# REGULATING MACHINE LEARNING BY DESIGN

BY

**MARCO ALMADA**

Researcher, EUI Department of Law. Email: Marco.Almada@eui.eu. The author's work on this paper was partially funded by a Fundación Carolina doctoral grant. The author would also like to thank Felipe Roquete for comments in related work, which helped shape this paper.

TechREG CHRONICLE
**FEBRUARY 2023**

## REGULATING MACHINE LEARNING BY DESIGN

By Marco Almada

The regulation of digital technologies around the world draws from various regulatory techniques. One such technique is regulation by design, in which regulation specify requirements that software designers must follow when creating any systems. This paper examines the suitability of regulation by design approaches to machine learning, arguing that they are potentially useful but have a narrow scope of application. Drawing from EU law examples, it shows how regulation by design relies on the delegation of normative definitions and enforcement to software designers, but such delegation is only effective if a few conditions are present. These conditions, however, are seldom met by applications of machine learning technologies in the real world, and so regulation by design cannot address many of the pressing concerns driving regulation. Nonetheless, by-design provisions can support regulation if applied to well-defined problems that lend themselves to clear expression in software code. Hence, regulation by design, within its proper limits, can be a powerful tool for regulators of machine learning technologies.

# 01 INTRODUCTION

Machine learning ("ML") is a new frontier for regulation. Little more than a decade ago, ML-based technologies were a niche concern even in the field of technology regulation, as the field of artificial intelligence ("AI") lingered on at a low point of investment. Move forward a decade, and the situation could not be more different. The risks and opportunities associated with AI technologies have become a problem not only in domains typically associated with digital technologies, such as privacy or intellectual property, but they now permeate the most varied dimensions of social life. For example, the early months of 2023 have seen intense debates about the impact of large language models such as ChatGPT. Will these systems facilitate the spread of online misinformation? Did the creators of these systems breach intellectual property rights as they assembled the massive datasets powering them? Are the capabilities of these systems enough to transform the work of lawyers — or even replace them altogether? To answer these and other questions, regulators need to engage with the technical aspects of ML technologies.

When it comes to the governance of the technical side of ML and other digital technologies, regulators worldwide are increasingly reliant on Regulation by Design ("RbD"). In general lines, RbD operates by incorporating legal requirements into software design:[2] the law specifies requirements that a computer system must meet, and the designers of any computer system subject to that law must choose the technical arrangements that ensure the system always meets those requirements. Provisions laying down RbD requirements are common in data protection law,[3] and, since ML systems require huge data sets for their training, these requirements encompass most applications of ML directed at natural persons. However, by-design provisions are not restricted to the field of personal data. In the EU, the Digital Markets Act ("DMA") encourages those undertakings designated as gatekeepers to adopt by-design measures to foster fairness and market contestability,[4] and the proposed AI regulation ("AI Act")[5] imposes various design measures to be adopted by AI systems deemed to pose a high risk to fundamental rights and other public interests. Whenever such provisions are adopted, the technical decisions made in constructing and deploying ML systems become directly relevant for regulatory compliance. Therefore, the use of RbD amounts to a technology-sensitive approach to regulation.

This paper argues that RbD offers a powerful tool for regulators, but one with a narrow scope of application. The following section characterizes RbD as a modality of meta-regulation, in which the designers of ML systems are required to give effect to legal requirements through code. After this high-level overview, Section III argues that RbD provisions can lay down standards and offer guidance to designers as they seek to implement legal requirements in ML. But every regulatory tool comes with its drawbacks, and Section IV shows that the limits of RbD render it unsuitable to address many of the pressing challenges driving AI regulation, even being harmful to some of these regulatory goals. However, one should not throw the baby out with the bathwater, so the paper concludes with a discussion of the proper scope for RbD provisions in the regulation of ML systems.

# 02 REGULATION BY DESIGN AS META-REGULATION

A key aspect of RbD approaches is that they afford considerable discretion to software designers.[6] Consider Article 25 GDPR, titled *data protection by design and by default*.

[2] Pieter Van Cleynenbreugel, *EU By-Design Regulation in the Algorithmic Society: A Promising Way Forward or Constitutional Nightmare in the Making?*, *in* Constitutional Challenges in the Algorithmic Society 202, 202 (Amnon Reichman et al. eds., 2021).

