CENTER FOR DATA INNOVATION

The Impact of the EU’s New Data Protection Regulation on AI

By Nick Wallace and Daniel Castro | March 27, 2018

The EU’s new data privacy rules, the General Data Protection Regulation (GDPR), will have a negative impact on the development and use of artificial intelligence (AI) in Europe, putting EU firms at a competitive disadvantage compared with their competitors in North America and Asia. The GDPR’s AI-limiting provisions do little to protect consumers, and may, in some cases, even harm them. The EU should reform the GDPR so that these rules do not tie down its digital economy in the coming years.

The GDPR will come at a significant cost in terms of innovation and productivity. EU policymakers need to recognize that a failure to amend the GDPR to reduce its impact on AI will all but consign Europe to second-tier status in the emerging algorithmic economy.

EXECUTIVE SUMMARY

The EU’s new data privacy rules, the General Data Protection Regulation (GDPR), will go into effect on May 25, 2018. The GDPR regulates EU organizations that use or processes personal data pertaining to anyone living in the EU—regardless of where the data processing takes place. These new regulations will kick in at a time when companies around the globe are fiercely competing to develop and use artificial intelligence (AI)—a set of technologies that allows computers to perform tasks much like a human—as a means of boosting productivity through its more efficient processes and higher-quality outputs.

While a substantial number of AI's uses do not involve personal data, the many others that do will be subject to the GDPR. Consumers who routinely interact with AI-enabled services such as personal assistants that respond to spoken queries, robo-advisors that provide automated financial advice, and movie recommendations on streaming services will be significantly affected, as will virtually every European company that processes personal data—such as payroll—and can use AI to make their operations more efficient.[1] As such, by both indirectly limiting how the personal data of Europeans gets used and raising the legal risks for companies active in AI, the GDPR will negatively impact the development and use of AI by European companies.

Despite different jurisdictions having different goals when it comes to privacy, policymakers and citizens in the EU should understand that the GDPR will come at a significant cost in terms of innovation and productivity. At a time when two major world powers, the United States and China, are vying for global leadership in AI, EU policymakers need to recognize that a failure to amend the GDPR to reduce its impact on AI will all but consign Europe to second-tier status in the emerging AI economy.

HOW THE GDPR WILL INHIBIT AI DEVELOPMENT AND USE IN EUROPE

There are at least nine specific aspects of the GDPR that will have a negative effect on the development and use of AI in Europe:

1. Requiring companies to manually review significant algorithmic decisions raises the overall cost of AI.

The most direct restriction in the GDPR that specifically targets the use of AI is the requirement in Article 22 that companies must have humans review certain algorithmic decisions. This restriction significantly raises labor costs and thus creates a strong disincentive from using AI—as a main reason for developing AI in the first place is to automate functions that would otherwise be much slower, costlier, and more difficult to complete if performed by humans.

2. The right to explanation could reduce AI accuracy.

Articles 13–15 of the GDPR create an obligation for companies to provide either detailed explanations of individual algorithmic decisions or general information about how the algorithms make decisions—which remains a point of contention.[2] However, the former would undermine the accuracy of algorithms and, perversely, lead to unfair decisions, as there is inherently a trade-off between accuracy and transparency in algorithmic decisions.[3]

3. The right to erasure could damage AI systems.

The "right to erasure" in Article 17(1) will also harm AI in Europe. All AI systems that operate using unsupervised machine learning—those that improve themselves, without outside help, by learning from the data they process—will be required to "remember" all the data they used to train themselves in order to sustain rules derived from that data.[4] However, erasing data that underpins key rules in an AI system's behavior can both make it less accurate and limit its benefit to other data subjects—or even break it entirely.[5]

4. The prohibition on repurposing data will constrain AI innovation.

Like its predecessor, the Data Protection Directive, Article 6 of the GDPR imposes a general prohibition on using data for any purposes other than that for which it was first collected, thus making it difficult for firms to innovate using data. This restriction will limit the ability of companies developing or using AI in the EU to experiment with new functions that could improve their services. As a result, EU consumers and businesses will be slow to receive the benefits of the latest innovations in AI.

5. Vague rules could deter companies from using de-identified data.

Although the GDPR rightly allows exemptions for de-identified data, the lack of clarity in the GDPR about precisely which standards of de-identification are acceptable will deter companies from attempting to de-identify data—lest they face harsh enforcement by regulators. This will undermine companies' incentives to process and share de-identified data that could be used to improve AI systems, while at the same time driving some firms to process personal data when de-identified data would suffice, and as a result incur unnecessary compliance costs and restrict their range of legal uses.

6. The GDPR's complexity will raise the cost of using AI.

The GDPR is a very complex piece of legislation, which can make it difficult to follow.[6] Companies developing or using AI will need specialized personnel and technology to make sure they comply with the GDPR, thus raising the cost of AI and deterring its use.

7. The GDPR increases regulatory risks for firms using AI.

There is a growing body of evidence that a large proportion of companies, especially small and medium-sized businesses, do not understand the regulation or what it will mean for them—and are likely to find themselves the unwitting targets of legal action, thus adding further costs and disincentives from using AI.[7] The problem of complexity is compounded by the extraordinarily exorbitant fines the GDPR imposes: up to 4 percent of a company's global turnover, or €20 million (whichever is greater). The highest fines apply to breaches of the provisions most problematic for AI—namely, those identified in this report.[8] Because smaller firms typically generate less income than their larger counterparts, the maximum GDPR fines end up being proportionally costlier for small companies, who as a result will be even less likely to adopt AI.

8. Data-localization requirements raise AI costs.

Chapter 5 of the GDPR outlines its tight controls on flows of personal data outside of the EU, such as requiring companies to use data centers inside EU countries—which reduces competition between cloud-service providers and thereby raises the cost of data processing.[9] Although the GDPR does ban national governments from using privacy as a justification for forcing companies to store personal data in a particular country, it fails to recognize that the physical location of data has no inherent bearing on privacy or security.[10]

9. Data portability will stimulate AI competition, albeit at a cost.

The right to data portability in Article 20 of the GDPR is one of the few provisions in the regulation that could have a positive impact on the use of AI in the EU, as data portability will make it easier for consumers to share their data with companies that leverage AI, thereby fueling competition. However, Article 20 does not adequately account for the cost and feasibility of providing extremely large, complex data sets accumulated over many years.

OVERALL IMPACTS

The GDPR will limit both the emergence of European companies that develop and sell AI solutions globally, and the use of AI itself in European companies in a wide array of industries. While many companies outside of Europe will also be obligated to comply with the GDPR, the greatest impact will be on European companies because, in most cases, European data will be more important to them as they seek to use or develop AI than to companies whose presence is stronger in foreign markets. Because of these restrictions, firms in the EU developing or using AI will be at a competitive disadvantage compared with their competitors in North America and Asia.

The GDPR will also impact the behavior of foreign companies in ways that will hurt the EU's economy. For one, many foreign firms will be discouraged from offering their AI-driven services in the EU, thus leaving EU consumers and businesses unable to access beneficial services that are available to their counterparts and competitors elsewhere—and making the EU market for AI less competitive and innovative.

It is unnecessary for Europe to force these trade-offs to be made between data protection and innovation. The GDPR’s AI-limiting provisions do nothing to actually protect consumers—and may, in some cases, even harm them. For example, as there is a trade-off between algorithmic transparency and accuracy, a requirement to explain algorithmic decisions would force companies to use more transparent, less accurate algorithms, which would make for unfair decisions with important concomitant effects for consumers. Similarly, the general prohibition on solely-automated decisions that have legal or significant effects will lead to humans making unfair and unreasonable decisions, without the benefit of algorithms, that ultimately end up hurting consumers. This will also prevent companies from using rational algorithms that can be adjusted over time to account for unintended biases.

RECOMMENDATIONS

Policymakers in the EU should take the following measures to address these problems and create a better regulatory environment for AI in Europe, without reducing consumer protections:

- The EU should simplify the GDPR to focus exclusively on preventing harm to consumers, instead of needlessly limiting the use of data at the expense of data innovation.

- The EU should remove the right to human review of algorithmic decisions because such a policy will force companies to use less accurate AI systems that fail to protect consumers from unfair decisions. Due process and scrutiny should always be appropriate to the nature and seriousness of the decision at hand, and not be based on whether the decision was made by a human or an algorithm.

- The EU should make any requirements for transparency, evidence, oversight, or explanation technology-neutral, instead of basing these requirements on whether a particular decision was made by a human or an algorithm.

- The EU should amend the GDPR to allow companies to meet the obligation to provide “meaningful information” about algorithmic decisions by simply providing a basic description of how an algorithm works and what data it uses.

- The EU should revise the right to erasure to ensure companies are able to delete or anonymize data in ways that do not impact the functioning of algorithms for other customers.

- The EU should amend the GDPR to allow repurposing of personal data without additional consent, except when doing so would pose a significant risk to the data subject or transfer data to another controller—mergers and buyouts notwithstanding.

- The EU should amend data-portability rights to account for costs, as they could be considerable for requests to port exceptionally large and complex data sets that may be of limited value. In such extreme scenarios, the law should allow for a limited contribution toward such costs by the customers making the requests.

- The EU should amend the GDPR to make fines for breaching GDPR rules proportional to the level of harm to the data subjects and the company’s level of culpability for the breach. This would incentivize companies to focus on protecting their customers’ privacy, as opposed to just protecting themselves.

- Article 29, Working Party (WP29), should identify clear and practical guidelines for de-identifying data, such as those used in privacy rules for U.S. health data, so companies are clear about what they must do to avoid regulators taking action against them for their de-identification practices.

- The GDPR exempts certain types of data processing from its rules when a country deems the data processing to be in the public interest. National governments should use this authority broadly, such as to allow the use of AI in public services.

