Roadmap on AI technologies & applications for the Media Industry

Section: “Privacy-preserving AI”

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911

info@ai4media.eu www.ai4media.eu

Authors	**Patrick Aichroth** (Fraunhofer Institute for Digital Media Technology - IDMT) **Thomas Köllmer** (Fraunhofer Institute for Digital Media Technology - IDMT)

This report is part of the deliverable D2.3 - "*AI technologies and applications in media: State of Play, Foresight, and Research Directions*" of the AI4Media project.

You can site this report as follows:

F. Tsalakanidou et al., Deliverable 2.3 - AI technologies and applications in media: State of play, foresight, and research directions, AI4Media Project (Grant Agreement No 951911), 4 March 2022

This report was supported by European Union´s Horizon 2020 research and innovation programme under grant number 951911 - AI4Media (A European Excellence Centre for Media, Society and Democracy).

The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf.

Copyright

Privacy-preserving AI

Current status

"If you give me six lines written by the hand of the most honest of men, I will find something in them which will hang him."— Cardinal Richelieu[1]

Privacy can be understood as the ability of individuals to control their personal information, the right to not be observed, to be left alone, and to keep relationships and personal matters secret. It helps us making our own decisions, without being observed or disturbed, keeping social pressure at bay.

AI has created new challenges for privacy: Machine learning requires large datasets for training, creating fantastic new possibilities, but also pushing an increased desire for collecting data, including personal data to be used for targeted advertisement or service improvements. We create and share large amounts of personal data e.g. by using smartphones but also passively while living in an environment that collects more and more data – think about CCTV, online payments, location tracking. The problem is, thanks to the technologies being developed and applied within recently, we are neither fully aware of the kind and amount of data being collected and used, nor can we predict how data will be used by AI in the future.

Does that mean that privacy is not a concern for people anymore? Probably not. A survey in the EU indicated that 41% of respondents were not willing to share personal data with private companies, only 5% were willing to share facial images or fingerprints with private companies, and 55% were afraid of criminals or fraudsters accessing their personal data.[2] Similarly, a 2022 survey in the US indicated that a majority of respondents was concerned about how much data is collected about them by companies and the government (79%/64%), believing that much of what they do online and on their cellphone is being tracked by companies or the government (72%/47%). More than 80% said that they feel having very little or no control over the data collected about them and that the potential risks outweigh the benefits when it comes to companies collecting data.[3] Such concerns could also explain some of the worries regarding AI, for instance in Germany, where public skepticism about AI is already considered a key burden to innovation by German SMEs.[4]

However, the view that privacy and AI are mutually incompatible is both wrong and dangerous: if we are willing to give up privacy because we (wrongly) believe we have to, we are paying a tremendous price. And if we bluntly reject technologies such as AI because we (wrongly) believe we have to give up privacy, we give up on designing and improving technologies, ending up with

[1] Fischer, D. H. (2009). Champlain's Dream (Reprint Edition). Simon & Schuster, p. 704

[2] How concerned are Europeans about their personal data online? (2020, June 15). European Union Agency for Fundamental Rights. https://fra.europa.eu/en/news/2020/how-concerned-are-europeans-about-their-personal-data-online

[3] Auxier, B., & Rainie, L. (o. J.). Key takeaways on Americans' views about privacy, surveillance and data-sharing. Pew Research Center. Abgerufen 4. February 2022: https://www.pewresearch.org/fact-tank/2019/11/15/key-takeaways-on-americans-views-about-privacy-surveillance-and-data-sharing/

[4] Rammer, C. (2021). Herausforderungen beim Einsatz von Künstlicher Intelligenz, Ergebnisse einer Befragung von jungen und mittelständischen Unternehmen in Deutschland. Mannheim: Bundesministerium für Wirtschaft und Energie (BMWi)

a self-fulfilling prophecy, leaving markets to those who do not care about privacy. Instead, we should aim at solutions to build privacy into AI, and invent and use technologies that allow us to do that. But which technologies could that be?

