GIANLUIGI M. RIVA - SIMONA TIRIBELLI

MORAL AND LEGAL AUTONOMY IN THE ERA OF ARTIFICIAL INTELLIGENCE

ABSTRACT: *MORAL AND LEGAL AUTONOMY IN THE ERA OF ARTIFICIAL INTELLIGENCE*

The ethical principle of autonomy is one of the core principles in the field of applied ethics, from bioethics to ethics of artificial intelligence (AI). In this paper, we pursue an ethical inquiry into how AI systems can affect human autonomy according to a moral and legal perspective, that is, both in its moral dimension (as implicit endorsement) as well as in its legal one (as explicit consent). More specifically, after having defined the concept of individual autonomy from a moral and juridical standpoint as the human normative power of self-determination both in the moral sphere and in the legal one, and thus, better substantiating the AI ethics principle of autonomy as currently adopted in the field, we show how the design of novel AI systems, such as machine-learning and deep-learning algorithms, that widely rule the functioning of digital information and communication technology (ICT), can negatively affect both the preconditions of our moral and legal autonomy, by suspending intrinsic consent that individuals can express in order to endorse external information as a true motive of their choices and actions, as well as bypassing the legal conditions for a valid consent, as the external manifestation of the individual legal autonomy. We conclude by highlighting the ethical implications and risks of such phenomenon and proposing ethical and legal design practices to prevent or mitigate them.

1. *Introduction*

Artificial intelligence (AI)’s influencing power on human behavior is an idea that has been outlined in many fields a long time ago. However, it since the Cambridge Analytica case, which unveiled the huge misuse of AI providers of users’ sensitive data for third-party goals, that academic and political institutions started to raise serious and robust ethical and legal concerns around the potential influence of algorithmic techniques, and more specifically of online micro-targeting algorithms, on individuals' autonomy and decision-making processes. In February 2019, for example, the European Union (EU) adopted the Declaration on the invisible and manipulative algorithmic capacities[1], inviting the EU Member States to monitor the use of algorithm-based technologies, drawing specific attention to the commercial players' deployment of AI systems and algorithms to gather personal data to predict and reshape individuals' personal preferences up to manipulate their autonomy over choices and decision-making[2]. Such concerns have made the issue of human autonomy, and specifically the respect of human autonomy, an ethical and legal issue of paramount importance in scholarship and initiatives on the ethics, policy, and law of AI and algorithms – this is visible, for example, by considering how much frequent is the reference to the protection of human autonomy as informational control and self-determination over decision-making as a core principle in benchmarking ethics frameworks[3] as well as legal regulations[4] on AI and ICTs developed over the last few years. In the meanwhile, in the academic world, the legal scholar Brett Frischmann and the philosopher Evan Selinger describe this huge algorithmic influencing power on individuals and societies via the

[1] See the European "Declaration on the Manipulative Capabilities of Algorithmic Processes" (EU Decl [13/02/2019]).
[2] European Data Protection Supervisor (EDPS) Opinion No. 3/2018 on online manipulation and personal data.
[3] See, for example, High Level Expert Group on Artificial Intelligence [HLEGAI] – European Commission (2019 April, 8). *Ethics Guidelines for Trustworthy AI* (available at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai); see also A. Jobin, M. Ienca, E. Vayena, *The global landscape of AI ethics guidelines*, in «*Nature Machine Intelligence*», 1, 2019, pp. 389-399.
[4] See, for example, 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 [2016/679 (EU)].

macro-concept of "techno-social engineering"[5]. Such concept aims to convey the idea of those processes in which «technologies and social forces align and impact on how we think, perceive, and act»[6] by monitoring our lives at a micro-, meso-, and macro-level and thus reshaping the way in which we live and navigate them. Accordingly, influential philosopher Shoshana Zuboff describes the algorithmic ICTs as tools that treat human experience as free raw material used as a mean to obtain forecast product through which triggering behavior modification and change[7]. Since Cambridge Analytica, the corpus of literature committed to systematically analyze the impact of AI systems and algorithms on autonomy has expanded significantly[8]. However, the majority of such contributions mainly adopt a philosophical-psychological perspective on the issue[9], while the issue of how AI and algorithms influence human autonomy and therefore violate the ethical and legal principle of autonomy has been poorly explored from an interdisciplinary perspective rooted in in Moral Philosophy and Privacy & Data Protection EU laws and principles. Such a perspective sounds instead to be of particular importance especially in relation to AI, for which regulation such disciplines are increasingly called today to work jointly in order to be really effective, as well as if we consider the deep relationship between legal autonomy as informed consent and moral autonomy as reflective endorsement – as it will be clearer later on in this article.