INTRODUCTION

In the emerging data economy, innovation in many industries will be largely driven by what companies do with data.[11] This trend has made artificial intelligence—a set of technologies that allows computers to perform tasks much like a human—one of the most valuable tools available to businesses. AI enables organizations to use data to create new services, improve existing ones, and make many existing processes more efficient. AI can analyze large data sets much faster than humans, allowing it to make quick and accurate observations of trends in the data and draw conclusions from those observations; make predictions; automate machinery and interactions between machines; and help humans interact with machines in new ways.[12] AI can not only carry out many tasks more efficiently than humans, it can also do things humans cannot, such as process quantities of data too large for a human to comprehend, and spot things in data a human would miss. Companies must have access to data—and often large amounts of it—to successfully use AI. Regulations that control the use of data therefore have serious implications to AI.

As businesses in virtually every industry begin to improve their productivity using AI, the European economy will only be able to stay competitive if its own firms do the same. European tech companies, for example, have a huge market opportunity to develop AI for a multitude of different use cases. EU policymakers have recognized AI's economic significance and committed hundreds of millions of euros to AI research.[13] The consulting firm PwC estimates AI could add up to $2.5 trillion to the European GDP by 2030.[14] However, the EU's General Data Protection Regulation (GDPR)—new data privacy rules which will come into force in May 2018—) remains a significant roadblock to the widespread development and use of AI in Europe.

The GDPR imposes strict rules on how companies may use the personal data of anyone living in the EU—a constraint that will undoubtedly impede the development and use of AI in Europe. These restrictions affect virtually all European companies, as most every firm processes personal data about its workers, such as payroll information. And while not all uses of AI involve personal data, many do. For example, companies use AI to automate financial advice, process credit applications, and analyze medical test results.

The GDPR contains rules that both directly and indirectly limit the development and use of AI:

- Companies must have humans review certain algorithmic decisions, which raises the labor costs of using sophisticated AI systems.
- Companies having to explain the logic behind their algorithmic decisions is an ambiguous requirement that could compel companies to make trade-offs between accuracy and interpretability of their computer models.
- The right to erase data could reduce the accuracy of some algorithmic models.

- Noncompliance with the GDPR's extremely complex set of rules results in stiff penalties that make advanced data processing a legally and financially risky endeavor. Because the GDPR’s requirements would be impractical—and in some cases impossible—to fulfill, many companies will ultimately limit their use of AI.

Some aspects of the GDPR remain open to interpretation, both by regulators and the courts, such as what technical measures might satisfy the requirement to “erase” data and what a right to “meaningful information about the logic involved” in an algorithmic decision really entails. Such vagueness presents both an opportunity and a problem. On one hand, it creates leeway for policymakers to limit the most harmful side effects of the GDPR without amending it—as amending the regulation would be a tall order given the EU’s complex legislative process. For example, regulators can interpret poorly defined provisions like the right to “meaningful information” about algorithms in a way that would not necessarily chill investment in AI. But on the other hand, ambiguous legislation allows for capricious enforcement, which is a problem the EU can only address by amending the GDPR. Many companies will likely operate according to the strictest interpretations of the GDPR, lest they find themselves subject to the GDPR’s severe fines, especially because the EU’s data-protection advisory group, the Article 29 Working Party (WP29), tends to use prohibitive interpretations of the law in its guidelines, even when it comes at the expense of innovation. This uncertainty will likely hinder the development and use of AI.

BACKGROUND ON AI

Artificial intelligence is a field of computer science devoted to the pursuit of computer systems that perform operations analogous to human learning and decision-making. AI systems emulate various human functions such as learning, understanding, reasoning, and interacting with people, machines, and the environment.[15] Certain forms of artificial intelligence that can learn to carry out particular tasks far better than humans already exist, but their capacity for learning and self-improvement is always limited to an extremely narrow range of possibilities.

Although the field of AI research dates back to the end of the Second World War, and despite remarkable achievements in so many other domains of computer science, progress in AI has failed to keep up with expectations. Throughout the 1950s, 1960s, and 1970s, experts in AI predicted a rise of machines with human-like intelligence would occur within a few decades, or in at least one case, within a few months.[16] Progress has finally begun to accelerate during the last few years due to recent advances in algorithmic design, improvements in data processing capabilities, and the rise of cloud computing and the economies of scale cloud computing enables. The key breakthrough has been machine learning, wherein algorithms use data to automatically and iteratively build new analytical models, thus allowing them to learn how to solve problems within narrowly defined contexts without being explicitly programmed for a particular solution.[17] Consumers frequently encounter applications that use machine learning, such as personal assistants that respond to spoken queries, automatic language-translation services, and movie recommendations on streaming-media services.

PROBLEMS WITH THE GDPR FOR AI

The GDPR poses three main problems for businesses using AI: higher costs, practical limitations, and legal hazards. Higher costs and legal hazards will deter the use of AI altogether, while the practical limitations will make it difficult to use and undermine its effectiveness.

Several provisions of the GDPR impose direct or indirect costs on the use of AI. For example, the requirement to have a human review certain algorithmic decisions directly imposes significant costs on businesses using AI, as going through every detail of an algorithmic decision is complex and time-consuming work that demands particular skills. The right to data portability does not target AI directly, but it does impose indirect costs by creating an obligation for firms using AI to process and supply large and complex data sets in a reusable format.

Depending on how regulators and courts interpret it, the GDPR could impose significant practical limitations. For example, a right to explanation is problematic because, as research has shown, there is a trade-off between accuracy and transparency in algorithms. If companies are compelled to explain their decisions to customers, they will end up using algorithms that are designed to be transparent and result in less accurate and unfair decisions. Purpose limitation is also impractical for AI because it requires companies to get each data subject's permission before doing anything new with their data using AI, regardless of whether the repurposing would have any impact at all on privacy or consumer welfare.

The GDPR also makes AI a legally risky endeavor, which will turn some companies away from using it at all. The GDPR's complexity means there are a huge number of potential points of failure where companies could inadvertently breach the GDPR and thus face heavy fines.

THE RIGHT TO HUMAN REVIEW OF ALGORITHMIC DECISIONS RAISES THE COST OF AI

Article 22 confers a right for individuals "not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her, or similarly significantly affects him or her."[18] In other words, if the decision is necessary to complete a contract with the customer, or if the customer has given consent to the controller to make such a decision, then the GDPR requires data controllers to give the customer "at least the right to obtain human intervention on the part of the controller."[19] That means wherever use of AI has legal or similarly significant effects—such as in deciding whether to offer a loan—the data subject has the right to have a human review that decision.

The human review is no rubber-stamp exercise. The WP29 guidelines on GDPR compliance in algorithmic decisions state that "to qualify as human intervention, the controller must ensure that any oversight of the decision is meaningful, rather than just a token gesture. It should be carried out by someone who has the authority and competence to change the decision."[20] The WP29 guidelines also state that the human reviewer "should consider all the relevant data" involved in the decision.[21]

The Right to Human Review is Costly and Potentially Impractical, and Will End Up Reducing Investment in AI

Having humans review an algorithmic decision is costly—and only more so as the complexity of the algorithm increases. This is because there is a trade-off between the representational capacity of a model and the ease with which a human can review the calculations it makes.[22] In other words, the more sophisticated the algorithmic model, the more time and expertise (and, therefore, resources) is needed for a human to make sense of the model's decisions. Indeed, the main point of artificial intelligence is to processes large quantities of data much more efficiently and accurately than humans—if it were no costlier for humans to repeat these calculations, there would be little economic incentive to use AI. The right to human review is essentially a tax on AI systems that are capable of making calculations that would be impractical for humans.

Faced with the cost of paying a qualified human to review an algorithmic decision and all the data it involves, companies will respond in one of two ways, depending on how critical AI is to their business. They will either limit the quantity and complexity of the data and the sophistication of the algorithm in order to minimize the cost of compliance, or simply forgo the use of AI altogether. Either outcome is ultimately bad for European competitiveness, as it ends up limiting the valuable contribution AI makes to European industry in improved efficiency.

A RIGHT TO EXPLANATION WOULD DISCOURAGE COMPANIES FROM USING AI

Articles 13–15 of the GDPR confer a right to receive "meaningful information about the logic involved" in an algorithmic decision covered by Article 22—that is, one with legal or similarly significant effects. The phrase appears once in each of the three articles: Article 13 concerns personal data obtained from the subject, Article 14 addresses data obtained by other means, and Article 15 deals with data subjects' right to know whether somebody is processing their information, and if so, how.[23] Some scholars, such as Bryce Goodman and Seth Flaxman of Oxford University, argue that the right to "meaningful information" amounts to a "right to explanation" of algorithmic decisions, citing Recital 71's (a recital is a non-binding paragraph intended to help judges interpret the law) assertion that a data subject should have the right to "obtain an explanation of the decision" and challenge it after a human has reviewed it.[24]

The GDPR provides relatively little clarification as to how it defines "meaningful information," which could lead to legal battles over the extent of the right. The articles themselves do not specify whether "meaningful information about the logic involved" refers to an explanation of how a particular algorithm generally reached decisions, or to a precise explanation of exactly how the algorithm arrived at a particular conclusion. Recital 71 seems to imply the latter when it says the data subject should be able to "obtain an explanation of the decision," but it does not specify what information would constitute an explanation or whether information about the "logic involved" should pertain to the algorithm or to the decision. Were a regulator or court to interpret the law to mean data controllers must be able to explain precisely how each individual decision was reached, it would severely inhibit AI because there is a trade-off between an algorithm's sophistication and its explicability—arising from the fact that AI systems are designed to carry out data processing tasks that would be more difficult or time-consuming for a human.[25]

WP29's interpretation of a right to "meaningful information" is useless and vague. It recommends companies inform data subjects of what data their algorithms use and then explain in general terms how the algorithm makes decisions. Although objectively reasonable, that interpretation is not very reassuring, as it does not rule out stricter interpretations. WP29 also says companies do "not necessarily" have to "provide a complex explanation of the algorithms used or disclosure of the full algorithm." But "not necessarily" implies by corollary that there could be a circumstance in which a data subject might demand the most detailed explanation possible. The WP29 guidelines go on to say the information provided should "be sufficiently comprehensive for the data subject to understand the reasons for the decision."[26]

Even if WP29 were less equivocal, its guidelines would remain nonbinding. And the GDPR creates plenty of space for a lawsuit seeking a legal precedent that interprets the law in a different, more innovation-hampering way. That is how the "right to be forgotten" came into being, when the European Court of Justice (ECJ) extrapolated the right from various parts of the Data Protection Directive, even though no such right was explicitly described in the law.

