

Dilemmas in Legal Governance

Pompeu Casanovas*, Pablo Noriega[◦]

** UAB Institute of Law and Technology (IDT-UAB), Universitat Autònoma de Barcelona, Spain; LawTech Research Group, Law School, La Trobe University, Australia*

◦ Artificial Intelligence Research Institute, Spanish National Research Council (IIIA-CSIC)

Abstract. This article develops the issue of power and language in Artificial Intelligence governance through three dilemmas that should be considered in the design of regulatory systems. Some previous uses of this figure in the literature are described. The article inspects some design choices that determine how legal governance in AI can be engineered. Stemming from the conundrum that enactment of rights and the implementation of good or better government can be facilitated at the cost of increasing control over citizens through AI-based systems, the article formulates three dilemmas—normative autonomy, legal isomorphism, and modes of regulation. Beyond deliberative and epistemic democracy, it expands then the notion of legal governance, connecting the three dilemmas with the ethical and political stances of the concept of linked democracy.

Keywords: Artificial Intelligence Governance, Legal Governance, Power, Dilemmas, Hybrid Intelligence, Collective intelligence, Legal isomorphism, Rules as Code

1. Introduction

Artificial intelligence has burst into our lives in a way that was unforeseen just a decade ago. Not only because of the pandemic but also because of the extraordinarily rapid transformation that industry, commerce, administration, government, and the law are also experiencing due to the massive adoption of the internet and AI enabled interactions. The growing swarm of Big Data, the notorious successes of machine learning technologies—in language processing and image recognition— together with the pervasive use of a variety of AI technologies to support all sorts of web-based social interactions are a constant reminder of the disruptive impact of AI. There are huge opportunities to make good use of Artificial Intelligence; however, since AI is a disruptive technology affecting existing systems in society, there are equally formidable challenges that should also be addressed.

As in other disruptive activities, there are things in AI that we can do, some things that we should do, and more significantly there are also things that we should not do, even if we could. The problem is that, for now, it is not clear where the red line should be drawn. To address this ambiguity, we advocate a regulatory perspective. In fact,

we propose to develop what we can call *legal governance* in Artificial Intelligence.

By legal governance we are not understanding the set of tools used in the corporate field to monitor business processes in the so-called LGCR—Legal Governance, Risk Management, and Compliance—but the mindset and regulatory toolkits which are necessary to face legal implementation, application, and enforcement in complex “*hybrid*” social environments that may contain both analogic and digital elements at the same time (Pagallo et al. 2019a, 2019b; Casanovas et al. 2022; Noriega et al. 2021). That is, we will refer both to the use of artificial intelligence for regulation of hybrid social environments, and to the regulation of artificial systems that operate within those hybrid social environments.

Obviously, this is an issue that holds a political dimension, but we think this is also a matter of ‘engineering’ a particular type of governance in artificial systems and for that purpose. We propose to inspect some design choices that determine how legal governance in AI can be engineered. Note that these questions do not apply to the actual development of a system: they are prior to it, they guide, determine both its content and its application, and therefore are embedded into the general governance process. In this sense, they are part of the proposed regulation system and are by no means expendable.

We will articulate these challenges in the form of three dilemmas related to hybrid intelligence which can help us understand the complexity of the problem. In applied ethics the formulation of dilemmas tends to be avoided in favour of moral problematisation (Camps 2013, 399 and ff.). Some moral philosophers prefer Aristotelian *phronesis* and the ethics of virtue over choices. But this is a difficult issue for AI governance. For the sake of simplicity, we will refer to *dilemmas* as situations in which a difficult choice must be made between two *or more* alternatives. The word ‘dilemma’ will be used in a wide way to point out the depth of the challenges. It will be linked to hybrid intelligence and self-awareness (conscientious design).

Hybrid intelligence has been recently defined as “the combination of human and machine intelligence, augmenting human intellect and capabilities instead of replacing them and achieving goals that were unreachable by either humans or machines” (Akata et al. 2020, 18).

Conscientious design has been proposed by some researchers to face the ‘Collingridge dilemma’ (Noriega et al. 2016), underpinning the ethical and social values at stake in building normative multi-agent systems. Before the Internet, David Collingridge (1980) posited the quandary that when a technology is new it is extremely hard to predict its negative consequences, but by the time one can figure those out,

it's too costly to do much about it. Time is a crucial variable in this dilemma. On the Internet of Things, information processing occurs in real time, and solutions should operate in the same way.

Conscientious design starts by developing an awareness of the concerns manifest in the current landscape and understanding how multi-agent techniques can be applied as an effective means to operationalise systems to ameliorate such concerns, and bring them to bear upon our everyday scientific and technological activity. (Noriega et al. 2016)

Thus, there is a need to rethink theories and models of norms, roles, relationships, languages, architectures, governance, and institutions for such systems, and do so in interdisciplinary research. How could we deal with such challenges in regulatory environments, and more specifically, in legal domains?

This paper is a first contribution to start answering this question for a 'design that is responsible, thorough and mindful', focusing on *Hybrid Online Social Systems* (HOSS) from a legal governance approach. The remaining of this paper is divided into four more sections. The first one (2) raises the issue of power and language as a conundrum for AI legal governance. The third section (3) includes some recent uses of dilemmas and formulates briefly the three dilemmas of legal governance to be discussed (normative autonomy, legal isomorphism, and the modes of regulation). Section 4 contains the discussion within the linked democracy framework. A final Section (5) wraps up the issue and introduces some possible follow-ups.

2. Conundrum

Let's start with the classic question of power and language. It is worth remembering Lewis Carroll's *Through the Looking Glass* ([1871] 2010):

'When I use a word', Humpty Dumpty said in rather a scornful tone, 'it means just what I choose it to mean – neither more nor less.'

'The question is', said Alice, 'whether you can make words mean so many different things.'

'The question is', said Humpty Dumpty, 'which is to be master. That's all.'

