

Winter December 10, 2019

Natural Law and its Implications for AI Governance

Yueh-Hsuan Weng, *Tohoku University*
Takashi Izumo, *Asahi University*

Natural Law and its Implications for AI Governance

Yueh-Hsuan Weng and Takashi Izumo*

With the recent emergences of AI technologies, our societies are facing regulatory challenges in terms of their design, manufacture, sale and use. In addition to the existing norms, many new 'AI laws' will be needed for early stage AI governance. However, when it comes to AI, there is a significant gap between hard laws and soft laws. Although we have witnessed the development of soft law from both public institutions and organisations like the EU and the IEEE in recent years, hard law has been less forthcoming. Answering the question of why this gap exists and whether or not 'natural law' can narrow is the chief purpose of this paper. To do so we will draw on two supplemental principles from the natural law tradition.

I. Introduction

When it comes to effectively governing emerging AI technologies in the short run, a critical issue relates to the establishing relevant norms. Drafting hard law to regulate unknown new technologies is difficult when the public believe that the 'law shall always fall behind the technology'.¹ Furthermore, regulators usually have a tendency to over-regulate emerging technologies because of their lack of knowledge in a particular technological field. One famous example is the UK's 1861 Locomotive Act or 'Red Flag Laws', which is now regarded as an excessively stringent regulation of what was then an emerging technology; the steam engine powered vehicles.² In addition, when we take a legal viewpoint we can easily analyse conflicts between existing AI laws, but it is more difficult to imagine legal issues emerging from future laws for technologies which do not, as of yet, exist.

On the other hand, soft law for AI governance, such as ethical principles or legislative guidelines, has become much more prominent in recent years. Let us take Europe as an example. In 2014 the EU introduced guidelines for regulating robotics, the so called FP7 Robolaw Project.³ In 2016 the EU Parliament published the Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics⁴ and in 2019 the AI HLEG, an expert group set up by the EU Commission, published the Ethics Guidelines for Trustworthy Artificial Intelligence (AI).⁵ There have also been many soft law initiatives outside of Europe, such as the IEEE's Ethically Aligned Design (v1, v2, and e1),⁶ the Future of Life's Asilomar 23 AI Principles,⁷ and the Japanese Society for Artificial Intelligence's Ethical Guidelines.⁸

In terms of the regime of positive law, there is a clear distinction between hard law and soft law. For example, Abbot and Snidal define soft law as any law

* Yueh-Hsuan Weng. Assistant Professor, FRIS, Tohoku University, Japan. Visiting Scientist, RIKEN-AIP, Japan.; Dr. jur. Takashi, Izumo. Associate Professor, Faculty of Law, Asahi University, Japan. For correspondence: <y.weng@srd.mech.tohoku.ac.jp>

1 Mark Fenwick, Wulf Kaal and Erik Vermeulen, 'Regulation Tomorrow: What Happens When Technology Is Faster than the Law?' (2017) 6 American University Business Law Review 3

2 Red Flag Laws were regulations made by the UK Parliament for the use of steam engine powered vehicles on public highways in the 19th century. <<https://www.legislation.gov.uk/ukpga/Vict/24-25/70/contents>> accessed 05 September 2019.

3 Erica Palmerini et al, 'Guidelines for Regulating Robotics' (2014) EU FP7 RoboLaw Project

4 Mady Delvaux-Stehres et al, 'Recommendations to the Commission on Civil Law Rules on Robotics, European Parliament' (2016) Draft Report

5 The AI HLEG, 'The Ethics Guidelines for Trustworthy Artificial Intelligence' (2019) European Commission

6 The IEEE A/IS Initiative, 'Ethically Aligned Design (EAD)' <<https://ethicsinaction.ieee.org>> accessed 5 September 2019

7 Future of Life Institute, 'Asilomar 23 AI Principles' <<https://future-of-life.org/ai-principles>> accessed 5 September 2019

