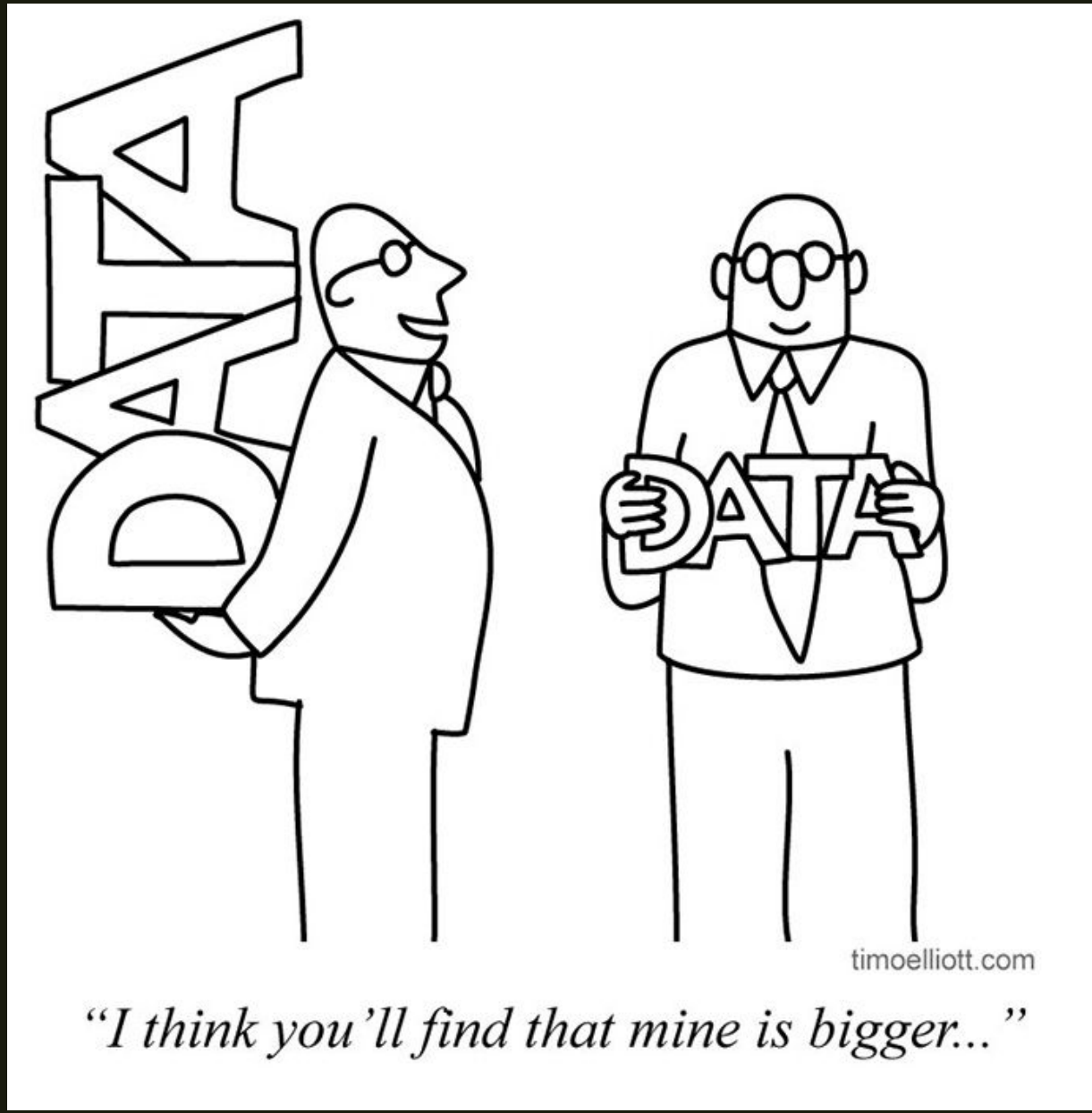
LEGAL PROTECTION BY DESIGN FOR ML



Law and CS: an Agenda

- Black boxing the 'other'
- From interdisciplinary soup to cross-disciplinary intercourse
- Law and CS as architectures that shape our world
- By design approaches
- Legal by Design and Legal Protection by Design
- Design principles for ML

Black Boxing the Other

- **CS Folk** make their own definitions of privacy and fairness, based on disparate input from lawyers they happen to know
- **Lawyers** buy into unsubstantiated AI narratives, 'legel tech' and ML that supposedly 'outperforms' experts
- **CS Folk** joining the naïve mantra that 'regulation' is always behind the facts
- **Lawyers** advise clients based on 'beliefs' about 'economic forces' or 'all the good things that AI achieves', as if these are facts

