



Privacy-driven speech and text transformation tools



© Inria / Photo G. Scagnelli

The joint efforts of the COMPRISE consortium are bearing fruit as its first privacy preservation tools have been released. The COMPRISE Voice Transformer and the COMPRISE Text Transformer protect both the voice of the users and their personal information. These software tools can be easily integrated into existing voice technologies, and provide developers with open source, validated, private-by-design tools. The development of privacy-enhancing technologies (PETs) in Europe is a crucial step forward in ensuring the resilience of critical information infrastructure and in supporting the expansion of digital businesses.

Protecting users' privacy

The key to deploying voice technology and expanding the range of languages and use cases is to collect massive amounts of speech data from users. This compromises the users' privacy in two ways. First, speech is a type of biometric data, which can be used to identify users against their will. Second, the amount of information that can be extracted from it is simply astonishing: it may reveal information about the users' personality, general traits (gender, age, ethnic origin, nativeness, etc.), states (health condition, intoxication, sincerity, etc.) and preferences that they may not want to share with a company. In the case of a security breach, this information may even be misused by third parties. One of the main objectives of COMPRISE is to develop private-by-design voice technology.

Now, after months of hard work, the COMPRISE consortium led by Inria is moving closer to this goal as the project's first privacy preservation tools have recently seen the light. This exemplifies Inria's ambition to develop flagship digital

technologies in France and Europe based on cutting-edge, often multi-disciplinary research. Via its recent strategic plan «Inria Ambition 2023», Inria committed to supporting Europe's digital sovereignty, and fostering the development of numerous software products able to achieve a global impact, e.g., via the open source model.

COMPRISE Voice Transformer

The COMPRISE Voice Transformer aims to prevent biometric identification of the users by converting their voice to another random person's voice.

It is based on cutting-edge deep learning and speech processing technology and provides a measurable increase in privacy, as validated through state-of-the-art biometric protocols.

This tool can be applied in the broad context of voice technology and artificial intelligence (AI), where privacy is key to build user trust and collect the increasing amounts of data required to improve user experience and develop innovative services. It would also be valuable in situations where speech is to be stored, transmitted, released, or processed remotely.

As such, it will be of interest to large and small voice technology / AI companies, open source developers, call centers, online game studios, hospitals, journalists, etc.

COMPRISE Text Transformer

The COMPRISE Text Transformer aims to identify potentially privacy-threatening words or phrases in a piece of text and to replace them by harmless alternatives preserving the text's structure.

It relies on state-of-the-art deep learning and natural language processing technology. The main innovation lies in our scientifically founded word and phrase replacement strategy which provides formal privacy guarantees while minimising the negative impact on downstream users or readers. Users can set the trade-off between privacy and utility themselves within the established framework of Differential Privacy.

This tool allows individuals and companies in various application sectors to mask out critical information in documents that would otherwise threaten their privacy or that of third parties. This includes sectors such as finance, medicine, and law, among others.

Please check the comprehensive infographic on this new technology on our blog: <https://www.compriseh2020.eu/comprise-features-two-main-branches/>

Inria is the French national research institute for digital science and technology. World-class research, technological innovation and entrepreneurial risk are its DNA. In 200 project teams, most of which are shared with major research universities, more than 3,500 researchers and engineers explore new paths, often in an interdisciplinary manner and in collaboration with industrial partners to meet ambitious challenges.

As a technological institute, Inria supports the diversity of innovation pathways: from open source software publishing to the creation of technological startups (Deeptech).

The COMPRISE project is a Research and Innovation Action (RIA) funded by the European Union's Horizon 2020 program (Grant Agreement 825081) which creates software tools that ensure data privacy, reduce the cost and increase the inclusiveness of voice technology.

Further information related to the COMPRISE project is available on the project's website

<https://www.compriseh2020.eu/>

Additionally, you can reach us via our social media accounts:

LinkedIn: <https://www.linkedin.com/company/comprise-h2020>

Twitter: <https://twitter.com/compriseh2020>

MEDIA CONTACTS

Inria - Laurence Goussu - + 33 (0)6 81 44 17 33
laurence.goussu@inria.fr

Inria – Véronique Poirel - +33 (0)6 84 82 23 15
veronique.poirel@inria.fr

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