[5] B. Frischmann, E. Selinger, *Re-Engineering Humanity*, Cambridge University Press, Cambridge 2018, p. 4.

[6] *Ibid.*

[7] S. Zuboff, *The Age of Surveillance Capitalism. The fight for a human future at the new frontier of power*, Public Affairs, New York 2019.

[8] B.D. Mittelstadt *et al.*, *The Ethics of Algorithms: Mapping the Debate*, in «Big Data & Society», 3, 2016.

[9] D. Susser et al., *Technology, autonomy, and manipulation*, in «Internet Policy Review», VIII, 2, 2019. J. Williams, *Stand out of our Light: Freedom and Resistance in the Attention Economy*, Cambridge University Press, Cambridge 2018. F. Jonjepier, M. Klenk, *The Philosophy of Online Manipulation*, Routledge, London 2022.

The present paper aims at filling this gap by highlighting the impact of AI systems and specifically of algorithmic ICTs on individual autonomy conceived in both its legal and philosophical dimension. After defining from a moral and legal standpoint the concept of autonomy as human normative power of self-determination at the core of our analysis, the paper offers a multidisciplinary argument on how AI can affect human autonomy with a specific focus on the ethical and legal consequences that this impact implies. The paper first offers an ethical inquiry rooted in moral philosophy on how AI can suspend the intrinsic consent (or moral endorsement) as crucial precondition for individuals in order to be considered authors of their choices by exercising a certain degree of control or participation over the information steering their choices and actions; put it differently: on the normative reflective power that individuals can exercise to adopt certain information as an authentic motive of their choice and actions. Taking into account the concept of autonomy as emerging from such an ethical inquiry, the paper then argues around the impact of algorithms on the legal conditions for a valid consent (as the external manifestation of the individual legal autonomy) in the EU Privacy regime. The study grants an in-depth focus on the forms of AI influence and how they can be reconducted to the current regulatory paradigms. Finally, the study draws the conclusive remarks outlining the direction for a set of practical approaches to tackling the issue from an ethical and legal self-regulative (design) perspective.

2. *On Moral and Legal Autonomy: methodology and taxonomy*

Considering the interdisciplinary approach to the issue of autonomy we endorse in this paper, a few methodological remarks need to be preemptively added. First, from the philosophical and specifically ethical standpoint, we are called to clarify the specific concept of autonomy we refer to in this article. Indeed, autonomy is one of the most debated issues in moral and socio-political philosophy as well as in applied ethics, from bioethics to medical ethics; therefore, it is worth underlining that we will not take into consideration or compare the different connotations that such ethical concept assumes in diverse ethical branches and in the long-standing and heterogeneous debate developed in Western and non-Western moral philosophy, as it is beyond of the scope of this article. The goal of this paper is indeed to shed light on a broad range of ethical and legal implications that can arise if we consider current AI ethics and legal principle of human autonomy as informational self-determination from both a legal and moral perspective and show how such ethical-legal inquiry allows to unpack further dimensions of the principle of autonomy, as currently adopted, that ought to be considered if we are serious in preserving or promoting human autonomy in the era of AI. In this sense, in our analysis, we refer to autonomy as currently emerging mainly conceptualized in benchmarking legal and ethical frameworks on AI and algorithms, that is, to autonomy as human self-determination, control, and self-governance[10], and we limit the scope of our analysis to better unpack it – i.e. clarify its plural dimensions – and analyze it in relation to algorithmic techniques from an ethical and legal standpoint. In this sense, we acknowledge that current widespread conceptualizations of autonomy in the field are mainly rooted in Western Philosophy, and that crucial future works on the AI ethics and legal principle of autonomy should also consider both non-mainstream and non-Western philosophical accounts of autonomy to be, beyond properly adequate and exhaustive, truly ethical and inclusive.