From Interdisciplinary Soup …

- Both disciplines **watering down** their content to accommodate the other
- Translating own vocabulary and grammar into what they think the other will understand
- Resulting in a 'shared vocabulary' that **means nothing in either discipline**
- **Cherry picking** findings of the other discipline without understanding its contested status:
  - *CS using unsubstantiated theory of psychometrics or nudge theory*
  - *Law using unsubstantiated theory of neo-classical economics or behavioural economics*

… to Cross-Disciplinary Intercourse

- Understanding the web of meaning that informs the other discipline
- *Web: **terms are always defined in terms of other terms***
- *Web: the underlying structure (assumptions, beliefs, aims)*
- *Including controversies and contestation*

1. *E.g. term like 'legal effect' is explained in terms of 'legal status', 'legal rights', etc.*
   - Term like 'loss function' is explained in terms of 'mathematical optimisation'
2. *E.g. in CS we have assumptions re formalisation, logical operators, disambiguation*
   - Term like 'legal subjectivity' does not entail consciousness
3. *E.g. in CS there is controversy around the issue of bias in ML (incorrect or 'merely' unfair)*
   - Term like consent in relation to cookies, used for 'additional processing' is contested in law

… to Cross-Disciplinary Intercourse

- Speaking one's own language while hearing the other
- *Webs: **clarity comes from language usage** (intra-liguistic coherence)*
- *Webs: the internal usage co-determines the extra-linguistic reference (speech act theory)*
- *Discovering conflicting 'makings' of our shared world*

1. E.g. to understand private law liability requires study of case law
- *The term 'optimisation' in ML must be understood as approximation of target function*
2. E.g. to understand tort law you need experience of real life events that cause harm
- *Code and compiler ultimately refer to instructions at the level of computer hardware*
3. The formalisation that is inherent in CS may be in conflict with the values embedded in the use of natural language
- *The legal conditions imposed on data controllers may be hard or impossible to implement in backend systems*

Law and CS as Architectures that Shape the World

- Law shapes our world by way of speech acts
  - *Constitutive nature of positive law*
  - *Concepts such as legal effect, legal competence, legal subjectivity*
  - *Shape our institutional world, determine our 'decision space'*

- CS shapes our world by way of disambiguated code
  - *Constitutive of our choice architecture*
  - *Does not depend on our understanding*
  - *Does not necessarily afford contestation*

By Design Approaches

- **Naïve:** let's design the ideal world (ideal according to whom, for whom?)
  - *Naïve understanding of technology, of humans, and their interaction*
  - *Techno-solutionism (techno-regulation)*
  - *Legal by Design*

- **Informed:** let's acknowledge that technology matters, makes a difference
  - *How can we protect fundamental values, norms, power balance, individual rights*
  - *Responsible AI*
  - *Legal Protection by Design*

