European Journal of Risk Regulation (2023), 1–13
doi:10.1017/err.2023.37

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Regulation by Design and the Governance of Technological Futures

Marco Almada

European University Institute, Florence, Italy
Email: Marco.Almada@eui.eu

Abstract

Regulation by design is an increasingly common approach in the governance of digital technologies. Under this approach, the developers of digital systems must adopt technical measures that implement the specific requirements mandated by law in their software. Some jurisdictions, notably the European Union (EU), have turned to regulation by design as a mechanism to automatically enforce legal requirements, but this article argues that such an approach can have important implications for long-term governance. Drawing from examples of regulation by design in EU law, it shows that by-design provisions delegate rule-making power to software designers, whose interpretations of the law become entrenched in digital systems. This delegation process suffers from legitimacy deficits, which are compounded whenever digital systems continue to enforce the designer-made rules as they operate for years and, sometimes, decades. Yet, these legitimacy deficits are not unavoidable, as regulation by design can be used to force designers to adopt technical and organisational practices that mitigate the risks of rule entrenchment to future generations. By mapping the long-term risks of regulation by design and potential solutions to them, this article provides a first step towards approaches to regulation by design that do not sacrifice the future for present interests.

Keywords: digital infrastructure; futureproof regulation; legitimacy; regulation by design; technological normativity

I. Introduction

Regulation by design (RbD)[1] is a widespread approach to regulating digital technologies.[2] Under this approach, regulators oblige the designers[3] of digital systems[4] to ensure that their system design incorporates specific legal requirements.[5] Designers, in turn, comply with such a requirement by turning the legal requirements into technical requirements,

[1] Also referred to as "by-design regulation" throughout the text.

[2] Throughout this article, RbD refers to the regulation by design of digital technologies. Design requirements may also be used to regulate other kinds of technologies, but this article makes no claims about the effectiveness or limits of by-design approaches beyond the digital domain.

[3] The term "designer", in this article, refers to all actors involved in the process of creating a digital system, from its conception to its implementation, such as programmers and project managers.

[4] I use "digital system" to refer to any system implemented using digital technologies, such as artificial intelligence systems, social networks or database management systems.

[5] P van Cleynenbreugel, "EU By-Design Regulation in the Algorithmic Society: A Promising Way Forward or Constitutional Nightmare in the Making?" in H-W Micklitz et al (eds), *Constitutional Challenges in the Algorithmic Society* (Cambridge, Cambridge University Press 2021) p 202.

which embed the contents of the RbD provision into the code of the digital system.[6] Designers thus ensure the automatic enforcement of the goals laid down by the regulator, as each task carried out by a compliant digital system necessarily follows the requirements embedded in its code. It is no surprise, therefore, that various jurisdictions – notably, but not exclusively,[7] the European Union (EU)[8] – have incorporated RbD provisions into their laws dealing with digital technologies.

This article examines whether this regulatory turn towards design has long-term implications for law and policy. At first glance, the answer might seem negative: since RbD produces its effects through the design of digital systems, these effects would cease once a system is modified or ceases to operate. However, some digital systems remain in operation for many decades, performing irreplaceable services as part of society's digital infrastructures.[9] And, even when specific digital systems cease to operate, the design decisions made in their creation can influence the development of future technologies.[10] Whenever one of these mechanisms operates, legal requirements embedded by design into a digital system might remain in force for longer than designers or regulators anticipated.

The persistence of digital systems over time can, I argue, hamper the ability of future generations to respond to the societal challenges they will face. To make this point, the remaining sections proceed as follows. Section II presents an overview of RbD as a regulatory modality that is heavily dependent on the technical work of designers and, as such, might struggle to impose legal requirements that cannot be fully represented in software code. Section III then shows how the decision to regulate through code impacts the legitimacy of RbD approaches, as by-design requirements entrench the designer's interpretation of the legal requirements laid down by the regulator. Section IV, in turn, examines the measures regulators adopt to forecast potential long-term risks of regulation and suggests how they can be extended to deal with the risks stemming from RbD. Finally, Section V summarises the argument and presents its implications for policymaking and the interpretation of RbD requirements in law.

II. Regulation by design as technological co-regulation

RbD is a form of regulation; as such, it is intended to influence human behaviour according to the aims of regulators.[11] In the case of by-design regulation, this influence happens in two stages. First, an RbD provision in the law is directed at the designer of a digital system, who becomes obliged to ensure that the digital system meets the requirements stipulated by law.[12] Once designers comply with this requirement, RbD produces effects in the

[6] E Bayamlıoğlu and R Leenes, "The 'Rule of Law' Implications of Data-Driven Decision-Making: A Techno-Regulatory Perspective" (2018) 10(2) Law, Innovation and Technology 295, 298.

[7] See, eg, Art 46 of the Brazilian Data Protection Law (Lei 13.709/2018) and Art 29 of the South Korean Personal Information Protection Act (Act No. 10465, 29 March 2011, as amended by Act No. 16930, 4 February 2020).

[8] The remainder of this article draws mostly from EU law examples.

[9] See, eg, the role of COBOL mainframes developed in the 1970s – or even earlier – in current infrastructures in businesses and the public sector: M Hicks, "Built to Last" (2020) (11) Logic Magazine.

[10] On path dependency in the development of digital technologies, see PN Edwards, "Platforms Are Infrastructures on Fire" in TS Mullaney et al (eds), *Your Computer Is on Fire* (Cambridge, MA, MIT Press 2021); S Hooker, "The Hardware Lottery" (2021) 64(12) Communications of the ACM 58.

[11] See, eg, P Drahos and M Krygier, "Regulation, Institutions and Networks" in P Drahos (ed.), *Regulatory Theory* (Canberra, ANU Press 2017) pp 7–8.

