CEUR-WS.org/Vol-3289/paper4.pdf

Legality, Legitimacy, and Instrumental Possibility in Human and Computational Governance for the Public Sector

Vanja Skoric[1], Giovanni Sileno[1] and Sennay Ghebreab[1]

[1]*Socially Intelligent Artificial Systems (SIAS), Informatics Institute, University of Amsterdam, the Netherlands*

Abstract

Artificial Intelligence (AI) can have a significant beneficial or harmful impact when used in public policies and services. This paper provides a reflection on the EU regulatory initiatives towards regulating governance and risk management of AI, elaborating on the entrenchment of the dimensions of legality, legitimacy and instrumental possibility. We argue that only against this conceptual backdrop we can possibly unpack what a legal, legitimate, trustworthy, fair and accountable approach to designing and implementing AI in public service is. As an outcome, we identify a few crucial architectural requirements, both at normative and at computational level.

Keywords

AI Governance, Legality, Legitimacy, Intervention points, Oversight, Engagement spaces, Public sector

1. Introduction

The public sector is increasingly developing, procuring, or integrating digital and emerging technologies, including artificial intelligence (AI) systems, for a variety of decision-making and service provision purposes [1]. While the benefits of AI use in the public sector are expected to be significant, attaining them is not a straightforward task. AI's potential and already manifested harms in the public sector have both evident and implicit impacts on human rights, democracy and the rule of law, which demand multi-pronged interventions, primarily from the state.[1] Technological innovation is known to have the potential to deeply modify power relationships, therefore, when deciding to use AI in public and e-government services, issues of both *legality* and *legitimacy* of its use must be considered and addressed, to strive for societal sustainable solutions. For the scope of the present paper we focus also on the fact that both legality and legitimacy relies on an implicit third dimension, that of *technical* or *instrumental possibility*, which concerns enabling of normative and material (including computational) interventions. We argue that only against this conceptual backdrop we can possibly fully unpack what a legal, legitimate, trustworthy, fair and accountable approach to designing and implementing AI in public service is.

Whereas most policy and technical contributions today focus on specific inferential algorithms (taking what we may call internal views on the problem), we converge our attention on *human*

IFDaD 2022 INTERNATIONAL FORUM on DIGITAL and DEMOCRACY, November 17-18, 2022, live and digital event in Rome, Italy

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

[1] Due to the public sector being maintained primarily by the state.

and computational governance layers directing and regulating any form of algorithms (and in doing so, we switch to external views). From the human side, we examine what are the minimal legality and legitimacy requirements for a meaningful governance structure. The issue here is how we can attain AI benefits and mitigate risk and harm, ensuring consistent and inclusive governance roles to relevant social participants (section 3). On the computational side, we briefly elaborate on what are the infrastructural components required to regulate algorithms in alignment with the directives provided by humans. The associated challenge concerns how we can set up at design phase forms of continuous development and integration, meant to adapt the behaviour of the computational system to dynamic and plural contexts, directives, and understandings (ie. forms of *continuous governance*) (section 4). Section 2 provides a more general background and motivation to our contribution, elaborating on the most relevant current (proposals of) regulations in Europe.

2. Legality, Legitimacy, and Instrumental possibility

2.1. In theory...

Legality is at the core of the domain of law; formal institutions define criteria of validity and for qualifying behaviour, and operationalize procedures that apply those for promoting normative order in society according to the source of law. Legitimacy is generally seen as a more complex category, drawing both from morality (eg. moral imperatives, legitimate interests, principles of justice, ...) and political science (power and distribution of powers, ...). Simplistically, one can see legality as being expressed by rules, and with a *top-down* characterization: what is legal or illegal depends on some given source of law. Legitimacy on the other hand has a *bottom-up* nature; people who are producing or disabled to produce of an act (more in general, entities holding *moral agency*), or that are suffering the occurrence or the omission of a certain act (*moral patiency*) are the ones that are primarily driving or experiencing such an act, *a priori* of circumstantial legal arrangements. Yet, very few scholars would argue that the two dimensions are strictly separated, and, in fact, one can find in the very functioning of the legal activity traces of their interaction.

