DOI:10.1145/3378061

How design and operation of modern cloud-scale systems conflict with GDPR.

BY SUPREETH SHASTRI, MELISSA WASSERMAN, AND VIJAY CHIDAMBARAM

GDPR Anti-Patterns

THE GENERAL DATA Protection Regulation (GDPR)[26] is a European privacy law introduced to offer new rights and protections to people concerning their personal data. While at-scale monetization of personal data has existed since the dot-com era, a systemic disregard for privacy and protection of personal data is a recent phenomenon. For example, in 2017, we learned about Equifax's negligence[17] in following the security protocols, which exposed the financial records of 145 million people; Yahoo!'s delayed confession[21] that three years ago, a theft had exposed all three billion of its user records; Facebook's admission[33] that their APIs allowed illegal harvesting of user data, which in turn influenced the U.S. and U.K. democratic processes.

Thus, GDPR was enacted to prevent a widespread and systemic abuse of personal data. At its core, GDPR declares the privacy and protection of personal data as a fundamental right. Accordingly, it grants new rights to people, and assigns companies that collect their personal data, new responsibilities. Any company dealing with the personal data of European people is legally bound to comply with all the regulations of GDPR, or risk facing hefty financial penalties. For example, in January 2019, Google was fined[6] €50M for lacking a customer's consent in personalizing advertisements across their different services.

In this work, we investigate the challenges that modern cloud-scale systems face in complying with GDPR.

Specifically, we focus on the design principles and operational practices of these systems that conflict with the requirements of GDPR. To capture this tussle, we introduce the notion of GDPR *anti-patterns*. In contrast to outright bad behavior, say storing customer passwords in plaintext, GDPR anti-patterns are those practices that serve their originally intended purpose well but violate the norms of GDPR. For example, given the commercial value of personal data, modern systems have naturally evolved to store them without a clear timeline for deletion, and to reuse them across various applications. While these practices help the systems generate more revenue and thereby value, they violate the storage and purpose limitations of GDPR.

Building on our work analyzing GDPR from a systems perspective,[30–32] we identify six GDPR anti-patterns that are widely present in the real world. These include storing personal data without a timeline for deletion; reusing personal data indiscriminately; creating black markets for personal data; risk-agnostic data processing; hiding data breaches; and making unexplainable decisions. These anti-patterns highlight how the traditional system design goals of optimizing for performance, cost, and reliability sit at odds with GDPR's goal of data protection by design and by default. While eliminating these anti-patterns is not enough to achieve overall compliance under GDPR, ignoring these will definitely violate its intents.

We structure the rest of this article as follows: First, we provide a brief primer on GDPR, then describe the six GDPR anti-patterns, discussing how they came to be, reviewing the conflicting regulations, and chronicling their real-world implications. Finally, we ruminate on the challenges and opportunities for system designers as societies embrace data protection regulations.

GDPR

On May 25, 2018, the European Parliament rolled out the General Data Protection Regulation.[26] In contrast with targeted privacy regulations like HIPAA and FERPA in the United States, GDPR takes a comprehensive view by defining *personal data* to be any information relating to an identifiable natural person. GDPR defines three entities that interact with personal data: *data subject*, the person whose personal data is collected; *data controller*, the entity that collects and uses personal data; and, *data processor*, the entity that processes personal data on behalf of a data controller. Then, GDPR designates supervisory authorities (one per EU country) to oversee that the rights and responsibilities of GDPR are complied with.