[12] On RbD as a source of obligations of result, see L Jasmontaite et al, "Data Protection by Design and by Default: Framing Guiding Principles into Legal Obligations in the GDPR" (2018) 4(2) European Data Protection Law Review 168, 173.

behaviour of those who interact with the system; for example, by preventing them from using the system for certain purposes.[13] Therefore, any by-design approach to regulation is a form of delegation in which regulators empower designers to enforce the stipulated legal requirements against those who use or are otherwise affected by the digital system.

1. Designers as rule-makers

The task delegated to designers is not, however, the mere enforcement of regulatory goals specified in law. Designers cannot comply with their legal obligations without interpreting them: by necessity, legal requirements are formulated in general terms, and designers need to identify how these general formulations apply to the specific contexts in which the system is expected to operate.[14] Once the content of the applicable legal requirements is determined, designers must decide the technical means that address the demands of the law. Some legal requirements might force designers to refrain from using certain techniques,[15] while others might demand hardcoding a specific rule into the system.[16] As a result of these two processes, RbD becomes a form of co-regulation, in which the regulator delegates not just the executing power but also the power to determine the actual content of the regulatory requirements and the mechanisms used to enforce them.[17]

Consider the accuracy requirement introduced by the EU's proposed regulation for artificial intelligence (AI) systems ("AI Act"[18]). Under this RbD provision, any high-risk AI system must achieve a level of accuracy appropriate to its function.[19] Such a requirement influences the behaviour of third parties: public and private actors will only be able to acquire and use AI systems that meet the accuracy standards, and the general population is reassured that any lawfully developed AI system will be accurate enough for its purpose.[20] Still, designers have ample leeway to choose between techniques that meet the accuracy standard: they might choose to implement the most accurate system they can create, or they might create a system that is just accurate enough to meet the design requirements but can be more easily understood by its users.[21] Designer decisions may produce different systems even if they start from the same requirements.

[13] L Diver, "Law as a User: Design, Affordance, and the Technological Mediation of Norms" (2018) 15(1) SCRIPTed 4.

[14] The need for interpretation comes, at least in part, from the fact that the programming languages used to create software are much less tolerant than natural languages when it comes to the ambiguity and open texture of legal provisions written in text: M Hildebrandt, "The Adaptive Nature of Text-Driven Law" (2022) 1(1) Journal of Cross-disciplinary Research in Computational Law 1.

[15] See, eg, how the legal framework for passenger name records in the EU precludes the use of machine learning techniques for the automated processing of these records: Case C-817/19 *Ligue des droits humains ASBL v Conseil des ministres* [2022] ECLI:EU:C:2022:491, para 194.

[16] See, eg, how EU law encourages – and in some cases requires – online platforms to use automated tools to filter certain types of content, such as audio-visual files shared in breach of copyright: G Sartor and A Loreggia, "The Impact of Algorithms for Online Content Filtering or Moderation. Upload Filters" (European Parliament, EPRS 2020).

[17] van Cleynenbreugel, supra, note 5, s 10.2.

[18] 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" COM (2021) 206 final.

[19] Art 15(1) AI Act.

[20] On the role of trust in the adoption of AI systems, see, from an EU perspective, C Derave et al, "The Risks of Trustworthy Artificial Intelligence: The Case of the European Travel Information and Authorisation System" (2022) 13(3) European Journal of Risk Regulation 389.

[21] Art 13(1) AI Act also establishes transparency as a design requirement for high-risk systems. On the potential trade-offs between transparency and accuracy, see A Bell et al, "It's Just Not That Simple: An Empirical Study of the Accuracy–Explainability Trade-off in Machine Learning for Public Policy" (ACM Conference on Fairness, Accountability, and Transparency, Seoul, 2022).

Since designers shape the digital system used to enforce RbD's legal requirements, a by-design approach is unlikely to achieve the intended results if designers fail to carry out the requirements imposed by the regulator. In many cases, such a failure might rest solely on the shoulders of designers; for example, if they misinterpret the legal requirements or if their programming work is inadequate or fails to solve software defects. In other cases, failure might result from broader socio-technical issues: in the field of AI, for example, most designers are reliant on tools developed by a handful of providers, which they lack the power to alter in response to legal requirements.[22] While both kinds of factors are relevant, the following discussion focuses on another potential source of RbD failures: the limits of expressing the law through software.

2. *Expressing legal requirements in software*

Over the past few decades, computer scientists have developed various techniques to represent legal requirements and instruments in software.[23] Some of these approaches have been used in practice, powering digital systems that address various problems, such as the automation of rules on taxes and social security benefits[24] or automated verification of compliance with standardised trade rules.[25] These approaches suggest that, at least under some circumstances, legal requirements can be faithfully translated into software code, which then enforces the encoded requirements. Whenever that is the case, RbD might be a feasible approach for regulators.

Translating legal requirements into software can be relatively straightforward in some cases. For example, legal rules can be presented in conditional structures – "if [such and such] then [so and so]"[26] – very similar to the conditional logic in programming languages.[27] If a rule's conditions and consequences can be encoded in the computer system, as is the case of the rules above, it is amenable to programming. However, such an encoding is not always possible if the conditions of the rule refer to aspects of humans and society that cannot be represented in the binary language or software, such as those concerned with the flourishing of human personality.[28] Encoding also becomes a problem if the automation of a rule would undermine the very objectives it is meant to pursue, as is the case with the principle that a defendant in a jury trial should be judged by their peers.[29] Consequently, the content of a legal rule might be an obstacle to its expression in software.

Further complications appear whenever RbD imposes other types of legal requirements. In many circumstances, RbD requirements do not oblige designers to implement legal

[22] On the phenomenon of AI as a service, see J Cobbe and J Singh, "Artificial Intelligence as a Service: Legal Responsibilities, Liabilities, and Policy Challenges" (2021) 42 Computer Law & Security Review 105573.

[23] For an introduction to these approaches, see T Bench-Capon, "Thirty Years of Artificial Intelligence and Law: Editor's Introduction" (2022) 30(4) Artificial Intelligence and Law 475.

