



Speech data privacy: a real challenge!



© Inria / Photo G. Scagnelli

Registration is now open to take part in the first international challenge on speech data privacy. The VoicePrivacy initiative is spearheading the effort to develop privacy preservation solutions for speech technology. It aims to gather a new community to define the most effective processes and metrics, while also benchmarking the existing privacy-enabling solutions and creating a baseline against which to evaluate future improvements. This groundbreaking initiative is now looking for forward-thinking and passionate participants.

Vocal access systems and assistants are almost ubiquitous, picking up, storing and analyzing the data routinely being "told" to them. In parallel, questions are increasingly being raised about the privacy of these data, and the security of service users. This is because speech data contains much more than just the spoken words. Speech encapsulates a wealth of personal data (like age and gender), health and emotional state, racial or ethnic origin, geographical background, social identity, and socio-economic status. Since machines can now decipher spoken language with ever-impressive accuracy, there is no reason why political orientations, religious and philosophical beliefs could not also be derived from speech data.

Two main solutions are under development to preserve privacy in speech data: homomorphic encryption and anonymization. Homomorphic encryption protects personal data by making it inaccessible to third parties, however it is computationally cumbersome. Anonymization, by contrast, suppresses personal information in the data; it can run in real time but the absence of clear definitions and standards raises questions about its efficiency.

The VoicePrivacy initiative is the first of its kind: it aims to bring together a community of experts to focus on this specific security challenge. It sets a standardized frame and will be a crucial step forward to define the currently missing metrics and processes to evaluate speech privacy, and the efficiency of the technologies developed to protect it. The participants will be asked to process the same dataset of speech signals to anonymize them, while protecting the linguistic content and speech naturalness.

The co-organizers of the VoicePrivacy initiative are:

- University of Avignon - LIA, France - <http://univ-avignon.fr/international-english-version-3508.kjsp>
- Inria, France - inria.fr
- National Institute of Informatics, Japan - nii.ac.jp/en/
- EURECOM, France - eurecom.fr/en
- University of Edinburgh, UK - ed.ac.uk/

The VoicePrivacy initiative is supported by:

- Agence Nationale de la Recherche (projects Deep-privacy, Harpocrates and VoicePersonae) - anr.fr
- European Union's Horizon 2020 Research and Innovation Program (project COMPRISE) - compriseh2020.eu
- Japan Science and Technology Agency (project VoicePersonae) - jst.go.jp/EN/

The final results of the VoicePrivacy initiative will be published during the Interspeech conference in October 2020. Detailed information, as well as the registration information are available on the dedicated website: <https://www.voiceprivacychallenge.org>

The Comprise project aims to define a fully private-by-design methodology and tools that will reduce the cost and increase the inclusiveness of voice interaction technology through research advances on privacy-driven data transformations, personalised learning, automatic labelling, and integrated translation. It has received funding from the European Union's Horizon 2020 Research and Innovation Programme.

MEDIA CONTACTS

Inria - Zaineb Chelly-Dagdia - + 33 (0)3 54 95 84 10
zaineb.chelly-dagdia@inria.fr

Inria - Isabelle Kling - +33 (0)6 43 38 72 64
isabelle.kling@inria.fr