There Is a Trade-Off Between the Accuracy and Interpretability of AI

The main problem with the concept of a right to explanation is the more variables an algorithm represents in its model, and the more complex the links between those variables, the harder it is to for a human to assess how the algorithm arrived at any particular decision.[27] That means there is a trade-off between accuracy and interpretability—even if the interpreter understands the systems well.[28] This problem becomes even more complicated when the algorithm must be interpretable by a customer with a limited understanding of computer science or statistics—or a specialist human reviewer has to concisely explain the decision to the customer. Ergo, a right for data subjects to an explanation of algorithmic decisions necessarily implies a limitation on the feasible sophistication of the algorithm.

Another issue involves defining the nature of the explanation, as a decision may be empirically falsifiable even if the decision-maker's rationale is unknown.[29] For example, a doctor can check an AI system's determination that a mole on a patient's arm is cancerous by performing a biopsy, even if the system is a "black box." Exactly how the algorithm arrived at that conclusion is a separate question. It is possible to program algorithms to make their decisions transparent, but a human would still need to interpret those outputs—and the trade-off with accuracy persists, as algorithms designed to be more transparent tend to be less accurate, and vice-versa.[30] Therefore, the importance of explaining an algorithm depends not just on the impact of the decision, but on the extent to which the algorithm's conclusion is falsifiable.

Human Decisions Are Held to a Lower Standard Than Algorithmic Decisions

Because the right to explanation only applies to algorithmic decisions, some companies might choose to use humans instead of algorithms as a way of sidestepping the requirement to explain their actions. But such a decision is not without its own set of problems, not least of which is human decisions are often less accurate and more susceptible to bias than algorithmic decisions—which is the reason behind many organizations choosing to adopt AI systems in the first place.

Humans are also far more like "black boxes" than are algorithms, which heightens the folly of subjecting human decisions to lesser scrutiny than algorithmic decisions. Humans often lie to each other deliberately, and are prone to misunderstanding and misremembering their own subjective experience of the world, and of their own consciousness. The fact that humans can construct—for themselves or others—plausible explanations for how they arrived at particular decisions does not necessarily prove the explanations are accurate, even when the people in question really believe the explanations.

Human systems—or for that matter, complex algorithms—essentially being scientific black boxes is not a counsel of despair. The observation merely avers that for high-stakes decisions that demand a particular kind of explanation or evidence, the decision-making has to follow a carefully defined process that is appropriate to the nature of the decision at hand. Whether it is a human or an algorithm that must follow that given process is a separate issue. The need for transparency, evidence, oversight, or explanation is entirely dependent on the nature and severity of the decision, not the tools used to make it.

THE RIGHT TO ERASURE COULD NEGATIVELY IMPACT THE VALIDITY OF AI SYSTEMS

Article 17(1) of the GDPR states data subjects have the right "to obtain from the controller the erasure of personal data concerning him or her without undue delay" which could prove problematic for AI because some types of machine learning systems use data to improve themselves by generating new rules for processing future data, and removing that data could impact the algorithm's effectiveness for other users, or even break it completely. In simple terms, such algorithms need to remember the data used to train them.[31] This right will also raise the labor costs involved in managing AI systems. The EU should amend the right to erasure to provide exceptions where erasure is infeasible or likely to impact the service other users receive.

The right to erasure applies under the following circumstances:

- The data is "no longer necessary in relation to the purposes for which it was collected";
- The legal basis for processing the data was consent and the data subject has decided to withdraw their consent;
- The data subject objects to processing on the legal basis of public interest or overriding legitimate interest, and the controller is unable to demonstrate that there are overriding legitimate grounds for processing;
- The data is being used for direct marketing and the data subject objects to this;
- The data has been "unlawfully processed"—meaning some aspect of the data processing has breached one or more of the many complex rules in the GDPR;
- Another EU or member-state law requires the data to be erased;
- The data was collected to provide services to a child (on the basis of either the child's consent or their parents', depending on the child's age).[32]

There are some limitations to the right to erasure. The above provisions do not apply under these conditions:

- The data is necessary to uphold the right to freedom of speech and information;
- The data is necessary to comply with EU or member-state law;
- The data is necessary for reasons of public health;
- Erasing the data would "seriously inhibit" the objectives of processing for "archiving purposes in the public interest, scientific or historical research purposes, or statistical purposes" that are protected under Article 89.[33]

In most commercial uses of AI where companies collect data on the basis of consent in order to provide the customer with a service, the data subject may withdraw that consent and demand erasure at any time, provided no other laws (such as financial regulations that require banks to retain some data for anti-fraud purposes) obligate the company to keep it.[34]

Erasing Data Could Break Certain AI Systems

In computing terms, there is a difference between deleting data—defined as instructing a system to treat the space where the data is stored as blank space—and erasing data, which simply means destroying or overwriting it.[35] Once data is deleted, considerable time may pass until the system overwrites (and thus erases) it—and depending on the nature of the system and its needs, this delay can be indefinite. Forcing the system to erase the data earlier can be labor-intensive, time-consuming, and costly—and depending on the circumstances, can cause damage to the system's integrity.[36] In some cases, the data may have to remain present as long as the storage medium is still physically intact. For example, only rarely are individual pieces of data ever erased from a tape backup.

Although the distinction between deletion and erasure applies in just about any computing context, it is especially important for AI because many types of machine learning algorithms improve themselves by generating new rules based on the data used to train them. The impact of erasing this data within any fixed period of time is thus difficult to predict, as erasure could undermine algorithmic rules that depend on the data in question, along with subsequent rules derived from earlier ones.[37] This makes the phrase "undue delay" in Article 17 troublesome, as it is extremely difficult to determine at precisely which point data can safely be erased from an AI system's knowledge base (the database used to train it) without changing the system, including in ways that might lead to suboptimal outcomes.

If several data subjects in a particular data set exercise their right to erasure, the cumulative impact of the removals could affect the validity of the model itself. In one experiment, a team of four computer scientists found that removing a single, randomly selected data point from an algorithm's knowledge base did not have a significant effect on the model; but in another study, one of those computer scientists, accompanied by a policy analyst and a legal scholar, pointed out that the impact could be more severe in real-life scenarios where "people that want their data to be removed share some commonalities that are then missing from the data set."[38] Therefore, an unusually high number of individuals exercising their right to erasure could undermine important rules in the algorithm's model, thus impairing its effectiveness for other users, or even breaking it altogether.

Imagine a credit agency using AI to analyze subtle trends in much larger and more complex data sets than it had dealt with previously, and discovering some individuals were not as risky to lend to as the agency first thought, while others were much riskier than traditional models would suggest. If one customer traditionally thought of as posing a bad credit risk were found by the AI system to be a trustworthy borrower because of a subtle but rare constellation of particular characteristics, and this individual exercised their right to erasure, there may not be any discernable impact on the AI system. But if several other people with a similar combination of characteristics did the same thing, it could undermine the rule that caused the algorithm to treat those people as low-risk. If this does not stop the algorithm from working altogether, it may lead to unfair decisions for future customers, who would be treated according to the older, less accurate model and not the newer, fairer model created on the strength of those previous customers' data. For regulators and businesses concerned about accuracy and fairness, a right to erasure makes it exceedingly difficult to validate the performance of algorithms.

THE PROHIBITION ON REPURPOSING DATA WILL CONSTRAIN AI INNOVATION

Like the Data Protection Directive the GDPR will replace, Article 6 of the new regulation requires “purpose limitation”: a general prohibition on reusing data for purposes that are not “compatible” with those for which it was first collected.[39] This constraint will inhibit the development and use of AI by preventing companies from experimenting with their algorithms or trying out new uses for existing data. The EU should amend the GDPR to allow repurposing of personal data in general, but specify uses that should be deemed illegal without consent, such as transferring data to another controller.

Article 6(1) of the GDPR provides a narrow and exclusive range of legal bases for processing personal data. These are consent; a contract with the data subject; compliance with a legal obligation; protecting the vital interests of the data subject or another person; performing a task carried out in the public interest or in the exercise of official authority (such as delivering government-mandated public services); and overriding legitimate interests (such as fraud prevention).[40] For most commercial uses involving AI, the legal basis for data processing will be either via consent from, or a contract with, the data subject. In general, a data controller in such circumstances cannot use the data for any purposes the data subject did not consent to or that were not stipulated in the contract. To do this legally, the controller must ask for consent, establish a new contract, or, if appropriate, anonymize the data (see section on “Vague Rules Could Deter Companies From Using De-identified Data for AI”).

The GDPR does create some limited space for reusing data for purposes that are “compatible” with the original purpose. Whether a use is "compatible" depends on factors such as a link between the new purpose and the old one, the consequences to the data subject, and the existence of safeguards such as pseudonymization.[41] The GDPR lacks a concise definition of compatibility. The criteria it does provide were absent in the equivalent section of the Data Protection Directive, and were derived from older WP29 guidelines on the directive.[42] The GDPR may permit a data controller to switch from human to algorithmic decision-making if doing so ends up serving the same purpose—unless that purpose involves making decisions with effects that regulators might view as legal or similarly significant, because the GDPR treats such decisions differently depending on whether they are made by humans or algorithms (see section on “A Right to Explanation Would Discourage Companies From Using AI”). However, the GDPR would almost certainly not allow a controller to use the data to add completely new functionality to an AI system without consent.

Requiring Consent for Every New Use Raises Costs and Chills Innovation

This general prohibition on repurposing is a direct restriction on AI—and on data innovation in general—because it means companies cannot find serendipitous uses for data, even when there are no privacy implications in doing so. Instead, companies must seek permission from data subjects every time they want to try a new use for the data. For example, if a company wants to begin automating aspects of a service it already provides to an existing customer, it may have to ask permission to use that customer's data, especially if this automation adds new functions such as recommendations based on the customer's use of the service. This requirement will be costly and impractical in many cases.