8 The Japanese Society for Artificial Intelligence, 'The JSAI Ethical Guidelines' <<http://www.ai-elsi.org/wp-content/uploads/2017/05/JSai-Ethical-Guidelines-1.pdf>> accessed 5 September 2019

that lacks one or more of the three critical features of hard law: obligation, precision, and delegation.⁹ These soft laws are not only created in a flexible way, but are also difficult to be implemented. With this definition in mind, we can define soft laws in AI governance ('AI soft laws') as laws that meet the following three criteria: (1) Not formally binding (ie in regard to robot developers, vendors, users or other stakeholders) (2) Not precise, so that the judgment of the norm implementation is left to each party, and (3) Merely private, so that no third party has the authority to monitor implementation. In other words, AI soft laws is a set of weakened rules that help the parties to self-regulate their research and commercial activity on. However, the relative weakness of soft law makes it difficult to handle conflict between different values and norms. When testing for the real world, it is sometimes not clear on how to decide the hierarchy between current existing hard laws and a universal moral principle from AI ethics without a meta guiding principle. If we cannot ensure a corresponding rise in hard law to an explosion in soft law in this area, are there any other alternative way to cover the gap of AI soft laws in real governance as mentioned above?

II. Applying Natural Law as a Regulator in Early-Stage AI Governance

The classification of hard law and soft law is straightforward to modern governance system based on positive law. However, a boundary distinguished from 'hard' and 'soft' seems not clear to the other perspective called 'natural law' which is a kind of moral theory in defining legal authority in relation to different norms. As an alternative solution to above mentioned gap of governance in AI soft laws, we propose to consider natural law as a regulator in early-stage AI governance. In other words, we assume that the natural law traditions can be used as supplementary principles for soft laws in dealing with some difficult situations like to handle conflict between different values and norms. But sometimes it can also be ambiguous and difficult for the layman to understand what the concept of natural law is. Hence, at the beginning we want to clarify the exact scope of natural law and its connection with the context of AI soft laws.

As the concept of natural law itself has evolved over a long history, it is hard to find a uniformed de-

finition that can be applied to all natural law theories. Thus, in this study, we will focus only on several jurists and philosophers who had particular influence in the early-modern period and classify the concept of natural law according to their writings as mentioned below. The first criterion comes from a fundamental distinction between the fact and morality, ie between the being and ought-to-be. There are some laws that were regarded as a part of natural law in the 17th and 18th century but should be excluded from our study because they are treated by economists or natural scientists today. Such laws can be subdivided to the laws that are applied also today and the others that are false according to modern science. An example of the former laws is the law of supply and demand. Christian Thomasius (1655 - 1728), who is called 'Father of German enlightenment' (in German: Vater der deutschen Aufklärung),¹⁰ found that the price of merchandise depends on its supply and demand.¹¹ This idea was formulated independently by Alfred Marchall in the 19th century as so-called supply and demand.¹² In other words, his concept of natural law was not restricted to the so-called human morality, but it included also economic analysis. By contrast, Thomasius had a false prejudice and called it natural law by stating that humans are permitted to eat meat because of his belief that no other animal apart from human beings would have sense.¹³ This prejudice derived, as he named explicitly in another place,¹⁴ from the thought of the famous French philosopher Rene Descartes (1596 - 1650) that a dog, a cat, or a bird has no spirit

9 Kenneth Abbott and Duncan Snidal, 'Hard and Soft Law in International Governance' (2000) 54 International Organisation 3, 421 - 422

10 About the question of how the name of 'Father of German Enlightenment' (in German: Vater der deutschen Aufklärung) was given to Thomasius, see Markus Meumann, Diskursive Formationen zwischen Esoterik, Pietismus und Aufklärung: Halle um 1700, in Monika Neugebauer-Wölk (ed), *Aufklärung und Esoterik: Rezeption - Integration - Konfrontation* (Walter de Gruyter 2009) 78

11 Christian Thomasius, *De aequitate cerebri: l. 2. C. de rescind. vendit. et ejus usu practico* (Johann Christian Hendel 1749) 52

12 The Law of Supply and Demand 'states that in a free market the forces of supply and demand generally push the price toward the level at which quantity supplied and quantity demanded are equal'. W. Baumol et al, *Microeconomics: Principles and Policy* (11th edn, Cengage Learning 2008) 66