[3] See, e.g. Article 25 of the EU General Data Protection Regulation ("GDPR": *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, OJ 2016 L 119/1), Articles 46, § 2, and 49 of the Brazilian Data Protection Law (*Lei 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais (LGPD))*, (2018)), and Article 10 of the Council of Europe's modernized Convention 108 on personal data.

[4] See Recital 65 of Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) (Text with EEA relevance), OJ 2022 L 265/1 (2022).

[5] European Commission, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts* (2021) arts. 10-15.

[6] That is, the ones responsible for the various activities involved in the creation of a system, such as the definition of overall goals and requirements for that system, the stipulation of software architectures, and the actual programming.

According to Article 25(1) GDPR, any data processing must be accompanied by technical and organizational measures that ensure personal data is processed in conformity with the data protection principles enshrined in the regulation. While the principles are determined by legislation, their interpretation in a specific data processing context, the choice of which measures will be adopted, and the implementation of these measures all remain in the hands of the actors who determine the purposes and means of data processing.[7] The effectiveness of these principles will, accordingly, hinge on the technical (and, in this case, organizational) decisions made by designers as they pursue compliance with the legal requirement.

Designer discretion is, at least in part, a consequence of the broad scope of some RbD provisions. In the GDPR example above, the category "data protection principles" covers various legal interests. Since it is unlikely that any particular measure will succeed in protecting all such interests in all circumstances, the regulation does not specify any action that must be adopted in all cases.[8] But some forms of RbD are much narrower. For example, Article 12(1) of the proposed EU AI Act stipulates that the design and development of high-risk AI systems must create the conditions for automated logging of events during the system's operation. The remaining paragraphs of Article 12 AI present various requirements that must be met by any acceptable solution for logging: any high-risk AI system that does not provide the capabilities listed there is not in compliance with the AI Act. Still, designers are free to adopt any technical approach to produce the event logs as long as they meet these requirements.

This is not to say RbD is always a "hollow" approach to regulation,[9] devoid of any enforceable substance. Courts and data protection authorities can rely on those criteria when adjudicating particular cases[10] and thus subject designer decisions to external scrutiny. Furthermore, technical decisions can be compared to external benchmarks, such as those provided by administrative guidelines,[11] technical standards, and certification procedures. By-design provisions might not be sufficient to eliminate all designer discretion, but they offer constraints to its exercise. In doing so, they set up a meta-regulatory regime, in which the state delegates to designers the specification of norms in the contexts in which a system is meant to operate while establishing mechanisms to supervise the use of delegated power.

> ***A key aspect of RbD approaches is that they afford considerable discretion to software designers***

The meta-regulatory character of RbD follows from its targeting of design decisions. When designers make technical decisions that respond to a legal requirement, they effectively hardcode an interpretation of that requirement as a rule — or, more likely, a set of rules — in a computer system. Some of these rules bind the future behavior of system users: for example, an eBook might be accompanied by Digital Rights Management mechanisms that prevent unauthorized users from reading its contents.[12] But the encoded rules might also affect third parties without direct interaction with the system, as is the case when a public sector authority relies on an ML system to assess fraud risks.[13] In both cases, the impact of legal requirements on the outcome will be mediated by how these requirements influence system outputs. So, by regulating system design practices, RbD governs the role of designers in giving force to regulation.

Scholarship on meta-regulation points out that such approaches tend to emerge when the state lacks direct reg-

[7] In GDPR parlance, the "controllers" of the personal data being processed: Lina Jasmontaite et al., *Data Protection by Design and by Default: Framing Guiding Principles into Legal Obligations in the GDPR*, 4 Eur. Data Prot. L. Rev. 168 (2018).

[8] It does, however, provide criteria that must be considered when determining the measures to be adopted, such as the risks posed by the processing, the state of the art, and the cost of implementing such measures.

[9] Aurelia Tamò-Larrieux, Designing for Privacy and its Legal Framework: Data Protection by Design and Default for the Internet of Things 209 (2018).

[10] For an overview of administrative cases on Article 25 GDPR, see Marco Almada, Juliano Maranhão & Giovanni Sartor, *Article 25. Data protection by design and by default*, *in* General Data Protection Regulation. Article-by-article commentary (Indra Spiecker gen. Döhmann et al. eds., 2023).

[11] See, e.g. EDPB, *Guidelines 4/2019 on Article 25 on Data Protection by Design and by Default*, (2020).

[12] On computer systems as a source of rules for uses, see Laurence Diver, Digisprudence: code as law rebooted (2021).

