of the AI Act considers that biometric verification always entails “minimal risks”, except in the context of migration, asylum and border control management. In particular, AI systems used to verify the authenticity of travel documents and check their security features are considered high-risk (Annex III). This exclusion means that providers, users, and other third parties involved in the supply chain would, in principle, not be subjected to the obligations set out in Articles 16 to 29 (e.g., taking corrective actions in case of non-conformity, information and cooperation with national competent authorities, etc.).

Furthermore, only high-risk AI practices are required to comply with a set of requirements related to the establishment of a risk management system, data governance, technical documentation, record-keeping, transparency and provision of information to users, human oversight and accuracy, robustness and cybersecurity (Articles 9 to 15). These requirements would be applicable to biometric verification systems only on a voluntary basis, through the adoption of codes of conduct (Article 69).

From a technical perspective, linking the risks of biometric verification only to the number of individuals enrolled in a database is criticizable. Indeed, risks still arise even when the database contains a single individual. First, storing biometric identifiers in the cloud as

opposed to the user's device implies that they may be more easily stolen, or that the user might be identified in a situation when they don't want to. Second, the "vocal signature" has been shown to contain a lot more information than biometric identity, which might be inferred [3]. The same risk arises with, e.g., typing patterns associated with text. Third, the boundary between verification and identification is not always clear, e.g., when a smart speaker is used by 5 members of a family, running speaker verification against 5 "vocal signatures" could qualify as a form of identification. The risk should therefore be quantified depending on the usage context, the location where the identifiers are stored, and whether the user is willing to be identified.

4.3.2 Beyond identity

Speech and text snippets are complex sources of information conveying more than (biometric) identity. For example, they may reveal speakers' emotional states or health conditions. It is not always possible to dissociate and isolate different attributes captured from individuals, entailing the collection of a wide scope of sensitive personal data. Over time, such collections may also enable the constitution of extensive (e.g., personality) profiles.

Many technical and legal distinctions may be drawn to determine the sensitivity of the collection and processing of speech and text. For example, the collection of a single instance or aggregates of emotional states would have different impacts on concerned individuals. Similarly, the use of aggregates of speech and text snippets for profiling would have distinct risks and benefits depending on the context (e.g., commercial or medical uses). Accordingly, a blanket prohibition of the extraction of specific attributes of speech and text snippets may not be desirable.

At the same time, the entanglement of different attributes within snippets raises important challenges from a legal perspective. For example, it is unclear how speech or text snippets should be defined from a legal perspective or how to apply existing legal definitions. This difficulty was well illustrated in recent legislative debates over the legal concept of "biometric data" under the upcoming AI Act. In particular, the European Parliament is discussing the opportunity to distinguish the concept of "biometric data" and "biometric-based data" to account for processing beyond biometric recognition (e.g., emotion recognition).¹⁰

Similarly, this entanglement implies considerable contradictions with data protection principles, such as data minimization and purpose limitation. In other words, snippets may reveal more data than is necessary for a given purpose (e.g., text and typing patterns in language processing).

The coexistence of these different attributes is important when determining the sensitivity of speech and text snippets and determining the legal basis to be used. In particular, it would require taking into account overlapping legal categories of data (e.g., data concerning health, biometric data). In turn, this overlap may mandate the performance of risk assessments that consider the complex nature of speech and text snippets, and the different attributes revealed (e.g., biometric and health attributes).

This challenge has become even more relevant after the Court of Justice of the European Union's ruling *OT v Vyriausioji tarnybinės etikos komisija*¹¹. In previous years, the question to what extent data protection laws, and in particular the GDPR, offer protection against

¹⁰ See for example the following study commissioned by the European Parliament: "Biometric Recognition and Behavioural Detection" (2021) p.96.

