Influenceable Autonomy and Predictable Freedom in the IoE - A Research Proposal

Maximilian Gartner⋆[0000−0001−9601−122X]

University of Bologna, Bologna 40126, ITA

Abstract. The following paper outlines the doctoral research proposal that was presented at the 2019JURIX Doctoral Consortium. The undertaken dissertation aims to be a treatise exploring the intersection of autonomy and freedom of individuals with increasingly capable and autonomous (multi-agent) systems. The project is meant to encompass an interdisciplinary analysis of the concept of autonomy and areas of influence which agents of the IoE can be expected to exercise. Ultimately, a framework to explain exercise autonomy of individuals within a system constrained by agents of the IoE will be established.

Keywords: Autonomy · Freedom · AI Governance.

1 Introduction

The undertaken dissertation aims to be a treatise exploring the intersection of autonomy and freedom of individuals with increasingly capable and autonomous (multi-agent) systems. The project is meant to encompass an interdisciplinary analysis of the concept of autonomy and areas of influence which agents of the IoE can be expected to exercise. Ultimately, a framework to explain exercise autonomy of individuals within a system constrained by agents of the IoE will be established. Coming from both the increasing use of such technology and the increasing capability of the technology so deployed, thorough analysis of the impact of agents on human autonomy and freedom is vital. This is exacerbated by the high value society tends to attest to these principles. As this currently remains an emerging issue, legal and policy research on this topic remains at the frontier of understanding the defining challenges of society in the coming decades.

2 Problem Description and State of the Art

2.1 General Remarks

This work proposes an inquiry into the concepts of freedom and autonomy as they are aided and undermined by characteristics of and agents within the Internet of Everything. The word autonomy is derived from the ancient greek terms

"auto" (meaning "self") and "nomos (meaning rule or law). This term is used colloquially and therefore comes with considerable opaqueness; it can arguably be used to denote the characteristic of free will, the capacity or action of self-control or control over one's life, the state of freedom of coercion, or a right to or a value based on the above; all of which are dogmatically different.[1]

2.2   Law from Ethics from Morals from Ends - Tracing the concept of Autonomy

For the purpose of this inquiry, namely the analysis of autonomy as a target of influence and prediction, recourse will be had to three sources of meaning: (1) the general scholastic consensus and relevant theories in the field of philosophy and epistemology with respect to autonomy [25, 6][2], (2) moral and ethical frameworks to autonomy [25][3], and (3) legal frameworks, specifically human rights regimes, pertaining to autonomy.

   The reasoning for this is best explained when working backwards from the concrete and factual to the abstract. Legal frameworks, in short laws, structure the relations between people and other entities and objects, (e.g. other people, animals, material and immaterial objects, rights, etc.) of the system (i.e. society, the world) the people are acting in. By imposing enforceable rules and non-enforceable but authoritative endorsements, the legislative process steers individual system-members' behavior. Compliance is enforced through governmental sanctions or incentives. These legal boundaries are set out to satisfy certain societal functions that are deemed important based on or at least influenced by a set of values adopted by some part of the society by the legislative entity [24][4]. Not accounting for the ephemeral concept of natural law, the legal domain ends here [11, 16].[5] In other words, law is an inherently neutral set of tool, that is

---

[1] The term autonomy can also be charged with (unambiguous) meaning in other specialized domains, such as computer science, robotics, etc. This section deals with autonomy as pertaining to human agents in a broad setting.

[2] The terms "ethics" and "morals" are not interchangeable. Ethics are value systems of a certain group. Morality is then a subset of such ethics, namely a value system dealing first and foremost with the notions of "right" and "wrong". However, this statement is undermined by the fact that the term ethics is commonly (and colloquially) used interchangeably with morality.

[3] The terms "ethics" and "morals" are not interchangeable. Ethics are value systems of a certain group. Morality is then a subset of such ethics, namely a value system dealing first and foremost with the notions of "right" and "wrong".

[4] Indeed, the social functions of law might be so closely interwoven with the underlying moral or political imperatives that they are of "no use to anyone who does not completely and exclusively endorse them", according to Raz.