[13] See, e.g. David Hadwick & Shimeng Lan, *Lessons to Be Learned from the Dutch Childcare Allowance Scandal: A Comparative Review of Algorithmic Governance by Tax Administrations in the Netherlands, France and Germany*, 13 World Tax Journal (2021).

ulatory capability.[14] In the case of ML technologies, the capability gap stems from several factors. First, and perhaps foremost, the sheer variety of potential ML applications prevents regulators from addressing in-depth the risks associated with every single use context.[15] Second, the technological complexity of ML systems means that their regulation requires considerable resources and technical know-how, which are not always available to regulators.[16] Using RbD as a meta-regulatory strategy thus allows public regulators to tap into the resources and domain-specific knowledge available to the actors that design ML systems and still rein in their regulatory power. The following section examines some of the contributions RbD approaches can give to ML regulation.

## 03 REGULATORY INTERVENTIONS IN MACHINE LEARNING DESIGN

RbD is, by necessity, a context-sensitive practice. Certain technical solutions might be adequate for some problems and not for others: for example, approaches that minimize the use of personal data are widely promoted as conducive to the protection of privacy, but they might create obstacles to the kind of statistical analyses needed to detect algorithmic discrimination.[17] Still, all ML systems share a few traits: they rely on large data sets for training and operation, are opaque to untrained observers and technical experts, and rely on a somewhat narrow set of technical approaches and technological infrastructure.[18] Therefore, some kinds of design requirements will likely benefit various regulatory goals.

One of the primary objectives of ML regulation is to avoid, or at least mitigate, the harms produced by algorithms. In recent years, various such harms have been identified. Some of these harms have been detected early in software adoption, as is the case of the recent news on AI-powered search delivering wrongful results.[19] In other cases, the harmful impact of algorithmic systems can be more difficult to detect: in the Dutch Childcare Benefits scandal, a risk assessment system produced outputs that led to discriminatory enforcement of anti-fraud mechanisms against minoritized groups.[20] RbD approaches can contribute to avoiding such incidents by forcing designers to address *ex ante* known risks stemming from their systems.

The contribution of RbD to the quality of ML outputs comes from its binding force. If designers are mandated to achieve certain goals, or even required to use certain technical approaches, the legal obligation can lead them to use techniques that would otherwise be seen as too complex or expensive. For example, statistical metrics such as conditional demographic disparity can be used to detect whether an ML system discriminates against protected groups,[21] and accuracy benchmarks can be used to evaluate whether the system actually delivers the promised results.[22] Compliance with well-designed RbD rules

[14] Peter Grabosky, *Meta-regulation*, *in* Regulatory Theory: Foundations and applications 149, 155 (Peter Drahos ed., 1st ed. 2017).

[15] See, e.g. the claims that AI is a general-purpose technology: Manuel Trajtenberg, *Artificial Intelligence as the Next GPT: A Political-Economy Perspective*, *in* The Economics of Artificial Intelligence: An Agenda (Ajay Agrawal, Joshua Gans, & Avi Goldfarb eds., 2019).

[16] This issue is particularly salient in developing countries: Cecil Abungu, *Algorithmic Decision-Making and Discrimination in Developing Countries*, 13 Case W. Res. J.L. Tech. & Internet 39 (2022). However, even developed countries can struggle to cultivate and retain technical expertise, especially when research in ML is concentrated in a few corporate actors: Daniel Zhang et al., *The AI Index 2021 Annual Report*, (2021).

[17] Marvin van Bekkum & Frederik Zuiderveen Borgesius, *Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR need a new exception?*, 48 Comput. Law Secur. Rev 105770 (2023).

[18] For an overview of the technical convergence in AI for non-technical audiences, see Matthias Gallé, *Foundation Models in AI: what impact for policies and law?* (2022), https://www.eui.eu/Documents/Research/Clusters/Techcluster-memos/20220530-Foundation-Models-in-AI-memo.pdf (last visited Sep 29, 2022).

[19] Chloe Xiang, *Bing's ChatGPT-Powered Search Has a Misinformation Problem*, Vice (2023), https://www.vice.com/en/article/3ad3ey/bings-chatgpt-powered-search-has-a-misinformation-problem (last visited Feb 20, 2023).

[20] Amnesty International, *Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal* (2021).

[21] Sandra Wachter, Brent Mittelstadt & Chris Russell, *Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI*, 41 Comput. Law Secur. Rev 105567 (2021).