[24] L Huttner and D Merigoux, "Catala: Moving towards the Future of Legal Expert Systems" (2022) Artificial Intelligence and Law 10.1007/s10506-022-09328-5.

[25] J Mao and AY Dawod, "Legal Informatics of HS Code Automatic Compliance Translation Based on Cross-Border Trade Digitization" (International Electrical Engineering Congress, Khon Kaen, 2022).

[26] S Brewer, "Interactive Virtue and Vice in Systems of Arguments: A Logocratic Analysis" (2020) 28(1) Artificial Intelligence and Law 151, 154–55.

[27] In fact, conditional logic is one of the building blocks of computational logic: MD Neumann et al, *Teaching Computational Thinking: An Integrative Approach for Middle and High School Learning* (Cambridge, MA, MIT Press 2021) p 9.

[28] See, eg, M Hildebrandt, "Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning" (2019) 20(1) Theoretical Inquiries in Law 83.

[29] See, eg, K Brennan-Marquez and S Henderson, "Artificial Intelligence and Role-Reversible Judgment" (2019) 109(2) Journal of Criminal Law and Criminology 137.

rules, but principles.[30] A legal principle creates duties on those bound by it – in this case, the designer – but it does not specify the content of these duties, which must be evaluated on a contextual basis.[31] Therefore, a designer bound to implement a legal principle must identify the possible issues that might emerge in each operational context for their system and propose technical solutions before they come to pass.[32] Such an anticipatory response might not be feasible in all cases.

Two examples of by-design requirements might help to identify the limits of anticipatory treatments of legal principles. In the accuracy-by-design example presented above, we have a potential conflict of values between accuracy and transparency, as some highly accurate systems are inscrutable to human observers. This conflict, however, can be resolved during the design process. So long as the system is accurate enough to meet the applicable accuracy requirements and transparent enough to comply with the transparency-by-design stipulation, the designer is free to choose how to weigh these values in their system. And, once they make a choice that complies with the law, it will remain an acceptable solution to the clash between values until the circumstances change.[33]

Contrastingly, designers are likely to struggle whenever value judgments do not remain stable over time. Consider a scenario in which a social media platform is required to automatically remove posts containing hate speech. An erroneous decision to remove a user's post can be a substantial violation of the user's rights, especially that of freedom of expression,[34] so accuracy is a relevant factor for this requirement. Many unlawful posts might be detected through automated filters,[35] but these filters might also produce wrongful results, especially when faced with parodies and other humoristic content.[36] The correctness of a removal decision depends not just on the context of the communication itself – was it a joke? A legal form of protest? – but also on what is culturally acceptable within a society. Designers are unlikely to capture all relevant factors beforehand, and, even if they do, the requirements they embed into software might become outdated if society becomes more (or less) receptive to certain kinds of discourse.

The examples above suggest that RbD might be inadequate if the legal requirements cannot be fully specified before implementation, or if relevant factors cannot be expressed in binary terms. In these circumstances, compliance with by-design requirements becomes a matter of risk management[37]: any design decision will be insufficient to cover all relevant aspects of some legal requirements; nonetheless, designers are still obliged to choose and adopt measures that mitigate known risks to the values at stake.[38] If these choices do not match the regulator's

[30] For example, EU data protection law imposes on designers a duty to adopt technical measures that implement data protection principles: LA Bygrave, "Article 25. Data Protection by Design and by Default" in C Kuner et al (eds), *The EU General Data Protection Regulation (GDPR): A Commentary* (Oxford, Oxford University Press 2020).

[31] See, among others, HB Ávila, *Theory of Legal Principles* (Dordrecht, Springer 2007).

[32] There are some techniques for the automation of principle-based reasoning; see, eg, J Maranhão et al, "A Dynamic Model for Balancing Values" (International Conference on Artificial Intelligence and Law, São Paulo, 2021). Still, all such techniques depend on the value assessments made by the designer while implementing them.

[33] For example, if the system's performance degrades to the point that it is no longer transparent or accurate enough, as discussed in Section III.1.

[34] Which is why the use of automated filters to moderate user speech requires the presence of extensive legal safeguards: Case C-401/19 *Republic of Poland v European Parliament and Council of the European Union* [2022] ECLI:EU:C:2022:297, para 67.

[35] Sartor and Loreggia, supra, note 16.

[36] J Taylor Rayz and V Raskin, "Fuzziness and Humor: Aspects of Interaction and Computation" in RB Kearfott et al (eds), *Fuzzy Techniques: Theory and Applications* (Cham, Springer 2019). On the legal protection of humour in Europe, see A Godioli et al, "Laughing Matters: Humor, Free Speech and Hate Speech at the European Court of Human Rights" (2022) 35(6) International Journal for the Semiotics of Law 2241.

[37] Jasmontaite et al, supra, note 12, 182.

[38] Or to refrain from using the system if no compromise is possible.

priorities, or if they produce unacceptable side effects,[39] the actual effects of the system on users and third parties will likely differ from those expected by the regulator.[40] Compliance with broadly defined RbD requirements might thus undermine, or at least fail to promote, the goals that prompted regulators to choose a by-design approach in the first place.

III. The legitimacy of regulation by design in the long run

The suitability of RbD approaches to regulation should not be evaluated solely from the technical perspective of effective design. Regulation is a social practice, in which a regulator seeks to influence the behaviour of the regulated actors. Sometimes this influence happens through the use of legally sanctioned coercion, or the threat of it,[41] but coercive approaches are not always effective, sometimes even backfiring against the regulator.[42] Any regulatory approach thus depends on its *legitimacy* from the perspective of regulated actors; that is, on their willingness to follow regulatory requirements even when said requirements go against their interests.[43]

RbD's reliance on software as an enforcement mechanism would seem to overcome these limits of coercion, as the rules embedded in the software are applied automatically whenever the conditions for their application are present. And, indeed, software systems do constrain human behaviour in novel and strong ways.[44] However, at least three factors ensure the relevance of legitimacy for RbD approaches. First, by-design regulation is the *product* of software design practices, so the effectiveness of RbD depends on whether designers comply with the requirements to which they are subject. Second, the people subject to rules embedded in software design can sometimes change the system's operation[45] or the role the system plays in society.[46] Finally, legitimacy might be relevant for moral and political reasons, such as the democratic ideal of ensuring people have a say in the rules that govern their lives.[47] In light of these factors, even the most technical approaches to regulation may be affected if they are not perceived as legitimate.