[5] A legal exclusive positivist position is assumed here. More precisely, it is assumed that morality does not affect validity of legal frameworks while it is acknowledged that morality factually influences both the creation and implementation of legal frameworks. In contrast, an inclusive positivist position would allow for invalidation of a legal framework based on moral considerations while a natural law position

structured along the lines of an underlying value system, to understand its purpose and enactment an analytical, teleological approach with knowledge of this underlying system is helpful.

Normative, i.e. rule-giving, frameworks which have not been legitimized by a legislative process are (pseudo-) ethical frameworks [21].[6] Compliance can be achieved through non-governmental means such as social or cultural pressure from members of the group applying the ethical framework and this compliance can be as widespread as compliance with a legal rule. Ethical frameworks can exceed the scope of a legal frameworks (i.e. a behavior is legal but unethical) or can fall short of it (i.e. a behavior is illegal but ethical). In the first case, the ethical framework fulfills a guiding function between multiple legal behavioral options, or might be used to reason about situations that are plainly not covered by law yet. In the second case, the legal framework is often characterized to be unfair and/or faulty. In both cases, if the divergence between the law and the underlying ethics is stark enough a societal pressure is likely to grow, potentially leading to changes in the legal framework. A different approach of conceptualizing this is the notion that within a democracy the legitimacy and content of a legal system originates from its people; because these people are not ethically blank prior to the establishment of a legal system, the legal system will inevitably be infused with the ethical system of the people. At the same times, ethical frameworks are not uniform. Ethical systems, such as cultural or religious frameworks, can contradict each other. The same is true for ethical systems in the narrow sense (that is explicitly ethical frameworks), which is evidenced not only by the fact of the sheer multitude and variance but also by the meta-ethical disputes about their general viability and validity[6, 26].[7] In order to be effec-

---

would require alignment with some sort of moral standard in order to be considered valid.

[6] The distinction between ethics, religion, social conventions, cultural practices, etc. is a question of definition. When accepting all of these value systems as "ethics", more weight is put on the underlying question of morality. This latter option is appealing because it coincides with wide usage of the term; for example there is little value in labeling professional ethic-standards as pseudo-ethical. Consequentially, one must allow that professional ethic-standards might imply immoral rules, potentially making them an "immoral ethic framework". This view is contested, see for just one account Paul and Elder, but will be utilized here.

[7] Difficult questions arise, when trying to assess the existence of morality by itself and moral facts, the capability to address moral facts by moral statements and if or to what extent such is accessible through human inquiry at all. Generally, the meta-validity of ethical systems can be considered mostly with respect to (1) metaphysical and (2) epistemological positions. When it comes to metaphysical inquiry, e.g. nihilism/error theory maintain that there are no moral facts on which to base an ethical theory on, while subjectivism proclaims any moral fact to be inherently subjective. Within an epistemological inquiry, non-cognitivism characterizes all ethical concepts all expressions of an emotional state rather than an object of knowledge and as such without objective ethical/moral value. Of course even when finding that moral facts on which to base an ethical system exist, there is no universal consensus neither on their content nor on how to find or approximate them.

tive in effectuating legal frameworks, an ethical framework must either be widely adopted or intellectually appealing, be it through internal logic, through offering a solution to a contemporary problem or through prohibitive cognitive costs for not abiding by it (e.g. deeply rooted beliefs, disavowal of which conflicts directly with the identity of the agent).[8] To assess the flow of imperatives between these systems, consideration needs to be given to some of the underlying values of the ethical approach propagated; the proposition being that certain values will give stronger (factual or moral) justification to the ethical systems that depend on them. In other words, it is proposed here that the (factual) persuasiveness of underlying values determines the (factual) adoption of ethical frameworks that build upon them.