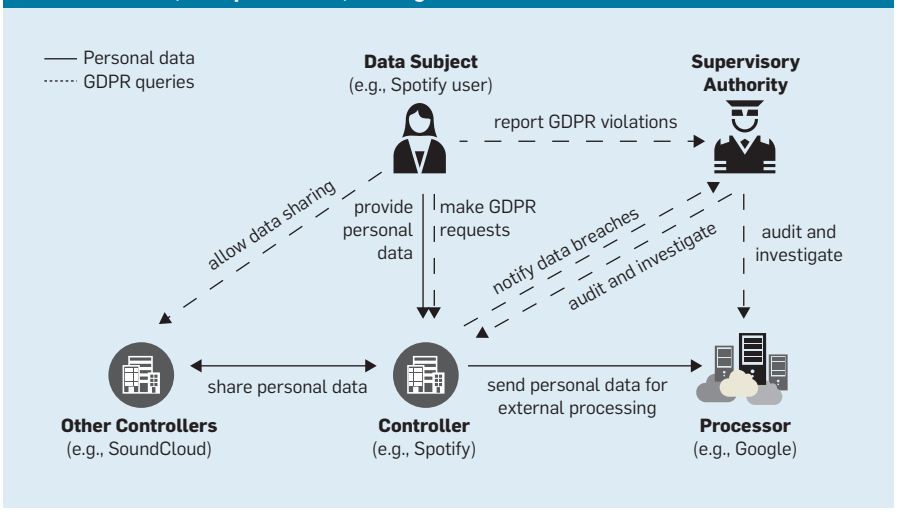
The accompanying figure represents how GDPR entities interact with each other in collecting, storing, processing, securing, and sharing personal data. Consider the music streaming company Spotify collecting its customers' listening history, and then using Google cloud's services to determine new recommendations for customers. In this scenario, Spotify is the data controller and Google Cloud is the data processor. Spotify could also engage with other data controllers, say SoundCloud, to gather additional personal data of their customers.

To ensure privacy and protection of personal data in such ecosystems, GDPR grants new rights to customers and assigns responsibilities to controllers and processors. Now, any person can request a controller to grant access to all their personal data, to rectify errors, to request deletion, to object to their data being used for specific purposes, to port their data to third parties and so on. On the other hand, the controller is required to obtain people's consent before using their personal data, to notify them of data breaches within 72 hours of finding out, to design systems that are secure by design and by default, and to maintain records of activities performed on personal data. For controllers failing to comply with these rights and responsibilities, GDPR regulators could levy penalties of up to €20M or 4% of their annual global revenue, whichever is higher.

Structure. GDPR is organized as 99 *articles* that describe its legal requirements, and 173 *recitals* that provide additional context and clarifications to these articles. The first 11 articles layout the principles of data privacy; articles 12–23 establish the rights of the people; then articles 24–50 mandate the responsibilities of the data controllers and processors; the next 26 articles describe the role and tasks of supervisory authorities; and the remainder of the articles cover liabilities, penalties and specific situations. We expand on the relevant articles later.

Impact. Compliance with GDPR has been a challenge for many companies that collect personal data. A number of companies like Klout and Unroll.me terminated their services in Europe to avoid the hassles of compliance. Few other businesses made temporary modifications. For example, media site *USA Today* turned off all advertisements, whereas the *New York Times* stopped serving personalized ads. While most organizations are working toward compliance, Gartner reports[13]

Flow of personal data and GDPR queries between the four GDPR entities: data subjects, data controllers, data processors, and regulators.



that less than 50% of the companies affected by GDPR were compliant by the end of 2018. This challenge is further exacerbated by the performance impact that GDPR compliance imposes on current systems.[30]

In contrast, people have been enthusiastically exercising their newfound rights. In fact, the EU data protection board reports[12] having received 144,376 complaints from individuals and organizations in the first year of GDPR. Surprisingly, even the companies have been forthcoming in reporting their security failures and data breaches, with 89,271 breach notifications sent to regulators in the same 12-month period. In 2019, several companies have been levied hefty penalties for GDPR violations: €50 million for Google,[6] £99M for Marriott International,[25] and £183M for British Airways.[24]

GDPR Anti-Patterns

The notion of anti-patterns was first introduced[19] by Andrew Koenig to characterize patterns of software design and behavior that superficially look like good solutions but end up being counterproductive in reality. An example of this is performing premature optimizations in software systems. Extending this notion, we define the term GDPR anti-patterns to refer to system designs and operational practices, which are effective in their own context but violate the rights and regulations of GDPR. Naturally, our definition does not include design choices that are bad in their own right, say storing customer passwords in plaintext, though they also violate GDPR norms. In this section, we catalog six GDPR anti-patterns, detailing how they came to be, which regulations they violate, and their implications in the real-world.

Genesis. GDPR anti-patterns presented here have evolved from the practices and design considerations of the post dot-com era (circa 2000). These modern cloud-scale systems could be characterized by their quest for unprecedented scalability, reliability, and affordability. For example, Google operates eight global-scale applications at 99.99% uptime with each of them supporting more than one billion users. Similarly, Amazon's cloud computing infrastructure provides on-demand access to inexpensive computing to over 1 million users in 190 countries, all the while guaranteeing four nines of availability. This unrelenting focus on performance, cost-efficiency, reliability, and scalability has resulted in relegating security and privacy to a backseat.