The more people a data controller must ask for explicit permission, the greater the costs and the less confidence developers have in the usefulness of the data set they will be left with in the end. Such constraints will make it harder to quickly test and evaluate different potential solutions to get AI projects off the ground, and then limit the potential of those that are successful.

For example, in 2017, the Royal Free Hospital in London found itself on the wrong side of a similar rule regarding existing data protection law when it used historical patient data to test an AI tool for diagnosing kidney injuries.[43] Because the hospital had originally collected that data to treat those patients—and not to test an AI tool that could improve outcomes for future patients—the United Kingdom's privacy regulator forced the hospital to seek the consent of the 1.6 million people in the data set. That ruling occurred under the auspices of the Data Protection Act, which gives U.K. legal effect to the GDPR's predecessor—the EU Data Protection Directive—and is somewhat more flexible than the GDPR, particularly with regard to processing data for medical purposes.[44] Replacing the Directive with the GDPR will therefore create even more problems like the one at the Royal Free Hospital, as the GDPR raises the threshold for consent and makes it even harder to repurpose data legally.[45]

VAGUE RULES COULD DETER COMPANIES FROM USING DE-IDENTIFIED DATA FOR AI

De-identification allows developers to use and share data in AI systems with fewer privacy constraints. There are two forms of de-identification that are relevant to the GDPR: anonymization and pseudonymization. Anonymization is a process for making it impossible to identify the data subject in a data set, while pseudonymization is a process for making it impossible to identify the data subject without additional information. Both processes, performed properly, can assure data privacy.[46] However, lack of clarity about when de-identified data fully satisfies the GDPR's terms may deter companies from using it, lest it cause them legal trouble. This deterrent effect will limit the supply of de-identified data for use in AI and drive companies serving customers in the EU to use personal data under full GDPR constraints wherever de-identified data would suffice—while companies serving customers in other markets will be able to experiment with de-identified data more freely. The EU should provide clear, practical guidelines companies can use to make sure they are in full compliance with the GDPR's rules for de-identification, lest uncertainty deter them from making use of the exemptions the GDPR provides.

The GDPR Relaxes the Rules for Anonymized and Pseudonymized Data to Enable Innovation

In principle, anonymized data is not subject to any GDPR restrictions, as the word “anonymization” does not appear anywhere in the GDPR. However, Article 2 states the regulation applies to personal data, while Article 4 defines personal data as “any information relating to an individual or identifiable natural person.” Recital 26 explains that the GDPR does not apply to “anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable.”

Recital 26 also states “personal data which have undergone pseudonymization, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person.” Article 4 defines pseudonymization as “the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided such additional information is kept separately and is subject to separate technical and organizational measures to ensure the personal data is not attributed to an identified or identifiable natural person.” Unlike fully anonymized data, pseudonymized data is still subject to the GDPR, although some restrictions are relaxed for pseudonymized data. In this sense, pseudonymization is an attempt at compromise intended to give companies flexibility in how they can process data that has undergone different levels of de-identification. For example, Article 6 creates some leeway for companies to repurpose data without additional consent, provided they take certain specific factors into account, such as “the existence of appropriate safeguards, which may include encryption or pseudonymization.”

Vague Law Will Undermine De-identification

Recital 26 states the exemptions do not apply when it is “reasonably likely to be used” to reidentify the data subject “directly or indirectly.” But there is disagreement about when de-identification is reasonably likely, in part because academic studies on the subject are often misinterpreted.[47] For example, some commentators have claimed a 2013 study on de-identified spatial-temporal data had identified 95 percent of the individuals in the data set, when in fact it had done no such thing. The study proved that in 95 percent of cases, each individual’s mobility data was so unique that three randomly chosen data points were sufficient to distinguish them from other individuals—although the researchers did not discover the identity of any of the individuals in the data set.[48] As a result of the combination of mistrust of de-identification and the legal haziness in the GDPR, companies developing or using AI may decide not to take advantage of GDPR exemptions for de-identified data, lest they face a backlash from regulators for not de-identifying data properly.

De-identification Works, Contrary to Widespread Misconceptions

Due to widespread misunderstanding and misrepresentation of primary research on de-identification, a number of people mistakenly believe de-identification does not work.[49] For example, a widely cited 2013 study by researchers in Belgium showed that four pieces of time-stamped location data from cell towers were sufficient to distinguish between unique individuals in 95 percent of a sample of 1.5 million people.[50] But the researchers did not personally identify anyone, and though they suggested this might be possible with additional information (such as addresses or social-media data), they did not attempt to prove this or explain how others could prove it. The authors also only removed obvious identifiers, which would not even qualify as anonymization under the GDPR—instead falling under the rules governing pseudonymization.[51] Nevertheless, many commentators and policymakers persistently and wrongly portray reidentification as something that is simple and easy to do.[52]

THE GDPR'S COMPLEXITY WILL RAISE THE COST OF USING AI

The Economist quotes one privacy specialist at the Free University of Brussels as calling the GDPR the "most complex piece of regulation the EU has ever produced."[53] The GDPR contains so many rules that virtually all companies processing personal data will have to hire privacy professionals to keep them on the right side of the law, thereby diverting resources that could otherwise be spent on innovation.[54] Firms developing or using AI will likely face the greatest costs because of the special attention the GDPR pays to automated processing and the challenges provisions like the right to erasure pose to AI in particular. These costs will inhibit the advancement of AI in Europe by displacing investment in research and development, thus making it more difficult for European AI firms to take off, and discouraging foreign AI companies from entering the European market. The EU should revise and simplify the GDPR such that, at the very least, everyone will know how to follow it.

The GDPR Will Raise Labor Costs

EU companies developing or using AI will have to hire specialists in EU data protection law to give themselves the best chance of fully complying with the GDPR. The regulation itself requires firms to designate a data protection officer (DPO) to be responsible for monitoring compliance and liaising with the authorities, but it does not say the person must be dedicated solely to that role, or even that the individual should be a full-time member of staff.[55] However, in practice, companies will need somebody who knows the law inside out to make sure they are compliant. This means companies using personal data in AI will spend time and money protecting themselves from legal risks they could otherwise spend on innovation that benefits their customers and the advancement of AI generally.[56] The International Association of Privacy Professionals (IAPP) argues, "This is not the job of a low-level compliance manager. This is clearly a savvy operator within the business or public body, someone who can serve many constituencies, evaluate risk, and prioritize efforts."[57]

Hiring a DPO who is qualified to ensure GDPR compliance in a company processing personal data with AI systems on a large scale is a significant challenge. The right candidate needs to not only understand a variety of sophisticated uses of data-driven technologies and how they relate not only to provisions in the GDPR itself, but also to the various nuances in national law, such as how national laws on data retention impact the right to erasure. Professionals with this kind of expertise are hard to find, and demand for them will grow as the GDPR comes into force, thus creating a severe shortage of these workers in the near future. To put the problem in context, the IAPP, the largest association of privacy professionals in the world, has 30,000 members in 100 countries, not all of whom, presumably, specialize in European law. This number falls far short of the 75,000 data protection officers the IAPP estimates companies around the world will need in order to comply with the GDPR alone.[58]

THE GDPR IS LEGALLY HAZARDOUS FOR FIRMS USING AI

The GDPR's complexity also makes using AI legally hazardous. Companies developing or using AI are much more likely to find themselves the unwitting target of legal action after the GDPR goes into effect. With so many potential points of failure, many companies developing or using AI could find themselves subject to hefty fines simply because they do not adequately understand the GDPR's complexity—not because they are trying to do anything illegal or even harm consumers. Those fines will also be much larger than under the previous data protection regime. The cost and likelihood of inadvertent noncompliance could prove an even stronger disincentive to the use of AI in Europe than the costs of compliance, because some companies, especially small and mid-sized firms, may decide that investing in AI in the EU is not worth the risk. The EU should therefore amend the GDPR to make fines proportional to both the harm done and the firm's culpability.

More Companies Developing or Using AI Will Face Legal Trouble Under the GDPR Than They Do Currently

The number of companies prosecuted under data protection law will rise when the GDPR comes into force. Gartner, a research and consulting firm, estimates 50 percent of organizations will fail to comply with the GDPR as a result of being unprepared when it goes into effect in 2018.[59] Another consultancy, Vanson Bourne, found in a 2016 study commissioned by the security company Symantec that 96 percent of organizations did not fully understand the GDPR, 74 percent did not believe they were compliant, and 23 percent did not expect to become compliant by the time it comes into force.[60]

The GDPR Imposes the Highest Fines in Areas That Matter Most to AI

Compounding this heightened legal risk are higher stakes. The GDPR allows regulators to impose very large penalties for infringements, whereas the Data Protection Directive leaves that matter to national law.[61] For example, the Manchester-based cybersecurity company NCC Group found that if the GDPR had been in force in 2016, the £880,500 in fines the U.K.'s Information Commissioner's Office levied against British companies would have been

£69 million.[62] This is because rules created under the current Data Protection Act 1998 caps fines at £500,000, the GDPR will allow fines of up to 4 percent of global turnover, or €20 million (whichever is greater).[63] The "whichever is greater" clause means the GDPR's fines hit smaller companies harder than large companies, because while the latter will never pay more than 4 percent of turnover, a €20 million fine for a start-up with only a few employees could put them out of business.

Worse yet, the largest fines in the GDPR apply to the provisions that are most problematic for AI. Under Article 83 of the GDPR, regulators will be able to impose the maximum fine for violations such as repurposing data without explicit consent (see section on "The Prohibition on Repurposing Data Will Constrain AI Innovation"), or storing data outside the EU without a legal basis (see section on "Data Localization Raises AI Costs").[64] The GDPR also allows fines of up to 2 percent of turnover, or €10 million (again, whichever is greater), for lesser violations such as failing to complete adequate privacy-impact assessments.