As for protecting privacy for data analysis, ***anonymisation techniques*** and ***concepts*** have traditionally played a key role, including ***k-anonymity***, ***l-diversity***, and ***t-closeness***. Within a dataset, there are *direct identifiers* (attributes which directly identify an individual), *quasi-identifiers* (attributes which can identify an individual if combined with other quasi-identifiers, although the definition is not always used consistently[5]) and *sensitive attributes* (attributes that shouldn't be linkable to an individual, e. g. information about religion, politics, health, etc.). ***K-anonymity*** is about ensuring that there are at least k entries with the same attribute combination (e. g., 2-anonymity ensures there are at least two entries), by removing or altering data, e. g. by applying *suppression* (replacing values with standard values), or *generalisation* (replacing individual values with broader categories or ranges). ***L-diversity*** and ***t-closeness*** are extensions of *k-anonymity*, addressing some of its weaknesses, such as homogeneity and background knowledge attacks.[6] All of them aim at the goal of addressing ***re-identification risks,*** i.e. "*the potential that some supposedly anonymous or pseudonymous data sets could be de-anonymized to recover the identities of users*."[7], and the domain is often referred to as ***Privacy-Preserving Data Publishing***.

Privacy-Preserving Data Publishing remains useful in many domains, but with the advent of AI, it does not seem sufficient anymore: The mentioned approaches and statistical techniques are designed to consider a limited number of selected attributes, but AI is about processing large amounts of data with high dimensionality and complexity, resulting in several new challenges: much higher likelihood of sensitive information being included, much higher likelihood of models being able to reveal sensitive information, and significantly increased difficulty in protecting sensitive information.

Research challenges

Privacy-Preserving AI (PPAI) is about addressing the specific challenges related to AI and privacy, which can be split into four categories (see also section on "*AI robustness*"):[8]

- ***Training Data Privacy***, which is about preventing malicious actors from reverse-engineering the training data.
- ***Input Privacy***, which is about preventing that a user's input data can be observed by other parties, including the model creator.

[5] Bettini, C., Wang, X. S., & Jajodia, S. (2006). The Role of Quasi-identifiers in k-Anonymity Revisited. arXiv:cs/0611035. http://arxiv.org/abs/cs/0611035

[6] Machanavajjhala, A., Gehrke, J., Kifer, D., & Venkitasubramaniam, M. (2006). L-diversity: Privacy beyond k-anonymity. 22nd International Conference on Data Engineering (ICDE'06), 24–24.

[7] Chia, P. H., Desfontaines, D., Perera, I. M., Simmons-Marengo, D., Li, C., Day, W.-Y., Wang, Q., & Guevara, M. (2019). KHyperLogLog: Estimating Reidentifiability and Joinability of Large Data at Scale. Proceedings of the 2019 IEEE Symposium on Security and Privacy.

[8] Thaine, P. (2020). Perfectly Privacy-Preserving AI. Medium. https://towardsdatascience.com/perfectly-privacy-preserving-ai-c14698f322f5

- ***Output Privacy***, which is about preventing that the output of a model is visible to anyone except the user whose data is being inferred upon.
- ***Model Privacy***, which is about preventing that the model is stolen.

Some of the most relevant attacks in the context of privacy and security include ***inference attacks***, i.e. attacks that aim at analyzing data to gain knowledge about a subject, and ***model poisoning***, i.e. attacks that manipulate data in order to influence or corrupt the model.

Among inference attacks, ***input inference attacks*** (also referred to as *model inversion* or *data extraction*) are probably the most common and relevant from a privacy perspective. Such attacks aim at extracting data from the training dataset, e.g. obtaining attributes, or audio or image training data related to a person based on her name. Similarly, ***membership inference*** and ***attribute inference*** attacks aim at finding out whether a particular example was in the dataset. It is noteworthy that the latter can not only be used as an attack, but also to check whether privacy-preserving measures were applied for training.