This is one of *Alice's* most quoted paragraphs, certainly, but what follows, the issue of the *impenetrability of language*, is just as important and much less quoted:

[...] Impenetrability! That's what *I* say!

‘Would you tell me, please,’ said Alice, ‘what that means?’

‘Now you talk like a reasonable child’ said Humpty Dumpty, looking very much pleased. ‘I meant by “impenetrability” that we’ve had enough of that subject, and it would be just as well if you’d mention what you mean to do next, as I suppose you don’t mean to stop here all the rest of your life.’

Indeed, as the sad ending of Humpty Dumpty falling from the wall warns, the challenges and risks posed by the social application of artificial intelligence have to do with action, agency, the impenetrability of language, and the opacity of the methods used. Unfortunately, examples of human errors and hasty applications are multiplying. The case of ROBODEBT—the automatic and improper recovery of debts by the Australian administration in 2016¹—is paradigmatic. Privacy, security, and data encryption issues are also well known. The massive surveillance systems that PRISM revealed (since 2007 under the *Protect American Act* of the Bush Administration), the Cambridge Analytica methods on citizens’ private data based on machine learning techniques, the recent PEGASUS methods for harvesting private information, should warn us about the ease with which Artificial Intelligence can be applied without transparency for dubious reasons. All of this has been publicised and is well-known by now. *Impenetrability* of language also means impenetrability of power, always for the sake of some interests. Let’s forget about it and move on quickly, as Humpty Dumpty would say.

Some descriptions of some current regulatory proposals are surprising to say the least. For example, it has caught our attention that for the first time GARTNER, the North American market analysis agency that elaborates the hyper cycle of over-expectations for emerging technologies, has included coded legislation [Machine Readable Legislation, MRL] and Rules as Code [RaC]—that is, the web of legal data and the application of law through computer rules—in its hyper cycle for the digital government. We have been even more surprised by the inclusion as an emerging technology of what they have called *citizen twins*, the administrative version of digital twins. We can paraphrase it as the ‘digital citizen’, the replica of real citizens based on their data and metadata which is used for the control (security) and allocation of public resources:

A digital twin of a citizen is a digital representation of an individual. [...]. Governments are developing digital twins of citizens to monitor the environment citizens live in and address health, safety, travel

¹ See an updated summary of this story at https://en.wikipedia.org/wiki/Robodebt_scheme.

and social media impacts on society. [...]. Some governments, such as China's, are building a scoring methodology. Aggregated citizen twins can help map broad patterns and drive resource allocation. [...]. *By implementing MRL, the room for interpretation of legislative or executive intent is eliminated from the process, instead making the law that is passed the same as that which is implemented* [our emphasis]. [...] (Mendonça, 2021, pp. 22 and ff.)

Scoring methodology usages are growing, not only in China but in many other places under its influence (Rwanda, Ghana, Uganda and Zimbabwe...).² The trend of using formal coded legislation is expanding. It is very strong in the United States, Asia, and some Commonwealth countries. The use of online legal web services (often referred to as *RegTech*, *LawTech*, *Fintech* or *SupTech* has been growing exponentially for at least ten years and is changing the landscape of the legal professions and the business of law firms (which also are their main clients) (Casanovas, 2022, pp. 7 and ff.). Through applied AI, companies such as JUDICATA or LEX MACHINA can predict the results and duration of trials based on the specific behaviour profile of judges and officers in similar cases.³

Given this, the reaction of national states has not been homogeneous. There have been some harsh reactions to the impact of AI, mainly in Europe. France has banned, under a €300,000 fine and up to five years in prison, the use of statistical analysis methods ('legal analytics') based on the identity of judges, prosecutors and clerks to "evaluate, analyse, compare or predict their professional practices".⁴ And, at the Community level, the European Union, following in the footsteps of the Data Protection Regulation (GDPR), has already prepared another Regulation, still in a preliminary proposal stage, on the construction and application of data protection systems and Artificial Intelligence.⁵

Governments also tend to introduce corporate schemes (architectures such as TOGAF) in their public administrations, and some state agencies in New Zealand, Canada, UK, and France are actively developing the Rules as Code program since 2018, by which they impose "authorised" or "official" interpretations on citizens. This occurs with

² For a more detailed analysis, cf. Casanovas (2022, Chapter 1).

³ For an accurate description of the set of techniques used, cf. Ashley (2017).

⁴ *Loi n° 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice*. Art. 33 : *ii Les données d'identité des magistrats et des membres du greffe ne peuvent faire l'objet d'une réutilisation ayant pour objet ou pour effet d'évaluer, d'analyser, de comparer ou de prédire leurs pratiques professionnelles réelles ou supposées.* *ii*

⁵ *Artificial Intelligence Act Proposal* (2021). Document 52021PC0206.

the explicit objective of facilitating policy and administrative procedures, minimizing the role of intermediaries that law firms have traditionally had (sometimes qualified as 'modems'). It is still early to issue a judgement, although there are reasons both for praise and criticism (Governatori et al. 2020a, Casanovas and Hashmi et al. 2020).

On the other hand, it seems that the application of AI-based clustering and decision techniques is inevitable to manage the millions of cases that accumulate each year between citizens and the state—in the issuance of spent convictions certificates to obtain work, for example (Governatori et al. 2020b), or in visa, asylum seeking, and citizenship applications. It also seems that is most needed to spare time and personnel to decide on difficult cases, and it would be necessary to establish review mechanisms to correct errors and false positives or negatives as well. In other words, going along with this reasoning, we find the following

Conundrum: The enactment of rights, the reduction of waiting times, and good or better government can be facilitated at the cost of increasing control over citizens through AI-based systems.