DATA LOCALIZATION RAISES AI COSTS

Chapter 5 of the GDPR makes it illegal for EU member states to obstruct data flows to other member states on privacy grounds. This prohibition is a positive development because it would allow companies using AI to access the services of competing cloud services providers throughout the EU, instead of in just one country, thereby creating a larger, more competitive market for cloud services. Competitive pricing for cloud services makes it more affordable for companies to use AI to store and analyze large data sets.[65]

However, the GDPR maintains the existing general prohibition on transfers of personal data out of the European Union, only allowing out-of-union transfers in specific circumstances, such as to countries whose data protection laws the European Commission deems "adequate," or where there are specific safeguards or binding corporate rules that ensure the GDPR remains enforceable by EU authorities (the Privacy Shield framework, for example). Thus, it will still be difficult for EU companies developing or using AI to use competing cloud services outside of the EU.[66] Even though the GDPR's rules on data localization are preferable to the status quo, they still fail to address the fundamental problem of European privacy law unnecessarily restricting where organizations may store data. The EU should therefore abolish all data localization laws.

Data localization rules are unnecessary distortions that provide no privacy protection but raise the cost of AI by making cloud services less competitive.[67] If a company is legally answerable to European courts, then the privacy risks of storing Europeans' data in another country are no greater than those of storing it in the EU, because the company would still have to treat the data according to EU law, with EU courts holding the company accountable for any failures on its part or those of its contractors.[68]

DATA PORTABILITY WILL STIMULATE AI COMPETITION, BUT AT A COST

The right of data subjects to port their personal data to other service providers is one of the few provisions in the GDPR that is likely to have at least some offsetting positive impacts on AI, at least in the short term, because it will increase and diversify the amount of data available for new AI services. However, the cost of providing users access to extremely large, complex, and disparate data sets accumulated over many years could weaken incentives for companies to collect and store that much data in the first place. Whenever companies can prove to regulators that the cost of a particular data portability request is excessive, the law should allow for alternatives, such as having the customer pay a reasonable sum, as is common with costly freedom-of-information requests.

Data Portability Will Stimulate Competition in AI

Under Article 20(1), data subjects have the right to both receive—in a structured, commonly used, machine-readable format—all personal data they have provided to the data controller, and share that data with anyone they choose.[69] This means consumers can port their data from their existing service providers to competitors, including start-ups using AI to extract value from this information. Moreover, data portability will allow consumers to aggregate data from multiple service providers, thereby creating more comprehensive data sets. Consumers will then be able to share these data sets with both new and existing service providers.

Wherever feasible, companies must transmit personal data on behalf of their customers, at their request. Article 20(2) states that, wherever technically feasible, data subjects should have the right to transmit personal data directly from one controller to another. Direct transfers are easier and simpler for customers than receiving and sharing their data in a downloadable file. According to Article 29 Working Party, the right to transmit data from one controller to another does not amount to an obligation on companies to maintain compatible data processing systems, but it does mean they cannot deliberately obstruct the direct transfer of data to another provider.[70] Wherever system incompatibility—or other technical obstacles, such as the quantity of data—make such transfer infeasible, customers retain their right under Article 20(1) to receive and transmit the data themselves.

The right to data portability only covers data collected on the basis of consent or a contract. The data subject has no data portability rights on data collected according to other legal bases specified in the GDPR, such as compliance with a legal obligation. For example, the data subject would not be able to port data collected as part of anti-fraud checks on their activity.

Data Portability Does Not Threaten Intellectual Property

The right to data portability does not cover data that constitutes intellectual property or might reveal trade secrets, as the GDPR stresses that the right to data portability must not adversely affect the rights of others. For example, outputs from algorithms, such as personalized recommendations, are generally not covered by data portability because they could potentially reveal trade secrets of the algorithm.[71] Researchers at the Tilberg Institute suggest that clashes between data portability and intellectual property could give rise to companies porting data for specific purposes that do not threaten their legitimate interests, similar to certain sector-specific portability rules designed to address anticompetitive behavior, rather than the general-use data portability policymakers had envisaged.[72]

WP29 interprets the right to data portability as a right to all "observed" data about the user, but not to the "inferred" data.[73] This means the data subject has the right to data they have directly provided or the controller has collected about them (subject to the limitations mentioned above), but not to processed data that has been extrapolated from the original data. For example, the data portability requirement would force a movie-streaming service to allow its customers to download a list of all the movies they have already watched (i.e., "observed data"), but not the stream service's recommendations about movies its customers might enjoy ("inferred data"). In the event something about the "observed" data, such as its structure, exposes trade secrets or even constitutes intellectual property, WP29 guidelines state the company may provide the data in a form that does not do this. Although the GDPR only makes explicit allowances for feasibility with regard to direct transfers to other companies, cases where separating personal data from intellectual property is not feasible would likely qualify for an exemption, as the GDPR says data portability is subordinate to other rights.

Data portability will be costly, especially in cases involving large, complex, and disparate data sets. The anticipation of receiving a large number of such requests can negatively impact companies' incentives to collect and store such data in the first place—which could constrain the data available to fuel AI innovation. The WP29 guidelines state that in cases where customers demand an unwieldy amount of data, companies may provide the data in an appropriate format, such as physical storage media, and are not obligated to supply all of the data via the Internet, for example. But unlike the rules on transfers between data processors, the GDPR makes no allowances for technical difficulties in providing data to data subjects.

THE GDPR WILL HURT EUROPEAN COMPETITIVENESS BY LIMITING THE USE OF AI IN THE EU

The core economic value of AI lies in its ability to automate complex processes, which like previous waves of automation, promises to dramatically improve economic output, increase societal wealth, and raise living standards. But the GDPR's restrictions on AI will make it much harder for the EU to strengthen its economy using AI, thus letting other parts of the world race ahead. In addition to undermining the benefits of AI use to European industry, the EU's limitations on AI will also make it very difficult for European AI firms to become leaders in developing and supplying AI services.

AI makes business processes more efficient by automating tasks, which cuts the cost of production and frees up human labor. Besides the savings and efficiency gains of automating those tasks, in the longer term, AI also boosts productivity by diverting human labor to more economically useful activities they previously lacked the time to address, further improving economic output and wealth creation. The concomitant result of automation and technological advancement in European industry over the last two centuries has been to drastically raise the living standards of even the poorest Europeans. AI is another wave of industrial automation that promises to continue the beneficial impact of technology on the economy.

The potential value of AI to an advanced and high-skilled economy like that of the EU is therefore twofold: besides the overall competitive value boosting efficiency in European industries, the value of AI to those industries creates a market for firms that can develop and supply the AI tools needed globally for the next wave of industrial automation. Europe may be an economic powerhouse, but it has failed to produce any truly large Internet giants. AI development could have been another opportunity for Europe to produce some major players in this space and respond to worldwide demand for AI, but because of the GDPR, European AI firms have to compete with essentially one hand tied behind their back.

THERE ARE BETTER WAYS TO PROTECT EU CONSUMERS

The GDPR provisions targeting algorithms will not only inhibit AI without protecting users from unfair decisions, they will stifle AI development altogether in the EU and drive out foreign AI companies—resulting in fewer choices for European consumers and the businesses that serve them. There are, however, other ways the EU can protect consumers' data without stifling artificial intelligence.

Rather than putting direct restrictions on algorithms, policymakers could scrutinize decisions about consumers based on the seriousness—and context—of those decisions, not the technology used to make them. More broadly, data protection law should prohibit actions that are known to be harmful, and punish errors that have deleterious effects, without treating every repurposing of data as a potential privacy threat, as the GDPR does now. Indeed, by limiting the deployment of AI, the GDPR may actually erode consumer privacy, as AI has the potential to reduce the threat of other individuals having unauthorized access to personal information.[74]

Safeguards Against Unfair Decisions Should Focus on the Decisions Themselves, Not on Algorithms

Both the general prohibition on solely automated decisions that have legal or similarly significant ramifications and the requirement to provide “meaningful information” in such cases are meant to be safeguards against biased or unfair algorithmic decisions. These methods, however, have proven to be ineffective in preventing unfair decisions by algorithms—and they do not take into account humans being both far more susceptible to bias than AI systems and just as likely to make bad choices when given inaccurate data.[75] There is no question that when companies make important decisions about people, they should be held accountable. But the best way to achieve this accountability is to scrutinize the decisions regardless of whether they were made by software or a person.

The potential for algorithmic bias and unfairness arises from important variables, such as credit risk, correlating with characteristics that should not form the basis of a decision about a person, such as ethnicity. Although it is possible to program an algorithm to ignore sensitive characteristics, because indicators of those characteristics can be quite subtle, it is extremely difficult to account for all of them in advance.[76] This subtlety also causes human reviewers to easily miss the significance of such markers when analyzing an individual decision—not to mention their own personal biases potentially influencing them to deliberately disregard such information.

In the same way companies can require their workers to follow procedures that would be open to scrutiny in the event an employee's impartiality were ever to come into question, firms can control for algorithmic bias by both monitoring how an algorithm behaves over time and investigating the causes of any troubling correlations—which could as easily result from exogenous factors like poverty as unfair bias. Inaccurate data can equally lead to humans or algorithms making unfair decisions, which is why European law includes a right to amend inaccurate personal data—though humans are far more likely than algorithms to make errors because of inaccurate data.

When a decision has the potential to do the consumer harm—such as refusing a loan—the consumer's right to know exactly what information the decision was based on should be independent of whether AI was involved. Other, more serious decisions—such as demanding early repayment of a loan—may require a more specific explanation of the rationale used to make that decision, again regardless of whether algorithms had anything to do with it. With such a requirement, a company claiming it was unable to justify its decision because the algorithm it used was too complicated would be the equivalent of it admitting, "We do not know why our guy does the things he does, but we trust him, and we reckon he has it right." It is not only unnecessary to create a specific rule targeting algorithmic decisions, doing so weakens consumer protection by failing to identify the root causes of any potential problem.

RECOMMENDATIONS

There are nuances to the GDPR that could provide some leeway for regulators and courts to interpret the law in ways that could be less harmful to AI's development and use in the EU. However, those interpretations could also end up being more harmful to AI, so in reality this is as much a risk as an opportunity. The only way for these unambiguously damaging aspects of the GDPR to be eliminated is by policymakers amending the regulation. But given both the tremendous time and effort invested in creating the GDPR and the generally cumbersome nature of EU lawmaking, amending the GDPR would be a tall order. It is, however, the best—and perhaps only—way for Europe to eliminate this competitive disadvantage and become a leader in AI.