In moral philosophy, there is a widespread agreement on the ethical concept of autonomy (i.e. principle of self-determination) as emerging mainly in ethical and legal scholarship on AI. There

[10] See, for example, HLEGAI, *Ethics Guidelines for Trustworthy AI* 2019; A. Jobin, M. Ienca, E. Vayena, *The global landscape of AI ethics guidelines* 2019; GDPR 2016.

is indeed a consensus on autonomy as a key pre-condition of human freedom of choice and action, according to which autonomy refers to the capacity of the individuals to be in control of the reasons and motives underpinning their choices and actions: in this sense, an autonomous choice is a choice between potential courses of action that cannot be casual, that is, that cannot result from factors out of control of the agent – as in the case of deciding whether to make a certain action on the basis of the toss of a coin, or if coerced to make a certain choice and action[11]. Put it differently: autonomy as self-determination and self-governance refers to the capacity of the individuals to choose and act according to beliefs, values, motives, and reasons that are in a relevant sense their own, as reflectively endorsed (authenticity conditions) and developed and chosen in a context of a meaningful availability of alternative options[12]. In our ethical analysis of the impact of AI systems on autonomy from a moral perspective we will refer to human autonomy according the above-mentioned definition, aware of the possibility to devote in future works wider space to different philosophical conceptions and traditions on human autonomy for an ethical assessment of whether they can be fruitful to better understand the relationship between human autonomy and artificial or algorithmic entities[13].

The legal analysis focuses on Privacy consent *qua* the phenomenological element that legally connects one's manifest agreement with the legitimacy of AI's algorithmic activity related

[11] M. De Caro, *Il libero arbitrio. Una introduzione*, Laterza Editori, Roma-Bari 2004.

[12] J. Christman, J. Anderson, *Autonomy and the Challenges to Liberalism: New Essays*, Cambridge University Press, New York 2005; G. Dworkin, *The Theory and Practice of Autonomy*. Cambridge University Press, New York 1988; S. Killmister, *Taking the Measure of Autonomy: A Four-Dimensional Theory of Self-Governance*, Routledge New York 2017. C Mackenzie, N Stoljar, *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*. Oxford University Press, New York 2000.

[13] For a wider consideration of different accounts (e.g., liberal, communitarian, relational perspectives) of autonomy from within moral philosophy in relation to AI, see S. Tiribelli, *Moral Freedom in the Age of Artificial Intelligence*, Mimesis International, Milan-London 2022.

to personal data processing. Privacy is a special cross-discipline within the Law, and the legal relationships that it supersedes trace back to the private autonomy of negotiation (Contract Law). For this reason, the legal analysis of the consent-giving within the scope of Data Protection legal regime must ground on Contract Law paradigms, qua the regime that informs the notion of legal autonomy in private agreements and their effects. The lenses we adopt are those of the EU Civil Law system, in which private autonomy is codified in a body of norms, i.e. the civil code (according to each Member's legislation). However, the complete analysis of legal autonomy requires an in-depth investigation that overpasses the goal of the present article, which instead focuses on the specific elements related to the vices of the will, i.e. the conditions for which consent (volition) is free or influenced, and, as such, valid or not. In Europe, the General Data Protection Regulation[14] is the primary source that regulates Privacy rights and Data Protection and the related data subjects' consent-giving activities in any personal data processing carried out for all those activities that overpass domestic purposes. It is a general regulatory framework which is then coherently and autonomously developed in every Member State's national regulation. Therefore, the study must refer to a national body of law in order to investigate the specific regime of the consent validity. For the sake of this study, among the many Civil Law systems in Europe, we choose the Italian Law systems, in which the Civil[15] and Penal Code[16] provide cross-referenced rules for the regime of (legal) autonomy. Although taking national legislations as an example may appear too narrow, the European codifications share a common approach for the codes. Therefore, this analysis can easily adhere

[14] 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 [2016/679 (EU)].

[15] Royal Decree 16 March 1942 n. 242 "Italian Civil Code".

