Translation/Dialogue Tutorial: Purpose Limitation and Data Minimization in Data-Driven Systems*

Asia J. Biega
Max Planck Institute for Security and Privacy
Germany
asia.biega@mpi-sp.org

Michèle Finck
University of Tübingen
Germany
michele.finck@uni-tuebingen.de

KEYWORDS

GDPR, purpose limitation, data minimization, data governance

ACM Reference Format:
Asia J. Biega and Michèle Finck. 2018. Translation/Dialogue Tutorial: Purpose Limitation and Data Minimization in Data-Driven Systems. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/1122445.1122456

1 TUTORIAL OVERVIEW

Contemporary data-driven systems frequently process personal user data. As a result, they need to comply with data protection laws governing data processing for users from certain jurisdictions, such as the European Union's General Data Protection Regulation ('GDPR'). Purpose limitation and data minimization are two of the core GDPR principles. Unlike other principles, including fairness or transparency, they have not yet received as much attention from the FAccT community. As a result, their implementation poses a number of challenges and open research questions. This tutorial synthesizes the state-of-the-art knowledge about the two principles from across the (i) research literature in law and computer science, (ii) guidelines issued by data protection authorities, as well as (iii) relevant court rulings. We present recent advances in computational interpretations of the principles as well as highlighting future interdisciplinary research opportunities.

2 PRESENTING TEAM

Asia J. Biega: is a computer scientist and a tenure-track faculty member at the Max Planck Institute for Security and Privacy leading the Responsible Computing group. Her research centers around developing, examining and computationally operationalizing principles of responsible computing, data governance&ethics, and digital well-being. Before joining MPI-SP, Asia was a postdoctoral researcher at Microsoft Research Montréal in the Fairness, Accountability, Transparency, and Ethics in AI (FATE) Group. She completed her PhD in Computer Science at the MPI for Informatics and the MPI for Software Systems, winning the DBIS Dissertation Award of the German Informatics Society. In her work, Asia engages in interdisciplinary collaborations while drawing from her traditional Computer Science education and her industry experience, including at Microsoft and Google.

Michèle Finck: is a legal scholar and a professor of Law and Artificial Intelligence at the University of Tübingen, an Affiliated Fellow at the Max Planck Institute for Innovation and Competition in Munich and the Centre for Blockchain Technologies at University College London as well as a Visiting Professor at LUISS University in Rome. She previously worked at the University of Oxford and the London School of Economics. She is a member of a number of expert committees on digitalization, including the Council of Europe's Ad Hoc Committee on Artificial Intelligence ('CAHAI') and the European Commission's Blockchain Observatory and Forum. She has moreover advised national institutions as well as the European Commission and the European Parliament on different occasions. Her research focuses on artificial intelligence and the digital economy with a particular emphasis on data (protection) law and governance.

3 TUTORIAL DESCRIPTION

3.1 Purpose Limitation and Data Minimization

Purpose limitation (PL), a principle specified in Article 5(1)(b) of GDPR, requires that personal data shall be:

> *"collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes".*

Data minimization (DM), a principle specified in Article 5(1)(c) of GDPR, requires that data shall be

> *"adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed".*

PL and DM are GDPR's two core data protection principles. However, unlike fairness or transparency, they have thus far received less attention from the FAccT community. As a result, practitioners often struggle with adequate computational interpretations [12], highlighting a need and opportunity for interdisciplinary research in the spirit of FAccT.

*This document serves as a teaser overviewing the tutorial to be presented at FAccT '22. Please refer to Biega and Finck [1] if you're interested in the underlying material.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Woodstock '18, June 03–05, 2018, Woodstock, NY
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

3.2 Material Overview

This tutorial synthesizes the state-of-the-art knowledge about the two principles from across the (i) research literature in law [1, 5, 11, 13, 14], HCI and computer science [2, 6, 7, 10, 12], (ii) guidelines issued by data protection authorities [3, 4, 8, 9], as well as (iii) relevant court rulings. The presentation will be based on the presenters' recent techno-legal analysis of PL and DM published in Technology and Regulation [1], and will cover related work surveyed in the article (further examples beyond the papers listed in this paragraph are omitted in this tutorial overview for brevity).