SIMPLIFY THE GDPR

While several GDPR provisions pose specific problems for AI, the overall complexity of the regulation makes its harmful impact on AI worse than the sum of its most damaging parts. Research by consultancies and cybersecurity firms suggests companies generally do not understand their GDPR obligations, with many firms likely to find themselves on the wrong side of them.[77] When laws are too difficult to follow, regulators end up not enforcing the standards of behavior the laws were ostensibly written to create. Ironically, the GDPR calls for firms to publish clear and concise privacy notices, yet the law companies must comply with is over 250 pages of legalese. If the EU is serious about building a rules-based single market with high standards, it needs to drastically simplify the GDPR by reducing it to a set of easily comprehensible rules that focus exclusively on preventing consumer harm, rather than trying to tightly control how companies manage and use data, at the expense of innovation.

REMOVE THE RIGHT TO HUMAN REVIEW

The right to a human review of algorithmic decisions will make AI more expensive and force companies to use less accurate AI systems, without protecting consumers from harmful or unfair decisions—in large part because humans can be more biased and inscrutable than algorithms. Instead, policymakers should encourage the use of tools such as the disparate impact analysis that was developed to combat bias against protected classes of individuals.[78]

MAKE THE RIGHTS TO REVIEW AND EXPLANATION TECHNOLOGY-NEUTRAL

Any requirements for transparency, evidence, oversight, or explanation should be technology-neutral. The EU should ensure that an individual’s right to a review or an explanation of a particular decision should depend on the nature and seriousness of the decision in question, not simply whether the decision was made by a human or an algorithm. Applying these rights exclusively to decisions made by algorithms creates a disincentive for companies to use AI as it represents an additional compliance cost and makes using the technology less efficient. Moreover, such a requirement would allow unfair decisions made by humans, which tend to be more open to bias in the first place, to avoid similar levels of scrutiny and accountability. Finally, the EU should take into consideration that different rules may be necessary for different industries and avoid general requirements.

EXPLAIN “MEANINGFUL INFORMATION”

The EU should amend the GDPR to clarify that, in the context of algorithmic decisions, providing “meaningful information” means a data processor should present the data subject with a description of the data the algorithm uses and a basic explanation of how it makes decisions. Requiring every conceivable nuance of each individual decision to be explained would be costly and deter the use of AI. Moreover, complicated explanations are unlikely to be useful to the average consumer.

CURTAIL THE RIGHT TO ERASURE

Depending on how regulators implement the law, forcing companies to erase data from algorithmic models could damage the algorithm and undermine its benefits to users. Regulators should take as liberal an interpretation of “erasure” as they can and allow companies to delete information in the conventional way, without impairing algorithmic models. However, because the distinction between the computing terms “erasure” and “deletion” is so clear, the best solution is for the EU to amend the GDPR to replace the former word with the latter. Moreover, whichever word applies, customers' rights to delete their data should not usurp the algorithms' functionality in regard to working for other customers. To that end, rather than a right to erasure, consumers should have a right to anonymity, wherein they can require companies to either delete their information in such a way that does not interfere with an algorithm’s behavior, or anonymize their data before the additional processing. As "anonymization" means retaining only that which is not personally identifiable, fulfilling the right to erasure through anonymization may be legally permissible within the current wording of the GDPR. Amending the GDPR to clarify this would undoubtedly help remove uncertainty.

The right to erasure should also not apply to data that was put into the public domain lawfully—that is, the so-called “right to be forgotten” provision should be repealed altogether. Information in the public domain is a public good that is also a valuable component of algorithmic tools, such as search or translation.

ALLOW REPURPOSING OF DATA THAT POSES NO RISK TO THE DATA SUBJECT

The GDPR treats all repurposing of data beyond its original purpose as a serious offense, for which it imposes maximum fines that severely limit offending AI companies’ ability to experiment and innovate using personal data. But not all reuses of data are harmful. In fact, many are extremely beneficial to EU society. The EU should therefore amend the GDPR to make the repurposing of data without asking for consent legal, provided it neither poses a risk to the data subject nor transfers data to another controller. When such a transfer is merely the result of a merger or buyout, notification should be sufficient.

AMEND DATA PORTABILITY RIGHTS TO ACCOUNT FOR COSTS

Limited data portability rights can be useful for firms using AI by making it easier for them to access larger and more diverse data sets. But enforcing this right without reasonable limitations will impose tremendous costs that could disincentivize some forms of data collection, with consequent limitations for AI.

Customers should help cover a portion of the particularly high cost of porting—just as citizens exercising their right to information from the government often have to pay a small fee for freedom- of-information requests, which are extraordinarily costly to the taxpayer.

PROVIDE CLEARER GUIDELINES FOR DE-IDENTIFYING DATA

Policymakers were right to include exemptions in the GDPR for pseudonymized and anonymized data, as it incentivizes companies developing or using AI to do more with data that poses no threat to privacy. But the current rules on de-identification are vague, so clearer guidelines would make it easier for companies to de-identify data with confidence.

Data-protection regulators in the EU should come up with clearer guidelines for how companies may legally keep data anonymous, as this would encourage companies developing or using AI to de-identify more data, knowing their legal responsibility relates to their actions and not to eventualities beyond their control. Overall, they would anonymize more data for use in AI, thus allowing a wider range of data processing activities than the GDPR permits for personal data.

In the United States, Section 164.514 of the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule sets standards for de-identifying data, unlike the GDPR.[79] Although HIPAA's supporting definition of de-identification is essentially the same as the GDPR's definition of anonymization, as the IAPP points out, a company following the HIPAA guidelines in the EU cannot be reasonably certain it is meeting the GDPR's requirements for pseudonymization—let alone for anonymization.[80]

Admittedly, HIPAA deals with a narrow type of data, whereas the GDPR relates to all personal data—meaning it is easier for HIPAA to offer companies specific steps to anonymizing data satisfactorily. Nevertheless, guidelines that at least take these types of steps into account and, like HIPAA, allow for expert certification of anonymized data, would allow companies developing or using AI in the EU to confidently reap the benefits of anonymized data that can be processed outside of the GDPR.

MAKE FINES FOR BREACHING THE GDPR PROPORTIONAL TO HARM AND CULPABILITY

The GDPR imposes extraordinarily large fines for breaches to the provisions that are considered especially harmful to AI, such as the right to explanation and the general prohibition on repurposing data. Fines for data breaches should be proportional to the extent of both the harm caused by the breach and the firm's culpability for it.[81] Companies should not be fined for activities that do no harm.

For example, letting consumers' credit card details fall into the wrong hands can result in significant costs to both consumers and their credit card providers. Should the breach be a result of negligent cybersecurity practices, in addition to being liable for the costs, data controllers should face large fines as an incentive to behave more responsibly in the future. And whenever an unforeseeable cyber-attack succeeds in spite of the controller's best efforts, the firm should only be liable for costs, and not face additional fines. On the other hand, whenever a company uses personal data to test a new algorithm and doing so does not harm any consumers, it should not incur any penalties at all. Companies developing or using AI should be able to experiment with new algorithms without first having to seek permission from every data subject for fear of being fined—particularly €20 million or more—because this activity is not a threat to data subjects.

AUTHORIZE USES OF AI IN THE PUBLIC INTEREST

One of the legal bases for data processing in the GDPR is performing tasks in the public interest or in the exercise of official authority, and as such, uses of data that rely on this legal basis are not subject to all of the same restrictions as other uses. European national governments should liberally apply this authorization to exempt uses of AI that serve the public interest, including in areas such as health care, education, and the environment. For example, AI has important applications in the health care sector, where it can help doctors diagnose diseases early and identify potentially effective treatments.[82] This authorization creates an opportunity for European governments to both lead in AI and demonstrate its usefulness by deploying AI in the public sector. European national governments should use AI to make public services more efficient and deliver the best possible outcomes for the people that use them.

THE GDPR IS THE WRONG FRAMEWORK FOR AI

If EU policymakers want to accelerate European productivity and competitiveness through AI, then the Commission will need to submit a proposal to amend the GDPR. Unfortunately, given the amount of time and energy it originally took to finalize the GDPR (the European Parliament and the Council of the European Union finally adopted the regulation four years after the Commission submitted its original proposal), the argument that policymakers need to revisit it shortly after it goes into effect is unlikely to be a popular one. However, amending the GDPR is the only way to deal with all of the harmful effects the regulation is set to have on AI. Therefore, rather than taking a victory lap for an ultimately flawed piece of legislation, the Commission should immediately begin work on GDPR 2.0.

The GDPR has important lessons for policymakers in other parts of the world, particularly those in regional trade blocs. Like all EU single-market regulation, the GDPR rests on the principle that it is better for the EU to have one data protection law than 28, because regulatory fragmentation obstructs the flow of digital services. This is a sound principle, but it does not follow that the rest of the world has to copy the GDPR in order to engage in digital trade with the EU. Copying the GDPR would be an unmitigated mistake because the regulation includes unnecessary restrictions that would be as damaging anywhere else as they promise to be in Europe.

The GDPR puts pressure on other countries to copy EU regulations under the guise that doing so eases trade with the world's largest single market.[83] But countries do not have to adopt the GDPR to get free data flows with the EU, as Article 46 of the GDPR allows data to flow freely when there are mechanisms in place for enforcing EU law for EU data, such as the Privacy Shield arrangement with the United States. Insofar as governments resist adopting the GDPR, so too will they lessen pressure on companies to make the GDPR the de facto set of rules for their global activities.

The GDPR is the wrong framework for AI and the digital economy—both in Europe and everywhere else. The regulation in its current form will make it very difficult for the EU to compete with other regions in which businesses have a freer hand in the development and use of AI. The EU should simplify the GDPR and cut the restrictions that threaten to tie down its digital economy for years to come. At the same time, policymakers elsewhere who pay close attention to EU affairs should consider the GDPR a reminder of the EU's mistakes, as well as its successes, and learn from them.