[16] Royal Decree 19 October 1930 n. 1398 "Italian Penal Code".

to other Civil Law countries regimes. On the contrary, Common Law systems (such as the US and UK's) work according to different legal paradigms and, therefore, the rules that underpin the validity of consent and remedies may vary. The choice of Civil Law systems as reference for the legal analyses comes from the articulation of the regulatory framework around the concept of autonomy and consent-giving. For the legal analysis, we use the empirical hermeneutic reasoning based on the legal syllogism technique. The method grounds on the classic critical legal approach (legal qualitative/speculative analysis) that grounds on the so-called General Theory of the Law[17].

3. *How AI is affecting human autonomy from a moral perspective: a new [algorithmic] choice-architecture*

In the last few years, the pervasive deployment and use of digital ICTs in a wide array of different domains has allowed the conceptualization of what is known as the "fourth revolution"[18] to point out the fast digitalization and hybridization of our contemporary reality and stress that we live today in an increasingly information-based environments. Such pervasive presence of digital ICTs has indeed blurred the distinction between online and offline, leaving the space for a reality that is, beyond describable increasingly conceivable in informational terms. Indeed, today, each intangible and tangible piece of reality can be datafied, that is, codified in data (*datafication*) – from people's characteristics and physical movements to habits, intimate opinions, beliefs, interactions, values and affiliations – therefore almost everything can be captured and transformed in

[17] N. Bobbio, Teoria Generale Del Diritto, Collana Recta Ratio, Giappichelli Editore 1993. For the sake of clarity, when in the legal analysis we refer to "norms" and "normative", we intend the legal meaning of the terms, i.e., respectively, a set of prescriptive propositions, and the character of a rule that provide a binding legal precept.

[18] L. Floridi, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, Oxford University Press, Oxford 2014.

huge value-laden amount of information about people, their deep connections, and the world in its complexity and generality.

Such datafication process and the digitalized space (infosphere)[19] it envelops are today driven by new kind of AI techniques and algorithms increasingly regulating the functioning of digital ICTs, such as machine learning (ML) and deep learning (DL) algorithms, namely, probabilistic models fed and trained by data we directly enter online (i.e. *provided data*), the onlife trails[20] we leave indirectly behind us (i.e. *observable dat*a), and those *derived* and *inferred* by correlating such categories of data with other huge datasets already available, models pre-set to probabilistically learn how to harness such datafied material to achieve pre-defined tasks and very often third-party goals. Profiling ML is one of the most used techniques to rule ICTs and consists of a series of techniques to mine such data and discover them meaningful patterns and correlations in order to profile and categorize online users and predict a range of value-laden (often choice-driving)[21] elements, ranging from deep preferences, needs, and vulnerabilities up to religious, sexual, and political orientation, that are capturable directly or by association or affinity[22].

This kind of profiling and predicting power characterizes what can be defined as algorithmic "agency", which can result from the interconnection of (technically speaking) narrow tasks–oriented ML algorithms, and shows the potential to create what we may define as novel *choice-architectures* that can reshape the social, environmental, and structural conditions of our informational societies: the way we perceive our reality, we interact, consider, and understand with each other and ourselves, how we make our

[19] *Ibid.*

[20] *Ibid.*

[21] S. Tiribelli, *Moral Freedom in the Age of Artificial Intelligence*, Mimesis International, Milano-Udine 2022.

[22] S. Wachter, *Affinity Profiling and Discrimination by Association in Online Behavioral Advertising*, in «Berkeley Technology Law Journal», XXXV, 2, 2020.

tasks, choices, decisions, and actions, and to how we live and – by exercising our autonomy as normative power of self-determination – steer, self-govern, and self-determine our lives. This is because in our informational environment algorithms are informational gatekeepers[23], which decide – filter, classify and personalize – *what* information and *how* is shown to us. Let us think about how the information is classified in response to an online query on Google: the first results displayed, personalized *ad hoc* on the basis of the profiling of each individual, are probably what one would look at the first place or in absolute[24]. What does this mean for human autonomy? The issue lies in the fact that these information, informational contents or informational options pre-defined and chosen algorithmically are what will end to substantially inform directly or indirectly our knowledge and standpoint on the specific issue and subject searched, namely, the epistemological level of human autonomy, the epistemological dimension (i.e., level of knowledge) of our autonomy and decision-making. To put it differently: algorithms preselect and decide what is meaningful for individuals on the basis of profiling tasks often driven by third-party goals, that is, the specific range of informational options (meaningful availability of options as a condition for autonomy) on which the agent chooses in the decision-making and that can become potential motives for different courses of actions, and this algorithmic narrowing choice of "meaningful" options is based on parameters that are very often opaque to users[25] and that to prioritize economic interests can result to be disaligned and therefore to do not respect people's real goals and ends.

