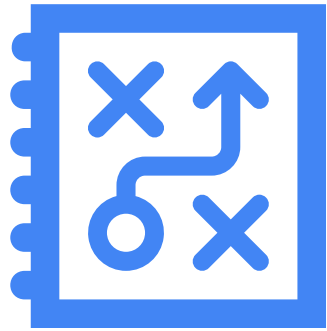


Christian Peukert

Some thoughts on the economics and regulation of data

CSTD Workshop on Data Governance Principles, Sep 2024

The AI value chain and data-driven business models



Algorithms



Data



Computation



Models



Insight



Behavioral information (→ online ads)

- actions, reactions, patterns, e.g. web browsing data, app interaction data

Sensor measurements (→ robotics)

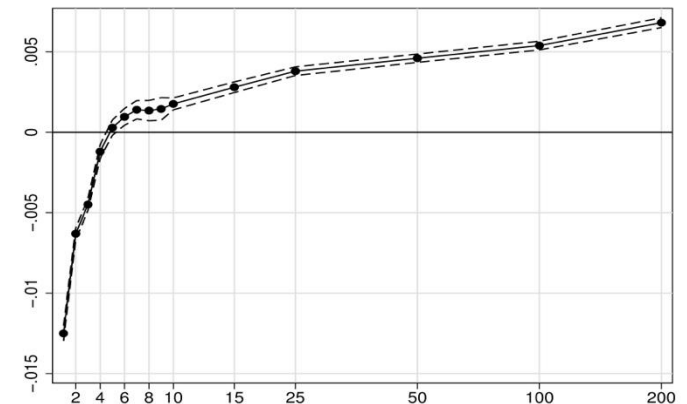
- physical devices with sensors, e.g. IoT

Human-generated content (→ genAI)

- text, images, sound, video

Economics of data

- Access to information can shift economic power
 - Easier for firms to charge (higher) individualized prices
 - Easier for consumers to find (low-priced) products that match their taste
- Sharing of information can have externalities
 - Helping to generate valuable insights for society as whole (e.g. health information)
 - Information disclosure of others can help to infer information about non-disclosing individuals (users versus non-users of platforms)
- Decreasing returns to data
 - More data helps to increase quality of prediction/models, but at a decreasing rate



Stock versus flow of data



- The potentially usable **stock of data** online is huge
- Much of the stock was created before it became commercially valuable
- In some cases, no need to update training data (e.g. cat images)



- Once the stock is exhausted, **flow of data** becomes important
- In dynamic settings, the stock of data can get stale
 - The **flow of data** is endogenously determined
 - Function of economic incentives (and regulation)

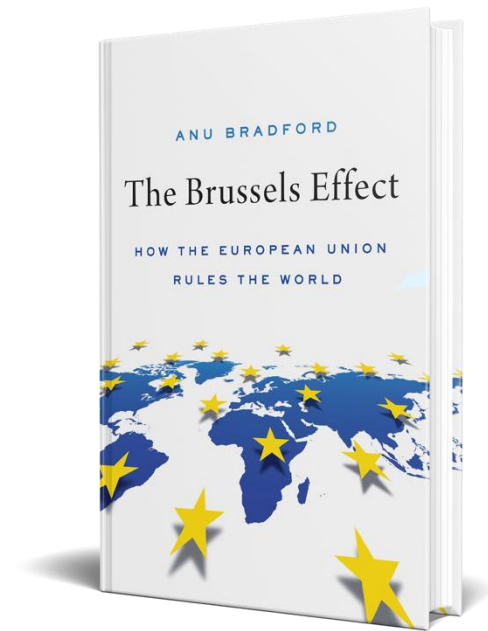
Incentive structures of data providers

- Article 10, EU AI Act:
“Training, validation and testing datasets shall be relevant, sufficiently representative, and to the best extent possible, free of errors and complete in view of the intended purpose.”
- But this can clash with the incentives structures of data providers
 - How can a dataset be “sufficiently representative” and “complete” when rightsholders can exercise exclusive rights (Levendowski, 2018)?
 - A large audit of ten thousands of web domains shows rising AI data restrictions, with 45% now limited, impacting the diversity and scalability of AI systems (Longpre et al., 2024)
 - After a stock photo website released images for AI training, photographers deleted accounts or uploaded less, and more “boring” content (Peukert et al. 2024)