While our GDPR analysis recognizes six anti-patterns, this list is not comprehensive. There are many other

unsavory practices that would not stand the regulator scrutiny. For example, the design and operation of consent-free behavioral tracking.[22] Our goal here is to highlight how some of the design principles, architectural components, and operational practices of the modern cloud-scale systems conflict with the rights and responsibilities laid out in GDPR. We present six such anti-patterns and summarize them in the accompanying table.

Storing data without a clear timeline for deletion. Computing systems have always relied on insights derived from data. However, in recent years, this dependence is reaching new heights with a widespread adoption of machine learning and big data analytics in system design. Data has been compared to oil, electricity, gold, and even bacon.[1] Naturally, technology companies evolved to not only collect personal data aggressively but also to preserve them forever. However, GDPR mandates that no data lives without a clear timeline for deletion.

ARTICLE 17: RIGHT TO BE FORGOTTEN. *"(1) The data subject shall have the right to obtain from the controller the erasure of personal data without undue delay ..."*

ARTICLE 13: INFORMATION TO BE PROVIDED WHERE PERSONAL DATA ARE COLLECTED FROM THE DATA SUBJECT. *"(2)(a) ... the controller shall provide the period for which the personal data will be stored, or the criteria used to determine that period;"*

ARTICLE 5(1)(E): STORAGE LIMITATION. *"kept... for no longer than is necessary for the purposes for which the personal data are processed ..."*

GDPR grants data subjects an unconditional right, via article 17, to request their personal data be removed from the system within a reasonable time. In conjunction with this, articles 5 and 13 lay out additional responsibilities for the data controller: at the point of collection, users should be informed the time period for which their personal data would be stored, and if the personal data is no longer necessary for the purpose for which it was collected, then it should be deleted. These simply mean that all personal data should have a time-to-live (TTL) that data subjects are aware of, and that controllers honor. However, the law makes exceptions for archiving data in the public interest, or for scientific or historical research purposes.

Deletion in the real world. While conceptually clear, a timely and guaranteed removal of data is challenging in practice. For example, Google cloud describes the deletion of customer data as an iterative process[8] that could take up to 180 days to fully complete. This is because, for performance, reliability, and scalability reasons, parts of data get replicated in various storage subsystems like memory, cache, disks, tapes, and network storage; multiple copies of data are saved in redundant backups and geographically distributed datacenters. Such practices not only delay the timeliness of deletions but also make it harder to offer guarantees.

Reusing data indiscriminately. While designing software systems, a purpose is typically associated with programs and models, whereas data is viewed as a helper resource that serves these high-level entities in accomplishing their goals. This portrayal of data as an inert entity allows it to be used freely and fungibly across various systems. For example, this has enabled organizations like Google and Facebook to collect user data once and use it to personalize their experiences across several services. However, GDPR regulations prohibit this practice.

ARTICLE 5(1)(B): PURPOSE LIMITATION. *"Personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes ..."*

ARTICLE 6: LAWFULNESS OF PROCESSING. *"(1)(a) Processing shall be lawful only if ... the data subject has given consent to the processing of his or her personal data for one or more specific purposes."*

ARTICLE 21: RIGHT TO OBJECT. *"(1) The data subject shall have the right to object ... at any time to processing of personal data concerning him or her ..."*

The first two articles establish that personal data could only be collected for specific purposes and not be used for anything else. Then, article 21 grants users a right to object, at any time, to their personal data being

GDPR anti-patterns, their real-world examples, and the GDPR articles that prohibit such behavior.