[23] L. Shapiro, *The Control Revolution: How the Internet is Putting Individuals in Charge and Changing the World We Know*, PublicAffairs, New York 2002.
[24] i.e., very often we stop our research online after having clicked the first two or three search results.
[25] F. Pasquale, *The Black Box Society: The Secret Algorithms that Control Money and Information*, Harvard University Press, Cambridge 2015.

If we consider this argument on a larger scale, we might rightly claim that the invisible action of ubiquitous algorithms redefines quantitatively and qualitatively our informational environment, i.e., our availability of informational contents and options, showing information filtered and classified on the basis of at least three key elements a. the pertinence of the research (content-based filtering), b. individuals' profiled preferences as profiled (and correlated with others' preferences and choices; i.e., collaborative filtering) and c. the maximization of preset goal that very often, especially in online service providers (OSPs) lies in the maximization of the click (CTR: click-through-rate) of the most sponsored informational contents. From this view, we might say that we daily experience an invisible, subtle, but pervasive action of AI techniques on our autonomy – on our capacity to self-determine our choices and action according to reasons and motives that are somehow our own as chosen in a context of meaningful options – and especially at the *epistemological level* of our autonomy: the level of autonomy as authenticity characterizing the process of formation of individuals' beliefs, reasons, values, and projects informing their decision-making which result as prechosen, reshaped, and thus influenced by AI systems. To sum up: by algorithmic choice-architecture we refer to this algorithmic capacity to preselect without users' active participation what is deemed as informationally meaningful for the individuals and redefine their epistemological space of alternative options accordingly. Such algorithmic choice-architectures, which reflect the design of ubiquitous digital ICTs and mainly act by filtering, classifying, personalizing informational contents, can be unveiled as raising a possible algorithmic interference to human autonomy at the *epistemological level*: pre-selecting information means influencing what can become motives of people's choices, actions, and identities, that is, those informational options that people can endorse as the motives of their choices, actions, and behavior. This algorithmic action is therefore more of a mere influence: it can be understood as a first algorithmic constraint on human autonomy, in the form of an *epistemological hetero-definition* of the availability of informational options characterizing users' choice-contexts. This epistemological impact exercised by the algorithms on human autonomy constrains human decision-making, binding the informational options and hence the possible motives that people may endorse for their choices and actions to a certain algorithmically predefined space of information based on predictive profiling.

4. *The Algorithmic Double-Level Impact on Human Autonomy*

To better understand this algorithmic action on human autonomy, it is helpful to consider Sunstein's and Thaler's renown theory of choice-architecture and nudges[26]. According such a theory, nudges are institutional suggestions or reinforcements that can be implemented by institutional actors (indeed, they always refer to institutional choice-architectures), basing on the confutation of the widespread idea that people choose in optimal conditions and therefore always rationally (i.e. *homo oeconomicus*' paradigm), in order to influence and predictably change individuals' behavior according to long-term desirable goals, without denying their freedom of choice. Nudges, in fact, operate at the informational level and consist of intentionally altering people's informational environment by modifying the order of presentation of informational options to people (as citizens and consumers). A classic example is to switch the place of food in a store and present to consumers the healthy options at the eye-contact level or close to the cash register, while placing junk food options to other parts less visible. This pre-set display of options aims to

[26] C. Sunstein, R. Thaler, *Nudge: Improving Decisions about Health, Wealth and Happiness*, Yale University Press, New Heaven 2008.