[22] On the various ways they may fail to do so, see Inioluwa Deborah Raji et al., *The Fallacy of AI Functionality*, *in* 2022 ACM Conference on Fairness, Accountability, and Transparency 959 (2022).

can lead to adopting ML systems that produce higher-quality outputs.

Output quality is not the only goal that design practices can foster. In fact, it has been argued that promoting increases in quality metrics, such as accuracy and completeness of training data sets, often happens at the expense of other human values, such as privacy and dignity.[23] To protect these values, RbD approaches can mandate the use of certain technical practices that foster them even if at the expense of some accuracy. For example, many of the most potent ML systems, such as ChatGPT, are unfathomably big and complex. Still, for some tasks, it might be possible to achieve similar results — or, at least, good enough performance — with simpler techniques that are more amenable to human scrutiny.[24] In that case, a requirement imposing the use of interpretable ML techniques, or the adoption of explainable AI techniques that reduce the complexity of the larger systems,[25] can bolster human oversight at a reduced cost to the system's fitness to purpose. Requirements such as these supply designers with solutions to value conflicts they would otherwise need to solve, thus easing compliance with general regulatory requirements.

By-design regulation is not limited to promoting a specific set of values, but it can direct designers towards values they would otherwise disregard or treat as secondary. For example, much of contemporary software development is guided by so-called agile methodologies, in which the continuing evolution of software systems is prioritized over the comprehensiveness of documentation.[26] Requirements such as the AI Act's demand for up-to-date documentation of various aspects of high-risk AI systems[27] counteract this tendency by forcing designers to evaluate whether their documentation-light approach is enough to meet legal requirements and, if not, drawing up additional documents.

The examples above show that RbD plays a triple role in shaping the regulatory activities of designers: it supplies those actors with quality standards, specifies solutions to potential value conflicts, and shifts the weight they must confer to the values considered. In doing so, RbD approaches can rely on practices that apply to all sorts of digital systems, such as those governing data protection by design. But, to address the specific challenges of ML systems, regulators will need to look into the technical arrangements that power those systems. Otherwise, they might fail to spot how the mechanisms through which ML can produce harmful or otherwise undesirable effects, as well as potential technical fixes for these problems.

# 04

# THE LIMITS OF MACHINE LEARNING REGULATION BY DESIGN

Despite its relative novelty, RbD has been extensively critiqued in technology regulation scholarship. Among other important points, it has been argued that the delegation of regulatory power to designers can suffer from legitimacy issues,[28] that software code cannot properly capture the ambiguity and value judgments that are inherently involved in legal interpretation,[29] and that rules hardcoded in software are difficult to change if the initial implementation was wrong or if the regulatory requirement changes after initial implementation.[30] These issues constrain the efficiency of by-design approaches not just in ML systems, but in any

[23] Paul Ohm, *Throttling machine learning*, *in* Life and the Law in the Era of Data-Driven Agency 214 (Mireille Hildebrandt & Kieron O'Hara eds., 2020).

[24] See, e.g. section 6 of Madalina Busuioc, Deirdre Curtin & Marco Almada, *Reclaiming Transparency: Contesting the Logics of Secrecy within the AI Act*, Eur. Law Open FirstView (2022).

[25] Adrien Bibal et al., *Legal requirements on explainability in machine learning*, 29 Artif Intell Law 149 (2021).

[26] Thomas Hoeren & Stefan Pinelli, *Agile programming – Introduction and current legal challenges*, 34 Comput. Law Secur. Rev 1131 (2018).

[27] Article 11 AI Act.

[28] Diver, *supra* note 12.

[29] Mireille Hildebrandt, *The adaptive nature of text-driven law*, 1 CRCL (2022).

[30] Lyria Bennett Moses & Monika Zalnieriute, *Law and Technology in the Dimension of Time*, *in* Time, Law, and Change: An Interdisciplinary Study 303, 317 (Sofia Ranchordás & Yaniv Roznai eds., 2020).

kind of digital system.[31] Until they are solved, if they are at all solvable, it follows that RbD approaches are at their most effective when they are deployed to implement requirements that are socially perceived as legitimate, which can be expressed in terms of "if–then" statements that can be implemented in computer code, and that are amenable to change in the future.