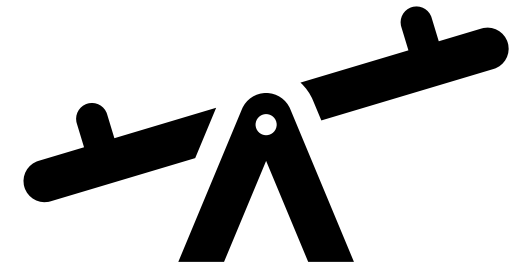
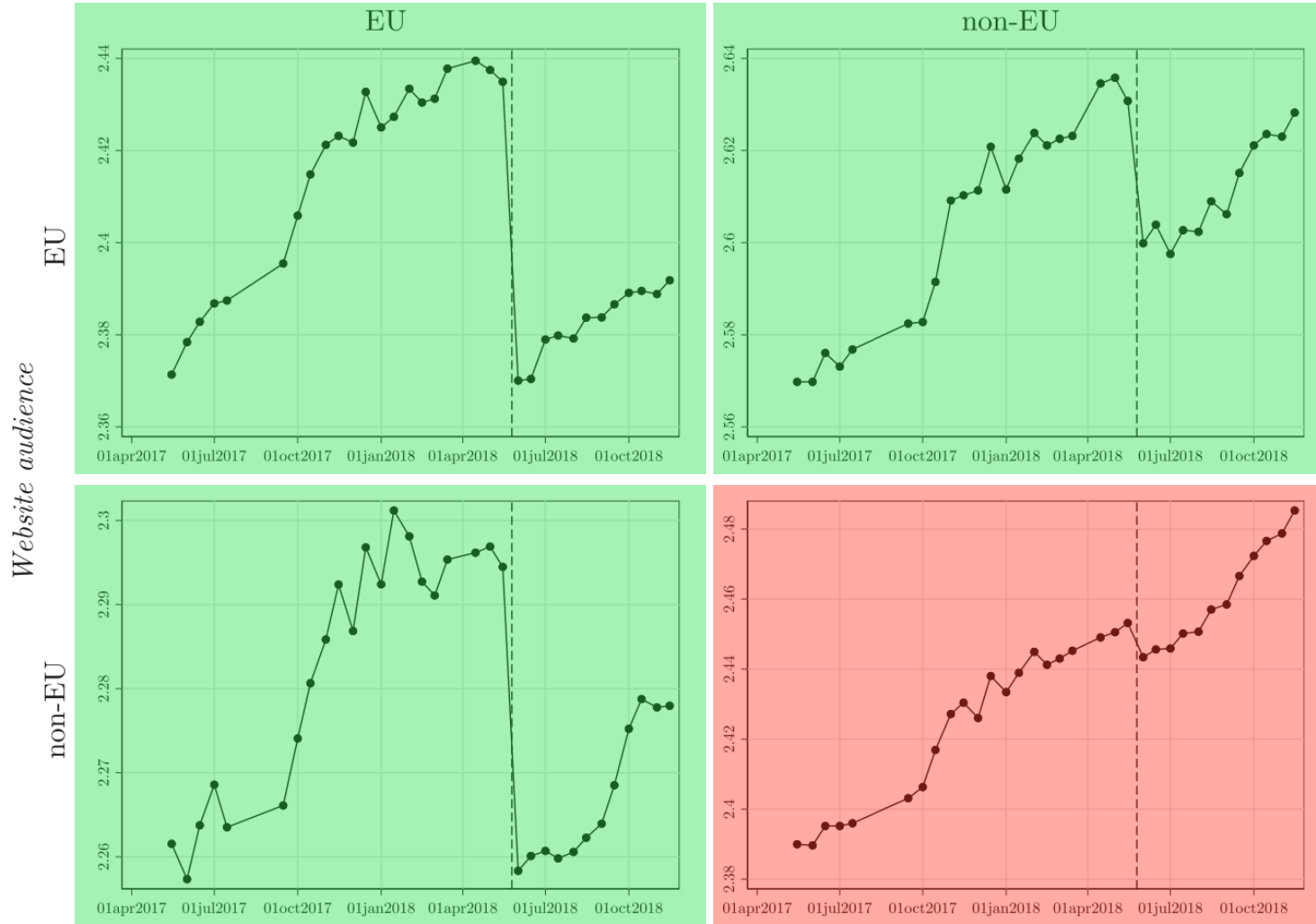
Global data governance - *Example: GDPR*