¹¹ Court of Justice of the European Union, Judgment of 1 August 2022, (*OT v Vyriausioji tarnybinės etikos komisija*), C-184/20, ECLI:EU:C:2022:601: "[...] Article 9(1) of Regulation 2016/679 must be interpreted as meaning that the publication, on the website of the public authority responsible for collecting and checking the content of declarations of private interests, of personal data that are liable to disclose indirectly the sexual orientation of a natural person constitutes processing of special categories of personal data, for the purpose of those provisions

sensitive inferences (Article 9¹²) or remedies to challenge inferences or important decisions based on them (Article 22(3)) has been discussed in legal scholarship. Wachter et al., for instance, have pointed to significant shortcomings in this regard and concluded that individuals are granted little control and oversight over how their personal data is used to draw inferences about them [4]. In the ruling *OT v Vyriausioji tarnybinės etikos komisija*, the Court had the opportunity to illuminate the question whether Article 9 of the GDPR applies in the situation where special categories of personal data are not explicitly made public (more notably, in online declarations of interests by persons working in the public service as required under Lithuanian anti-corruption law), but Internet users may nevertheless infer certain sensitive information about the declarants, including their political opinions or sexual orientation. In other words, the personal data that needs to be published according to the Lithuanian anti-corruption law are not, inherently, sensitive data in the sense of the GDPR. However, it was possible to deduce from the name-specific data relating to the spouse, cohabitee or partner of the declarant certain information concerning the sex life or sexual orientation of the declarant and his or her spouse, cohabitee or partner. The question to be answered by the Court was, consequently, whether data that are capable of revealing the sexual orientation of a natural person by means of thinking (e.g., involving comparison or deduction) fall within the special categories of personal data, for the purpose of Article 9(1) of the GDPR. The Court confirmed the Advocate General’s opinion from December 2021, namely that Article 9(1) must effectively be interpreted as meaning that the processing of special categories of personal data includes publishing the content of the declaration of interests on the website of the controller in question. In other words, the Court interprets the scope of Article 9 of the GDPR to include sensitive inferences, something advocated for by Wachter et al. [4].

Risk assessments may also need to be performed taking into account the impact of the processing on fundamental rights [5]. For example, under European data protection laws, controllers are obliged to carry out Data Protection Impact Assessments. Article 35 of the GDPR mandates such assessment where a type of processing is “likely to result in a high risk to the rights and freedoms of natural persons”. Similarly, Article 7 of the upcoming AI Act expects the European Commission to consider the risks to individuals’ fundamental rights when amending the list of high-risk AI systems. In relation to speech and language technologies, what would these obligations mean for data controllers when considering the principle of non-discrimination and the right to freedom of speech? Would new fundamental rights be necessary (e.g., right to freedom of emotions)?

4.3.3 Vulnerable groups

From a legal perspective, special attention must be given to the concept of vulnerability. Under the upcoming AI Act, vulnerability will be introduced under two key provisions. Firstly, the impact on vulnerable individuals or groups is a determining factor to qualify certain AI practices as unacceptable practices. For example, Article 5 prohibits the use of AI systems that exploit “any of the vulnerabilities” of a specific group of persons due

¹² Article 9(1) of the GDPR (previously Article 8(1) Directive 95/46) provides for the prohibition, inter alia, of processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of data concerning a natural person’s sex life or sexual orientation. According to the heading of those articles, these are special categories of personal data, and such data are also categorized as “sensitive data” in recital 34 of Directive 95/46 and Recital 10 of the GDPR.

to their age, physical or mental disability when such use would distort their behavior in a manner that causes or is likely to cause physical or psychological harm. Similarly, the use by public authorities of AI systems to evaluate or classify the trustworthiness of individuals based on social behavior, known or predicted personal or personality characteristics are also prohibited, under certain conditions, when it leads to detrimental or unfavorable treatment of certain individuals or groups.

Additionally, the concept of vulnerability is also used as a factor to be assessed by the European Commission when amending the list of high-risk AI systems. Under Article 7, the European Commission needs to consider:

- the extent of harm or adverse impact of AI systems in terms of intensity and ability to affect a plurality of persons and
- whether impacted persons would be in a vulnerable position, particularly due to an imbalance of power, knowledge, economic or social circumstances, or age.

At the same time, the concepts of vulnerability and vulnerable groups in relation to language technologies raise many questions from an inter-disciplinary perspective.