change towards a desired direction people' decision-making, without removing their free capacity to choose and act. The algorithmic design ruling ICT is methodologically informed by Sunstein-and-Thaler's theory of nudges (and by the methods of behavioral economics and psychology); nonetheless, algorithmic choice-architectures show clear differences if compared to institutional ones, as traditionally theorized. First, it is worth clarifying their affinities. To do so, we can consider one of the most widespread algorithmic techniques used to rule OSPs: recommender systems (RS). RS does not just redefine users' informational environment by classifying and filtering information, they are preset and designed to recommend (or nudge) specific contents to individuals. As it has been pointed out in the literature, RS nudge users with catching-attention ads to reshape predictably their economic, social, and political behavior, e.g. "you've seen this, you should see", or by pervading users' onlife environment with information that frequently interrupts onlife activities they carry on[27]. However, there are differences between the kind of institutional nudges proposed by Sunstein and Thaler and the algorithmic nudging and influencing action here in question. First, the agents of nudges are not institutional as in the traditional libertarian paternalistic theory of nudge, but are algorithms often driven by third-party interests – as those of advertising companies and political parties (as clearly unveiled by the case of Cambridge Analytica, when political parties have exploited the collection of people's personal data to micro-targeted informational ads preset to change users' political vote in the 2016 U.S election). Second, the degree of algorithmic nudges' impact is higher than that of institutional nudges and those of any previous technologies, given the fact that we are completely immersed in this informational

[27] K. Yeung, *Hyper-nudge: Big data as a mode of regulation by design*, in «Information, Communication and Society», 20, 2017.

environment permeated if not yet governed by algorithms, and therefore always exposed to their powerful but invisible design. Third, the purpose of the nudge: if the nudging actions operated by institutions have the purpose to influence short-term individuals' choices in order to improve their long-term well-being, the aim of algorithmic nudges is very often the short-term boosting of the click, namely, the economic interests and income of such companies (or political parties) that subsidize OSPs to show and privilege some informational options reflecting their economic and political long-term goals.

From the perspective of human autonomy, we claim that the nudging action exercised by such algorithms cannot just create a *soft constraint* on individuals' choice behavior, but can also raise a *strong constraint* that thus does not just affect the epistemological level of human autonomy by filtering informational options shown to users (i.e. the options that can become motives we can endorse in our choices and actions). This first algorithmic action is a *soft constraint* on individuals' behavior, insofar as, though profoundly affecting how people develop knowledge and perceive themselves, the others, and the reality broadly, it cannot force or coerce a person to choose and act in a certain way anyway. This is because we as persons and *moral agents* retain a certain degree of reflective power to decide to act also against a range of informational options, as well as against our preferences or needs, and/or choose not to choose. We own and deploy this *normative power* when after hesitating in front of some alternative options available as possible courses of actions we say yes (or no) to a certain option rather than to one another, by *endorsing* it as a *motive* into a practical choice. In this *reflective endorsement* (or intrinsic consent) we can find the ethical-normative value of our autonomy, that is, our power of self-determination, which allows us expressing and developing our *ought to*, our *moral posture* (that can be, depending on the individual, more or less defined), forming and endorsing our own values, ideas of good, moral ground-projects and/or moral horizons of meaning, that we bring into play as intrinsic acts or moral dispositions (moral disposition as *diàthesis*, at the root of moral behavior, as *héxis*). In other terms, this posture is the way we learn to respond, by taking a *moral stand*, to the reality, to world data, conveyed by our mediated perceptions and emotions: namely, how we develop over time our *moral identity as unique persons*.

However, this intrinsic endorsement, on which the normative value of human autonomy lies (*moral level or dimension stricto sensu*), can be affected too by algorithmic techniques, as currently designed. This phenomenon may occur when the information used to nudge individuals is morally sensitive (information about individuals' physical-psychological status, traumas, vulnerabilities, and weaknesses, just to mention a few). For example, AI could infer from a person's queries on Google about e.g. "how people with cancer feel" and from his/her past geo-localizations, e.g. at the hospital (see very recent Google's claims on suspending individuals' profiling/geo-localization of abortion clinics, after 2022 US supreme court overturning of Roe vs Wade), by correlating them with other sets of data presenting the same characteristics, that a person is interested (he/she can be ill, or just particularly sensitive to that topic, e.g., hypochondriac) on certain kind of information about cancer. Levering this inference, algorithmic techniques might [hyper]nudge the person ubiquitously (in all her inter-connected devices) with preselected (as tested) sponsored items linked to that pathology for maximizing its goal, that is, with the capacity to trigger individuals' behavior change toward the pre-set algorithmic goal: ads that show up in the webpages the user consults, videos he/she watches, or SNS he/she use. These subtle nudges – that for example can assume the form of strong pictures (e.g. a cancer patient under medical treatment) or sensitive information (e.g. the right of care's withdrawal) – hold the power to crucially jeopardize people's control and self-governance and *determine* people's life-changing choices (i.e. the renounce of a medical treatment) by exploiting people's sensitive weaknesses and vulnerabilities inferred via profiling techniques.