- GDPR applies to companies offering goods or services to individuals located in the EU, regardless of whether the companies are located in the EU or not

		Firm	
		EU	Non-EU
Customer	EU	Standard case Art. 3(1) GDPR	Effects doctrine Art. 3(2) GDPR
	Non-EU	Establishment principle Art. 3(1) GDPR	"Brussels effect"?



GDPR reduced cookies – also where it shouldn't have



Tradeoff between **data protection** (privacy, copyright, trade secrets, ...) and **innovation**

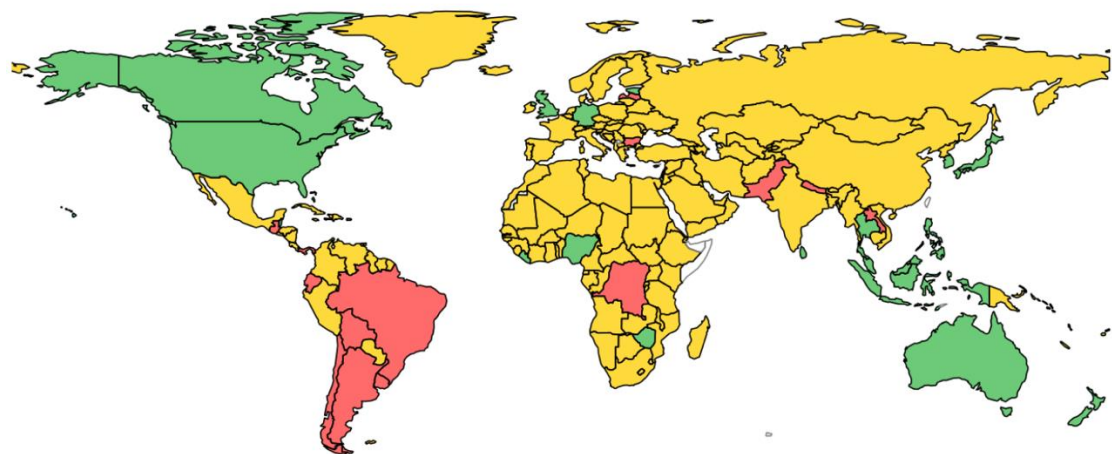
Restrictive data access can slow down innovation

- GDPR increased concentration in web tracking (Peukert et al., 2022; Johnson et al., 2024)
- GDPR reduced technology venture funding (Jia et al., 2021)
- GDPR increased exit in the mobile app market (Kircher & Foerderer, 2021; Janssen et al., 2022)
- GDPR led to a substantial shift from radical to incremental product innovation (Blind et al., 2023)
- Countries without copyright exceptions that allow text-and-data-mining see less AI R&D (academic papers and open source code) and less AI commercialization (patents, investments, startups) (Peukert, *forthcoming*)

DATA ACCESS

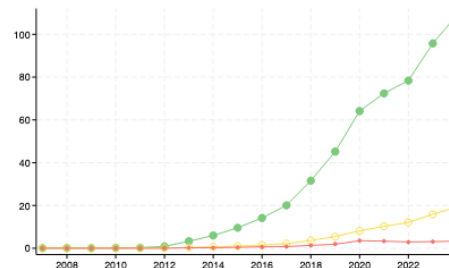
Legal reform to enhance global text and data mining research

Outdated copyright laws around the world hinder research

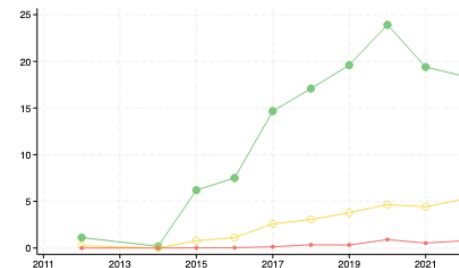


Fiil-Flynn et al. (2022)