Anti-Pattern	Real-world Examples	Governing GDPR articles
Storing data without a clear timeline for deletion	Search engines before Right-to-be-forgotten (circa 2014)	5(1e). Storage limitation 17. Right to be forgotten
Reusing data indiscriminately	Facebook collecting phone numbers for 2FA and using them for ads and marketing	5(1b). Purpose limitation 6. Lawfulness of processing 21. Right to object
Creating black markets	Illegal data harvesting by programmatic ad exchanges	14. Information to be provided[...] 20. Right to data portability
Risk-agnostic data processing	Strava global heatmap that revealed classified military bases	35. Data protection impact assessment 36. Prior consultation
Hiding data breaches	Uber paying off hackers to hide their 2016 data breach	5. Principles relating to processing 33. Notification of personal data breach
Making unexplainable decisions	Using software like COMPASS in courts to predict recidivism	15. Right of access by the data subject 22. Automated individual decisionmaking

used for any purpose including marketing, scientific research, or historical archiving, or profiling. Together, these articles require each personal data (or groups of related data) to have their own blacklisted and whitelisted purposes that could be changed over time.

Purpose in the real world. The impact of the purpose requirement has been swift and consequential. For example, in January 2019, the French data protection commission[6] fined Google €50M for not having a legal basis for their ads' personalization. Specifically, the ruling said the user consent obtained by Google was not "specific" enough, and the personal data thus obtained should not have been used across 20 services.

Walled gardens and black markets. As we are in the early days of large-scale commoditization of personal data, the norms for acquiring, sharing, and reselling them are not yet well established. This has led to uncertainties for people and a tussle for control over data among controllers. People are concerned about vendor lock-ins, and about a lack of visibility once their data is shared or sold in secondary markets. Organizations have responded to this by setting up walled gardens and making secondary markets even more opaque. However, GDPR dismantles such practices.

ARTICLE 20: RIGHT TO DATA PORTABILITY. *"(1) The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller. (2) ... the right to have the personal data transmitted directly from one controller to another."*

ARTICLE 14: INFORMATION TO BE PROVIDED WHERE PERSONAL DATA HAVE NOT BEEN OBTAINED FROM THE DATA SUBJECT. *"(1) (c) the purposes of the processing ..., (e) the recipients ..., (2) (a) the period for which the personal data will be stored ..., (f) from which source the personal data originate ... . (3) The controller shall provide the information at the latest within one month."*

With article 20, people have a right to request for all the personal data that a controller has collected directly from them. Not only that, they could also ask the controller to directly transmit all such personal data to a different controller. While that tackles the vendor lock-ins, article 14 regulates the behavior in secondary markets. It requires that anyone indirectly procuring personal data must inform the data subjects, within a month, about how they acquired it, how long would they be stored, what purpose would they be used for, and who they intend to share it with. The *data trail* set up by this regulation should bring control and clarity back to the people.

Data movement in the real world. When GDPR went live, a large number of companies rolled out[7] data download tools for EU users. For example, Google Takeout lets users not only access all their personal data in their system but also port data directly to external services. However, the impact has been less savory for programmatic ad exchanges[9] in Europe, many of which had to shut down. This was primarily due to Google and Facebook restricting access to their platforms for those ad exchanges, which could not verify the legality of the personal data they possessed.

Risk-agnostic data processing. Modern technology companies face the challenge of creating and managing increasingly complex software systems in an environment that demands rapid innovation. This has led to a practice, especially in the Internet-era companies, of prioritizing speed over correctness; and to a belief that *unless you are breaking stuff, you are not moving fast enough*. However, GDPR explicitly restricts such approaches when dealing with personal data.

ARTICLE 35: DATA PROTECTION IMPACT ASSESSMENT. *"(1) Where processing, in particular using new technologies, ... is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data."*

ARTICLE 36: PRIOR CONSULTATION. *"(1) The controller shall consult the supervisory authority prior to processing where ... that would result in a high risk in the absence of measures taken by the controller to mitigate the risk."*

GDPR establishes, via articles 35 and 36, two levels of checks for introducing new technologies and for modifying existing systems, if they process large amounts of personal data. The first level is internal to the controller, where an impact assessment must analyze the nature and scope of the risks, and then propose the safeguards needed to mitigate them. Next, if the risks are systemic in nature or concern common platforms, either internal and external, the company's data protection officer must consult with the supervisory authority prior to any processing.

Fast and broken in the real world. Facebook, despite having moved away from the aforementioned motto, has continued to be plagued by it. In 2018, it revealed two major breaches: first, that their APIs allowed Cambridge Analytica to illicitly harvest[33] personal data from 87M users, and then their new *View As* feature was exploited[28] to gain control over 50M user accounts. However, this practice of prioritizing speed over security is not limited to one organization. For example, in November 2017, fitness app Strava released an athlete motivation tool called global heatmap that visualized athletic activities of worldwide users. However, within months, these maps were used to identify undisclosed military bases and covert security operations,[27] jeopardizing missions and lives of soldiers.