As discussed in Section II, RbD regulates behaviour in two stages. In the first one, RbD operates as a traditional form of legal regulation, using the force of law to compel

[39] For example, a requirement that a system must authenticate its users through a facial recognition algorithm can be useful – at least in theory – to prevent unauthorised access, but it might discriminate against people whose facial traits are not recognised or are mischaracterised by the algorithm. See, among others, O Keyes, "The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition" (2018) 2(CSCW) Proceedings of the ACM on Human–Computer Interaction 1; R Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Medford, Polity 2019).

[40] More generally on regulatory effectiveness, see M de Benedetto, "Effective Law from a Regulatory and Administrative Law Perspective" (2018) 9(3) European Journal of Risk Regulation 391.

[41] On the role of coercion in legal regulation, see, among others, F Schauer, *The Force of Law* (Cambridge, MA, Harvard University Press 2015).

[42] J Braithwaite, "The Essence of Responsive Regulation" (2011) 44(3) UBC Law Review 475, 484.

[43] On the individual dimensions of legitimacy not covered here, see TR Tyler, "Psychological Perspectives on Legitimacy and Legitimation" (2006) 57(1) Annual Review of Psychology 375, 377.

[44] M Hildebrandt, "The Force of Law and the Force of Technology" in MR McGuire and T Holt (eds), *The Routledge Handbook of Technology, Crime and Justice* (London, Routledge 2017).

[45] See, eg, the use of adversarial attacks to distort the outputs of AI systems: B Biggio and F Roli, "Wild Patterns: Ten Years after the Rise of Adversarial Machine Learning" (2018) 84 Pattern Recognition 317.

[46] NF Doherty et al, "A Re-conceptualization of the Interpretive Flexibility of Information Technologies: Redressing the Balance between the Social and the Technical" (2006) 15(6) European Journal of Information Systems 569.

[47] See, eg, Hildebrandt, supra, note 28.

designers to implement legal requirements into software.[48] The distinctiveness of RbD appears only in the second stage, in which technology is used to enforce regulation against the users and third parties that interact, directly or indirectly, with a digital system. Since the regulatory effects of technology are shaped by the technical decisions made by designers as they create digital systems, an analysis of RbD's legitimacy must consider whether delegating regulatory power to designers can be a legitimate form of regulation.

1. The legitimacy of delegating regulation to designers

Various scholars have worked on the general challenges of what makes regulation legitimate[49] and the specific implications of technology for political legitimacy.[50] Drawing from this scholarship, this section focuses on two social mechanisms that can legitimise a regulatory approach. Regulations can derive *output legitimacy* from the outcomes they produce: if an individual or group sees the effects of regulation as desirable, they are more likely to acquiesce to regulatory demands, even if those demands clash with some of their own interests.[51] They can also gain *input legitimacy* by involving relevant actors in the regulatory processes, thus reassuring these actors that regulation accounts for their values and interests.[52] These legitimacy-building mechanisms are not self-excluding, as sources of input legitimacy may reinforce,[53] compensate for[54] or undermine[55] one another. Legitimacy must, therefore, be evaluated in terms of how potential sources of legitimation manifest themselves and interact with one another in practice.[56]

Since RbD approaches rely on digital systems as enforcement mechanisms, their legitimacy is affected by factors related to how those systems are built and operated.[57] From the perspective of output legitimacy, the discussion in Section II of this article suggests that encoding legal rules in software can be a double-edged sword. On the one hand, good RbD requirements can

[48] On the legitimacy of legal regulation in techno-scientific contexts, see, among others, Bayamlıoğlu and Leenes, supra, note 6, 304; A Volpato and A Offermans, "Lessons for Participation from an Interdisciplinary Law and Sustainability Science Approach: The Reform of the Sustainable Use of Pesticides Directive" (2023) European Journal of Risk Regulation 10.1017/err.2023.9.

[49] In the EU legal order, see, among others, G Majone, "The Regulatory State and Its Legitimacy Problems" (1999) 22(1) West European Politics 1; VA Schmidt, *Europe's Crisis of Legitimacy* (Oxford, Oxford University Press 2020); M Eliantonio and C Cauffman, *The Legitimacy of Standardisation as a Regulatory Technique: A Cross-Disciplinary and Multi-Level Analysis* (Cheltenham, Edward Elgar Publishing 2020).

[50] See, among others, S Borrás and J Edler, "The Governance of Change in Socio-Technical and Innovation Systems: Three Pillars for a Conceptual Framework" in S Borrás and J Edler (eds), *The Governance of Socio-Technical Systems* (Cheltenham, Edward Elgar Publishing 2014) pp 36–38; L Diver, *Digisprudence: Code as Law Rebooted* (Edinburgh, Edinburgh University Press 2021) ch 5; B Green, "The Flaws of Policies Requiring Human Oversight of Government Algorithms" (2022) 45 Computer Law & Security Review 105681, s 4.2.

[51] This point is often framed in terms of "government *for* the people": P Verbruggen, "Does Co-regulation Strengthen EU Legitimacy?" (2009) 15(4) European Law Journal 425, 431. With the formulation above, I adopt a more granular view of output legitimacy, acknowledging the possibilities of political divergences within a society with regard to interests and desirable outcomes.