This is a conundrum, a difficult issue stemming from a practical reasoning in the sense of G. H. v. Wright. Side, and disruptive effects in the introduction of technology and the use of AI seem to be unavoidable in the process of facilitating good or better government for citizens. This argument can also be understood as a contemporary version of the old doctrine of double effects, as described by McIntyre (2019), although we are not going to take it any further in this paper:

“According to the principle of double effect, sometimes it is permissible to cause a harm as a side effect (or “double effect”) of bringing about a good result even though it would not be permissible to cause such a harm as a means to bringing about the same good end”.

It has also been stated that as long as risks can be controlled and/or monitored and minimised, it is worth facing them in a proportioned strategy of checks and balances. The paradoxes of power and the fallacy of unexplained or biased agendas that can be paraphrased by the previous conundrum as well have been pointed out by many important works—e.g., O’Neil (2016) and Zuboff (2019). They have stressed that there is not a necessary and sufficient link between the implementation of AI and side effects, as technological innovation can lead to alternative consequences. What matters most in decision-making is the policy design behind it and the intended results in mind. Thus, what people (some people) are able to plan and materialise.

It is worth suggesting from our side that behind any prudential argument we can find some issues to be solved at a deeper level. Thus,

we will develop further the previous conundrum under the form of three dilemmas that are being faced in AI & Law approaches to legal governance.

3. Dilemmas in Legal Governance

3.1. SOME USES OF DILEMMAS

As already stated, we will use this figure in a broad way. Dilemmas have been described for a long time in legal theory, international relations⁶, and in Artificial Intelligence, to express conundrums, alternatives, quandaries, and controversial issues. In legal theory, dilemmas have been mainly used related to moral or logical dilemmas. An example of the former ones is Posner's positioning: "'Moral dilemma' is the term we use to describe a moral issue that moral theory cannot resolve" (Posner 1998, p. 1673). A classic example of dilemmas in deontic logic and legal theory is the so-called Jørgensen's dilemma on validity and imperative inferences (Jørgensen, 1937-1938), translated to legal theory by A. Ross (1944). On the one hand, inference gives expression to the fact that the conclusion has value as the premise or premises. On the other hand, it seems evident that it is still possible to draw inferences in which either or some components are imperatives, like in 'Keep your promises. This is a promise of yours. Keep this promise', Cf. Ross (1944, p. 32).

In classical AI, i.e., for Ford and Pylyshyn (1997), dilemmas have remained linked to the frame problem, "the challenge of representing the effects of action in logic without having to represent explicitly a large number of intuitively obvious non-effects" (Murray, 2016). In recent times, epistemology and ethics have opened its scope in AI, to include many practical effects.

Denning and Denning (2020) have singled out (also very broadly) ten AI dilemmas that have not been solved yet. In a Humpty-Dumpty mood, some of them would "arise from seemingly *impenetrable complexity* of the new technology" while the remaining ones "include strong social dimensions and arise from the difficulty of resolving emotional value conflicts to everyone's satisfaction" (Denning and Denning, 2020, p. 22).

Dilemmas of the first kind would be: (i) *explainability* (opacity of how the weights of artificial neural networks relate to the unexpected

⁶ For instance, the 'security dilemma' which exists "when many of the means by which a state tries to increase its security decrease the security of others". Cf. Jervis (1973), Glasser (1997).

output), (ii) *fragility* (sensitivity to small changes), (iii) *bias* (bias in the training data can skew outputs), (iv) *fakes* (how can we trust digital identifications when digitized forms of traditional identifications cannot be distinguished from fakes?), (v) *high cost of reliable training data* (getting properly labelled data is time consuming and expensive).

Dilemmas of the second kind would be: (vi) *military uses* (the dilemma is in how to balance the need for national defence with the desire of many employees to avoid contributing to war), (vii) *weapons and control* (should a drone be allowed to deploy its weapon without an explicit command from a human operator?), (viii) *employments and jobs* (the appearance of new jobs does not help the displaced by technological innovations), (ix) *surveillance capitalism* (“The dilemma for app developers is to find a way that provides the service without compromising individual user control over their data. The dilemma for citizens is how to effectively resist the trend to monetize their personal data and manipulate their behavior.”), (x) *decision making* (do we want machines to only make recommendations or machines that make and act on decisions autonomously?).

As a matter of fact, rather than as dilemmas—referring to linguistic or logical contrasting inferences—these points could be described as controversial or even conflictual areas of AI developments, in which some ethical decisions must be taken and made explicit.

Strümke et al. (2021) have depicted as dilemmas in AI and Ethics the tensions between individual and collective rationalities: “A social dilemma exists when the best outcome for society would be achieved if everyone behaved in a certain way, but actually implementing this behaviour would lead to such drawbacks for an individual that they refrain from it”, i.e. “when the interests of the collective conflict with the interests of the individual making a decision” (Strümke et al., 2021, p.3). Thus, the prisoner dilemma or the tragedy of the commons would fall under this category.

This is an interesting epistemic field, in which the dilemma between ‘I-intentions’ and ‘We-intentions’ has been qualified by Searle (1997) as a *false dilemma*, for the “‘I-intend’ occurs only as a part of the ‘We-intend’” and ‘the definition of a social collective consists in the fact that the participants *think* it is a collective”, hence “the collective consists of individuals with collective individuality” (Searle *ibid.* 1997, pp. 449-450). We cannot dwell on this interesting issue now. We will confine ourselves to observing that emergent collective properties might not exist only “in the minds of individual persons”. There is more to it than that.

Game and moral dilemmas—such as the trolley problem—have been taken up by Machine Ethics, where many AI systems have been built

reproducing them (Tolmeijer et al., 2020). In the inception of the field, Moor (2006) defined it by classifying four categories of moral agency: (i) *Ethical Impact Agents*: Any machine that can be evaluated for its ethical consequences; (ii) *Implicit Ethical Agents*: Machines that are designed to avoid unethical outcomes, (iii) *Explicit Ethical Agents*: Machines that can reason about ethics; (iv) *Full Ethical Agents*: Machines that can make explicit moral judgments and justify them. Moral dilemmas would typically be embedded into the three latter categories (the last one is still to be reached).⁷ However as pointed out by Winfield et al. (2019, p. 514):

Even if the technical problem of machines able to resolve real-world ethical dilemmas were solved, society-wide debate would then be needed to discuss and agree on the rules and protocols for such machines, not least because society as a whole needs to take responsibility for the human causalities of accidents caused by such machines.