One of the most important approaches within the realm of privacy-preserving AI is ***Differential Privacy*** (DP). First introduced in 2006[9], it provides a mathematical definition of privacy that ensures that no individual data entry (e.g. referring to a specific user) has significant influence on the overall output distribution and hence no significant influence on query results. This is typically achieved by adding noise to the input, the output, or by modifying the query algorithm itself (Figure 1). With DP, the amount of information that can be gained about a given individual is limited to a specific value, thereby also providing a way to measure privacy. At the same time, overall accuracy does not significantly decrease (the statistical properties of a dataset are preserved). However, the costs and benefits of DP depend on the specific case at hand: the smaller the datasets, the more accuracy tends to decrease due to the added noise, while for larger datasets, accuracy may even *increase*, as the introduction of noise can reduce overfitting. Due to its advantages, especially regarding application in the context of AI and cloud computing, DP has seen a significant increase in relevance and demand within recent years, with DP libraries being developed and used by many major companies and vendors.[10] Moreover, further adaptations and variants to DP have been developed e. g. for specific needs related to AI training, including *Differentially Private Stochastic Gradient Descent* (DPSGD)[11,12] or *Private Aggregation of Teacher Ensembles* (PATE)[13].

[9] Dwork, C. (2006). Differential Privacy. In M. Bugliesi, B. Preneel, V. Sassone, & I. Wegener (Hrsg.), Automata, Languages and Programming (S. 1–12). Springer. https://doi.org/10.1007/11787006_1

[10] Examples of Differential Privacy libraries: https://github.com/IBM/differential-privacy-library (IBM), https://github.com/OpenMined/PipelineDP (Google / OpenMined), https://github.com/opendp/smartnoise-core (Microsoft), https://github.com/pytorch/opacus (Facebook)

[11] Song, S., Chaudhuri, K., & Sarwate, A. D. (2013). Stochastic gradient descent with differentially private updates. 2013 IEEE Global Conference on Signal and Information Processing, 245–248.

[12] Wu, X., Li, F., Kumar, A., Chaudhuri, K., Jha, S., & Naughton, J. F. (2017). Bolt-on Differential Privacy for Scalable Stochastic Gradient Descent-based Analytics. arXiv:1606.04722 [cs, stat]. http://arxiv.org/abs/1606.04722

[13] Papernot, N., Song, S., Mironov, I., Raghunathan, A., Talwar, K., & Erlingsson, Ú. (2018). Scalable Private Learning with PATE. arXiv:1802.08908 [cs, stat]. http://arxiv.org/abs/1802.08908

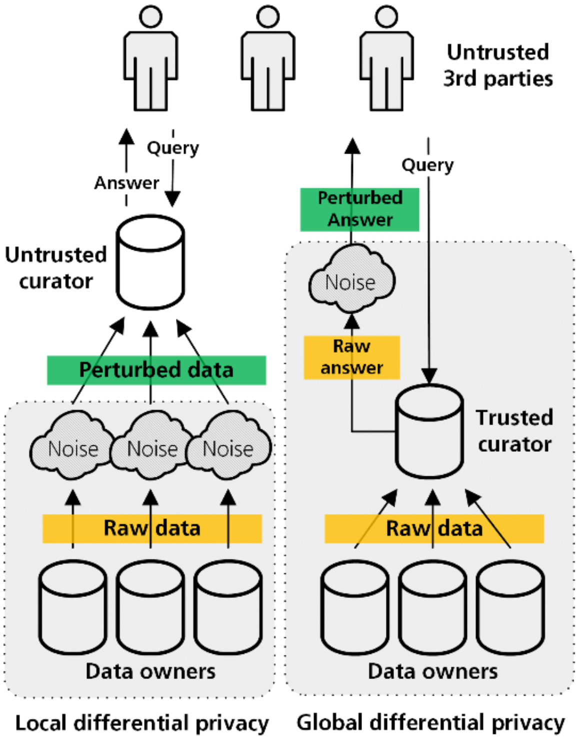

Figure 1: Local and global differential privacy.