[52] In a democratic society, this involvement happens, to a large extent, within the representative bodies that craft the legislation that empowers and constrains regulation: FW Scharpf, *Governing in Europe: Effective and Democratic?* (Oxford, Oxford University Press 1999) s 1.1.

[53] Eg, defenders of epistemic democracy argue that democratic participation can also lead to better political outcomes: RE Goodin and K Spiekermann, *An Epistemic Theory of Democracy* (Oxford, Oxford University Press 2018).

[54] Eg, citizens might decide to accept a regulation with little input legitimacy because it produces good results.

[55] See, eg, CT Nguyen, "Transparency Is Surveillance" (2022) 105(2) Philosophy and Phenomenological Research 331.

[56] On the grammar of legitimacy, see F Kratochwil, "On Legitimacy" (2006) 20(3) International Relations 302.

[57] In addition to more general legitimacy concerns, such as whether the regulator is entitled to impose the legal requirements they intend to pursue through design. On the general criteria for regulatory legitimacy, see R Baldwin et al, *Understanding Regulation: Theory, Strategy, and Practice* (Oxford, Oxford University Press 2012) ch 3.

lead to systems that apply regulation uniformly and automatically. On the other hand, an RbD approach that imposes requirements that cannot be fully expressed in software will not produce the outcomes expected by regulators. In the latter case, a by-design approach might lose some, or even all, of its claim to being an effective means to promote regulatory aims.

By-design approaches also affect a regulation's input legitimacy. Some scholars have argued that the embedding of legal rules into software could bolster regulatory legitimacy by eliminating the possibility of arbitrary decision-making by human enforcers,[58] while others have suggested that enforcement by design weakens the legitimacy of alternative approaches, as it allows legislators to avoid the delegation of rule-making powers to executive bodies.[59] Yet, as discussed in Section II.1, RbD necessarily entails delegating rule-making power to designers, who are not subject to the same procedural requirements that foster popular representation in democratic legislatures.[60] Consequently, RbD replaces the possibility of arbitrary decision-making by administrators with the possibility of arbitrary decision-making by designers, a switch unlikely to positively impact input legitimacy.[61]

2. *Long-term challenges to the legitimacy of regulation by design*

The legitimacy of by-design approaches may also be affected by the passage of time.[62] Given that RbD produces its effects by embedding legal requirements into software systems, these requirements are enforced for as long as the system remains in operation, unless its code is actively changed.[63] However, changing a digital system after its design can be difficult, as those systems are complex technical objects,[64] often built upon components that designers lack the power and the technical resources to change.[65] This difficulty in changing digital systems means, in turn, that such systems may lag behind the law whenever the latter changes, continuing to enforce outdated requirements for some time.[66]

If it is not feasible to adjust an existing digital system, its designers or regulators might decide instead to replace it with a new one.[67] But replacement is not always a simple task:

[58] See, eg, JB Bullock, "Artificial Intelligence, Discretion, and Bureaucracy" (2019) 49(7) The American Review of Public Administration 751; CR Sunstein, "Governing by Algorithm? No Noise and (Potentially) Less Bias" (2022) 71(6) Duke Law Journal 1175.

[59] R Calo and DK Citron, "The Automated Administrative State: A Crisis of Legitimacy" (2021) 70(4) Emory Law Journal 797.

[60] Some design approaches involve a broad range of stakeholders. However, their use is far from the rule in software development. See, eg, F Delgado et al, "An Uncommon Task: Participatory Design in Legal AI" (2022) 6(CSCW1) Proceedings of the ACM on Human–Computer Interaction 51.

[61] Such an effect might happen if designers – who, in most case, are private actors – are more trusted than the executive bodies who would be responsible regulatory alternatives. Yet, in many contexts, delegation to private actors is seen as particularly problematic, and rule-making private actors are increasingly subject to constraints that were previously extended only to the state: G De Gregorio, "The Rise of Digital Constitutionalism in the European Union" (2021) 19(1) International Journal of Constitutional Law 41.

[62] On the temporal persistence of technology, see more generally D Edgerton, "De l'innovation aux usages. Dix thèses éclectiques sur l'histoire des techniques" (1998) 53(4) Annales 815; LB Moses and M Zalnieriute, "Law and Technology in the Dimension of Time" in S Ranchordás and Y Roznai (eds), *Time, Law, and Change: An Interdisciplinary Study* (London, Hart Publishing 2020).

[63] In this sense, code is different from written law, as the content of the law can change with time if the text is interpreted differently from its original, authentic interpretations: Hildebrandt, supra, note 14.

[64] Modern digital systems are compositional objects, in the sense that they are built through the combination and extension of previous systems, which are used as platforms for further development: Edwards, supra, note 10; and, more generally, BK Sovacool et al, "Reconfiguration, Contestation, and Decline: Conceptualizing Mature Large Technical Systems" (2018) 43(6) Science, Technology, & Human Values 1066.

[65] Cobbe and Singh, supra, note 22.

[66] Moses and Zalnieriute, supra, note 62, 309.

[67] If the system is not directly under the control of the designer (eg in the case of a software sold to end-users), the adoption of the new system might require further action, such as product recalls: Art 65(2) AI Act.

some systems might be too big to rebuild or carry out sensitive tasks – or both, as is often the case with the systems used in domains such as banking, healthcare or the public administration.[68] Even if replacement is possible, the legal requirements embedded in the system by design might be replicated in other systems: designers sometimes need to ensure the replacement system behaves as its predecessor,[69] and, even more often, design choices in the new system are themselves influenced by what was done in previous systems.[70] As a consequence, the decision to embed legal requirements in a digital system might entrench these requirements by creating obstacles against future change.