In complex social environments, ethical and legal governance take place at the same time. Normative, organisational, and ethical dilemmas have been mentioned to formulate a legal governance approach, in which socio-technical and cyber-physical systems are generated within hybrid intelligence environments (Pagallo et al., 2018).

We can specify a bit more three of the main dilemmas in AI and legal governance developments. Legal governance, as it has been defined by Pagallo et al. (2019a, 2019b) and Casanovas et al. (2022), entails the use of an array of AI instruments of different nature—semantic web languages, natural language processing, ontologies, knowledge graphs, convoluted machine learning. . . — not necessarily in a joint way, entailing the emergence of sustainable legal ecosystems. These legal ecosystems encompass also different components according to the nature of the field to be regulated. For example, Connected Autonomous Vehicles (CAVs) in Mobility as a Service (MaaS) infrastructures require different middleware system designs than Industry 4.0 platform-driven manufacture management (in which quality control, augmented reality and digital twins are used to monitor mass production of goods).

Let's formulate the dilemmas that can arise in building AI legal governance ecosystems.

⁷ We borrow this description from Winfield et al. (2019).

3.2. DILEMMAS IN LEGAL GOVERNANCE

3.2.1. *First Dilemma: Normative Autonomy*

The first engineering position that needs to become explicit is what is the scope of the legal governance of AI. Our proposal is to clarify the need to engineer governance for hybrid systems that include: (a) artificial systems whose behaviour may be autonomous (no direct involvement of a human in the actions of the system) and (b) the use of AI as a means to enforce compliance within a regulated hybrid system.

The first dilemma has the following form:

First Dilemma: On the one hand, human intervention is necessary for the profitable application of intelligent information systems and the minimization of risks that their use entails. But, on the other hand, given that the learning and operation processes of intelligent systems and agents have an increasing degree of autonomy, it is possible to state that artificial intelligence systems should be able to be self-controlled, or controlled by other artificial intelligence systems.

This dilemma has to do with autonomy, compliance, and with the formal limitations of normative systems.

3.2.2. *Second Dilemma: Legal Isomorphism*

The second issue concerns the implementation of norms in real contexts and is complementary to the next one. The applications of artificial intelligence to law have been based almost from the beginning on the notion of *legal isomorphism*, that is, on the extraction of semantic rules with the general form 'condition (facts)—effects (normative) ' [if...then] from legal texts originally written in natural language.⁸ Legal texts are interpreted here in a broad sense (values and ethical principles, public policies, protocols, good practices, international agreements, etc.). They constitute the basis of the usual legal instruments, which accept a high degree of ambiguity and contradiction. More specifically, in practice, the same term (for example, what is a contract, or what does “justice” mean, or the conditions for the use of violence) can be interpreted in a contradictory way, both in the present and in the future. In other words: the conflictive basis of human interactions in natural languages is ineliminable from its logical representation.

Second Dilemma: How should we operate with formal representation languages? Do we interpret legal isomorphism as the basic

⁸ The first approach to legal isomorphism, after the formalisation of the articles of the British Nationality Act by the Logic Programming Group of Imperial College only considered the content of the laws. Cf. Sergot et al. (1986), Bench-Capon and Coenen (1992).

pillar of normative systems and argumentation—this is the most common practice—, or do we rather look for other solutions for the knowledge acquisition process that should be modelled?

This dilemma has to do with knowledge acquisition, representation languages, and modelling practices.

3.2.3. *Third Dilemma: The Modes of Regulation*

The third dilemma is similar to the first one, but it does not derive from the complexity of the systems and their actual deployment; it has a formal origin, and it influences the way regulation is made operational on hybrid systems. The question is to decide what artefacts are needed to build artificial intelligence legal governance. Should regulatory artificial systems be based only on rules, or should they adopt other elements that do not have this form? That is, should the systems to be implemented assume contradictions, elements, or exogenous variables that modify the results or preserve an open path for uncertainty?

The standard approach has been to reduce the problem to finding an appropriate way of understanding ‘norms’ and to accompany this understanding with a compatible notion of ‘consequence’. If this were enough, one could embed legal governance as a set of norms and a notion of inference that would serve to identify violations and contrary to duty actions. It is not clear whether this is enough, or we also need other means to make governance operational; but even if it were, making these two assumptions explicit still faces the representation issue.

In fact, the problem is more general: It comes from the very nature of the logical systems that underlie the way Artificial Intelligence is usually applied into legal systems:

Third Dilemma. Can we limit regulation simply to set off formal expressions (expressed in a rigorous formal and machine-readable language), and concomitantly use an automated inference mechanism to assess compliance and execute enforcement actions when appropriate? Or should we combine formal methods with some kind of heuristics and non-provable intuitions?

This dilemma has to do with social control, but also with the architecture of legal knowledge systems (expressed with rules + exceptions). The application of formal semantics to law depends on the answer we give. “Rules + exceptions” has been the formula most adopted by AI and Law researchers, with the application of non-standard and non-monotonic (‘defeasible’) deontic logic to argumentation and legal systems.⁹ Doing so, they have taken into consideration the openness

⁹ Cf. Bench-Capon and Gordon (2009).

of legal theory to argumentative forms and the criticism of Dworkin, Alexy, Habermas, Aarnio and Peczenik (among others) to rule modelling. Although not all logicians agreed to dispense with recursive models.¹⁰

4. Discussion: Legal Governance and Linked Democracy

The dilemmas of normative autonomy, legal isomorphism and the modes of regulation can be formulated in other ways. We might try to cluster them, for example, as a Fishkin-like trilemma on deliberative democracy to show the interdependence of responses —the inability to fully realise the three basic democratic principles (political equality, participation, and deliberation) in any single democratic institution (Fishkin, 2009). The trilemma of deliberative reform states that there are three basic principles internal to the design of democratic institutions: political equality (people’s views are counted equally), mass participation (we are all given the opportunity to provide informed consent), and deliberation (we are all given the opportunity to provide opinions and weigh competing arguments). Any effort to attain any of the two principles inevitably hinders the third.