We outline this connection for illustrative purposes in Figure 1. If a certain action is deemed licit by the current normative setting (eg. as described by national laws), this is presumptively to protect concerns deemed legitimate by law-makers at a certain point in time. Evidently, such *"internal" legitimacy* (as related to the will of the decision-maker) can not rule out that the normative action may be in some context inadequate, for several reasons. Subjects impacted by this inadequacy may find support in higher-level principles, that may be in turn organized in legal forms (eg. international laws, human rights frameworks, etc.), in the best case already setting up some proceduralized form of protection. This feedback cycle shows that legitimacy concerns (or better, *"external" legitimacy*, as related to the interests of others than the decision-maker) readjust, if not drive, what the law is, and does so from the bottom-up.

There is however an additional dimension which is for the most neglected in discussions on legality and legitimacy, but which becomes particularly evident in discussions on technological transformation: *instrumental possibility*. In both schemes, the lack or inadequacy of instrumental possibilities undermines the possibility of aligning legality with legitimacy. Said differently,

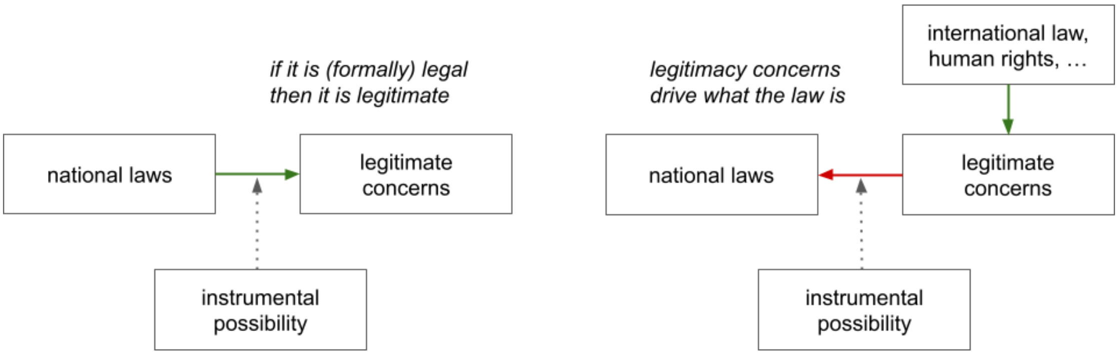

Figure 1: Interactions between legality, legitimacy, and instrumental possibility.

social participants need to have the possibility to have their legitimate concerns adequately promoted and protected, both when these concerns are properly identified by the law, and when they are inadequately treated by the law. This possibility can be translated as a matter of access-to and control-of *points of interventions*, both at normative, and at material (computational) level.

2.2. In practice...

Looking at the current debate, the absence of clear normative frameworks in the public sector, especially for safeguarding human rights, has prompted scholars to consider if this constitutes a violation of the international human rights obligations in itself (eg. [2]).[2] In this line of thoughts, there is a state's obligation to ensure that the use of AI in e-governance in particular is adequately addressed due to the risk it poses to human rights. Consequently, the absence of such safeguards may constitute a violation of the state's duty to protect.[3] A question then emerges: *to what extent can existing or proposed EU level normative frameworks facilitate legality within AI design, development and use to safeguard human rights, enhance democratic values and assist in building up civic participation?*

In parallel to the above, there are discussions about the need for increased legitimacy of the use of AI [3, 4, 5]. One potential venue to support legitimacy might include democracy by design in the context of AI (see eg. [6]), in line with the idea of a value-sensitive design[4] and frameworks of privacy by design[5]. There are increasing efforts to develop methods for assessing impact of AI early in the design and development process.[6] Therefore another question follows: *how to ensure that the AI system is effective for its purpose, as well as able to safeguard human rights, ensuring intrinsic value of human rights, democracy and participation are embedded in the*

[2] Also the UN Human Rights Committee requires states to put in place a framework that prevents human rights violations from taking place (General Comment 31, UN doc CCPR/C/21/Rev 1/Add 13, 26 May 2004).

[3] See for instance "The State duty to protect human rights", from the UN Guiding principles on business and human rights (2012), pp. 3–12.

[4] See eg. the IEEE P7000 standard [7].