Entrenchment is not a novel issue for legal scholarship. Certainty against arbitrary change is widely acknowledged as a desirable property of a legal system,[71] and one of the core ideas of modern constitutionalism is that modifications to constitutions should demand considerable technical and political efforts.[72] However, the entrenchment promoted by the law is the entrenchment of legal text, which, by necessity, is formulated in general and abstract terms.[73] Contrastingly, entrenchment in a digital system entails that any requirements embedded in the system's code will continue to operate as programmed—following the designer's interpretation of the original legal requirements. Whenever technical conditions create obstacles to technical change, an RbD approach can entrench not just the requirements imposed by the regulator, but the specific meaning given to them by the designer when the relevant technical decisions were made.

In the long term, the possibility of interpretive entrenchment through software can erode the output legitimacy of an RbD approach. If entrenchment prevents changes to a system, it will continue to enforce the requirements embedded into it during design. Yet, these requirements might become unacceptable as time goes by: the letter of the law or its interpretation can change,[74] societal change can make the original requirements irrelevant[75] or the system's operation might reveal biases that had escaped detection during design.[76] These and other developments might require adjustments to the system, which entrenchment prevents, or at least makes more expensive. As a result, the operation of the digital system will undermine these new regulatory aims to the extent that they clash with the interpretation of the original legal requirements embedded in the system.

The long-term effects of design decisions also create problems for the input legitimacy of RbD approaches. Since the occurrence of interpretive entrenchment leads to the enforcement of the designer's interpretation of the legal requirements at the moment

[68] These so-called *legacy systems* are subject to continuous maintenance processes to ensure their operation and to add new functionalities. Nonetheless, their operation is largely bound by their original design: M Bellotti, *Kill It with Fire: Manage Aging Computer Systems* (San Francisco, CA, No Starch Press 2021).

[69] See, eg, A Sundelin et al, "The Hidden Cost of Backward Compatibility: When Deprecation Turns into Technical Debt – An Experience Report" (ACM International Conference on Technical Debt, Seoul, 2020) 67; M Srivastava et al, "An Empirical Analysis of Backward Compatibility in Machine Learning Systems" (ACM SIGKDD Conference on Knowledge Discovery and Data Mining, San Diego, CA, 2020).

[70] On the path dependence of technological development, see Hooker, supra, note 10; and, beyond the domain of computing, EW Constant, "Why Evolution Is a Theory about Stability: Constraint, Causation, and Ecology in Technological Change" (2002) 31(8) Research Policy 1241.

[71] See, eg, J Braithwaite, "Rules and Principles: A Theory of Legal Certainty" (2002) 27 Australasian Journal of Legal Philosophy 47; P Popelier, "Five Paradoxes on Legal Certainty and the Lawmaker" (2008) 2(1) Legisprudence 47.

[72] A Sajó and R Uitz, *The Constitution of Freedom: An Introduction to Legal Constitutionalism* (Oxford, Oxford University Press 2017) p 45.

[73] F Pirie, "Beyond Pluralism: A Descriptive Approach to Non-State Law" (2023) 14(1) Jurisprudence 1.

[74] Moses and Zalnieriute, supra, note 62, 309.

[75] See, eg, ML Jones, "Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw" (2018) 2018(2) Journal of Law, Technology & Policy 101.

[76] Or that emerge during operation, such as those caused by feedback loops: European Union Agency for Fundamental Rights, "Bias in Algorithms – Artificial Intelligence and Discrimination" (2022).

of design, it precludes systems from accommodating changes in perspectives over time.[77] Entrenchment also creates obstacles to the representation of perspectives that were not accounted for in the original design, especially those regarding the rights and interests of future generations.[78] For example, if future generations come to an agreement regarding the need for a change in regulation, they might struggle to make changes to older software systems, as the people who are familiar with such systems' architecture and technologies might be already retired – or long dead.[79] RbD thus creates the risk that future generations end up governed by systems they did not shape and have little power to change.

IV. Addressing the long-term risks of regulation by design

Regulators are not unaware of the potential long-term effects of RbD approaches. In fact, some regulatory approaches use design requirements as a tool not only to govern present systems but also to foster specific paths of technological development.[80] But, as the previous sections suggest, any such future-shaping effects from an RbD design also entrench how designers interpret legal requirements at the moment of design, an interpretation that might lack input and output legitimacy. Additionally, regulators might lack the legitimacy to impose their legal requirements on future generations,[81] especially if these requirements prevent changes in the applicable law[82] or to the constitutional structures of society.[83]

This is not to say that entrenchment risks make design an inherently illegitimate tool for regulation. Section III.2 of this article suggests that the likelihood of entrenchment grows with the complexity of the digital system, and so a small system might not produce much in terms of entrenchment. Likewise, the severity of any adverse effects will grow if the entrenched requirements deal with sensitive political topics such as fundamental rights and the rule of law,[84] but entrenchment might not be too much of a problem in other domains. The stability provided by entrenchment might even be a source of legitimacy, as is often the case with constitutional norms.[85] At the end of the day, the decision of whether the positive effects of RbD are worth the risk of entrenchment – or whether such

[77] On the subject of changes in views, see T Grüne-Yanoff and SO Hansson, *Preference Change: Approaches from Philosophy, Economics and Psychology* (Cham, Springer 2009).

[78] BE Tonn, "Philosophical, Institutional, and Decision Making Frameworks for Meeting Obligations to Future Generations" (2018) 95 Futures 44; S Clarke and J Whittlestone, "A Survey of the Potential Long-Term Impacts of AI: How AI Could Lead to Long-Term Changes in Science, Cooperation, Power, Epistemics and Values" (AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society, Oxford, 2022).

[79] This phenomenon can already be seen in the present, as the public administration and the private sector struggle to address competence gaps created once the first generations of software engineers left the trade: Hicks, supra, note 9; ML Cohn, "Keeping Software Present: Software as a Timely Object for STS Studies of the Digital" in J Vertesi and D Ribes (eds), *DigitalSTS: A Field Guide for Science & Technology Studies* (Princeton, NJ, Princeton University Press 2019).