REFERENCES

1. Daniel Castro and Joshua New, *The Promise of Artificial Intelligence,* (Center for Data Innovation, October 2016), http://www2.datainnovation.org/2016-promise-of-ai.pdf.
2. Bryce Goodman and Seth Flaxman, "European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation,'" presented at ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, NY, June 2016, accessed December 15, 2017, http://adsabs.harvard.edu/cgi-bin/bib_query?arXiv:1606.08813; Innocent Kamwa, S. R. Samantary, Geza Jobs, "On the Accuracy Versus Transparency Trade-off of Data-Mining Models for Fast-Response PMU-Based Catastrophe Predictors," *IEEE Transactions on Smart Grid,* Volume 3, Issue 1, March 2012, accessed February 1, 2018, http://ieeexplore.ieee.org/abstract/document/6096427/; U Johannson, U Norinder, H Boström, "Trade-Off Between Accuracy and Interoperability for Predictive In-Silico Modelling," *Future Med Chem,* April 2011, 3(6):647-663, accessed February 1, 2018, https://www.ncbi.nlm.nih.gov/pubmed/21554073.
3. Ibid.
4. Eduard Forsch, Peter Kiseberg, and Tiffany Li, "Humans Forget, Machines Remember: Artificial Intelligence and the Right to be Forgotten," *Computer Security & Law Review* (forthcoming), August 15, 2017, accessed December 16, 2017, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3018186.
5. Ibid; For definitions of terms such as "personal data," "data subject," and "data controller," see Regulation 2016/679 (General Data Protection Regulation), Article 4, (see page L 119/82-33-35), accessed December 20, 2017, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.
6. Ludwig Siegle, "New EU Data Rules Will Get Tough on Privacy," *The Economist: The World in 2018,* accessed December 19, 2017, http://www.theworldin.com/edition/2018/article/14563/dodd-frank-data.
7. Rob van der Meulen, "Top 5 Priorities to Prepare for EU GDPR," (Gartner, June 20, 2017), accessed December 20, 2017, https://www.gartner.com/smarterwithgartner/top-five-priorities-to-prepare-for-eu-gdpr/; *Symantec state of European Data Privacy,* (Symantec/Vanson Bourne, October 18, 2016), see https://www.symantec.com/en/uk/about/newsroom/press-releases/2016/symantec_1018_01 and https://www.slideshare.net/symantec/symantec-state-of-european-data-privacy, (both accessed December 20, 2017).
8. Regulation 2016/679 (General Data Protection Regulation), Article 83, (see page L 119/82-83), accessed December 20, 2017, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.
9. Nigel Cory, *Cross Border Data Flows: Where Are the Barriers, and What Do They Cost?,* (Information Technology and Innovation Foundation, May 1, 2017), http://www2.itif.org/2017-cross-border-data-flows.pdf.
10. Nick Wallace, "European Commission Should Stand Firm on Free Data Flows," (Center for Data Innovation, March 8, 2017), https://www.datainnovation.org/2017/03/european-commission-should-stand-firm-on-free-data-flows/.

11. Daniel Castro and Travis Korte, *Data Innovation 101,* (Center for Data Innovation, November 3, 2013), https://www.datainnovation.org/2013/11/data-innovation-101/.

12. Daniel Castro and Joshua New, *The Promise of Artificial Intelligence,* (Center for Data Innovation, October 2016), http://www2.datainnovation.org/2016-promise-of-ai.pdf.

13. "Making the Most of Robotics and Artificial Intelligence in Europe," (European Commission, November 17, 2017), accessed December 22, 2017, https://ec.europa.eu/commission/commissioners/2014-2019/ansip/blog/making-most-robotics-and-artificial-intelligence-europe_en.

14. "Potential Global Regional Gains from AI," (PwC, June 27, 2017), accessed February 13, 2018, https://press.pwc.com/News-releases/ai-to-drive-gdp-gains-of--15.7-trillion-with-productivity--personalisation-improvements/s/3cc702e4-9cac-4a17-85b9-71769fba82a6, and https://press.pwc.com/Multimedia/News-releases/All/potential-global-regional-gains-from-ai/a/a070cb3b-fd32-4a38-a59a-468d7d378af5.

15. Daniel Castro and Joshua New, *The Promise of Artificial Intelligence,* (Center for Data Innovation, October 2016), http://www2.datainnovation.org/2016-promise-of-ai.pdf.

16. Ibid.

17. Ibid.

18. Regulation 2016/679 (General Data Protection Regulation), Article 22, (see page L 119/46), accessed December 19, 2017, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.

19. Regulation 2016/679 (General Data Protection Regulation), Article 22(3), (see page L 119/46), accessed December19, 2017, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.

20. *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679,* (Article 29 Data Protection Working Party, February 6, 2018), accessed March 14, 2018, http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053.

21. Ibid.

22. Bryce Goodman and Seth Flaxman, "European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation,'" presented at ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, NY, June 2016, accessed December 15, 2017, http://adsabs.harvard.edu/cgi-bin/bib_query?arXiv:1606.08813.

23. Regulation 2016/679 (General Data Protection Regulation), Article 13-14, (see page L 119/40-42), accessed December19, 2017, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.

24. Andrew D. Selbst and Julia Powles, "Meaningful Information and the Right to Explanation," *International Privacy Law,* Volume 7, Issue 4, November 1, 2017, pages 233-242, accessed January 30, 2018, https://doi.org/10.1093/idpl/ipx022; Ibid.

25. Bryce Goodman and Seth Flaxman, "European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation,'" presented at ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, NY, June 2016, accessed December 15, 2017, http://adsabs.harvard.edu/cgi-bin/bib_query?arXiv:1606.08813; Innocent Kamwa, S. R. Samantary, Geza Jobs, "On the Accuracy Versus Transparency

Trade-off of Data-Mining Models for Fast-Response PMU-Based Catastrophe Predictors," *IEEE Transactions on Smart Grid,* Volume 3, Issue 1, March 2012, accessed February 1, 2018, http://ieeexplore.ieee.org/abstract/document/6096427/; U Johannson, U Norinder, H Boström, "Trade-Off Between Accuracy and Interoperability for Predictive In-Silico Modelling," *Future Med Chem,* April 2011, 3(6):647-663, accessed February 1, 2018, https://www.ncbi.nlm.nih.gov/pubmed/21554073.

26. *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679,* (Article 29 Data Protection Working Party, February 6, 2018), accessed March 14, 2018, http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053.

27. Bryce Goodman and Seth Flaxman, "European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation,'" presented at ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, NY, June 2016, accessed December 15, 2017, http://adsabs.harvard.edu/cgi-bin/bib_query?arXiv:1606.08813.

28. Ibid; Innocent Kamwa, S. R. Samantary, Geza Jobs, "On the Accuracy Versus Transparency Trade-off of Data-Mining Models for Fast-Response PMU-Based Catastrophe Predictors," *IEEE Transactions on Smart Grid,* Volume 3, Issue 1, March 2012, accessed February 1, 2018, http://ieeexplore.ieee.org/abstract/document/6096427/; U Johannson, U Norinder, H Boström, "Trade-Off Between Accuracy and Interoperability for Predictive In-Silico Modelling," *Future Med Chem,* April 2011, 3(6):647-663, accessed February 1, 2018, https://www.ncbi.nlm.nih.gov/pubmed/21554073.

29. Zachary C. Lipton, "The Mythos of Interpretability," 2016 ICML Workshop on Human Interoperability in Machine Learning (WHI 2016), New York, NY, USA, accessed February 1, 2018, https://arxiv.org/pdf/1606.03490.pdf.

30. Ibid.

31. Regulation 2016/679 (General Data Protection Regulation), Article 17, (see page L 119/43-44), accessed December 19, 2017, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.

32. Ibid.

33. Ibid, Article 89, (see page L 119/84-85).

34. Ibid, Article 17 (see page L 119/43-44) and Ibid, Article 7 (see page L 119/37).

35. Bernd Malle et al, "Right to be Forgotten: Towards Machine Learning on Perturbed Knowledge Bases," Workshop on Privacy Aware Machine Learning (PAML), August 2016, due to become available at https://hal.inria.fr/IFIP-LNCS-9817/hal-01635002 from January 1, 2019, (cited in "Humans Forget, Machines Remember: Artificial Intelligence and the Right to be Forgotten").

36. Ibid.

37. Ibid.

38. Ibid.

39. Regulation 2016/679 (General Data Protection Regulation), Article 6, (see page L 119/36-37), accessed December 19, 2017, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf; Directive 95/46/EC (Data Protection Directive), Article 6(1)(b), (see page L 281/40), accessed January 3, 2018, http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en.

40. Regulation 2016/679 (General Data Protection Regulation), Article 6, (see page L 119/36-37), accessed December 19, 2017, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.

41. Ibid.

42. Directive 95/46/EC (Data Protection Directive), Article 6(1)(b), (see page L 281/40), accessed January 3, 2018, http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en; Opinion 03/2013 on purpose limitation, (Article 29 Data Protection Working Party, adopted April 2, 2013), accessed January 3, 2018, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf.

43. Nick Wallace, "UK Regulations Need an Update to Make Way for Medical AI," (Center for Data Innovation, August 12, 2017), https://www.datainnovation.org/2017/08/uk-regulations-need-an-update-to-make-way-for-medical-ai/.

44. Data Protection Act 1998, Section 4(3), Schedule 3, Paragraph 8, (see pages 53-54), accessed December 19, 2017, https://www.legislation.gov.uk/ukpga/1998/29/pdfs/ukpga_19980029_en.pdf; Directive 95/46/EC (Data Protection Directive), accessed December 19, 2017, http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en.

45. Regulation 2016/679 (General Data Protection Regulation), Article 7, (see page L 119/37), accessed December 19, 2017, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.

46. Anna Cavoukian and Daniel Castro, *Big Data and Innovation, Setting the Record Straight: Anonymization* Does *Work,* (Information and Privacy Commissioner Ontario, Canada and Information Technology and Innovation Foundation, June 16, 2014), http://www2.itif.org/2014-big-data-deidentification.pdf.