[5] Ses eg. the Privacy by design clause in the GDPR [8], Art. 25 section 1

[6] See eg. the list of initiatives in the recent EY survey [9].

design of AI itself? Indeed, AI impact is to a large extent dependent on choices made during design, development and use, implying the importance of governance and process that allows one to choose whenever needed whether to develop the technology further, and by which path and means to do so. This is a crucial requirement for societal viability.[7] If well functioning, the entrenchment and alignment of human governance structures with computational control structures would be an expression of "continuous governance". This brings us to the final question: *what is needed in terms of computational infrastructural components for effectively enabling continuous governance?*

2.3. Current regulatory frameworks

One of the key functions of an adequate normative framework for AI should be to properly guide and structure the AI life-cycle [10], from conception and design to deployment. Its main goal would be to establish appropriate checks and balances for the use of AI, particularly sensitive for the public sector. To that aim, the EU has in recent years taken several regulatory steps towards addressing governance and risk management in the digital ecosystem, including for the public sector. We will now review three main regulatory developments which are relevant for the public use of AI, addressing governance and process: the GDPR (2018), the EU AI Act (2021), and the Corporate Sustainability Due Diligence proposal (2022). We will observe that all three regulatory frameworks exhibit some human rights and democratic principles safeguards, but they also include significant limitations with respect to scope, methods, or inclusivity, which may hamper both legality and legitimacy aspects of AI use in public services.

GDPR The main goal of the General Data Protection Regulation (GDPR) is to protect the "fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data" [8]. GDPR involves a series of due diligence requirements for "controllers" or "processors" of personal data. These include conducting a data protection impact assessment on the personal data processing that could result in a high risk to the rights and freedoms of individuals. The main focus is on privacy, although the full scope of fundamental rights included in the EU Charter should be considered [11]. However, the GDPR does not explicitly require engagement of potentially impacted stakeholders. This is only required "where appropriate, considering the protection of commercial or public interests or the security of processing operations". In addition, requirements around transparency and communicating the findings of the assessment are lacking. Moreover, in practice, GDPR is rarely used to assess the impact on the full range of human rights [12]. Finally, GDPR allows individuals to be aware they are giving up their data to use a service, but this does not provide equal control over their data.

EU AI Act The proposed EU regulation on harmonized rules on artificial intelligence (EU AI Act) [13] has two principal aims: ensuring safety and respect for fundamental rights when AI systems are used, and stimulating the development and uptake of AI-based technology across all sectors. These become manifest in banning certain AI systems to protect fundamental rights,

[7] Note that the impact dimension is independent from the internal characterization of a certain computational system (eg. rule-based or machine-learning based), making much of the debate about what counts as AI irrelevant. What matters is the coupling that such a system forms with the (human) environment.

as well as imposing a series of due diligence obligations, requiring developers and users of 'high-risk AI systems' to identify and mitigate risks. The EU AI Act also requires establishing and implementing a lifecycle risk management process for high-risk AI systems, along with post-market monitoring that can serve to detect negative impacts after deployment. The proposal would create a process for self-certification of high-risk AI systems, but such self-conducted risk management entails a conflict between the developer's interests and the public interest. However, all of these safeguards are limited for high-risk AI systems only, and there is no clear methodology provided for categorizing risk levels for AI systems by the competent authorities. Another key element missing is stakeholder engagement, as there are no requirements to consult with stakeholders in general, or with those potentially impacted by AI systems.

CSDD The Proposal for a Directive on corporate sustainability due diligence (CSDD) [14] aims to enhance sustainable and responsible corporate behaviour as the addressed private sector actors will be required to identify and, possibly, prevent or mitigate adverse impacts of their activities on human rights. This is a framework relevant for public use of AI because the private sector largely supplies AI systems to public institutions. The proposed norms would apply only to a limited number of large companies, of substantial size and economic power; small and medium enterprises are not within the scope of the proposal. The addressed companies need to integrate due diligence into their policies; identify actual or potential adverse human rights and environmental impacts; prevent or mitigate those; bring to an end or minimize actual impacts; establish and maintain a complaints' procedure; monitor the effectiveness of the due diligence policy and measures; and publicly communicate on findings. However, despite positive process elements, the proposal falls short on the scope of obligation, as it would only apply to very large companies. In addition, the human rights scope covers an incomplete list of potential violations, although is includes a catch-all clause referring to UN instruments, that can be confusing for implementation.[8]