When speech recognition or natural language processing systems are utilized on a broad scale, these systems at some point will interact with individuals from the so-called vulnerable groups. This broad term typically includes humans with conditions that require special consideration, both in a technical and legal dimension. We can distinguish three types of vulnerable groups of relevance here:

1. individuals with special characteristics of their voice or language,
2. individuals that are not themselves able to utilize their human rights, and
3. individuals that belong to discriminated groups due to special personal characteristics like sexual orientation, ethnicity, or religious or political position.

In the first group, people with speaking issues like stuttering, aphonia, or amnesic aphasia clearly become relevant. The so-called “Doddington zoo” effect [6] also means that some people’s voices are more easily identifiable than others for reasons that cannot be traced back to a specific characteristic. As discussed previously, AI-based speech recognition works with training based on a large set of speech examples, which may or may not have contained people with these specific conditions. If present, the trained AI might be able to cope with (and hide) the specific type of speech characteristics, but if the training dataset did not contain such examples, it might work less well when confronted with speech or language examples from such individuals. Hence, one challenge lies in the proper and non-biased selection of training data, as inclusion of all possible speech- or language-specific abnormalities in the training dataset tends to raise discriminatory real-world issues in itself. As an example, consider an advertisement explicitly asking for stutterers to join a training dataset recording. The resulting dataset would be biased towards favoring stutterers to other speech issues, and the real-world discriminatory effects of such an advertisement could be socially challenging as well.

The second group requires close attention, especially from the legal point of view. Transfer of self-responsibility to another human is a severe and highly sensitive issue, and should only be done in cases that have no alternative. Children are especially vulnerable in this case, as they cannot oversee the consequences of their actions sufficiently, so their parents or legal guardians have to approve decisions or even make decisions themselves for the children. In terms of speech-based interaction technology, this dependency of a child towards its custodian makes the former especially vulnerable, as audio surveillance of sleeping babies is a common and mostly socially accepted scenario. However, this raises a lot of open issues when it comes to questions of secondary use of the voice data created by children, e.g., towards advertising or psychological analysis by third parties – especially in the long term, when these children grow up to be adults of the same personality.

Another example of the second group type is people with diseases like dementia or mental disorders. Even if these may at some point decide to e.g., utilize smart speakers in their homes, or consent to having their language in a social media chat app get analyzed by a research institution, this decision may not stay aware to them. Hence, subsequently, when confronted with the ongoing voice surveillance of the smart speaker, or receiving the feedback from the research institutions, such individuals may suffer from severe trauma. On the other hand, availability of such technical surveillance or assistance systems might be very beneficial towards these individuals, especially for those also suffering from physical deficiencies like inability to type or utilize other input devices for a computer.

The third group is special in a large variety of possible ways, ranging from sexual orientations that are considered illegal in some countries of the world to social discrimination or even physical frays based on skin color, nationality, or political opinions expressed. In all of those cases, speech and language processing systems to some extent may be able to identify such conditions, based on what was said or how it was said in specific contexts (e.g., lie detection when confronted directly).

In general, belonging to a vulnerable group is no explicit act, and the definitions of what substantiates a vulnerable group differ largely.

What is common to them is that speech and language processing systems have to be designed in a way that they are either reliably agnostic to these conditions or consider them appropriately in the design and behavior of the system in consideration. Here, privacy-enhancing technologies may help, and should be considered wherever possible.