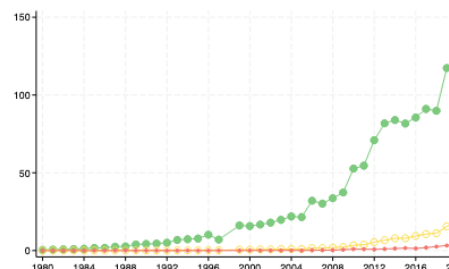
A: Papers (arXiv)



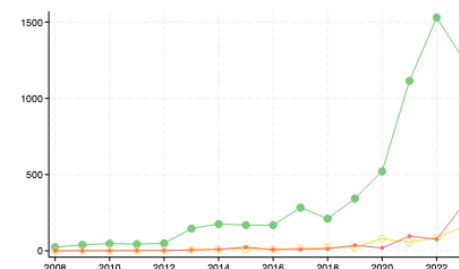
B: Code (GitHub)



C: Patents (USPTO)



D: Ventures (Crunchbase)



Peukert (forthcoming)

Data governance policy needs to find a balance

Research and
development;
business goals



Interests of society

Interests of
data holders
(privacy and IP)

References

- Blind, K., Niebel, C., & Rammer, C. (2023). *The impact of the EU General data protection regulation on product innovation*. Industry and Innovation.
- Fiil-Flynn, S.M., Butler, B., Carroll, M., Cohen-Sasson, O., Craig, C., Guibault, L., Jaszi, P., Jütte, B.J., Katz, A., Quintais, J.P. and Margoni, T., 2022. *Legal reform to enhance global text and data mining research*. Science.
- Janssen, R., Kesler, R., Kummer, M.E. and Waldfogel, J., 2022. *GDPR and the lost generation of innovative apps*. Working Paper.
- Jia, J., Jin, G.Z. and Wagman, L., 2021. The short-run effects of the general data protection regulation on technology venture investment. Marketing Science.
- Johnson, G.A., Shriver, S.K. and Goldberg, S.G., 2023. *Privacy and market concentration: intended and unintended consequences of the GDPR*. Management Science.
- Kircher, T. and Foerderer, J., 2021, December. *Does EU-consumer privacy harm financing of US-app-startups? Within-US evidence of cross-EU-effects*. Proceedings of the 42nd International Conference on Information Systems (ICIS).
- Levendowski, A., 2018. *How copyright law can fix artificial intelligence's implicit bias problem*. Wash. L. Rev., 93, p.579.
- Longpre, S., Mahari, R., Lee, A., Lund, C., Oderinwale, H., Brannon, W., Saxena, N., Obeng-Marnu, N., South, T., Hunter, C., Klyman, K., Klamm, C., Schoelkopf, H., Singh, N., Cherep, M., Anis, A.M., Dinh, A., Chitongo, C., Yin, D., Sileo, D., Mataciunas, D., Misra, D., Alghamdi, E., Shippole, E., Zhang, J., Materzynska, J., Qian, K., Tiwary, K., Miranda, L., Dey, M., Liang, M., Hamdy, M., Muennighoff, N., Ye, S., Kim, S., Mohanty, S., Gupta, V., Sharma, V., Chien, V.M., Zhou, X., Li, Y., Xiong, C., Villa, L., Biderman, S., Li, H., Ippolito, D., Hooker, S., Kabbara, J., and Pentland, S. (2024). *Consent in Crisis: The Rapid Decline of the AI Data Commons*. Working Paper.
- Peukert, C. (2025). *Copyright and the Dynamics of Innovation in Artificial Intelligence*. Proceedings of the 58th Hawaii International Conference on System Sciences (HICSS).
- Peukert, C., Bechtold, S., Batikas, M. and Kretschmer, T. (2022). *Regulatory Spillovers and Data Governance: Evidence from the GDPR*. Marketing Science.
- Peukert, C., Abeillon, F., Haese, J., Kaiser, F., and Staub, A. (2024). *Strategic Behavior and AI Training Data*. Working Paper.