But it is not necessary to do so to realise that there is no general theory nor institution that offers a unique solution to the three dilemmas formulated above. There are partial answers, adapted to the regulatory sector, to the regulatory requirements, and to the provisions of the environment (affordances) that must be considered to model the system.

Our intuition is that NP-complete limitations—the class of computational problems for which no efficient solution algorithm has been found—affect legal isomorphism approaches, and very likely all attempts to define legal validity as well. ‘NP-Complete’ means ‘nondeterministic polynomial-time complete’. According to Garey (1979), the foundations for the theory of NP-completeness were laid in a paper of Stephen Cook, presented in 1971, entitled “The Complexity of Theorem Proving Procedures”. Garey provided a long list of ‘intractable’ problems because of NP-completeness (among them, graphs and hypergraphs, and many issues related to planning design, such as various versions of the travelling salesman problem). In normative multi-agent system theory, Ågotnes et al. (2010) have shown that in games where

¹⁰ The late Alchourrón (1996) considered non-monotonic logics unsound for the representation of defeasible reasoning. He preferred a logic of belief revision, and he insisted on the identification of a normative system behind what he called ‘the Master Book’ (the set of all normative expressions, such as statutes, codes, etc.).

agents must determine whether to comply with the normative system or not, “the complexity of checking whether there exists a normative system which has the property of being a Nash implementation is NP-complete”. In legal theory and AI and Law, not much attention has been devoted to it. One exception is Governatori et al. (2014), in which it is contended that the problem of deciding what move (set of rules) to play at each turn in dialogue games is an NP-complete problem.

Compliance and legal compliance checking raise interesting issues, in which legal isomorphism can be challenged from conceptual and ontological approaches that require a previous theorisation, including legal concepts, architecture, and reasoning.¹¹ From Business Process Management and Service-Oriented Architecture, Hashmi et al. (2016) have set some normative requirements along with the formal limitations to express them in deontic logic languages (Hashmi et al. 2018).

The modes of regulation and the many ways to logically represent rules have to do with the human ability to understand different normative constraints to solve social problems. More than fifty years ago, Martin Gardner pointed out that the paradox of reflexivity that Russell faced with the theory of logical types and the notion of metalanguage could take many forms. One of them was: “All rules have exceptions” (Gardner, 1963). The notion of a system of norms (or rules) has formed the basis of regulatory systems (and in part of legal systems) for more than fifty years now and, in their conventional form, those rules have a conditional structure ‘if..., then...’ in a classical logical interpretation.

We think that legal validity and legal validation can be set apart from legal validity using empirical fined-grained causal loop models.¹² The distinction between Compliance *through* Design (CtD) (legal compliance) and Compliance *by* design (CbD) (in business and corporate design) deals with these different fields: Public law entails the construction of a semi-automated *public* space in which norms have several open interpretations and citizens’ rights should be enacted and protected.¹³

Poblet and Plaza (2017) observed that “if we conceptualize political equality in the classical sense [*isegoria* (equal voice) + *isonomia* (equality of political rights)] self-selection does not necessarily diminish the principle of equality (non-participation is an individual decision)”. They propose developing civic technologies for democracy and democratisation processes to mitigate the effects of the trilemma and

¹¹ Francesconi and Governatori (2022) have recently proposed a model based on modelling deontic norms in terms of ontology classes and ontology property restrictions.

¹² Cf. Casanovas, Hashmi, and de Koker (2021) for a three steps empirical methodology for validation checking.

¹³ Cf. for a survey, Hashmi, Casanovas, and de Koker (2018) for a survey.

to complement deliberative and epistemic accounts of democracy with the strength of technological innovation.

These types of resources are both essential for the correct construction of the legal system that uses AI, and for the prevention of unwanted collateral effects. On the contrary, if they try to avoid the preceding steps, we can incur in what we could call the paradoxes of the simplistic regulation of AI, as already described in our first conundrum, related to the ‘impenetrability’ of language and power.

In previous work we approached the design foundations we just mentioned and advocated three resources for approaching the legal governance of artificial intelligence legal systems: (i) We distinguished five levels of governance of autonomous systems from and through artificial intelligence, each level presenting distinguishing features whose adequate governance may be addressed through specific means (Noriega and Casanovas, 2022)¹⁴; (ii) we returned to some formulations of cybernetics and design sciences for the construction of normative systems adapted to their environment, e.g. for the construction of regulatory and legal ecosystems (Casanovas, de Koker and Hashmi, 2022) and proposed constructs and methodologies for their implementation; (iii) we combined normative multiagent systems with ethical design for the construction of hybrid online systems (involving both human beings and autonomous artificial systems) that may be provably aligned with human values (Noriega et al. 2021).

Our intuition is that regulation is not enough. The objectives and functioning of a hybrid system can be formally aligned with human values and rights (Sierra et al., 2021), and could also be harmonized with a theory of collective (or common) rights for the consistent construction of social communities (Poblet and Sierra, 2020; Sierra and Osman, 2022). Elinor Ostrom’s institutional analysis¹⁵ has already triggered several proposals of computer formalisation, mainly focused on a set

¹⁴ The five levels of governance correspond to decision-making by the system autonomously based on the complexity of the system. The last level points to general artificial intelligence, whose development we know that it will take time to come.