3. Filling the gap: right-based co-governance of AI

Addressing the issues observed above and streamlining the promotion and protection of human rights and democratic principles in AI design and development requires dedicated instruments. The architecture of governance for AI systems in public institutions should: (i) start from a general but clear normative framework, (ii) translate (contextualize) it in a social arrangement *holistically*, ie. by engaging with private and public sectors, as well as potentially affected communities (cf. [15]). Essentially, this structure provides a basis for a multi-layered form of governance: an interplay of international norms, national regulation, co-created technical standardization, frameworks for guidance, as well as spaces for civic oversight, monitoring and engagement (see eg. [16]).

[8] See Annex and para 25 of the proposal [14].

3.1. Legality and legitimacy benchmarks for AI governance

One of the key aspects of any impact assessment, prerequisite for making informative choices in the AI design and development, concerns the metrics of value constructs that are promoted and/or protected. Broadly accepted international human rights *a fortiori* provide benchmarks on what considerations are worthy of protection. However, an additional important feature of the human rights framework is the potential to balance competing principles and impacts in a holistic manner, for the multidimensional stance it takes. Starting from human rights standards, including the existing interpretative doctrine and case law on proportionality and the balancing of rights, we can weight consequences and competing interests of using AI systems. This task requires translating the "three part test" (legality, legitimacy, and proportionality and necessity) of international human rights case law, repurposing it in the AI design context, as follows:

- *Is the intended purpose of an AI allowed and legal?*
- *If yes, is this particular AI system effective in achieving its intended purpose?*
- *If yes, is it proportionate and necessary, considering its human rights impact?*

The three questions can be reframed respectively in the legality, internal legitimacy and external legitimacy categories we outlined in section 2.1. To address these, at architectural level, three fundamental dimensions need to be considered: *scope*, *process* and *participants*, and additional cross-concern requirements.

Scope The scope dimension concerns criteria and indicators that lead and motivate the assessment, in our framing associated to the human rights framework. These criteria are however expressed in a general legal terminology, and need to be operationalized, depending on the context of AI development and deployment.

Process On the process dimension, indeed, the matter is specifying how the criteria and indicators are measured (*assessment process*); however, the procedural meta-level is even more relevant in terms of governance: how to specify the assessment process in itself (*assessment policy process*), and how to determine whether assessment was performed correctly (*assessment oversight process*). Standard, accepted methodologies are required to adequately execute processes at each of the three levels of abstraction. As these require participants, each level activates specific stakeholders.

Participants Criteria are multi-dimensional, and so require multi-disciplinary participants, with diverse expertise and experiences. This implies that the "assessment team" cannot consist only of developers, but should involve (depending on the AI application) legal, ethical, privacy experts, social scientist, as well as representatives from the public and relevant communities. Deciding who is going to participate and how depends instead on a "policy team" that commissioned the development, based on the targeted AI use and balancing competing priorities. In the case of public sector, such team would generally consist of decision-makers within the organization deploying the AI system. The third process level concerns oversight of the assessment outcome and process. In order to be functional, it requires eventually a competent, well-resourced authority. Differently from the common perspective of auditing, which focuses mostly on process, here oversight also requires measuring impacts during AI use, possibly using different metrics and independently from the deploying organization. Furthermore, there needs to be an open channel for the public (citizens, civic organizations, vulnerable groups) to raise concerns about impact, otherwise the feedback loop necessary for external legitimacy would not be closed.

Cross-concerns Cross-concerns along all three dimensions include *transparency* (of decisions, of process, of participants, of assessment outcome) and *due process* (eg. privacy, non-discrimination). All these are key for achieving legitimacy; missing one of the components would make the resulting structure inadequate in systematic terms and pose substantial risks of misuse (see eg. [17]).