[80] See, eg, how Recital 5 AI Act states that the Act – and hence its design requirements – is meant to support the EU's ambitions of being a global leader in the development of "secure, trustworthy and ethical artificial intelligence".

[81] See, eg, MK Kołacz et al, "Who Should Regulate Disruptive Technology?" (2019) 10(1) European Journal of Risk Regulation 4.

[82] See, eg, F Peirone, "The Rule of the Present, Not the Past" (2021) 3(3) Jus Cogens 229; BA Greenberg, "Rethinking Technology Neutrality" (2016) 100(4) Minnesota Law Review 1495.

[83] See, eg, M Loughlin, *Against Constitutionalism* (Cambridge, MA, Harvard University Press 2022).

[84] See, among others, B-J Koops and R Leenes, "Privacy Regulation Cannot Be Hardcoded. A Critical Comment on the 'Privacy by Design' Provision in Data-Protection Law" (2014) 28(2) International Review of Law, Computers & Technology 159; E Aizenberg and J van den Hoven, "Designing for Human Rights in AI" (2020) 7(2) Big Data & Society 2053951720949566; M Zalnieriute et al, "The Rule of Law 'by Design'?" (2021) 95(3) Tulane Law Review 1063.

[85] Sajó and Uitz, supra, note 72, 45.

entrenchment is desirable – is a political decision made in the present by the regulator, but one that will have consequences in the future.

As the introduction to this special issue shows, regulators are increasingly called to consider the impacts of their actions on future generations.[86] In technology regulation, such reflections on the future are usually mediated by the ideal of "futureproof regulation".[87] According to this ideal, regulation should be constructed so that the regulatory framework continues to operate in the same way even if technologies change.[88] For example, the draft AI Act pursues futureproof regulation of AI systems by two mechanisms: on the one hand, its technical requirements are formulated with reference to the expected outcomes and not to specific technologies[89]; while on the other hand, it adopts mechanisms such as regulatory sandboxes[90] and periodic reviews of the effectiveness of the regulatory outcome[91] to evaluate whether existing provisions are still suitable as new technologies come into play. Mechanisms such as those allow regulators to use the same regulatory approach in a broad range of functionally equivalent technologies[92] and reduce the adjustments needed to incorporate new technologies into the existent approach. Futureproof regulation thus contributes to ensuring its relevance over time.[93]

Futureproofing an RbD approach creates two obstacles to dialogue with future generations. The first obstacle is shared with other forms of regulation: since futureproof regulation is expected to stay roughly in its current shape as time passes, it codifies the interests of present regulators. Current approaches to futureproofing mitigate this present-centric outlook through anticipatory methodologies, which identify future opportunities and risks associated with the regulated technologies.[94] Yet, identifying future risks is not enough to ensure the interests and values of future generations are properly accounted for. As an example, debates on nuclear waste management acknowledge the potential risks of waste to future generations but often frame the response to these risks as a zero-sum game between present and future interests.[95] To mitigate this reading of future interests and values in presentist terms, society will need to rely on mechanisms such as law-making oversight practices,[96] expanded mechanisms for participation in governance[97] or judicial pathways for addressing issues of intergenerational justice.[98]

[86] On the issue of duties to future generations, see, additionally, K Shrader-Frechette, "Duties to Future Generations, Proxy Consent, Intra- and Intergenerational Equity: The Case of Nuclear Waste" (2000) 20(6) Risk Analysis 771.

[87] For an overview of futureproofing as an approach to digital regulation, see S Ranchordás and M van 't Schip, "Future-Proofing Legislation for the Digital Age" in S Ranchordás and Y Roznai (eds), *Time, Law, and Change: An Interdisciplinary Study* (London, Hart Publishing 2020).

[88] Greenberg, supra, note 82, 1523.

[89] Arts 7–15 AI Act.

[90] Arts 53–55 AI Act.

[91] Art 84 AI Act.

[92] S Shadikhodjaev, "Technological Neutrality and Regulation of Digital Trade: How Far Can We Go?" (2021) 32(4) European Journal of International Law 1221, 1242.

[93] On regulatory resilience, see JB Ruhl et al, "Resilience of Legal Systems: Toward Adaptive Governance" in M Ungar (ed.), *Multisystemic Resilience: Adaptation and Transformation in Contexts of Change* (Oxford, Oxford University Press 2021).

[94] On the use of anticipation and forecasts in the management of the risks associated with digital technologies, see PAE Brey, "Anticipatory Ethics for Emerging Technologies" (2012) 6(1) NanoEthics 1; K Heo and Y Seo, "Anticipatory Governance for Newcomers: Lessons Learned from the UK, the Netherlands, Finland, and Korea" (2021) 9(1) European Journal of Futures Research 9; M-V Florin, "Risk Governance and 'Responsible Research and Innovation' Can Be Mutually Supportive" (2022) 25(8) Journal of Risk Research 976.

[95] Shrader-Frechette, supra, note 86, 773.

[96] See, eg, the use of parliamentary committees examined by V Koskimaa and T Raunio in this special issue.

[97] See, eg, the article by M Revel in this special issue.

[98] See, eg, the article by T Steinkamp in this special issue, as well as D Bertram, "'For You Will (Still) Be Here Tomorrow': The Many Lives of Intergenerational Equity" (2023) 12(1) Transnational Environmental Law 121.

However, futureproofing the law is not enough to futureproof an RbD approach. Even if the legal requirements imposed by RbD address future values and interests, designers might still comply with those requirements through approaches that favour the present over the future (eg by adopting technical configurations that are more likely to entrench said requirements). An approach to this second stage would need to combine two strands of work: participatory approaches to software design, which allow designers to engage with the perspectives of individuals and groups affected by the system[99]; and approaches to engaging with the future described above. Design methodologies for that purpose have begun to emerge,[100] but any such practices will need to deal both with the challenges of identifying future preferences and the limits of participatory design practices.[101] Between these technical limits and the legitimacy issues designers face as long-term rule-makers,[102] it seems unlikely that RbD approaches will manage to fully incorporate future perspectives into designer decisions.