Persuasion aiding the ethical system is rooted in the values of the system itself, specifically in values that determine if an action is "right" or "wrong". Such underlying systems of values and corresponding principles concerned with (absolute) rights or wrongs are moral frameworks. How exactly these values derive authority (and subsequently grant authority to ethical frameworks based upon them) is arguably one of the core questions of the scientific discipline of ethics. Explanation attempts are broadly of two camps: deontological/duty-based and teleological/end-based assessment of an action. With the former, an action is right if it is compliant with rules that have a self-evident character (i.e. Intuitionism) or supernatural endorsement (i.e. Divine Command Theory), legitimation through a (hypothetical) form of social cooperation that would lead to their adoption/finding, i.e. through the fairness of such a process (i.e. Contractarianism) or that they stem from pure reason (i.e. Formalism). With the latter, an action is right when it promotes the right end and it is the best action to promote said end [6]. While only with the latter, the question of right or wrong is fully subordinate to a question of ends, these overarching moral principles still impact the framework of the other approaches [7, 23, 12].[9]

---

[8] This distinction should not be construed to hide that intellectually appealing ethical frameworks are naturally suited to be or become widespread. However, it is conceivable that certain ethical guidelines might be transplanted into legal frameworks before they become widespread, or even despite the fact that they deal with a subset of situations that is just not considered widely enough to enter the general ethical discourse at all.

[9] From a contractarian viewpoint, an action is right or wrong depending on its compliance with a ruleset that is or would have been constructed through a fair process of people living together in fellowship and treating each other as equals. However, in that "constitutional" drafting process, the rules agreed upon will likely reflect the ends (e.g. happiness or pleasure of the individual or the collective). From a divine command theory-viewpoint, not only are actions wrong that violate the rules of the authoritative supernatural entity, but the supernatural entity is in itself the origin of ends; it seems to follows that relevant ends are congruent with (at least compliance with) the will or aim of the supernatural entity. From the viewpoint of formalism, pure reason leads to moral principles that satisfy the condition of being a universal law; all this to ensure that the moral agents are all individually an end in itself. Similar connections can be made for other ethical approaches.

The integrity of the research methodology should have now become clear. Personal Freedom and autonomy is inarguably considered and protected by legal frameworks on national and international level. This protection is reflecting a corresponding call to do so by ethical frameworks. The ethical frameworks in question are factually relevant due to the persuasiveness of their underlying moral principles, which mandate that actions undermining autonomy and personal freedom as morally wrong and protecting and fostering autonomy as morally right. Depending on the approach, this moral framework is more or less directly originating due to an end of which autonomy or freedom is a necessary or helpful requirement, feature or consequence. The validity of this approach, if not verifiable, can be made plausible by observing the prevalence of rules that pertain to autonomy and freedom in legal, ethical and moral frameworks.

This is expressively confirmed here: Many rules of international law, having achieved nearly universal status, address the issue of private autonomy of the individual. Inter alia, Article 19(1) of the ICCPR and Article 19 of the UDHR stipulate the freedom to hold opinions without interference as an absolute right. Article 8, 9 and 10 of the ECHR similarly set out the rule of freedom of thought and expression as well as broad protection of an individual's private life[5].[10]

Protection of autonomy and personal freedom against undue interference by technology is a relatively new field of (legal) inquiry; it follows that concrete legal countermeasures are lacking. However, the arising conflicts can be located within the protective scopes of existing human right frameworks; data-driven targeting of individuals might collide mainly with (1) the right to privacy and more specifically data protection, (2) the right to freedom of thought, conscience, religion and belief , or (3) the right to freedom of opinion, expression and information. (Connecting the concepts of privacy and autonomy is the notion that the right to privacy encompasses (partly) the concept of human autonomy, and concerns itself specifically with the human being as autonomous subject; in other words individual autonomy which its existence and field of actions do not touch upon the sphere of liberty of others is considered privacy by legal scholars[19]. To illustrate, the right to privacy (as a prerequesite for protection of human autonomy) is enshrined in inter alia Article 12 of the Universal Declaration of Human Rights, Article 17 of the International Covenant on Civil and Political Rights, Article 16 of the United Nations Convention on the Rights of the Child, Art. 14 of the International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families, Article 8 of the European Convention on Human Rights, Article 7 and 8 of the Charter of Fundamental Rights of the European Union and the Article 11 of the American Convention on Human Rights.)