Hiding data breaches. The notion that one is *innocent until proven guilty* predates all computer systems. As a legal principle, it dates back to 6th century Roman empire,[3] where it was codified that *proof lies on him who asserts, not on him who denies*. Thus, in the event of a data breach or a privacy violation, organizations typically claim innocence and ignorance, and seek to be absolved of their responsibilities. However, GDPR makes such presumption conditional on the controller proactively

implementing risk-appropriate security measures (that is, accountability), and notifying breaches in a timely fashion (that is, transparency).

ARTICLE 5: PRINCIPLES RELATING TO PROCESSING. *"(1) Personal data shall be processed with ... lawfulness, fairness and transparency; ... purpose limitation; ... data minimization; ... accuracy; ... storage limitation; ... integrity and confidentiality. (2) The controller shall be responsible for, and be able to, demonstrate compliance with (1)."*

ARTICLE 33: NOTIFICATION OF A PERSONAL DATA BREACH. *"(1) the controller shall without undue delay and not later than 72 hours after having become aware of it, notify the supervisory authority. ... (3) The notification shall at least describe the nature of the personal breach, ... likely consequences, and ... measures taken to mitigate its adverse effects."*

GDPR's goal is twofold: first, to reduce the frequency and impact of data breaches, article 5 lays out several ground rules. Controllers are not only expected to adhere to these internally but also be able to demonstrate their compliance externally. Second, to bring transparency in handling data breaches, articles 33 and 34 mandate a 72-hour notification window within which the controller should inform both the supervisory authority and the affected people.

Data breaches in the real world. In recent years, responses to personal data breaches have been ad hoc: while a few organizations have been forthcoming, others have chosen to refute,[11] delay,[16] or hide by paying off hackers.[18] However, GDPR's impact has been swift and clear. Just in the first eight months (May 2018 to Jan 2019), regulators received 41,502 data breach notifications.[12] This number is in stark contrast from the pre-GDPR era, with reports of 945 worldwide data breaches in the first half of 2018.[34]

Making unexplainable decisions. Algorithmic decision-making has been successfully applied to several domains: curating media content, managing industrial operations, trading financial instruments, personalizing advertisements, and even combating fake news. Their inherent efficiency and scalability (with no human in the loop) has made them a necessity in modern system design. However, GDPR takes a cautious view of this trend.

Given the importance of personal data, and the implications of misusing them, we believe that system designers should examine their systems for these anti-patterns, and work toward eliminating them with urgency.

ARTICLE 22: AUTOMATED INDIVIDUAL DECISION-MAKING. *"(1) The data subject shall have the right not to be subject to a decision based solely on automated processing ..."*

ARTICLE 15: RIGHT OF ACCESS BY THE DATA SUBJECT. *"(1) The data subject shall have the right to obtain from the controller ... meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing."*

On one hand, privacy researchers from Oxford postulate[14] that these two regulations, together with recital 71, establish a "right to explanation" and thus, human interpretability should be a design consideration for machine learning and artificial intelligence systems. However, another group at Oxford argues[37] that GDPR falls short of mandating this right by requiring users to demonstrate significant consequences, to seek explanation only after a decision has been made, and to have to opt out explicitly.

Decision-making in the real world. The debate over interpretability in automated decision-making has just begun. Starting 2016, the machine learning and artificial intelligence communities began exploring this rigorously: *The Workshop on Explainable AI at IJCAI*, and the *Workshop on Human Interpretability in Machine Learning at ICML* being two such efforts. In January 2019, privacy advocacy group NoYB has filed[23] complaints against eight streaming services including Amazon, Apple Music, Netflix, SoundCloud, Spotify, YouTube, Flimmit, and DAZN for violating article 15 requirements in their recommendation systems.

Concluding Remarks

Achieving compliance with GDPR, while mandatory for companies working

with personal data of Europeans, is not trivial. In this article, we examine how the design, architecture, and operation of modern cloud-scale systems conflict with GDPR. Specifically, we illustrate this tussle via six GDPR anti-patterns, that use patterns of system design and operation, which are effective in their own context but violate the rights and regulations of GDPR. Given the importance of personal data, and the implications of misusing them, we believe that system designers should examine their systems for these anti-patterns, and work toward eliminating them with urgency.