Homomorphic Encryption (HE) represents another key technology in the PPAI domain: such encryption schemes can perform different classes of computations over encrypted data and can be split into *partially* and *somewhat* homomorphic encryption (which are limited with respect to operation type / amount), and *fully homomorphic encryption* (which can perform addition and multiplication any number of times). In the context of PPAI, HE is a powerful tool e.g. in that it can support data processing and training performed by an aggregator, without the aggregator gaining access to the clear-text data (Figure 2). One of the most relevant HE approaches is *the Cheon-Kim-Kim-Song* (CKKS) scheme[14], which has been implemented within several libraries[15] and is subject to standardisation activities.[16] One key challenge for the practical implementation of HE is computational and memory overhead, which varies significantly depending on the scheme and implementation used, but challenges also include practical security challenges such as key management. *Secure Multiparty Computation* (SMPC) provides yet another relevant cryptographic technique for PPAI. It can be used to jointly compute a function over inputs while keeping inputs private, serving as an addition to the aforementioned techniques.

[14] Cheon, J. H., Kim, A., Kim, M., & Song, Y. (2017). Homomorphic Encryption for Arithmetic of Approximate Numbers. In T. Takagi & T. Peyrin (Hrsg.), Advances in Cryptology – ASIACRYPT 2017 (S. 409–437). Springer International Publishing.

[15] e. g. HElib (IBM): https://github.com/homenc/HElib, SEAL (Microsoft): https://github.com/microsoft/SEAL

[16] Homomorphic Encryption Standardization. https://homomorphicencryption.org/standard/

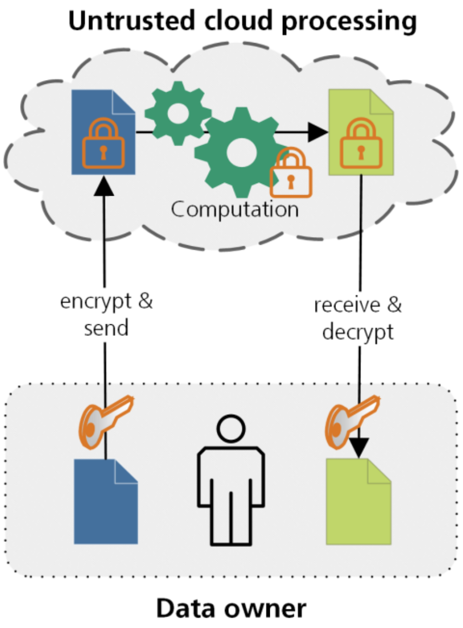

Figure 2: Homomorphic Encryption.

Another key tool for PPAI tools is ***Federated Learning*** (FL). First introduced in 2017[17], FL aims at conducting the training process among several participants, but without the need to exchange the training data: the data can remain "on prem", which means that FL provides great potential for many applications with respect to elevated security, copyright and privacy requirements (Figure 3). There are different variations to FL, ranging from centralised FL (with a central, orchestrating server) to decentralised FL (nodes / participants are able to organise themselves). As for the other PPAI techniques mentioned, practical application of FL is not trivial, and related challenges depend on the specific application context.[18] Also, for many applications, it is necessary to complement FL with HE and DP, as FL alone cannot prevent attacks such as inference attacks (to be addressed with HE and DP) or model poisoning (to be addressed with DP).

Finally, especially in the context of media, PPAI requires tools that can be used to remove person-related information from audio-visual material, e.g. by using source separation[19] or speech alienation[20] in the case of audio material.