4. Computational infrastructure for AI governance

Legality and legitimacy benchmarks do not operate in a technological *vacuum*. Even if we set up an adequate governance structure on the human side of the process, we cannot guarantee that the computational side will adapt accordingly. From what is known about the interactions between policy/legal and IT departments in administrative organizations (and more in general in business–IT alignment contexts), a clear-cut decomposition of concerns is a critical source of problems, due to intrinsic *misalignments* [18]. Typically, people at the design/development level (and even less at the policy level) are not aware of the problems observed at the operations level, while people at the operations level (eg. at the front-office) do not have the power to modify the functioning of the system, even when they recognize that the case they are handling is not being treated properly. If we consider the public as an additional level, the misalignment grows even further. In this context, failures accumulate, and for economic reasons, only the most frequent or critical types of failures will eventually be treated, while the rest will continue to stay unsolved in the long tail of "unfortunate" cases. These mechanisms of "bureaucratic alienation", which all of us experience at some point when dealing with public or private organisations, will only exacerbate by introducing AI components in the loop. As a result, legitimacy concerns on AI use will also quickly increase.

Settling upon a specific governance structure means specifying a series of *intervention points* used to modify, at request, the behaviour of the artificial system. Today, modifications are mostly done manually by developers, translating the directives provided by some stakeholder or actor with legal competence into some programming language, or by users/operators, providing labels useful for training, but this is not a scalable design choice. Even if we set up an adequate panel of stakeholders (experts in ethics, law, policies, community champions) interacting with AI developers, the time from conception to deployment of each update will be plausibly very long. As this was not enough, consequently to changes in the regulatory framework, governance structures may be modified in turn, therefore intervention points in themselves cannot be deemed static. The only way to reduce phenomena like "bureaucratic alienation", and improve the overall process (in terms of scalability, effectiveness, efficiency, ...), is to require a more direct coupling between governance constructs selected at the human side (as the one sketched in the previous section), with automated regulatory mechanisms running on the computational side. In other words, we need to pass from the classic/more common view of algorithms as individual devices/modules to that of a network of algorithms forming a digital social system operating in coupling with a human social systems, ie. to pass from "mechanical" to "institutional" approaches to computation [19]. At a very least, a sound system enabling computational normative control should satisfy the following requirements: (i) it should support the maintenance of a *pluralism* similar to the one occurring in human societies; (ii) any intervention should be by nature *incremental*, reproducing to some extent the constructivist nature of law; (iii) it should enable eg. private parties, intermediate bodies and public authorities to operate in a modular way on the system by means of adequate specifications.

4.1. The need for *continuous governance*

In design cycles concerning processes having legal relevance, selected normative sources or directives are interpreted (eg. by some legal expert) and specified in some computational form (eg. by some AI or software engineer). At the conceptual level, following legal theory (eg. [20]), a normative model consist of directives concerning obligation, prohibition, and permission (*deontic* dimension), and directives concerning who has the *power* to modify some normative relationships or normatively relevant conditions (*potestative* dimension), and various terminological definitions. The process of computational operationalization occurring during the design phase transforms these directives into computational behaviour, together with qualification functions about what currently holds and expectations about what may occur. However, existing technical solutions either leave all these interpretations implicit (eg. going directly from the normative sources and world knowledge to some program), or focus mostly on specifying the deontic dimension (eg. authorization and access-control systems), leaving the potestative dimension implicit or automatically treated by the operational model. On the light of the requirements sketched above, however, enabling a proper treatment of the potestative dimension is of uttermost importance. Power is central to implement governance structures (who has the power to do what, under which circumstances, eg. for what purpose), in particularly for legitimacy purposes. For instance, violation and conflict resolution procedures, distributing checks and balances across various actors, rely on specific instantiations of power. Enabling designers/programmers to make such *social operationalization* step explicit entails to improve the overall transparency of what the various computational components are doing and are meant to do, and provides the means to specify intervention points according to the governance structures agreed on the human side. In other words, bringing power constructs to the foreground at the computational level opens up to *continuous governance*[9] practices and tools (promoting instrumental possibility), ie. the integration at computational level of policies and regulations derived from different sources (legality), and of directives and requirements set by adequately empowered stakeholders (legitimacy).

[9] In software engineering, *continuous integration* (CI) is a term used to indicate practices and tools for automating the integration of code modified incrementally by different contributors into a single software project.