Open issues. While our list of GDPR anti-patterns is a useful beginning point, it is not exhaustive. Neither have we proposed a methodology for identifying a large number of such anti-patterns, nor do we prescribe any mechanisms toward eliminating them. The six anti-patterns highlighted here exist due to technical and economic reasons that may not entirely be in the control of individual companies. Thus, solving such deep-rooted issues would likely result in significant performance overheads, slower product rollouts, and reorganization of data markets. The equilibrium points of these tussles are not yet clear.

Future directions. While there have been a number of recent works analyzing GDPR from privacy and legal perspectives,[5,19,15,35,36,38] the systems community is just beginning to get involved. GDPR compliance brings several interesting challenges to system design. Prominently, addressing compliance at the level of individual infrastructure components (such as, compute, storage, and networking) versus achieving end-to-end compliance of individual regulations (such as, implementing right-of-access in a music streaming service) will result in different trade-offs. The former approach makes the effort more contained and thus, suits the cloud model better. Examples of this direction include GDPR compliant Redis,[30] Compliance by construction,[29] and Data protection database.[20] The latter approach provides opportunities for cross-layer optimizations (for example, avoiding access control in multiple layers). Google search's implementation[2] of Right to be forgotten is in this direction.

Another challenge arises from GDPR being vague in its technical specifications (possibly to allow for technological advancements). Thus, questions like *how soon after a delete request should that data be actually deleted* could be answered in several compliant ways. The idea that compliance could be a spectrum, instead of a well-defined point gives rise to interesting system trade-offs as well as the need for benchmarks that quantify a given system's compliance behavior.

While GDPR is the first comprehensive privacy legislation in the world, several governments are actively drafting and rolling out their own privacy regulations. For instance, California's Consumer Privacy Act (CCPA)[4] went into effect on Jan 1, 2020. We hope that this paper helps all the stakeholders in avoiding the pitfalls in designing and operating GDPR-compliant personal-data processing systems. C