47. Ibid.

48. Ibid.

49. Ibid.

50. Yves-Alexander de Montjoye et al, "Unique in the Crowd: The Privacy Bounds of Human Mobility," *Scientific Reports* 3, Article No. 1376, March 25, 2013, cited in Cavoukian and Castro, accessed December 18, 2017, http://dx.doi.org/10.1038/srep01376.

51. Ibid.; Gabe Maldoff, "Top 10 Operational Impacts of the GDPR: Part 8—Pseudonymization" (International Association of Privacy Professionals, February 12, 2016), accessed December 18, 2017, https://iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-8-pseudonymization/.

52. Ann Cavoukian and Daniel Castro, *Big Data and Innovation, Setting the Record Straight: Anonymization* Does *Work,* (Information and Privacy Commissioner Ontario, Canada and Information Technology and Innovation Foundation, June 16, 2014), http://www2.itif.org/2014-big-data-deidentification.pdf.

53. Ludwig Siegle, "New EU Data Rules Will Get Tough on Privacy," *The Economist: The World in 2018,* accessed December 19, 2017, http://www.theworldin.com/edition/2018/article/14563/dodd-frank-data.

54. Nick Wallace, "Overzealous EU Data Protection Regulations More Likely to Take Your Job Than a Robot," *City A.M.*, March 2, 2017, accessed December

20, 2017, http://www.cityam.com/260087/overzealous-eu-data-protection-regulations-more-likely-take.

55. Regulation 2016/679 (General Data Protection Regulation), Chapter IV, Section 4, (see page L 119/55-56), accessed December 20, 2017, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.

56. Nick Wallace, "Overzealous EU Data Protection Regulations More Likely to Take Your Job Than a Robot," *City A.M.*, March 2, 2017, accessed December 20, 2017, http://www.cityam.com/260087/overzealous-eu-data-protection-regulations-more-likely-take.

57. "The GDPR Demands 75k DPOs: Where Will They Come From?," (International Association of Privacy Specialists, November 19, 2016), accessed December 19, 2017, https://iapp.org/media/pdf/DPA-Whitepaper.pdf.

58. Ibid.

59. Rob van der Meulen, "Top 5 Priorities to Prepare for EU GDPR," (Gartner, June 20, 2017), accessed December 20, 2017, https://www.gartner.com/smarterwithgartner/top-five-priorities-to-prepare-for-eu-gdpr/.

60. "Symantec State of European Data Privacy," (Symantec/Vanson Bourne, October 18, 2016), see https://www.symantec.com/en/uk/about/newsroom/press-releases/2016/symantec_1018_01 and https://www.slideshare.net/symantec/symantec-state-of-european-data-privacy, (both accessed December 20, 2017).

61. Regulation 2016/679 (General Data Protection Regulation), Article 83, (see page L 119/82-83), accessed December 20, 2017, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.

62. "GDPR Impact Analysis," (NCC Group, April 28, 2017), https://www.nccgroup.trust/uk/landing-pages/gdpr-impact-analysis/, (accessed December 20, 2017).

63. Data Protection Bill (HL Bill 66), §150 (see pages 23-34), accessed December 20, 2017, https://publications.parliament.uk/pa/bills/lbill/2017-2019/0066/lbill_2017-20190066_en_1.htm; Data Protection Act 1998 §55E, accessed March 20, 2018, https://www.legislation.gov.uk/ukpga/1998/29/section/55E; The Data Protection (Monetary Penalties) (Maximum Penalty and Notices) Regulations 2010, accessed March 20, 2018, http://www.legislation.gov.uk/uksi/2010/31/regulation/2/made.

64. Regulation 2016/679 (General Data Protection Regulation), Article 83, (see page L 119/82-83), accessed December 20, 2017, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.

65. Nick Wallace, "How to Guarantee the Free Flow of Data In Europe," (Center for Data Innovation, December 7, 2017), https://www.datainnovation.org/2017/12/how-to-guarantee-the-free-flow-of-data-in-europe/.

66. Regulation 2016/679 (General Data Protection Regulation), Chapter V, (see page L 119/60-65), accessed December 20, 2017, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.

67. Nigel Cory, “Cross Border Data Flows: Where Are the Barriers, and What Do They Cost?,”(Information Technology and Innovation Foundation, May 1, 2017), http://www2.itif.org/2017-cross-border-data-flows.pdf.

68. Daniel Castro, “The False Promise of Data Nationalism,” (Information Technology and Innovation Foundation, December 2013), http://www2.itif.org/2013-false-promise-data-nationalism.pdf.

69. Regulation 2016/679 (General Data Protection Regulation), Article 20, (see page L 119/45), accessed December 20, 2017, http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf.

70. “Guidelines on the Right to Data Portability,” (Article 29 Data Protection Working Party, December 13, 2016), 16/EN WP 242, accessed December 19, 2017, http://ec.europa.eu/information_society/newsroom/image/document/2016-51/wp242_en_40852.pdf.

71. Ibid.

72. Inge Graef, Martin Husovec, and Nadezhda Purtova, “Data Portability and Data Control: Lessons for an Emerging Concept in EU Law,” November 15, 2017, accessed December 15, 2017, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3071875; Daniel Castro, “Blocked: Why Some Companies Restrict Data Access to Reduce Competition and How Open APIs Can Help,” (Center for Data Innovation, November 6, 2017), http://www2.datainnovation.org/2017-open-apis.pdf.

73. “Guidelines on the Right to Data Portability,” (Article 29 Data Protection Working Party, December 13, 2016), 16/EN WP 242, accessed December 19, 2017, http://ec.europa.eu/information_society/newsroom/image/document/2016-51/wp242_en_40852.pdf.

74. Daniel Castro and Alan McQuinn, “AI Offers Opportunity to Increase Privacy for Users,” (IAPP, January 12, 2018), https://iapp.org/news/a/ai-offers-opportunity-to-increase-privacy-for-users/.

75. Joshua New, “It’s Humans, Not Algorithms, That Have Been a Bias Problem,” (Center for Data Innovation, November 16, 2015), http://www.datainnovation.org/2015/11/its-humans-not-algorithms-that-have-a-bias-problem/.

76. Nick Wallace, “Comments to the Article 29 Working Party on Guidelines for Automated Decisions,” (Center for Data Innovation, November 28, 2017), http://www2.datainnovation.org/2017-article-29-automated-deicions.pdf.

77. Rob van der Meulen, “Top 5 Priorities to Prepare for EU GDPR,” (Gartner, June 20, 2017), accessed December 20, 2017, https://www.gartner.com/smarterwithgartner/top-five-priorities-to-prepare-for-eu-gdpr/; “Symantec State of European Data Privacy,” (Symantec/Vanson Bourne, October 18, 2016), see https://www.symantec.com/en/uk/about/newsroom/press-releases/2016/symantec_1018_01 and https://www.slideshare.net/symantec/symantec-state-of-european-data-privacy, (both accessed December 20, 2017).

78. Travis Korte and Daniel Castro, “Disparate Impact Analysis is Key to Ensuring Fairness in the Age of the Algorithm,” (Center for Data Innovation, January 20, 2015), https://www.datainnovation.org/2015/01/disparate-impact-analysis-is-key-to-ensuring-fairness-in-the-age-of-the-algorithm/.

79. “Guidance Regarding Methods for De-Identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule,” (U.S. Department for Health and

Human Services, last updated November 6, 2015, accessed December 20, 2017, https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html.

80. Matt Wes, “Looking to Comply With GDPR? Here’s a Primer on Anonymization and Pseudonymization,” (International Association of Privacy Professionals, August 25, 2017), accessed December 20, 2017), https://iapp.org/news/a/looking-to-comply-with-gdpr-heres-a-primer-on-anonymization-and-pseudonymization/.

81. Daniel Castro and Alan McQuinn, “How and When Regulators Should Intervene,” (Information Technology and Innovation Foundation, February 2015), accessed February 8, 2018, http://www2.itif.org/2015-how-when-regulators-intervene.pdf.

82. Nick Wallace, “5 Q’s for Ben Marthappu, Co-Founder of Cera,” (Center for Data Innovation, January 19 2018), https://www.datainnovation.org/2018/01/5-qs-for-ben-maruthappu-co-founder-of-cera/; Nick Wallace, “UK Regulations Need an Update to Make Way for Medical AI,” (Center for data Innovation, August 12, 2017), https://www.datainnovation.org/2017/08/uk-regulations-need-an-update-to-make-way-for-medical-ai/; Nick Wallace, “5 Q’s for Matteo Carli, Chief Technology Officer and Founder of Xbird,” (Center for Data Innovation, June 29, 2017), https://www.datainnovation.org/2017/06/5qs-for-matteo-carli-chief-technology-officer-and-founder-of-xbird/; Nick Wallace, “5 Q’s for Eyal Toledano, Co-Founder of Zebra Medical Vision,” (Center for Data Innovation, June 8, 2017), https://www.datainnovation.org/2017/06/5qs-for-eyal-toledano-co-founder-of-zebra-medical-vision/.

83. Alan Beattie, “Why the Whole World Feels the ‘Brussels Effect,’” *Financial Times,* November 16, 2017, accessed January 3, 2018, https://www.ft.com/content/7059dbf8-a82a-11e7-ab66-21cc87a2edde.

ABOUT THE AUTHORS

Nick Wallace is a senior policy analyst at the Center for Data Innovation. He has a master's degree in public policy, jointly awarded from the Central European University in Budapest and the Institut Barcelona d'Estudis Internacionals, and a bachelor's degree in politics from Liverpool John Moores University

Daniel Castro is the director of the Center for Data Innovation and vice president of the Information Technology and Innovation Foundation. He has a B.S. in foreign service from Georgetown University and an M.S. in information security technology and management from Carnegie Mellon University.

ABOUT THE CENTER FOR DATA INNOVATION

The Center for Data Innovation is the leading global think tank studying the intersection of data, technology, and public policy. With staff in Washington, D.C. and Brussels, the center formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. It educates policymakers and the public about the opportunities and challenges associated with data, as well as technology trends such as predictive analytics, open data, cloud computing, and the Internet of Things. The Center is a nonprofit, nonpartisan research institute affiliated with the Information Technology and Innovation Foundation.

contact: info@datainnovation.org

datainnovation.org