¹⁵ The eight institutional principles for self-government design identified by Ostrom are: (i) Clearly defined boundaries; (ii) Congruence between appropriation and provision rules and the state of the prevailing local environment; (iii) Collective-choice arrangements: those affected by the provision and appropriation rules participate in selection and modification of those rules; (iv) Monitoring by accountable agencies; (v) Graduated sanctions; (vi) Access to fast, cheap conflict-resolution mechanisms; (vii) Existence of and control over their own institutions not challenged by external authorities; (viii) Systems of systems (or nested enterprises: appropriation, provision, enforcement, conflict resolution, and governance activities are organized in multiple layers of nested enterprises). We have reproduced here the summary offered by Pitt et al. (2015) following the metareview by Cox et al. (2010).

of autonomous, heterogenous actors needing to collectivise and distribute resources without a centralised decision-making authority (Pitt et al. 2014), and on interactions based on rights and mutual obligations (Montes et al. 2022).

Poblet et al. (2019, 27) called *linked democracy* (LD) the distributed, technology-supported collective decision-making process, where data, information and knowledge are connected and shared online by citizens. They aligned CPR principles with LD properties.¹⁶ Likewise, the toolkit for SMART AI governance proposed by Pagallo et al. (2019a) is based on a middle-out approach in which some of LD properties (such as scalability, modularity and reusability) are considered as key within a more general AI regulatory landscape (Pagallo 2019b). This might change the way of looking at Open Access publishing on the Web, as on the one hand, conflicts, concurrence, and competition in the marketplace are also reflected into OA platforms. But, on the other, the way to handle them properly is working and fleshing out the principles and ethical values of legal governance that we have been contending for in this article. We believe that a conceptual, theoretical, and balanced approach to AI regulatory design can overcome the three dilemmas that frame hybrid intelligence, at the crossroads of human and machine behaviour.

Digital platforms, legal web services, and political crowdsourcing in a platform-driven economy could benefit from AI legal governance, even if not all problems can be solved at the same time. Democratic procedures and methods can be dealt with separately.¹⁷ Information flows on the Internet of Things are intertwined with web of data developments. The regulatory problems that arise in workplaces with sensors and the control of producers and machines must be solved at the micro-level. We are facing complex situations in which layers and dimensions of social behaviour should be carefully distinguished according to the service models.¹⁸

¹⁶ Contextuality, openness, blending, distribution, (technological) agnosticism, modularity, scalability, knowledge reusability and archivability; alignment (Poblet et al. 2019, p. 83).

¹⁷ Cf. Fishkin's late developments of his Deliberative Poll model towards the ideal of "deliberative democracy by the people themselves" (Fishkin 2018, 2019). It is worth noting that this proposal does not consider technology (and AI) as an independent party. On the contrary, it is our contention that legal governance is specifically based on the centrality of AI languages and means to represent, gear and operate with normative and legal knowledge.

¹⁸ Fog computing and Cloud ecosystems, Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS), etc. Cf. NIST 500-325, Iorga et al. (2018).

5. Final Comments

The change in legal culture brought about by the development of the network and information systems reflects the change from a mainly analogical to a mainly digital culture. However, they still coexist. But there is no longer a clear-cut separation between the digital and the physical world. Artificial Intelligence provides a wealth of tools and practices that allow the autonomy of machines in decision-making, and the production of synthetic data (data inferred through data). It constitutes a complex hybrid reality between humans and machines with overlapping levels of self, co- and hetero-regulation.

Automatic validation of regulatory compliance in real time is already a need in legal ecosystems related to augmented reality, digital replicas, hybrid online systems, as well as in the construction of infrastructures in manufacturing, commerce, and social communities. Very soon, in the administration and the government.

The conundrum of automatic social control and the dilemmas we have briefly exposed—related to the degree of autonomy, legal isomorphism, and reflexivity—do not have a single solution. They require further reflection and the formulation of a prior design framework through the explicit assumption of ethical, legal, and political assumptions. We cannot know with certainty what the new economic and regulatory framework will be for the new scenarios of a culture dominated by information platforms and processes. What we are beginning to realise is that the process is irreversible and that, whatever it may be, we should be able to take part directly in it and not blindly delegate our position to third parties, be they corporations or governments, or an autonomous artificial system.

Acknowledgements

This article is a partial result of the EU H2020 Project *Optimizing Manufacturing Processes through Artificial Intelligence and Virtualization* (OPTIMAI), Grant Agreement No. 958264. A shorter dissemination article in Spanish on this subject will be published in *Nuevos Diálogos*, Num 2, octubre-diciembre, 2022 (*Publicación de la Dirección General de Divulgación de la Ciencia / Dirección General de Divulgación de las Humanidades, UNAM, México*).

