Natural Language Processing

Contissa et al., (2018) presented a detailed report of preliminary results from a study aiming to automate the legal evaluation of privacy policies under the GDPR using ML. They outline the legal requirements the policies should meet. They analyzed 14 policies and tested the extent to which the analysis can be automated. None of the policies fully met the standards, with many containing vague language and potentially problematic processing. Their findings show promising results for automating the task, but more data is still needed. Martinelli, Marulli et al. (2020) discussed a methodology for enhancing privacy and data protection through NLP and AI. The methodology is aimed at building domain-specific knowledge datasets from documents that are unlabeled and trained by AI models to help identify sensitive information for obfuscation and to support human experts. It uses NLP techniques such as word embedding and topic modeling on corpora from different domains like healthcare and justice. The preliminary tests on 10,000 documents from these domains achieved promising precision, recall, and F1-scores when validated by domain experts. The authors consider the possibility of producing annotated corpora in a semi-automatic way, where humans' expertise is required only at the end of the loop, for validation and refinement of the data. Xing et al. (2023) compared privacy concerns related to AI between the USA and China, by reviewing data from Twitter and Weibo, and analyzing the textual data in these networks. They found that Americans are more concerned about privacy breaches through AI applications, while Chinese individuals are more optimistic about AI's role in privacy protection. Differences in opinion are attributed to cultural factors, with security, technology, and algorithms driving polarization in both countries. Their study offers insights into developing AI privacy policies. Zarifis et al. (2021) explored how trust and privacy concerns influence the acceptance of AI-driven health insurance services. They utilized the fact that in the insurance field natural language processing is extensively applied by virtual agents that interact with the consumer and AI is also used to detect fraudulent claims. They compared two scenarios: one where AI was not apparent in the interface and another where AI was explicitly integrated and visible. The study findings show that visible AI significantly reduces trust among consumers and increases privacy concerns, though not statistically significantly so. The findings suggest that the visibility of AI can influence user acceptance and trust in technology-driven health services. Shahriar, Allana, Hazratifard, and Dara (2023) explored privacy risks associated with the AI life cycle and proposed a framework for mitigating these risks through various privacy-enhancing technologies (PETs) and strategies. The focus of the paper is on the AI lifecycle, encompassing various branches of AI, including ML, expert systems, and NLP. The authors categorize privacy risks into four main groups: risk of identification, risk of making inaccurate decisions, risk of non-transparency, and risk of non-compliance with privacy regulations.

Each category encompasses a range of issues that could potentially compromise the integrity and confidentiality of personal data throughout the different phases of AI development. The survey emphasizes the importance of integrating privacy considerations from the onset of AI system development, advocating a 'privacy by design' approach. This includes the application of PETs at each stage of the AI life cycle and adherence to stringent regulatory frameworks such as the GDPR to ensure comprehensive data protection. Shahriar et al. also highlight the need for AI systems to maintain transparency and accountability, particularly in processes involving automated decision-making. They recommend that future research should focus on refining PETs, enhancing regulatory practices, and developing methods to improve the explainability and transparency of AI systems.

References

Contissa, G., Docter, K., Lagioia, F., Lippi, M., Micklitz, H.-W., Palka, P., . . . Torroni, P. (2018). Claudette meets GDPR: Automating the evaluation of privacy policies using artificial intelligence. *Available at SSRN 3208596*. doi:https://dx.doi.org/10.2139/ssrn.3208596

Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy. *IEEE Access*. doi:https://doi.org/10.1109/ACCESS.2023.3300381

Li, H., Guo, D., Fan, W., Xu, M., Huang, J., Meng, F., & Song, Y. (2023). Multi-step jailbreaking privacy attacks on chatgpt. *arXiv preprint arXiv:2304.05197*. doi:https://doi.org/10.48550/arXiv.2304.05197

Martinelli, F., Marulli, F., Mercaldo, F., Marrone, S., & Santone, A. (2020). Enhanced privacy and data protection using natural language processing and artificial intelligence. *2020 International Joint Conference on Neural Networks (IJCNN)* (pp. 1--8). IEEE. doi:https://doi.org/10.1109/IJCNN48605.2020.9206801

Mylrea, M., & Robinson, N. (2023). Artificial Intelligence (AI) trust framework and maturity model: applying an entropy lens to improve security, privacy, and ethical AI. *Entropy, 25*(10), 1429. doi:https://doi.org/10.3390/e25101429

Peres, R. S., Manta-Costa, A. a., & Barata, J. (2023). Implementing Privacy-Preserving and Collaborative Industrial AI. *IEEE Acces*. doi:https://doi.org/10.1109/ACCESS.2023.3296143

Shahriar, S., Allana, S., Hazratifard, S. M., & Dara, R. (2023). A survey of privacy risks and mitigation strategies in the Artificial Intelligence life cycle. *IEEE Access, 11*, 61829--61854. doi:https://doi.org/10.1109/ACCESS.2023.3287195

Wei, W., & Liu, L. (2024). Trustworthy distributed ai systems: Robustness, privacy, and governance. *ACM Computing Surveys*. doi:https://doi.org/10.1145/3645102

Xing, Y., He, W., Zhang, J. Z., & Cao, G. (2023). AI privacy opinions between US and Chinese people. *Journal of Computer Information Systems, 63*(3), 492--506. doi:https://doi.org/10.1080/08874417.2022.2079107

Zarifis, A., Kawalek, P., & Azadegan, A. (2021). Evaluating if trust and personal information privacy concerns are barriers to using health insurance that explicitly utilizes AI. *Journal of Internet Commerce, 20*(1), 66--83. doi:https://doi.org/10.1080/15332861.2020.1832817