5. Conclusion

The public should take an active role in designing and implementing AI systems to instil values, based on international human rights framework, and ensure that democratic processes and outcomes are realised through its use. Democratic principles, operationalised through good governance and the human rights framework, can definitively serve as core design elements for AI systems in public use, weaving in both legality and legitimacy dimensions. Even if all the steps above function correctly, however, we still lack adequate technologies supporting a continuous alignment between human and computational realms, necessary to avoid informational and operational *cul-de-sacs*, often resulting in 'bureaucratic alienation' scenarios, and consequently erosion of trust and legitimacy. A key point of our contribution is that instrumental possibility requires further attention from policy-makers and researchers and cannot be separated from discussion on legality and legitimacy. The present paper sketched in an organic whole the main challenges, elaborating on possible solutions, at normative and computational level. Future work will further examine and continue elaborating on these aspects.

Acknowledgments Giovanni Sileno was partly funded by the Dutch Research Council (NWO) for the HUMAINER AI project (KIVI.2019.006).

References

[1] W. G. de Sousa, E. R. P. de Melo, P. H. D. S. Bermejo, R. A. S. Farias, A. O. Gomes, How and where is artificial intelligence in the public sector going? a literature review and research agenda, Government Information Quarterly 36 (2019).

[2] J.-M. Bello y Villarino, R. Vijeyarasa, International human rights, artificial intelligence, and the challenge for the pondering state: Time to regulate?, Nordic Journal of Human Rights (2022).

[3] P. D. König, G. Wenzelburger, The legitimacy gap of algorithmic decision-making in the public sector: Why it arises and how to address it, Technology in Society 67 (2021) 101688.

[4] S. Grimmelikhuijsen, A. Meijer, Legitimacy of Algorithmic Decision-Making: Six Threats and the Need for a Calibrated Institutional Response, Perspectives on Public Management and Governance 5 (2022) 232–242.

[5] K. Martin, A. Waldman, Are algorithmic decisions legitimate? the effect of process and outcomes on perceptions of legitimacy of ai decisions, Journal of Business Ethics (2022).

[6] P. Nemitz, Constitutional democracy and technology in the age of artificial intelligence, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 376 (2018).

[7] IEEE Standard Model Process for Addressing Ethical Concerns during System Design, IEEE Std 7000-2021 (2021).

[8] European Commission. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016. URL: CELEX:32016R0679.

[9] G. Ezeani, A. Koene, R. Kumar, N. Santiago, D. Wright, A survey of artificial intelligence risk assessment methodologies, Technical Report, Ernst & Young LLP, 2021.

[10] C. Djeffal, AI, Democracy and the Law, in: The Democratization of Artificial Intelligence: Net Politics in the Era of Learning Algorithms, Verlag, 2020, pp. 255–284.

[11] European Commission. Charter of Fundamental Rights of the European Union, 2012. URL: CELEX:12012P/TXT.

[12] M. E. Kaminski, G. Malgieri, Algorithmic impact assessments under the GDPR: producing multi-layered explanations, International Data Privacy Law 11 (2020) 125–144.

[13] European Commission. Regulation of European Parliament and of the council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts, 2021. URL: CELEX:52021PC0206.

[14] European Commission. Proposal for a Directive of the European Parliament and of the Council on Corporate Sustainability Due Diligence and amending Directive (EU) 2019/1937, 2022. URL: CELEX:52022PC0071.

[15] Ministry of Economy, Trade and Industry of Japan (METI). Governance Innovation: Redesigning Law and Architecture for Society 5.0, 2020.

[16] A. S. Gill, S. Germann, Conceptual and normative approaches to ai governance for a global digital ecosystem supportive of the un sustainable development goals (sdgs), AI and Ethics 2 (2022) 293–301.

[17] B. Nonnecke, P. Dawson, Human Rights Implications of Algorithmic Impact Assessments: Priority Considerations to Guide Effective Development and Use, Technical Report, Carr Center for Human Rights Policy Harvard Kennedy School, Harvard University, 2021.

[18] A. Boer, T. van Engers, Agile: a problem-based model of regulatory policy making, Artificial Intelligence and Law 21 (2013) 399–423.

[19] G. Sileno, Of duels, trials and simplifying systems, European Journal of Risk Regulation 11 (2020) 683–692.

[20] G. Sartor, Fundamental legal concepts: A formal and teleological characterisation, Artificial Intelligence and Law 14 (2006) 101–142.