Legal by Design

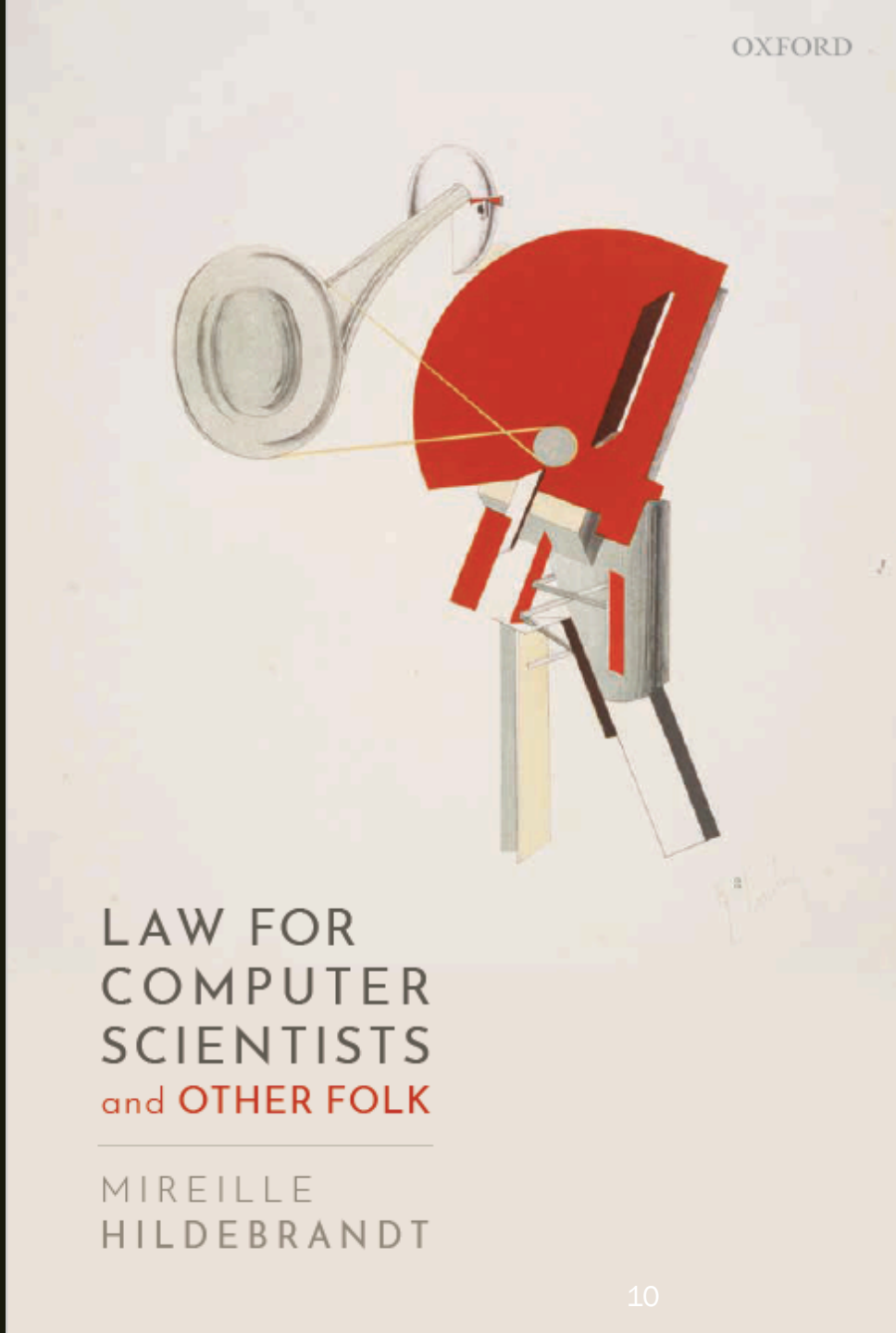
- Code-driven:
  GOFAI IFTTT (decision trees, assuming complete foresight, evoking appeal or excluding contestability)

- Data-driven:
  personalisation of law (micro-targeting law, increasing complexity, reducing legal certainty)

Legal Protection by Design

- Orientation towards constitutive goals of the law: justice, legal certainty and purposiveness
- Embedding checks and balances of the rule of law in computational architectures
- Reconfiguration of backend systems checking affordances, forward engineering

https://global.oup.com/academic/product/law-for-computer-scientists-and-other-folk-9780198860884?facet_narrowbybinding_facet=Paperback&lang=en&cc=pt (open access)



LPbD Principles for ML

- **Work in progress:**
  - *Data minimisation (e.g. Throttling ML, Paul Ohm)*
  - *Purpose limitation (e.g. Agonistic ML, Hildebrandt)*
  - *Integrity and confidentiality (e.g. Defense against Fingerprinting, Miro)*
  - *Transparency (e.g. transparency design and post hoc explanation, Xu et al.)*
  - *Reliability (e.g. new frontiers in transparency, Cabitza et al.)*

- Cp. the CONSORT and SPIRIT Guidelines for reporting and protocols of clinical trials involving AI

references

Cabitza, Federico, Andrea Campagner, and Davide Ciucci, 'New Frontiers in Explainable AI: Understanding the GI to Interpret the GO', in *Machine Learning and Knowledge Extraction*, ed. by Andreas Holzinger, Peter Kieseberg, A Min Tjoa, and Edgar Weippl, Lecture Notes in Computer Science (Cham, 2019), 27–47 <http://link.springer.com/10.1007/978-3-030-29726-8_3> [accessed 29 September 2020]

Hildebrandt, Mireille, 'Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning', *Theoretical Inquiries in Law*, 20/1 (2019) <http://www7.tau.ac.il/ojs/index.php/til/article/view/1622> [accessed 4 February 2019]