If the concerns of future generations cannot be addressed at the moment of design, regulators can still adopt measures that weaken the effects of interpretive entrenchment on future generations. Some of these measures might have an organisational character. For example, regulators might oblige designers to adopt quality management processes to identify and address risks stemming from the operation of the digital system,[103] thus ensuring that the system is constantly updated to cope with legal requirements. Another potential measure would be forcing the recall of software systems that impose unacceptable risks,[104] a practice that removes from circulation systems that cannot be repaired by their maintainers. Furthermore, regulators can also reduce the barriers to change in old, still-functional systems by establishing knowledge pools that prevent the loss of knowledge about old technologies as their creators retire or die. None of these measures is a design requirement,[105] but they remove or mitigate some potential causes of interpretive entrenchment identified in Section III.2. In doing so, they provide safeguards against the long-term risks of RbD approaches.[106]

Finally, RbD itself might provide some tools for mitigating the risk of interpretive entrenchment through software. Over the last few decades, software engineers have developed techniques that simplify the management of the complexity of software systems.

[99] See, among others, MJ Muller et al, "Participatory Practices in the Software Lifecycle" in MG Helander et al (eds), *Handbook of Human-Computer Interaction* (Amsterdam, North-Holland 1997); C Andersson et al, "Unpacking the Digitalisation of Public Services: Configuring Work during Automation in Local Government" (2022) 39(1) Government Information Quarterly 101662; A Birhane et al, "Power to the People? Opportunities and Challenges for Participatory AI" (ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, Arlington, VA, 2022).

[100] Eg, S Bødker and M Kyng, "Participatory Design That Matters – Facing the Big Issues" (2018) 25(1) ACM Transactions on Computer–Human Interaction 1; RC Smith and OS Iversen, "Participatory Design for Sustainable Social Change" (2018) 59 Design Studies 9.

[101] With an AI focus, see S Robertson and N Salehi, "What If I Don't Like Any of the Choices? The Limits of Preference Elicitation for Participatory Algorithm Design" (ICML Participatory Approaches to Machine Learning Workshop, Online, 2020); M Sloane et al, "Participation Is Not a Design Fix for Machine Learning" (ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, Arlington, VA, 2022).

[102] See Section III, supra.

[103] See, eg, Art 9 AI Act: J Schuett, "Risk Management in the Artificial Intelligence Act" (2023) European Journal of Risk Regulation 10.1017/err.2023.1. On the technical tasks of quality management, see I Sommerville, *Software Engineering* (Harlow, Pearson 2016) ch 24.

[104] See, eg, Art 65(2) AI Act.

[105] At least in the narrow, technical sense of design used in this article.

[106] Data protection approaches to RbD tend to adopt a broader definition of RbD, which encompasses organisational measures in addition to technical ones: Bygrave, supra, note 30. However, current practices in the EU emphasise technical solutions, as shown by the AI Act and even in the interpretation of data protection law: European Data Protection Board, "Recommendations 01/2020 on Measures that Supplement Transfer Tools to Ensure Compliance with the EU Level of Protection of Personal Data" (2020).

For example, modular software architectures allow designers to make changes to parts of the system without needing to alter everything else,[107] while automated registers of a system's operation allow designers to trace the sources of defects or otherwise undesirable outcomes,[108] and reliance on established design patterns might bolster the legitimacy of designers and render the system's technical architecture more intelligible.[109] If RbD requirements oblige designers to follow such design practices, the ensuing systems become more amenable to future changes. This malleability, in turn, allows future generations to make changes to existing digital systems if they deem such changes necessary. In doing so, an RbD approach might help future generations to exercise control over the rules embedded in future systems rather than subjecting them to the will of the past.

V. Concluding remarks

In the previous sections, I have argued that RbD approaches can have long-term implications, as they turn software designers into rule-makers whose decisions are enforced by digital systems that often operate for decades. Given the ubiquity of software in modern societies, the effects of RbD requirements are not confined to digital environments such as the Internet, as the systems built in compliance with such requirements are used to make decisions, create recommendations and perform other tasks that affect the lives of people. In a digitalised society, software design is a horizontal concern for governance.

The entrenchment risks associated with RbD should not lead us to discard the approach itself. Instead, they suggest the need for mechanisms to measure these risks – such as indicators of entrenchment – and tools for mitigating it (eg through the reinterpretation of existing RbD provisions and new legislative instruments). By accounting for the potential effects of RbD, regulators might be able to leverage the potential gains from software without sacrificing the interests of future generations. Design cannot save us from the tyranny of the past by itself – but it can be a tool for empowering future generations to make their own political choices.

Acknowledgments. The author would like to thank Zak Kallenborn, Alberto Alemanno, Martin Ulbricht, Natalia Menéndez González, Nicola Hargreaves, Sara Guidi and Bas Heerma van Voss for feedback on previous drafts of this paper.

Competing interests. The author declares none.

[107] B Liskov, "Modular Program Construction Using Abstractions" in D Bjørner (ed.), *Abstract Software Specifications* (Berlin, Springer 1980).

[108] JJ Bryson and A Theodorou, "How Society Can Maintain Human-Centric Artificial Intelligence" in M Toivonen and E Saari (eds), *Human-Centered Digitalization and Services* (Singapore, Springer Singapore 2019) p 317.

[109] See, eg, E Dickhaut et al, "Lawfulness by Design – Development and Evaluation of Lawful Design Patterns to Consider Legal Requirements" (2023) European Journal of Information Systems 10.1080/0960085X.2023.2174050.

Cite this article: M Almada, "Regulation by Design and the Governance of Technological Futures". *European Journal of Risk Regulation.* https://doi.org/10.1017/err.2023.37