13 Christian Thomasius, *Institutiones jurisprudentiae divinae* (7th edn, Christoph Salfeld 1730) 184

14 *ibid* 8

Table 1. Source: Authors' elaboration. The traditional concept of 'natural law' can be classified as follows:

World	Modern discipline	Subject	Strength
Natural world in the narrow sense	Natural Science	Factual rules of physical entities	Neither hard nor soft because these laws cannot be violated
Social world concerned with facts	Sociology and economics	Factual rules of human society	Normally inviolable but social customs are changeable
Social world concerned with morality	Jurisprudence	Moral rules of human society which can be sued in court	Normally hard but sometimes soft
	Policy Making	Interdisciplinary field of discussions on moral rules of human society	From hard to soft Hard: eg official guideline Soft: eg recommendation
	Ethics	Moral rules of human society which cannot be sued in court	Soft
Religious world	Theology	Conscience = moral rules of human society which is controlled by God	Religiously hard but secularly soft

and thus are merely 'automata'.¹⁵ Today this theory has been denied by biologists and is often criticised in the context of animal rights.¹⁶ These laws, or 'natural law developed with the concern of facts', will be excluded from discussion in this paper regardless of the question of whether they are true or false, because they are controlled by natural force. In other words, we have no choice whether or not to follow natural science.

If the context world of natural law is one which is developed with a concern for morality, then we should refer not to natural science but to jurisprudence and ethical norms. Furthermore, jurisprudence and ethical norms should be distinguished from each other in regard to the existing problems we described in the early stages of AI governance.

Hence, the scope of natural law traditions we will investigate in this paper will tentatively exclude its connection with hard law. However, the concept of natural law we will investigate, namely the one at the level of ethical norms, should be further sub-classified into three groups, (1) conscience, or people's inner voice, (2) legislative policy and (3) ethics in the narrow sense.

First of all, the conscience can be regarded as one recognition method of natural law given the assumption that God will punish a person who committed a crime against his or her conscience, even if a court overlooked the crime.¹⁷ The assumption that God is omniscient and omnipotent is useful when it is not clear whether a Christian committed a crime. One of the merits of this belief is the restraint of sanctions ie the church entrusts ambiguous cases to God. Therefore religious natural law, which is ordered by God, is indeed ethical because when it is violated it is not punished in this world, but in the background. On these grounds we should also exclude the conscience in this paper.

Finally, it is clear that legislative policy, and ethics in the narrow sense remain as the subject of our study.

¹⁵ Norman Smith, *New Studies in the Philosophy of Descartes* (Macmillan 1952) 136, 140

¹⁶ Tom Regan and Peter Singer, *Animal Rights and Human Obligations* (2nd edn, Prentice Hall 1989) 13-19

¹⁷ For example, in *Summa theologiae*, Part 1-2, Question 87, art 1, Thomas von Aquin wrote that there are three sorts of punishment, ie punishment from oneself, from other people, and from God.

Legislative policy is different from ethics in the narrow sense in that it suggests policies which are enforceable in future. In other words, natural law includes some rules which should be incorporated into positive laws by legislation even if not everyone agrees on doing so¹⁸ (one pertinent example is the question of whether or not old people should return their car licenses once their driving skills deteriorate). By contrast there is no such enforceable obligation in the case of ethics in the narrow sense (eg people should return greeting to each other).

III. How Natural Law Can Support AI Soft Laws?

From a natural law viewpoint, the concept of AI soft law at the very least should include legislative policy (eg the draft report from European Parliament) and ethics in the narrow sense (eg the Future of Life's Asilomar 23 AI Principles). However, some AI soft laws like the IEEE's 'Ethically Aligned Design' have dual characteristics in that they are not merely an ethical guideline, but also include law and policy sections for legislators to use in developing laws related to AI governance. The question, then, is how natural law can support AI soft law in the early stage AI governance? To answer this, we found two supplemental principles from the natural law tradition, namely, shareability and complementarity. First, the concept of shareability means that natural law must be acceptable to all nations as a common rule regardless of their cultural background. This feature was required by Hugo Grotius by stating that international customary law that he called *ius gentium* has also high shareability but only natural law is perfectly shareable and can be accepted to all nations.¹⁹ Specifically, AI soft law must become acceptable to all parties and stakeholders. Secondly, complementarity is the notion that multiple rules should support each other because it is impossible for a single rule to regulate everything flawlessly. This second requirement was suggested also by Grotius, and he wrote that there are multi kinds of law in this world, namely, natural law, international customary law, which was called *jus gentium*, civil law, and divine law, which God revealed. Grotius believed that if a rule of natural law is contrary to another rule of civil law, then the latter should be applied, in other words, in his legal system, natural law is regularly soft whereas