2.3   Interconnectivity as a Catalyst for Paradigm Change

Within the last decades, society has undergone significant changes due to the increasing prevalence and importance of the Internet. Due to the nature of this

---

[10] The ECHR has found that the terminus "private life" is not accessible to exhaustive definition, and therefore generally also covers the psychological integrity of a person.

emerging domain, interactions that have traditionally taken place between humans now often occur with or are influenced by non-human agents. Algorithms and similar agents not only directly interact with human users but can also select and curate the information most humans will reasonably access. Selection and curation of information for a target audience is not a new phenomenon, however due to the scalability of the underlying employed technology, automated processing individualized for a single human exacerbates the potential impact of such endeavors. In addition, such selection is not always obvious to the user, leading to a lack of critical investigation with respect to the information received.

2.4   Impact on Human Autonomy

Due to the nature and somewhat intelligent adaptiveness of the underlying technology, respective agents can also cause the human user to change their behavior, and by induction their underlying thought processes, as part of the optimization of their employment strategy. In effect, this would constitute a transformative process, leading to ethical concerns even if all epistemic and normative concerns have been satisfied previously [15]. This issue is different from change of behavior based on pure exposure, as the behavior change in this case is not strictly a goal of the deployed agent. By means of indiscriminate goal-setting for such agents, the accessible variables to optimize can then include the human user. To illustrate this, the following situation can be imagined as an example: the agent, an adaptive information algorithm, has the task to select content to be displayed on the starting page of a social media side. Based on the information about the user, the agent might only select to present newspaper articles aligning with the political view of the user, thus creating what has been called a "bubble" [3, 9] . Alternatively, the adaptive algorithm might learn that one of the "tweakable variables" is indeed the human user, who can be influenced to fulfil the reward function of the algorithm more reliably. Such actions might include showing increasingly more controversial information to provoke the user to engage. Another area of conflict arises if the user becomes aware of the impact or reach of such agents; he or she might change their behavior pre-emptively. Knowing that certain actions might prompt seemingly unrelated analysis, human users might proactively abstain from a catalogue of actions, thereby limiting their autonomy [20]. The influence or potential manipulations of such agents can therefore conflict in multiple ways with the principles of autonomy, freedom or free will which are deemed to be of high importance in contemporary society. Finally, the question arises if such agents might exercise neutral or beneficial influence, such as when counteracting harmful influence of other agents or nudging individuals towards beneficial behavior in a system of limited choices.

2.5   Venues of data-driven human machine interactions

Influence by IoE-agents can be exercised on different interaction-venues. Among those that are of potentially high impact are in particular (1) Location-Based Services, (2) Social Credit Systems, (3) Consumer Interactions, (3) Personal

Health, (4) Journalism / Information Aggregation, (5) Financial Services, and (6) Individual Pricing.

2.6   Legal Framework

Traditionally, rapid advancement of technology is difficult to address with legislative means in due time. For example, as privacy legislation is enacted to address gathering and interpreting sensitive information and linking them to a certain individual, new technologies does not have to rely on a specific identity of a human user, but allows micro targeting by focusing on the individual's relationships and adherence to a defined group [28, 30, 13]. The lag between technological progress and legislative action creates a gap that requires research in order to guide later legal actions and to evaluate the impact of actions already taken. This holds especially true for the essential prerequisite question of accountability when utilizing autonomous agents in the IoE [15, 2, 14].

2.7   Current Academic Activity

Academic inquiries in the nature of human autonomy or freedom, mostly from the viewpoint of psychology or philosophy are legion. With respect to the research objectives of this project, as far as can be seen, only initial research has been conducted [29]. Research in the field of privacy, as relevant, is usually conducted in the legal field (where it follows privacy-related legislation and its scope and gaps) and the information science field [18]. On the contrary, there is intense academic activity when it comes to devising ethical frameworks for artificial intelligence [1, 10, 17]. Attribution of responsibility in multi-agent systems, especially mixed systems, i.e. systems in which humans and IoE-agents both are part of the structure, have been identified as an area that warrants further research. Problem awareness of the compounded issues outlined above is growing. On the 13th of February 2019, the Committee of Ministers of the Council of Europe adopted a declaration on the manipulative capabilities of algorithmic processes, acknowledging the risk of "fine grained, subconscious and personalized levels of algorithmic persuasion [4]. Similarly, the United Nations Special Rapporteur on freedom of opinion and expression has reflected on the negative impact of Artificial Intelligence on individual's rights to expression and opinion in its Report dated 29th of August 2018 [27]. More broadly, the European Commission has established a High-Level Expert Group on Artificial Intelligence to address, among others, these emerging issues. Similar analysis has been conducted in Non-European countries as well [22]. Similarly, academic research has been starting to focus on this situation. Efforts have started to be steered towards better analysis of certain aspects the problem outlined above, in particular towards the conflict between commercial use of such agents and consumer protection legislation [8].