These morally sensitive and relentless nudges, continuously finetuned by techniques such as RS on individuals to make recommendations increasingly tailored and effective, might therefore affect human autonomy in-depth by suspending or bypassing people's reflective *endorsement* and transforming the main informational option recommended – algorithmically pre-chosen to achieve certain goals – from being a *motive* (that a person might reflectively endorse) to being the *cause* (strictly determining) of a person's choice and behavior, from epistemologically informing human choices (i.e. advice people) to decide them at their place (i.e. determine people). From this view, human autonomy would become, beyond *epistemologically* influenceable, also *morally* re-shapeable, insofar as the degree of meaningful and reflective control individuals can exercise over their choices would result deeply undermined. Therefore, this algorithmic impact on human autonomy might not just affect the *epistemological dimension* of human autonomy, but also the *moral* one *stricto sensu*. As a consequence, this impact would not just create a soft constraint on human self-determination, but it might raise a *stronger one*, by binding, hetero-directing and hetero-defining the formation of people's identity toward third-party goals, rather than according to their deliberated reasons, intrinsic values, and meaningful long-term goals. Thus, this algorithmic choice-architectures could not only entail by default a fine-grained form of manipulation on individuals; they also show the inner capacity to turn this influence in an unprecedented and invisible form of subtle coercion that would not only negatively affect and erode human autonomy but even completely suspend and override it.

5. *AI and the influence spectrum for legal paradigms*

Under the AI's umbrella, scholars identify a vast range of technologies and informatics techniques, such as Machine and Deep Learning, Neural Nets and Algorithms. The legislation always adopts a neutral approach towards technologies and regulates them, or their effects, with a standardized approach. It means that the Law cares about the technology meeting certain level of requirements, without caring about the technical specificities of how it reaches these levels. For instance, the Law requires engines for urban vehicles to meet particular technical parameters to limit particular emissions but leaves producers free to match the requirements with any (legitimate and secure) technical solution Therefore, for the Law, it can be acceptable to consider all the mentioned different IT technologies as one (the AI) as long as the regulation addresses their impact and effects.

When it comes to considering how AI can influence the individual's decision-making process, interpreters must bear in mind that, typically, AI is developed by private entities for commercial purposes and, thus, commercial players perform this kind of activity for profit. This means that publicly available AI services remain connected with the revenue models that underpin the commercial purposes behind the services provided. This data economy drives how commercial players design and develops algorithms for gathering personal data[28]. We already discussed predictive profiling and its role in shaping the autonomy sphere of the individual. However, it also influences the relational paradigm between the individual and the service provider (i.e. the data controller). Since personal data nourish the data economy

[28] WP29 Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, Adopted on 3 October 2017 as last Revised and Adopted on 6 February 2018, 17/EN WP251rev.01.

based on advertisement and selling costumer's profiles/personas to third parties, data controllers have a profit interest in attracting and maintaining users online to let them keep interacting with their services and algorithms for profiling purposes[29]. AI algorithmic systems govern this kind of interactive loop through a complex mechanism of notifications, haptic stimuli (vibrations, sounds), and neuromarketing techniques that affect the interactive design of online services or platforms[30]. These sets of nudges aim at attracting users' attention and exploits all those psychological mechanisms that trigger the same brain area involved in pleasure and addiction[31]. Such mechanisms are, among many, echo chambers, random reward, sense of scarcity and Neuro-Linguistic Programming techniques. Besides, many of these triggers can be used to artificially create habits by exploiting the Pavlov's conditioned reflexes[32], which elicit automatic responses to certain stimuli (e.g. apps notifications) and bypass the conscious decision-making process. When these 'psycho-design' techniques are embedded into the design of a service platform, they are usually referred in scholarship as to "dark patterns"[33].