References

- Ágotnes, Thomas, Wiebe Van der Hoek, and Michael Wooldridge. “Robust normative systems and a logic of norm compliance”. *Logic Journal of IGPL* 18.1 (2010): 4-30.
- Akata, Zeynep, et al. “A research agenda for hybrid intelligence: augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence.” *Computer* 53.08 (2020): 18-28.
- Alchourrón, Carlos. “On law and logic”. *Ratio Juris* 9.4 (1996): 331-348.
- Artificial Intelligence Act. Document 52021PC0206. *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.
<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>
- Ashley, Kevin D. *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*. Cambridge: Cambridge University Press, 2017. Print.
- Bench-Capon, Trevor, and Coenen, Frans P. “Isomorphism and legal knowledge-based systems.” *Artificial Intelligence and Law* 1.1 (1992): 65-86.
- Bench-Capon, Trevor, and Gordon, Thomas F. “Isomorphism and argumentation”. In *Proceedings of the 12th ICAIL*, ACM, 2009, 11-20.
- Booker, Don M. “Are societies Turing machines? Some implications of the cyclical majority problem, an NP complete problem, for cybernetic models of social systems.” *Sociocybernetics*. Springer, Boston, MA, 1978, 105-127.
- Carroll, Lewis. *Through the Looking-Glass* [1871]. London: Penguin, 2010. Print.
- Casanovas, Pompeu, Hashmi, Mustafa, Barnes, Jeffery, de Koker, Louis, Ho-Pun Lam, Governatori, Guido, Zeleznikow, John. *Comments on Cracking the Code: Rulemaking For Humans And Machines (August 2020 draft) Comments on the draft OECD White Paper on Rules as Code, submitted on 27 August 2020 to the authors. Second Version, 2020* <https://zenodo.org/record/5797205#Yteg-HZBw2w>
- Casanovas, Pompeu. “The double implosion of the legal profession and Web Services in the digital age”, in Casanovas, P. *Law as Knowledge: Anthropological Bases for the Regulation of the Web of Data*. PhD in Legal Anthropology, University of Barcelona, June 2022, 73-113. <https://www.tdx.cat/handle/10803/674696?locale-attribute=es\#page=1>
- Casanovas, Pompeu, Hashmi, Mustafa, de Koker, Louis. “A Three Steps Methodological Approach for Legal Governance Validation”, *AICOL@JURIX 2021*.
- Casanovas, Pompeu, de Koker, Louis, Hashmi, Mustafa. “Law and Sociolegal Governance on the Internet of Things and Industry 4.0”. Special Issue, “The Impact of Artificial Intelligence on Law”, Open Access, *J. Multidisciplinary Scientific Journal* 5.1 (2022): 64-91. Open Access. Web.
- Camps, Victoria. *Breve Historia de le Ética*. Barcelona: RBA, 2013. Print.
- Collingridge, David. *The Social Control of Technology*. New York: St. Martin’s Press. Print.
- Cox, Michael, Arnold, Gwen, and Villamayor, Tomás. “A review of design principles for community-based natural resource management.” *Ecology and Society* 15.4 (2010). n. pag. Web.
- Denning, Peter J., and Denning, Dorothy E. “Dilemmas of artificial intelligence.” *Communications of the ACM* 63.3 (2020): 22-24.

- Fishkin, James S. *When the people speak: Deliberative democracy and public consultation*, Oxford: Oxford University Press, 2009. Print.
- Fishkin, James S. *Democracy when the people are thinking: Revitalizing our politics through public deliberation*. Oxford: Oxford University Press, 2018. Print.
- Fishkin, James S. "Democracy When the People Are Thinking: Deliberation and Democratic Renewal 1." *Proceedings of the American Philosophical Society*, 163.2 (2019): 108-121.
- Ford, Kenneth M., and Pylyshyn, Zenon W. eds. *The robot's dilemma revisited: The frame problem in artificial intelligence*. Westport, Greenwood Publishing Group, 1996.
- Gardner, Martin. "Logical Paradoxes". *The Antioch Review*, 23.2 (1963): 172-178.
- Garey, Michael R. and Johnson, David S. *Computers and Intractability. A Guide to the Theory of NP-Completeness*, San Francisco: W.H. Freeman, 1979. Print.
- Glaser, Charles L. "The security dilemma revisited." *World politics* 50.1 (1997): 171-201.
- Governatori, Guido, Olivieri, Francesco, Scannapieco, Simone, Rotolo, Antonino, and Cristiani, Matteo. "Strategic Argumentation is NP-Complete." In *European Conference on Artificial Intelligence, ECAI 2014*, Association for the Advancement of Artificial Intelligence, 399-404.
- Governatori, G., Barnes, Jeffrey, Zeleznikow, John, de Koker, Louis, Poblet, Marta, Hashmi, Mustafa and Casanovas, Pompeu. "Rules as Code will let computers apply laws and regulations. But over-rigid interpretations would undermine our freedoms." *The Conversation*, 2020a, n. pag. Web. <https://theconversation.com/rules-as-code-will-let-computers-apply-laws-and-regulations-but-over-rigid-interpretations-would-undermine-our-freedoms-149992>
- Governatori, Guido, Casanovas, Pompeu, de Koker, Louis. "On the Formal Representation of the Australian Spent Conviction Scheme", In *International Joint Conference on Rules and Reasoning, RuleML+RR 2020*, Cham: Springer, LNCS 12173, 177-185.
- Hashmi, Mustafa, Governatori, Guido, and Wynn, Moe T. "Normative requirements for regulatory compliance: An abstract formal framework." *Information Systems Frontiers* 18.3 (2016): 429-455.
- Hashmi, Mustafa, Governatori, Lam, Han Pun, Guido, and Wynn, Moe T. "Are we done with business process compliance: state of the art and challenges ahead." *Knowledge and Information Systems* 57.1 (2018a): 79-133.
- Hashmi, Mustafa, Casanovas, Pompeu. and de Koker, Louis. *Legal Compliance through Design: Preliminary results of a literature survey. TEREKOM2018@ JURIX, Technologies for Regulatory Compliance* 2018b, <http://ceur-ws.org>, 2309, p.06. <http://ceur-ws.org/Vol-2309/06.pdf>
- Jervis, Robert. "Cooperation under the security dilemma." *World politics* 30.2 (1978): 167-214.
- Jørgensen, Jørgen. "Imperatives and logic." *Erkenntnis* 7 (1937-38): 288-296.
- Iorga, Michaela, Larry Feldman, Robert Barton, Michael J. Martin, Nedim S. Goren, and Charif Mahmoudi. *NIST Special Publication 500-325 Fog Computing Conceptual Model. Recommendations of the National Institute of Standards and Technology*. (2018)
<https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.500-325.pdf>