4.3.4 Confidentiality vs. duty to rescue

In some situations, the users' right to privacy may conflict with the voice technology company's legal requirements. For example, if the voice technology company collects speech or text data suggesting that a crime (e.g., child abuse) or a life-threatening danger (e.g., heart attack) has taken place, should it report it to the relevant authority, thereby violating the user's privacy? Is it enough to report cases that have been incidentally found or should the company be required to automatically analyze the data to find all possible cases and have them screened by a human operator, which is a form of systematic surveillance? When answering these questions, it is important to realize that legal requirements regarding "duty to rescue" vary from one jurisdiction to another.¹³ In most jurisdictions under civil law (Europe, Latin America) and in some US states, it is a legal duty for citizens to assist in such cases unless this would put them in danger, with some exceptions (e.g., if the citizen is a priest or a lawyer hearing a person confess a crime, the confidentiality obligation is stronger). The duty to rescue does not apply to companies in these jurisdictions nor to citizens or companies in other jurisdictions, which implies that such cases can be reported but it's not an obligation. Nevertheless, some companies have been requested by law enforcement agencies to automatically screen for, e.g., child pornography in personal image data. This raises three open questions. From a societal point of view, should companies be requested, allowed, or forbidden to perform large-scale automatic screening in the speech and text data they collect? If this is requested or allowed, what should be the territorial extent (e.g., would it apply to a European company processing data from an American citizen) and which legal safeguards should be put in place to preserve fundamental human rights regarding censorship (what can or cannot be uploaded) and massive surveillance? Also, from a technical point of view, could this screening be performed on-device in a privacy-preserving way?

¹³https://en.wikipedia.org/wiki/Duty_to_rescue

References

- 1 Catherine Jasserand. Legal nature of biometric data: From generic personal data to sensitive data. *European Data Protection Law Review*, 2(3):304, 2016.
- 2 Els Kindt. Having yes, using no? About the new legal regime for biometric data. *Computer Law & Security Review*, 34(3):523–538, 2018.
- 3 Desh Raj, David Snyder, Daniel Povey, and Sanjeev Khudanpur. Probing the information encoded in x-vectors. In *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pages 726–733. IEEE, 2019.
- 4 Sandra Wachter and Brent Mittelstadt. A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review*, 2019(2), 2019.
- 5 Dara Hallinan and Nicholas Martin. Fundamental rights, the normative keystone of DPIA. *European Data Protection Law Review*, 6(3):178–193, 2020.
- 6 George Doddington, Walter Liggett, Alvin Martin, Mark Przybocki, and Douglas Reynolds. Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation. Technical report, DTIC Document, 1998.

4.4 Privacy attacks

Abdullah Elbi (KU Leuven, BE) abdullah.elbi@kuleuven.be

Anna Leschanowsky (Fraunhofer IIS – Erlangen, DE) anna.leschanowsky@iis-extern.fraunhofer.de

Pierre Lison (Norsk Regnesentral – Oslo, NO) plison@nr.no

Andreas Nautsch (Avignon Université, FR) andreas.nautsch@univ-avignon.fr

Olga Ohrimenko (University of Melbourne, AU) oohrimenko@unimelb.edu.au

Laurens Sion (KU Leuven, BE) laurens.sion@kuleuven.be

Marc Tommasi (University of Lille, FR) marc.tommasi@univ-lille.fr

License © Creative Commons BY 4.0 International license

© Abdullah Elbi, Anna Leschanowsky, Pierre Lison, Andreas Nautsch, Laurens Sion, Marc Tommasi

4.4.1 Context and motivation

One way to assess the strength of privacy-enhancing techniques (and the data protection they provide) is to conduct so-called *privacy attacks*. In our context, a privacy attack is a process which, given a particular input or model, seeks to uncover personal data that should be or should have been concealed. Privacy attacks can be employed as part of privacy risk assessments (including Data Protection Impact Assessments) or as an evaluation method in the development of privacy-enhancing techniques.

It is, however, important to stress that privacy attacks can usually only provide lower bounds when it comes to assessing the privacy risk associated with a given output or model. Privacy attacks are by construction not exhaustive and can only explore a limited region of the risk space. In other words, they can only demonstrate the presence of a privacy risk and not their absence. Although we can make assumptions about possible attackers and the background knowledge those attackers may have access to, those assumptions may very well turn out to be invalid. Attackers may also rely on other attack strategies than the ones that have been explicitly tested.

Although the present section focuses specifically on privacy attacks (i.e., attacks designed to uncover personal data), it is worth noting that security attacks (i.e., attacks targeting the confidentiality, integrity, or availability of an IT system) may also lead to privacy breaches. In particular, it has been shown that one can infer the hidden values of a black-box machine