References

1. Alexander, F. Data is the new bacon. IBM Business analytics blog. 2016; https://www.ibm.com/blogs/business-analytics/datais-the-new-bacon/.
2. Bertram, T. et al. trait, A., Thomas, K., and Verney, A. Five years of the Right to Be Forgotten. ACM CCS. 2019.
3. Buckland, W. and Stein, P. A Textbook of Roman Law: From Augustus to Justinian. Cambridge University Press, 2007.
4. California Consumer Privacy Act. California Civil Code, Section 1798.100 (Jun 28, 2018).
5. Casey, B., Farhangi, A., and Vogl, R. Rethinking explainable machines: The GDPR's right to explanation debate and the rise of algorithmic audits in enterprise. Berkeley Technology Law J. 34 (2019), 143.
6. CNIL. The CNIL's restricted committee imposes a financial penalty of 50 million euros against Google LLC, 2019; https://www.cnil.fr/en/cnils-restricted-committee-imposesfinancial-penalty-50-million-euros-against-google-llc.
7. Conger, K. How to download your data with all the fancy new GDPR tools. Gizmodo. 2018; https://gizmodo.com/how-todownload-your-data-with-all-the-fancy-new-gdpr-t-1826334079.
8. Data Deletion on Google Cloud Platform, 2018; https://cloud.google. com/security/deletion/.
9. Davies, J. GDPR mayhem: Programmatic ad buying plummets in Europe. Digiday; 2018; https://digiday.com/media/gdpr-mayhemprogrammatic-ad-buying-plummets-europe/.
10. Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F., and Holz, T. We value your privacy ... Now take some cookies: Measuring the GDPR's impact on Web privacy. NDSS, 2019.
11. Doshi, V. 2018. A security breach in India has left a billion people at risk of identity theft. Washington Post, 2018; https://wapo.st/3389hOn.
12. European Data Protection Board. EDPB: First Year GDPR—taking stock. EDPB News; https://edpb.europa.eu/news/news/2019/1- year-gdpr-taking-stock_en.
13. Forni, A., and Meulen, R. Organizations are unprepared for the 2018 European Data Protection Regulation. Gartner, 2017.
14. Goodman, B. and Flaxman, S. European Union regulations on algorithmic decision-making and a right to explanation. AAAI AI Magazine 38, 3 (2017).
15. Greengard, S. Weighing the impact of GDPR. Commun. ACM 61, 11 (2018), 16–18.
16. Grothaus, M. Panera Bread leaked millions of customers' data. In Fast Company, 2018; https://bit.ly/3cG5jQk.
17. Haselton, T. Credit reporting firm Equifax says data breach could potentially affect 143 million US consumers. CNBC, 2017.
18. Isaac, M., Benner, K., and Frenkel, S. Uber hid 2016 breach, paying hackers to delete stolen data. New York Times, 2017; https://www.nytimes.com/2017/11/21/technology/uber-hack.html.
19. Koenig, A. Patterns and antipatterns. J. Object-Oriented Programming 8, 1 (1995), 46–48.
20. Kraska, T., Stonebraker, M., Brodie, M., Servan-Schreiber, S. and Weitzner, D. DATUMDB: A data protection database proposal. In Proceedings of Poly'19 co-located at VLDB.
21. Larson, S. Every single Yahoo! account was hacked—3 billion in all. CNN Business, 2017.
22. Lomas, N. Even the IAB warned adtech risks EU privacy rules. TechCrunch, 2019; https://techcrunch.com/2019/02/21/even-theiab-warned-adtech-risks-eu-privacy-rules/.
23. Lomas, N. Privacy campaigner Schrems slaps Amazon, Apple, Netflix, others with GDPR data access complaints. TechCrunch, 2019.
24. Lunden, I. UK's ICO fines British Airways a record £183M over GDPR breach that leaked data from 500000 users. TechCrunch, 2019.
25. O'Flaherty, K. Marriott faces £123 million fine for 2018 mega breach. Forbes. 2019.
26. OJEU. General Data Protection Regulation. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46. Official Journal of the European Union 59 (2016), 1–88.
27. Quarles, J. An update on the global heatmap, 2018; https://blog. strava.com/press/a-letter-to-the-strava-community/.
28. Rosen, G. Security Update, 2018; https://newsroom.fb.com/news/ 2018/09/security-update/.
29. Schwarzkopf, M., Kohler, E., Kaashoek, F., and Morris, R. GDPR compliance by construction. In Proceedings of Poly'19 co-located at VLDB.
30. Shah, A., Banakar, V., Shastri, S., Wasserman, M., and Chidambaram, V. Analyzing the impact of GDPR on storage systems. USENIX HotStorage, 2019.
31. Shastri, S., Banakar, V., Wasserman, M., Kumar, A., and Chidambaram, V. Understanding and benchmarking the impact of GDPR on database systems. In Proceedings of the VLDB Endowment 13, 7 (2020).
32. Shastri, S., Wasserman, M., and Chidambaram, V. The seven sins of personal-data processing systems under GDPR. USENIX HotCloud, 2019.
33. Solon, O. Facebook says Cambridge Analytica may have gained 37M more users' data. In The Guardian, 2018; https://bit.ly/2S9xVYF.
34. Targett, E. 6 Months, 945 Data Breaches, 4.5 Billion Records. Computer Business Review, 2018; https://www.cbronline.com/news/globaldata-breaches-2018.
35. Tesfay, W., Hofmann, P., Nakamura, T., Kiyomoto, S., and Serna, J. I read but don't agree: Privacy policy benchmarking using machine learning and the EU GDPR. In Companion Proceedings of the Web Conference, 2018, 163–166.
36. Utz, C., Degeling, M., Fahl, S., Schaub, F., and Holz, T. (Un) informed consent: Studying GDPR consent notices in the field. ACM CCS, 2019.
37. Wachter, S., Mittelstadt, B., and Floridi, L. 2017. Why a right to explanation of automated decision-making does not exist in the general data protection regulation. Intern. Data Privacy Law 7, 2 (2017), 76–99.
38. Wachter, S., Mittelstadt, B., and Russell, C. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. Harvard J. Law & Technology 31 (2017), 841.

Supreeth Shastri is an assitant professor of computer science at the University of Iowa, IA, USA

Melissa Wasserman is the Charles Tilford McCormick Professor of Law at the University of Texas at Austin, TX, USA.

Vijay Chidambaram is an assitant professor of computer science at the University of Texas at Austin, TX USA.