3   Research Questions

3.1   Primary Research Questions

Research will be guided along the lines of two primary research questions:

- What is the level of freedom an individual can maintain in situations where multiagent systems decide on behalf of the individual?
- What kind of autonomy can an individual develop if all the options are previously detected and addressed?

3.2   Secondary Research Questions

In addition, and only to the extent that this serves to build a necessary fundament of understanding and supports the above, a set of supplementary research question will be examined:

- How does influence of agents on the autonomy of humans manifests itself in the IoE?
- How is the type of effectiveness of influence predicated by the underlying technology or method of an agent?
- How is such influence in contrast or to the benefit of the principle of human autonomy?
- How do different interaction-venues facilitate or inhibit influence and what is their effect on the resulting consequences of effective influence?
- How can humans become part of the optimization process of adaptive agents and what is the impact of such adjustment?
- How are existing legal and ethical frameworks equipped to deal with such influence?

4   Research Objective - Contribution to Current Knowledge in the Domain

The main contributions of the research project to the academic body of knowledge is expected to be in the following areas:

- Attribution and Responsibility of Agents of the IoE
- Exercise of Autonomy in Systems constrained by Agents of the IoE
- Ethical Frameworks for Agents of the IoE

5   Methodology

The nature of the research proposed is interdisciplinary, consequently the methods undertaken will be heterogeneous. An overview is given in the following paragraphs.

5.1   Law and Policy

With respect to questions of law and policy, the research methodology will be dictated by a (qualitative) black-letter approach to analyze the statutes and jurisprudence of relevant jurisdictions. Due to language abilities, focus will be laid on jurisdictions where legal sources are available in English or German. Where appropriate, comparative analysis of the legal situations within the jurisdictions in question will be conducted. Resources used will be both primary (i.e. laws, court cases) and secondary, where available. Where the legislative umbrella fails to cover a situation of interest, recommendations lege ferenda will be given where appropriate. Similarly to the outlined above, instruments of soft law or similar (guidelines, reports, etc.) will be referenced and analyzed.

5.2   Philosophy / Ethics

Moral and Ethical frameworks with respect to the area of inquiry will be analyzed, interpreted and compared. Creation of an ethical framework will be attempted with respect to the means of influence of agents of the IoE. The research method will be qualitative and interpretivist.

5.3   Computer Science / Information Science

Extrapolation from current scholarship with respect to agents deployed in the IoE and their potential to impact private autonomy and algorithmic influence shall be attempted. Starting from a consolidation of fundamental principles of function, uses and limitations of such agent, their potential impact will be assessed based on their technological capabilities; areas of risk with respect to autonomy / freedom originating from the underlying technical approach of such agents will be determined and highlighted. Agents will be assessed abstractly; no model building will be attempted.

5.4   Social and Cognitive Science

Current findings in the field of the effects of algorithmic manipulations will be consolidated, compared and interpreted. Similarly, findings in the field of private autonomy, free will and personal freedom will be consolidated and reviewed for their applicability to the topic of interest as laid out above. Out of this research, a working model of human autonomy will be synthesized to be used in the dissertation, as to avoid ambiguities when elaborating on the subject. Where research in the other areas leads to the conclusion that specific promising theories have not been made plausible yet, original research will be conducted as far as the provided resources allow for it.