[17] McMahan, B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. y. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 1273–1282

[18] A Comprehensive Survey of Privacy-preserving Federated Learning: A Taxonomy, Review, and Future Directions: ACM Computing Surveys: Vol 54, No 6. (o. J.). Abgerufen 4. February 2022

[19] Hennequin, R., Khlif, A., Voituret, F., & Moussallam, M. (2020). Spleeter: A fast and efficient music source separation tool with pre-trained models. Journal of Open Source Software, 5(50), 2154.

[20] Liang, D., Song, W., & Thomaz, E. (2020). Characterizing the Effect of Audio Degradation on Privacy Perception and Inference Performance in Audio-Based Human Activity Recognition. In 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services (S. 1–10). Association for Computing Machinery.

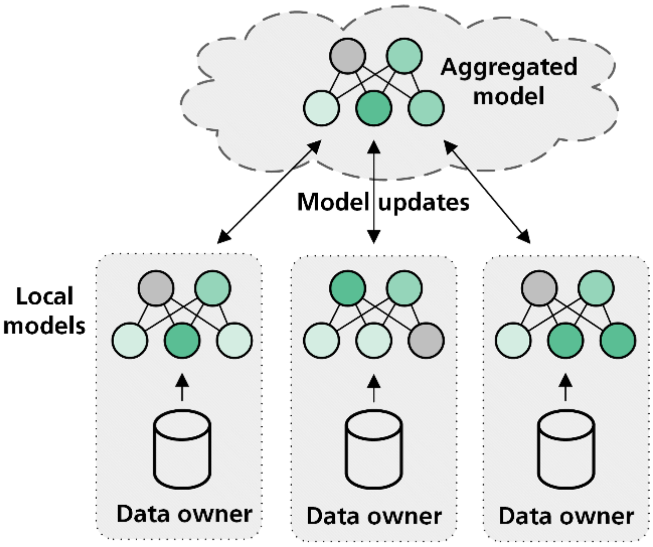

Figure 3: Federated Learning.

Societal and media industry drivers

Vignette 1: Privacy-enhanced news recommendations

Linda is very interested in politics and news. She is constantly looking for new tools that can help her receive news about the topics relevant to her, while covering a broad range of political perspectives: she is aware of the dangers of echo chambers and likes to read "other" opinions and news sources from time to time, even (or especially) if she does not agree with them, also to reflect her own arguments – she is active in a political party and therefore often involved in discussions. Finally, she finds a cross-vendor recommendation service for this, which uses her feedback to provide personalised news recommendations. It is a great service, providing high-quality information. However, Linda fears that the service learns a lot about her political views and preferences. She is afraid what could happen if such information is leaked or published, and checks information and reviews about the service. She learns that the service applies a range of privacy-preserving technologies including federated learning, homomorphic encryption, differential privacy and various standard security protocols, to ensure that no one except herself learns about her choices and preferences, and that no information can be connected to her real identity. She also learns that all these privacy and security claims have been tested and certified by well-known, independent security and privacy companies. She has finally found what she has been looking for.

Vignette 2: Privacy-enhanced speech transcription

Joseph is the owner of a small software company and has been using an app- and cloud-based speech transcription service from a big non-European vendor service since years, using it for a lot of purposes that involve highly personal information. He has recently grown increasingly concerned about what data is stored by the service, after reading about potential attacks and recent hacks of other services. He knows that the service uses speech recordings to continuously improve its performance – he agreed to do that before installing it, as it significantly improved performance. But apart from that, he does not know much about the service, and did not understand the terms of use in detail but confirmed them anyway, because it was the only working app for this purpose and he urgently needed it. Now, he is completely unsure whether speech recordings from him were uploaded and used, whether and how they were stored, and whether and how the created transcripts are stored and protected.

Actually, he uses many cloud-based services, e.g. for collaborative editing, issue management, backup and other means, without really understanding all legal details and technical risks involved. He is not happy at all about that, because he has a feeling that this comes with significant business and personal risks, but he gave up on the topic because he found that even legal or technical expert friends were hardly able to do that assessment. He would be willing to pay for more trustworthy alternatives with comparable performance and usability, but for some kinds of applications, he is simply not aware that any such alternatives exist.