These conditions rarely hold in real-world ML applications. When it comes to legitimacy, the use of ML systems has a mixed track record: recommender systems are part of everyday experiences in social media, but the automation of sensitive tasks such as grading high school leaving exams has been met with protests and other forms of political contestation.[32] Furthermore, these systems produce their outputs by applying statistical models to their input model, and this reliance on statistics introduces a degree of uncertainty into any decisions relying on ML techniques. As a result, a change to regulation that seems relatively small for a human observer might require considerable rework if it is to be implemented in a ML system.[33]

> *Despite its relative novelty, RbD has been extensively critiqued in technology regulation scholarship*

Beyond these conceptual challenges, the economics of ML introduce additional constraints to RbD. The construction of ML systems requires large amounts of data and computing resources,[34] which are not accessible to most private actors — and even to many public actors. To access those capabilities, most designers rely on ML-as-a-service solutions, hiring timeshares in platforms offered by large providers such as Amazon or Google. Such arrangements not only incentivize market concentration, but they create a regulatory conundrum: the actors who use ML-as-a-service lack the power to effect change into the large systems they rely on, but the providers of the general-purpose systems offered as a service lack the context-specific knowledge they need to address the risks associated with each possible application of their tools. Design requirements for ML systems walk a thin line between imposing impossible obligations to designers and creating obligations that are too general to have any binding content.

In light of the issues presented above, regulators should be wary of turning RbD into a pillar of their regulatory strategies. If designers cannot implement measures that contribute to the overall goals of the strategies but are nonetheless obliged to do *something* to fulfil a legal requirement, they might be pushed towards a Procrustean solution: pursing the regulatory goals only to the extent said goals can be expressed in terms of design requirements. For example, it has been argued that the EU AI Act's approach of protecting fundamental rights through product safety standards overlooks systemic violations of rights and dignity harms that cannot be described in terms of quantified risk.[35] Applying RbD approaches in contexts they are not suited to handle may lead to the construction of ML systems that produce undesirable side-effects or even undermine the goals that drive RbD in the first place.

Still, RbD provisions can be part of a well-calculated delegation strategy. Some technical goals and solutions are pretty much universal, and so their stipulation by design would face little opposition. Consider the human oversight requirements from Article 14 AI Act, which respond to widespread calls for tools that support human control over high-risk ML systems. RbD is also unlike to raise further issues if it is applied to a well-defined problem, as is the case of the logging requirement discussed in Section II. Finally, RbD provisions can also be useful if they remove some of the barriers to the effectiveness of other RbD provisions outlined above. For example, a requirement that ML systems must be designed with modularity and long-term maintenance in mind would reduce the costs involved in adapting software to cope with changes in the regulations it must comply to. The effectiveness of

[31] Indeed, some of these critiques are older than the current wave of ML technologies prompting calls for regulation: Bert-Jaap Koops & Ronald Leenes, *Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design' provision in data-protection law*, 28 Int. Rev. Law, Comput. Technol. 159 (2014).

[32] See, among others, Geoffrey Mead & Barbara Barbosa Neves, *Contested delegation: Understanding critical public responses to algorithmic decision-making in the UK and Australia*, Sociol. Rev. 00380261221105380 (2022).

[33] For a broad comparison between rules in code and roles in machine-learning systems, see Reuben Binns, *Analogies and Disanalogies Between Machine-Driven and Human-Driven Legal Judgement*, 1 CRCL (2022).

[34] Andrew Lohn & Micah Musser, *AI and Compute: How Much Longer Can Computing Power Drive Artificial Intelligence Progress?* (2022).

[35] See, among others, Marco Almada & Nicolas Petit, *The EU AI Act: Between Product Safety and Fundamental Rights*, (2022), https://papers.ssrn.com/abstract=4308072 (last visited Dec 21, 2022); Nathalie Smuha et al., *A Response to the European Commission's Proposal for an Artificial Intelligent Act*, 64 (2021).

any such measures must, evidently, be evaluated in light of the context in which a particular ML system is used and of the techniques available for their implementation. But, if seen as a supporting tool rather than a full-blown approach to regulation, RbD can help regulators in addressing the complexities of ML. ■

> *In light of the issues presented above, regulators should be wary of turning RbD into a pillar of their regulatory strategies*

# CPI
# SUBSCRIPTIONS

CPI reaches more than **35,000 readers** in over **150 countries** every day. Our online library houses over **23,000 papers**, articles and interviews.

Visit **competitionpolicyinternational.com** today to see our available plans and join CPI's global community of antitrust experts.

CPI COMPETITION POLICY® INTERNATIONAL