civil law is a sort of hard law. In the context of AI governance, this concept requires us to create a situation where AI soft laws does not violate hard law and does not disturb private business among citizens.

1. Shareability: A Universal Principle and a Golden Rule

In regard to shareability, we immediately face the problem that it is difficult to find a common rule that can be accepted by all reasonable persons regardless of their cultural background. And indeed it seems, at first glance, that this is also true when it comes to human-robot co-existence. However, if we look to legal history, there was a rule that was globally accepted, ie Golden Rule (hereinafter the GR).²⁰ Thomas Hobbes, one of the fathers of modern natural law theory, argued that every conclusion of natural law can be intuitively drawn from the GR.²¹ He believed that natural law must be understood by all humans without difficulty and must consist of a simple normative proposition. In short, the GR, according to Hobbes, is a law based on human nature, which by definition all humans have in common regardless of intellectual ability or cultural background.

The GR has at least two different forms, ie 'treat others as you would like others to treat you' (positive form) and 'do not treat others in ways that you would not like to be treated' (negative form). Its existence is not merely limited in Western traditions, but the GR also plays an important role in Eastern societies, such as Confucian from ancient China.²² Scholars have argued for a long time about the difficult question of whether the GR, especially in the positive

18 Takashi Izumo, *Die Gesetzgebungslehre im Bereich des Privatrechts bei Christian Thomasius* (Peter Lang Verlag 2016) 54-58

19 Hugo Grotius, *De jure belli ac pacis* (Joan Blaeu 1660) 7

20 In regard to the popularity of the GR, see Harry J Gensler, *Ethics and the Golden Rule* (Rout 2013) 34 seqq

21 Thomas Hobbes, *Leviathan: Reprinted from the Edition of 1651* (Oxford University Press 1929) 121

22 Fumihiko Takahashi, 'The Confucian Golden Rule: Chu Hsi's Neo-Confucian Interpretation and the Critical Arguments by Japanese Confucianists in the Seventeenth and Eighteenth Centuries', in B. Sharon Byrd, Joachim Hruschka und Jan C. Joerden (eds), *Jahrbuch für Recht und Ethik, Band 8 (2000), Themenschwerpunkt: Die Entstehung und Entwicklung der Moralwissenschaften im 17. und 18. Jahrhundert* (Duncker & Humblot 2000) 315 seqq

form, really leads to shareable conclusions. Thomasius convincingly argued that everyone must agree to the following negative form of GR: ‘You ought not to harm others because you do not want to be harmed’ (in original Latin: *Quod tibi non vis fieri, alteri ne feceris*; translated in English by T. Izumo).²³ In other words, he believed that the prohibition of harm is a universal rule for human beings.

Surprisingly this old wisdom is consistent with the following key principle within the ‘Ethics Guidelines for Trustworthy Artificial Intelligence’: namely ‘Develop, deploy and use AI systems in a way that adheres to the ethical principles of: respect for human autonomy, *prevention of harm*, fairness and explicability’.²⁴

The GR also partially coincides with the first rule of Asimov’s ‘Three Laws of Robotics’. This first rule says that ‘a robot may not injure a human being or, through inaction, allow a human being to come to harm’. Connected with the GR, the first law needs to be extended and to be rewritten as follows: A robot may injure neither human beings nor other entities nor, through inaction, allow them to come to harm. This extended version of the first rule is a minimum rule of AI and robot regulation and it mandates that AI and robots ought to maintain the safety of humans and human property (thus indirectly also the safety of AI and robots themselves as human property). In order to satisfy this order, various measures must be taken. These can be divided into three groups, ie damage prevention (eg setting of fire in prohibited areas), damage containment (eg firefighting equipment) and damage compensation (eg fire insurance).