6   Acknowledgement

References

1. Allen, Collin, Wallach, Wendell and Smit, Wendell, "Why machine ethics?", Intelligent Systems, IEEE 21(4) (2006)
2. Ashrafian, Hutan. "Artificial Intelligence and Robot Responsibilities: Innovating Beyond Rights." Science and Engineering Ethics. Kluwer Academic Publishers (2015)
3. Bakshy Eytan, Messing Solomon, and Adamic Lada, "Exposure to ideologically diverse news and opinion on Facebook" Science 348(6239) (2015)
4. Committee of Ministers – Council of Europe.: Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes. (2019) Available at: `https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b`
5. ECHR: Nicolae Virgiliu Tănase v. Romania [GC], § 128 (2019)
6. Deigh, John, 'Ethics', The Cambridge Dictionary of Philosophy (Cambridge University Press, 2015), p.284f.
7. Hampton, Jean, 'Contractarianism', The Cambridge Dictionary of Philosophy, 2nd edn (Cambridge University Press, 1999)
8. Jablonowska, Agnieszka et al., "Consumer law and artificial intelligence Challenges to the EU consumer law and policy stemming from the business' use of artificial intelligence - Final Report of the ARTSY Project" (2018), Available at: `http://cadmus.eui.eu//handle/1814/57484`
9. Kaplan Andreas, Haenlein Michael "Users of the world, unite! The challenges and opportunities of social media", Business Horizons 53(1) (2010)
10. Kraemer Felicitas, van Overveld, Kees, and Peterson Martin, "Is there an ethics of algorithms?"Ethics and Information Technology 13(3) (2011)
11. Kramer, Matthew, Where Law and Morality Meet (Oxford University Press, 2004) p.3f.
12. Kuhn, Steven, 'Formalism', The Cambridge Dictionary of Philosophy, 2nd edn (Cambridge University Press, 1999)
13. Leese, Matthias, "The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union" Security Dialogue 45(5) (2014)
14. Matthias, Andreas, The responsibility gap: Ascribing responsibility for the actions of learning automata. Ethics and Information Technology 6(3) (2004)
15. Mittelstadt, Brent et al. "The ethics of algorithms: Mapping the debate", Big Data & Society 121 (2016)
16. Moka-Mubelo, Willy, Reconciling Law and Morality in Human Rights Discourse (Springer, 2017), p.53f, 86f
17. Moor James, "The nature, importance, and difficulty of machine ethics" Intelligent Systems, IEEE 21(4) (2006)

18. Nissenbaum, Helen, "Privacy as Contextual Integrity", 79(1) Washington Law Review (2004)
19. Nowak, Manfred U.N. Covenant on Civil and Political Rights - CCPR Commentary, 2nd edn (Kehl: N.P. Engel, Publisher, 2005) p.377f
20. Pan, Sheri, "Get to Know Me: Protecting Privacy and Autonomy under Big Data's Penetrating Gaze", Harvard Journal of Law& Technology 30-1 (2016)
21. Paul, Richard, Elder, Linda, The Miniature Guide to Understanding the Foundations of Ethical Reasoning, 3rd edn. (Critical Thinking, 2005) p.7f
22. President's Council of Advisors on Science and Technology, Big Data and Privacy: A Technological Perspective xiii, (2014) Available at: `https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf`
23. Quinn, Philip 'Divine Command Ethics', The Cambridge Dictionary of Philosophy, 2nd edn (Cambridge University Press, 1999)
24. Raz, Joseph, The Authority of Law: Essays on Law and Morality (Oxford University Press, 1979), p.166
25. Andrews Reath, 'Ethics', Routledge Encyclopedia of Philosophy (Routledge, 1998) Available at `https://doi.org/10.4324/9780415249126-L007-1`
26. Singer, Marcus, 'Moral Epistemology', The Cambridge Dictionary of Philosophy, 2nd edn (Cambridge University Press, 2015)
27. United Nations Special Rapporteur on freedom of opinion and expression: Report on the promotion and protection of the right to freedom of opinion and expression. (2018) Available at: `https://undocs.org/A/73/348`
28. Van Wel, Lita and Royakkers Lmbèr, "Ethical issues in web data mining", Ethics and Information Technology 6(2) (2004)
29. Vedder, Anton, Naudts, Laurens "Accountability for the Use of Algorithms in a Big Data Environment", International Review of Law, Computers & Technology; 2017; 31(2) (2017)
30. Vedder, Anton, "Why data protection and transparency are not enough when facing social problems of machine learning in a big data context." in "Being profiled: Cogitas ergo sum" (Amsterdam University Press, 2018)