3.3 Details and Timeline

The tutorial is planned for 90 minutes according to the following timeline and subtopics:

(10 mins) Introduction: Definitions of purpose limitation and data minimization in GDPR. Overview of the legal debates about the desirability of the principles. Overview of the computational evidence that data minimization could be implemented to a larger extent than it currently is in practice.

(20 mins) Purpose limitation: Legal theory and components of PL (specificity, explicitness, legitimacy, compatible use). Evidence of current implementations by online platforms. Recent computational approaches that define purpose as service improvement. Computational challenges and open questions.

(10 mins) Repurposing data: The means and conditions under which service providers can repurpose data (scientific research, statistical purposes, consent). Overview of the associated practical challenges (beyond the computational and legal challenges, research has uncovered, for instance, the potential relationship between organizational structure of a company and their ability to repurpose data).

(20 mins) Data minimization: Legal theory and components of DM (relevance, adequacy, necessity). Different types of minimization: data quantity, data quality. Minimization of special categories of data. Evidence of current implementations by online platforms. Recent computational approaches, challenges, and open questions.

(10 mins) Trade-offs in data protection: An in-depth analysis of PL and DM reveals a number of data protection challenges and trade-offs. We will discuss (i) the tension between the generality of legal principles and the need for computationally operational interpretations, (ii) the unacknowledged trade-offs between various GDPR principles (e.g., data minimization and fairness), (iii) the economic and environmental costs of enforcing data subject rights, (iv) the cost of compliance and the (un)likelihood of enforcement.

(5 mins) Outlook: Short-term recommendations for practitioners. Identified long-term research questions.

(15 mins) Q&A.

REFERENCES

[1] Asia J. Biega and Michèle Finck. 2021. Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems. *Technology and Regulation* (2021).

[2] Asia J. Biega, Peter Potash, Hal Daumé, Fernando Diaz, and Michèle Finck. 2020. Operationalizing the Legal Principle of Data Minimization for Personalization. In *ACM(43) SIGIR '20*. 399–408.

[3] Reuben Binns and Valeria Gallo. 2019. *Data minimisation and privacy-preserving techniques in AI systems.* https://ico.org.uk/about-the-ico/news-and-events/ai-blog-data-minimisation-and-privacy-preserving-techniques-in-ai-systems/

[4] Norwegian Data Protection Authority Datatilsynet. 2018. *Artificial Intelligence and Privacy.*

[5] Imane Fouad, Cristiana Santos, Feras Al Kassar, Nataliia Bielova, and Stefano Calzavara. 2020. On compliance of cookie purposes with the purpose specification principle. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 326–333.

[6] Gemma Galdon Clavell, Mariano Martín Zamorano, Carlos Castillo, Oliver Smith, and Aleksandar Matic. 2020. Auditing Algorithms: On Lessons Learned and the Risks of Data Minimization. In *AAAI-ACM-AIES*. 265–271.

[7] Abigail Goldsteen, Gilad Ezov, Ron Shmelkin, Micha Moffie, and Ariel Farkash. 2021. Data minimization for GDPR compliance in machine learning models. *AI and Ethics* (2021), 1–15.

[8] Information Commissioner's Office ICO. 2017. *Big Data, Artificial Intelligence, Machine Learning and Data Protection.*

[9] UK Information Commisioner's Office: ICO. 2018. *Guide to the General Data Protection Regulation (GDPR). Principle (c): Data minimisation.* Retrieved Jan 22, 2020 from https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/

[10] Bashir Rastegarpanah, Krishna Gummadi, and Mark Crovella. 2021. Auditing Black-Box Prediction Models for Data Minimization Compliance. *Advances in Neural Information Processing Systems* 34 (2021).

[11] Cristiana Santos, Nataliia Bielova, and Célestin Matte. 2020. Are cookie banners indeed compliant with the law?: Deciphering EU legal requirements on consent and technical means to verify compliance of cookie banners. *Technology and Regulation* 2020 (2020), 91–135.

[12] Awanthika Senarath and Nalin Asanka Gamagedara Arachchilage. 2018. Understanding Software Developers' Approach towards Implementing Data Minimization. *arXiv preprint arXiv:1808.01479* (2018).

[13] Bart van der Sloot. 2013. From data minimization to data minimummization. In *Discrimination and Privacy in the Information Society*. Springer, 273–287.

[14] Tal Z Zarsky. 2017. Incompatible: The GDPR in the Age of Big Data. *Seton Hall Law Review* 47, 4 (2017), 2.