2. Complementarity: Harmonisation of Multiple Norms in Governance

New regulations for ensuring the safety of human-robot interactions will be needed when service robots that perform close interactions with humans become more popular in our society. However, current safety regulations for industrial robots, based on the ‘human-robot separation’ safety policy, contradict the original purpose of service robots.²⁵ Although the first AI soft law to ensure the safety of service robots has emerged in the form of the 2014 ISO 13482,²⁶ the law is not sufficient to regulate tasks involving more complex daily human-robot interactions. Therefore, the Japanese Ministry of Trade and Economy (METI) developed another new standard in July 2019 (JIS Y1001) which aims to act as supplemental measure with a focus on requirements for safe management and operations of robot service providers.²⁷ From this, we can see a trend that the type and content of AI soft laws is going to keep growing. More importantly, the larger issue is about the harmonisation of multiple norms in governance or as we call it: ‘complementarity’.

At present in this very early stage of human-robot co-existence, we do not yet have ‘AI hard laws’. However, we do have hard laws which already impact the AI field including civil law, criminal law, data protection law. These can be labelled ‘non-AI hard laws’. Problems arise in norm complementarity when the implementation of the GR conflicts with these non-AI hard laws. How should we address this?

As mentioned before, the GR is a universal moral principle from the natural law tradition and future AI legislation should respect the GR. Therefore, conflicts between the GR and AI hard laws should not happen. The GR can be regarded as a kind of meta rule for both AI hard law and soft law, but it does not mean that the GR has the same moral force of constraint on non-AI hard laws. Because those currently existing hard laws have their own values, AI governance might not be their priority. The importance of solving this norm conflict in the near future is about a unique regulatory sandbox or ‘Tokku’ for intelligent robots and safety critical autonomous systems. The ‘Tokku’ special zone for robotics empirical testing and development (RT special zone) originated in Japan, which is not only a shock buffer for emerging technology and society, but also a deregulation system for AI’s last mile testing into the real world.²⁸ Hence, conflicts between the GR and non-

23 Christian Thomasius, *Fundamenta juris naturae et gentium* (4th edn, Christoph Salfeld 1718) 183-184

24 European Commission, *Ethics Guidelines for Trustworthy AI* <<https://ec.europa.eu/futurium/en/ai-alliance-consultation>> accessed 16 August 2019

25 Yueh-Hsuan Weng, ‘Robot Law 1.0: On Social System Design for Artificial Intelligence’ in Woodrow Barfield, Ugo Pagallo et al (eds), *Research Handbook of Law and Artificial Intelligence* (Edward Elgar Publishing 2018)

26 Theo Jacobs, Gurvinder Singh Virk, ‘ISO 13482 – The New safety Standard for Service Robots’, *ISR/Robotik* 2014; 41st International Symposium on Robotics, 2014

27 New JIS as Safety Standards for Robot Services Established, see https://www.meti.go.jp/english/press/2019/0520_003.html

28 Yueh-Hsuan Weng, Yusuke Sugahara, Kenji Hashimoto, Atsuo Takanishi, ‘Intersection of “Tokku” Special Zone, Robots, and the Law: A Case Study on Legal Impacts to Humanoid Robots’ (2015) 7 *International Journal of Social Robotics* 5, 841-857

AI hard laws are more likely to happen during testing scenarios inside the special zone. One alternative solution for the potential conflicts between the GR and non-AI hard laws is to reserve some core value issues in legislative process as an upper meta rule for the early stage transition period.²⁹

Even if we ensure no conflict with current existing hard laws, some original rules or so called 'expedient measures' may come up during the process of rule implementation in governance. In regard to AI soft law a top-down institutional approach should also be avoided and a societal consensus should be sought. For example, even if no hard law enforces self-driving car owners to buy property damage liability insurance, it should not be forbidden for parking lot owners to provide the clause that self-driving cars are not insured in their lots. If such a clause becomes popular in a community, then the insurance would function as a kind of soft law. In this situation, the concept of 'complementarity' is required for the validity of AI soft laws.