- Loi n° 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice.* <https://www.legifrance.gouv.fr/dossierlegislatif/JORFDOLE000036830320/>
- McIntyre, Alasdair. “Doctrine of Double Effect”, *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), 2019, <https://plato.stanford.edu/archives/spr2019/entries/double-effect/>
- Montes, Nieves, Nardine Osman, and Carles Sierra. “A computational model of Ostrom's Institutional Analysis and Development framework.” *Artificial Intelligence* 311 (2022): 103756.
- Mendonsa, Alia. *Hype Cycle for Digital Government Technology*. Gartner Report. Published 21 July 2021, 110 p., ID G00747518.
- Moor, James H. “The nature, importance, and difficulty of machine ethics.” *IEEE intelligent systems* 21.4 (2006): 18-21.
- Noriega, Pablo and Casanovas, Pompeu. “La gobernanza de los sistemas artificiales inteligentes”, en O. Velarde y M. Martín Serrano (Eds.), *Mirando hacia el futuro. Cambios sociohistóricos vinculados a la virtualización*. Centro de Investigaciones Sociológicas (CIS), Madrid, [Ministerio de la Presidencia del Gobierno de España], 2022, 115-143 (in press).
- Noriega, Pablo, Verhagen, Harko, Padget, Julian, and d’Inverno, Mark. “Ethical Online AI Systems through Conscientious Design.” *IEEE Internet Computing* 25.6 (2021): 58-64.
- O’Neil, Cathy. 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy*. NY: Broadway Books, 2016.
- Pagallo, Ugo, Corrales, Marcelo, Fenwick, Mark, and s Forgó, Nikolaus. “The rise of robotics and AI: Technological advances and normative dilemmas.” In Corrales, Fenwick, M., and Forgó eds., N. *Robotics, AI and the Future of Law*, Singapore: Springer, 2018, 1-13.
- Pagallo, Ugo, Aurucci, Paola, Casanovas, Pompeu, Chatila, Raja, Chazerand, Patricia, Dignum, Virginia, Luetge, Christoph, Madelin, Robert, Schafer, Burkhard, and Valcke, Peggy. *On good AI governance: 14 priority actions, a SMART model of governance, and a regulatory toolbox*. AI4People, 2019a, n. pag. Web. https://www.eismd.eu/wp-content/uploads/2019/11/AI4Peoples-Report-on-Good-AI-Governance_compressed.pdf
- Pagallo, Ugo, Casanovas, Pompeu, and Madelin, Robert. “The middle-out approach: Assessing models of legal governance in data protection, artificial intelligence, and the Web of Data.” *The Theory and Practice of Legislation*, 7.1 (2019b): 1-25.
- Posner, Ronald A. “The problematics of moral and legal theory” *Harvard Law Review* 111.7 (1997): 1637-1717.
- Pitt, Jeremy, Busquets, Dídac, and Macbeth, Sam. “Distributive justice for self-organised common-pool resource management.” *ACM Transactions on Autonomous and Adaptive Systems (TAAS)* 9.3 (2014): 1-39.
- Pitt, Jeremy, Busquets, Dídac, and Riveret, Régis. “The pursuit of computational justice in open systems.” *AI and Society* 30.3 (2015): 359-378.
- Poblet, Marta and Plaza, Enric. “Democracy Models and Civic Technologies: Tensions, Trilemmas, and Trade-offs”, *Linked Democracy*, LINKDEM@IJCAI2017, n. pag. Web, 2017, <http://ceur-ws.org/Vol-1897/paper4.pdf>
- Poblet, Marta, Casanovas, Pompeu, and Rodríguez Doncel, Víctor. *Linked Democracy. Foundations, Tools, and Applications*. Springer Brief in Law No 750, Cham: Springer Nature, Open Access, 2019, <https://link.springer.com/book/10.1007/978-3-030-13363-4>

- Poblet, Marta and Sierra, Carles . “Understanding Help as a Commons”, *International Journal of the Commons* 14.1 (2020): 481–493. DOI: <http://doi.org/10.5334/ijc.1029>
- Ross, Alf. “Imperatives and Logic”, *Philosophy of Science*, 11.1 (1944): 30-46.
- Searle, John R. “Responses to critics of the construction of social reality.” *Philosophy and Phenomenological Research* 57.2 (1997): 449-458.
- Sergot, Marek J., Sadri, Fariba, Kowalski, Robert A., Kriwaczek, Frank, Hammond, Peter, and Cory, Terese H. “The British Nationality Act as a logic program.” *Communications of the ACM* 29.5 (1986): 370-386.
- Shanahan, Murray. “The Frame Problem”. *The Stanford Encyclopedia of Philosophy*, 2016: n. pag. Web. E.N. Zalta (ed.), <https://plato.stanford.edu/archives/spr2016/entries/frame-problem/>
- Sierra, Carles., Osman, Nardine, Noriega, Pablo, Sabater-Mir, Jordi, and Perelló, Antoni. *Value alignment: a formal approach. arXiv preprint arXiv:2110.09240*, 2021.
- Strümke, Inga, Marija Slavkovik, and Vince Istvan Madai. “The social dilemma in Artificial Intelligence development and why we have to solve it.” *AI and Ethics* (2021): 1-11,
- Tolmeijer, S., Kneer, M., Sarasua, C., Christen, M., & Bernstein, A. “Implementations in Machine Ethics: A survey”. *ACM Computing Surveys (CSUR)*, 53.6 (2020): 1-38.
- Wikipedia. “Robodebt scheme“ https://en.wikipedia.org/wiki/Robodebt_scheme
- Wikipedia. “TOGAF. The Open Group Architecture Framework”. n. pag. Web, Wikipedia, https://en.wikipedia.org/wiki/The_Open_Group_Architecture_Framework
- Winfield, Alan F., Michael, Katina, Pitt, Jeremy, and Evers, Vanessa. “Machine ethics: The design and governance of ethical AI and autonomous systems [scanning the issue].” *Proceedings of the IEEE* 107.3 (2019): 509-517.
- Zuboff, Shoshana. *The age of surveillance capitalism: The fight for a human future at the new frontier of power: Barack Obama's books of 2019*, NY: Profile Books.