One day, while spending his holidays with his family, Joseph learns that there was a phone scam that tricked his colleagues to transfer and effectively lose 150,000 € based on a replayed voice recording or deepfake, similarly to what happened in other cases.[21] At the same time, he learns that he is blackmailed with sensitive business information to be published if he does not agree to transfer another 350,000 € within the next days. He is not sure about where the respective leak came from but rejects to transfer 350,000 €, resulting in a publication of information about his clients and his private life that results in a very tough period for him and his family, but also for his company, which almost goes bankrupt in the months after the incident.

After many months of introspection, Joseph decides to turn the terrible experience into an opportunity, resulting in a strategic shift within his company: it starts developing software for transparent, privacy-aware speech transcription and note management services. After two years of very intense research and development in a strategic collaboration with other organisations being specialised in AI-based transcription, recommendation and privacy enhancing technologies, Joseph's company is now offering such services with good usability and performance, easy-to-understand terms of use, and supporting multiple languages. The services are continuously audited with respect to their privacy/security promises by 3rd party companies, and turn out to be a great market success, especially among security-/privacy-aware companies.

Future trends for the media sector

Privacy will be key for many media applications, including the following domains:

Recommendation will play an increasingly important role in the media sector, considering the ever-increasing amount of data and the need to deliver relevant information to audiences. Considering an increasing awareness about privacy, users will demand more transparency about which of their data is used and how, requiring that state-of-the-art technologies are applied to

[21] J. Damiani, “A Voice Deepfake Was Used To Scam A CEO Out Of $243,000”, Forbes (2019): https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/

protect their information and that technical audits are performed to ensure that such technologies are used (and used properly).

The ***processing of audio-visual data*** will frequently require the need to remove person-related information before content is stored and processed, using appropriate technologies.

The ***fabrication of synthetic audio and video*** material will bring an entirely new challenge for privacy – after all, being confronted with fabricated information about oneself that others wrongly consider authentic is even worse than losing control over authentic personal information. New technologies aiming at the detection and localisation of synthetic audio-visual material are currently being developed, but more awareness and modified processes within the media industry to deal with this problem are also required.

Services and offers in the media domain, as in any other domain, will only be successful if they combine good performance / usability etc. and privacy requirements. In other words, privacy and other trust aspects can become key features for commercial success, but only if we combine regulation (which is great, but not sufficient) with innovation.

In order to develop new AI-based tools and services for the media sector especially in Europe, ***"clean" datasets*** will be needed. It is good that Europe emphasises the need for privacy protection, but what is also needed is support and investment in creating and providing the appropriate alternative datasets.

AI does not only pose challenges. It can also be used to ***support and protect privacy***. For instance, increased automation and the use of AI can reduce the risk of data loss due to human error, it can improve auditing of privacy weaknesses in systems, it can improve transparency about the use of personal information, and it can support humans in becoming much more aware and rational about making cost-benefit considerations about privacy, including assessment of long-term costs and benefits (which human bias makes us especially unfit for).

Goals for next 10 or 20 years

Anonymisation techniques will be used to remove person-related data from audio-visual material, and differential privacy will provide metrics and means to guaranteed levels of privacy. Secure multiparty computation, homomorphic encryption and federated learning will allow AI training and data processing without the risks of unintended data loss and privacy violations. Media companies and society will have developed the organisational, educational and technical means to deal with manipulated, fabricated and decontextualised information as a potential threat to privacy. Principles and regulation regarding privacy will always be intertwined with respective innovation and investments in technology, and privacy and AI will not be treated as mutually incompatible anymore, but instead, privacy will become an integral feature of AI-based products.

AI4media

ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911

info@ai4media.eu www.ai4media.eu