Such complementarity was required in early-modern natural law theories because natural law could not cover all political or economic issues and it was necessary to mandate legislators in each country to make appropriate but not exhaustive positive laws.³⁰ However, there are still several fields where the role of hard law would become so important that soft law would not have space to be applied. In addition, the liability immunity clause that exempts the responsibility of robot developers and vendors should only be restricted by hard law, as in the case of product liability in the United States. Otherwise it might cause unfair outcomes for robot users.

There are also some overlaps between shareability and complementarity in Roman law, where the principle of complementarity can be a supportive tool in the implementation of the GR. Here the most important principle is that humans should be treated as free and equal by nature. Justinian the First, who was an Eastern Roman emperor and reigned from 527 to 565, confirmed that humans are free by nature and the slavery has been artificially introduced.³¹ This contradicts the Aristotelian believe that there are inferior people who suitable for slavery by their own nature.³² The Justinian claim can be easily justified by the GR, ie the explanation of Aristotle is not shareable because it cannot be accepted by the people who are regarded as inferior. No one thinks that he or she should serve others as a slave because

he or she wants to be served by others as slaves. In addition, this natural law satisfies the requirement of the complementarity because natural freedom can be restricted by positive laws and natural equality can be changed too. For example, in ancient Rome, Roman citizens enjoyed more freedom than foreigners in several fields.³³ Ulpian, who was one of the most famous Roman jurists, explained it clearly as follows: 'The *jus civile* is that which neither wholly diverges from the *jus naturale* and *jus gentium* nor follows the same in every particular. And so whenever to common law we add anything or take anything away from it, we make a law special to ourselves, that is, *jus civile*, civil law'.³⁴ This overlaying of multi-rules is applicable to modern society too. One possible application is that we will distinguish the professional residents who are involved with robotics from the others and restrict the freedom of the latter group (eg by setting up restricted areas) or give more freedom to the former (eg by giving permission to develop some technologies that are normally prohibited).

IV. Conclusion

The 'Regulation of the Unknown' is a kind of general problem for regulators when they are faced with emerging technologies like AI and robotics. It reflects the importance of the role of soft laws in AI governance at its initial stage. In this paper, we investigated the connection between the natural law tradition and AI soft laws with a focus on two supplemental

29 Yueh-Hsuan Weng, 'Beyond Robot Ethics: On a Legislative Consortium for Social Robotics, Advanced Robotics' (2010) 24 *Advanced Robotics* 13, 1919-1926

30 For example, Hugo Grotius and Samuel von Pufendorf thought that price control should be left to the judgment of each legislator or judge even though natural law requests it in several cases. Grotius (n 19) vol 2, cap 12, § 26, 237; Samuel von Pufendorf, *De jure naturae et gentium*, vol 5, cap 3, § 9 (Adam Junghans 1672) 630

31 Rolf Knütel et al, *Corpus Iuris Civilis: Die Institutionen: Text und Übersetzung* (4th edn, C.F. Müller 2013) 5-6

32 For an overview of Aristotle's defense of the slavery in ancient Greek see Fred Miller, 'Aristotle's Political Theory' in *The Stanford Encyclopedia of Philosophy* (First published Wed Jul 1, 1998; substantive revision Tue Nov 7, 2017) <<https://plato.stanford.edu/entries/aristotle-politics/>> accessed 23 August 2019.

33 Max Kaser, Rolf Knütel and Sebastian Lohsse, *Römisches Privatrecht: Ein Studienbuch* (21th edn, C.H. Beck 2017) 94

34 D 1.1.6 pr ('D' is the abbreviation of 'Digesta', 'pr' is 'principium') translated by Alan Watson, *The Digest of Justinian*, vol 1 (University of Pennsylvania Press 1998) 2

principles: shareability and complementarity. We found that the shareability principle is important for easing value conflicts between different AI soft laws when they are applied to the same target. On the other hand, the complementarity principle can help to solve conflicts between different kinds of norms in

terms of AI governance. Finally, we demonstrated the feasibility of natural law to be a policy tool to support AI governance in the early stage (for example to deal with potential conflicts happening in special zone deregulation systems for testing safety critical AI and autonomous products).