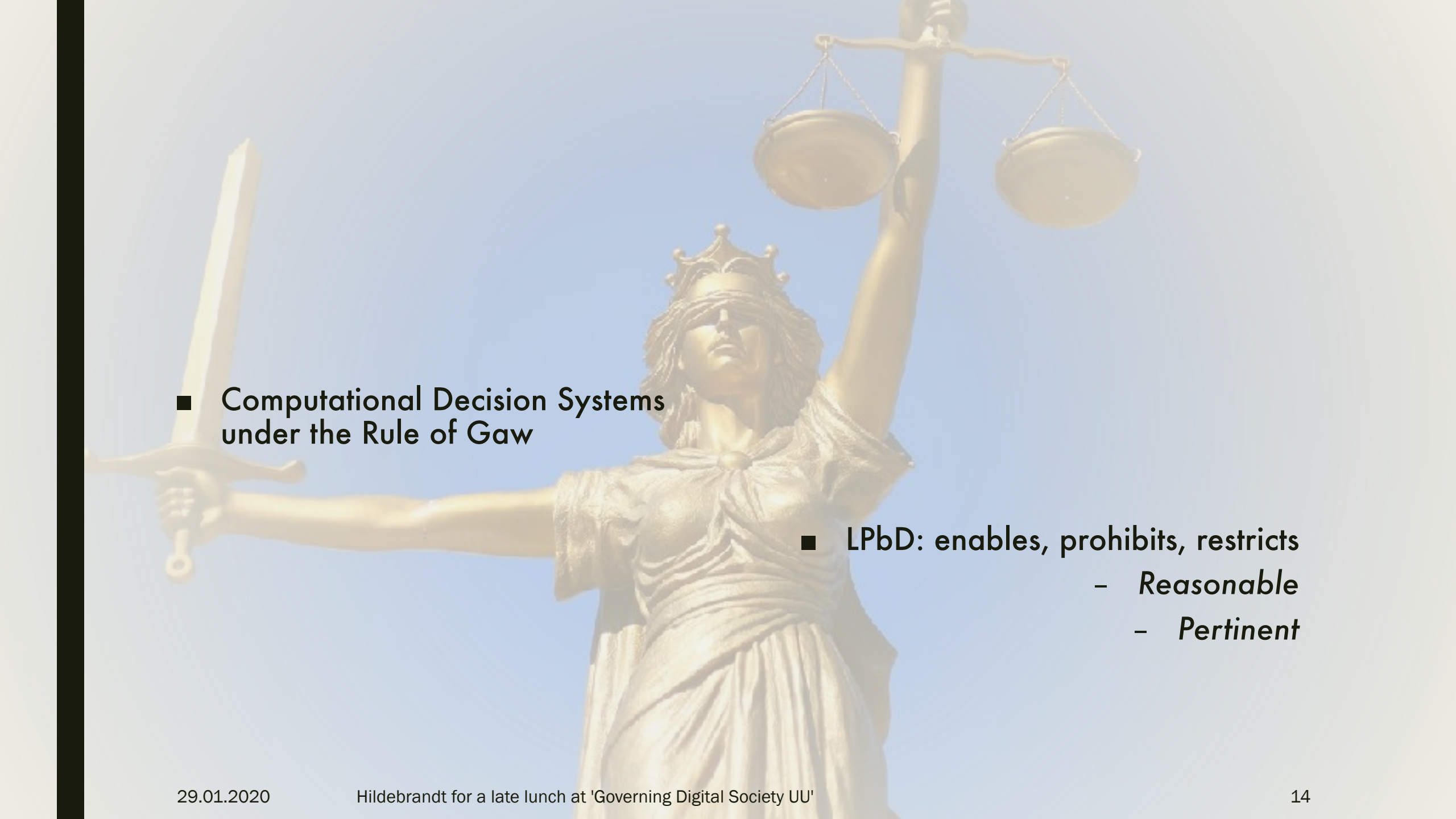
Juarez, Marc, Mohsen Imani, Mike Perry, Claudia Diaz, and Matthew Wright, 'Toward an Efficient Website Fingerprinting Defense', in *Computer Security – ESORICS 2016*, ed. by Ioannis Askoxylakis, Sotiris Ioannidis, Sokratis Katsikas, and Catherine Meadows, Lecture Notes in Computer Science (Cham, 2016), 27–46

Ohm, Paul, 'Throttling Machine Learning', *Life and the Law in the Era of Data-Driven Agency*, 2020 <https://www.elgaronline.com/view/edcoll/9781788971997/9781788971997.00019.xml> [accessed 11 September 2020]

Xu, Feiyu, Hans Uszkoreit, Yangzhou Du, Wei Fan, Dongyan Zhao, and Jun Zhu, 'Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges', in *Natural Language Processing and Chinese Computing*, ed. by Jie Tang, Min-Yen Kan, Dongyan Zhao, Sujian Li, and Hongying Zan, Lecture Notes in Computer Science (Cham, 2019), 563–74

references

- Cruz Rivera, Samantha, Xiaoxuan Liu, An-Wen Chan, Alastair K. Denniston, and Melanie J. Calvert, 'Guidelines for Clinical Trial Protocols for Interventions Involving Artificial Intelligence: The SPIRIT-AI Extension', *Nature Medicine*, 26/9 (2020), 1351–63
- Liu, Xiaoxuan, Samantha Cruz Rivera, David Moher, Melanie J. Calvert, and Alastair K. Denniston, 'Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: The CONSORT-AI Extension', *Nature Medicine*, 26/9 (2020), 1364–74



- Computational Decision Systems under the Rule of Gaw

- LPbD: enables, prohibits, restricts
  - *Reasonable*
  - *Pertinent*