The set of these techniques involves the nature of the legal relationship between users and service providers directly, as the latter grounds on parties' agreement and, therefore, on consent-giving.

[29] T. Wu, *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*, A. Knopf ed., Toronto 2016.

[30] C. Spence, *Neuroscience-Inspired Design: From Academic Neuromarketing to Commercially Relevant Research*, in «Organizational Research Methods», 22, 2019. See also C. R. Sustein, *Nudging: A Very Short Guide*, 37:4 in «Journal of Consumer Policy», 583, 2014.

[31] V.I. Maiorov, *The Functions of Dopamine in Operant Conditioned Reflexes*, in «Neuroscience and Behavioral Physiology», 49, 2019, pp. 887-893

[32] I.P. Pavlov, *Lectures on Conditionet Reflexes* (Vol. II), Conditioned Reflexes and Psychiatry, Lawrence & Wishard, London 1941.

[33] See infra § 5.3

6. *On Legal autonomy*

As mentioned, the legal regime that governs the relationship between users and service providers falls into the discipline of Contract Law. Nonetheless, as it relates to online services or interaction, it typically involves another species of legal regimes, which fall into Privacy Law and it is specifically regulated by Data Protection under the provision of the GDPR. Privacy and Contract Law share certain legal paradigms, as both refer to a legal relationship that occurs between private parties. Besides, both ground on the parties' agreements as the bonding of the relationship. The agreement itself represents the intersection of the parties' consent, which in turn can be seen as the external manifestation of the will. Indeed, the Law grants particular attention to the consent in the private sphere, as it represents the maximum expression of the individual right of self-determination[34], which embeds the concept of legal autonomy, i.e. the ability to autonomously dispose of one's rights and legal positions (patrimony, obligations, statuses).

6.1 *Private autonomy*

Private autonomy can be defined as the power that the Law recognises to individuals through which they regulate their interests thanks to the manifestation of their will. Private autonomy includes the concepts of negotiating and, therefore, contractual autonomy. In turn, negotiating autonomy represents the subject's power of self-determination, i.e. the power to dispose of one's own legal sphere. It includes the hypothesis in which autonomy is expressed through the fulfillment of a contract. As such, the contract is a mean of implementing the autonomy of negotiation as a source of obligations or transfer of rights. The Law establishes limits to this kind of autonomy, as it must derive from - or be recognised by - norms that attribute specific legal

[34] Charter of Fundamental Rights of the European Union 2012/C 326/02.

effects to the manifestation of the will[35]. It is possible to find a recognition of this private autonomy in the different EU States' Constitutions, each according to their specific paradigms and principles. Nevertheless, the EU itself grants to consumers that protection that counterbalances the potentially harmful effects of autonomy in a global market economy in which big commercial players hold dominant and influential positions. According to the EU Consumer Law[36], those contractual clauses that are vexatious towards consumers are not valid (so-called selective nullity), even if consumers accepted them. Indeed, the Law recognizes that consent can be, in particular situations, the weaker form of protection for those parties that stands in a sensitive position, e.g. workers, minors or, precisely, consumers. With the principles of general - and paternalistic - protection embedded in Consumer Law, the EU aims to protect the weaker parties from their very autonomy. The reason comes from the jurisprudential orientation to protect *"the contractual party economically weaker and legally less experienced than the professional counterparty"*[37]. However, from this reason, derives that the rationale for protecting the weaker parties would not deal with the ethical aspects of the relationship but should deal instead with patrimonial equality. However, the EU Law considers the weaker contracting parties those whose freedom of choice between several comparable options is compromised or excluded by a market structure that favors abusive, collusive, non-transparent, or discriminatory behavior by a

[35] Other limitations derive from the regime of unavailability or inalienability of personhood rights, such as dignity, freedom and so forth, which can be disposed in limited way only, according to the jurisprudential and doctrinal reconstructions. P. Perlingieri, in *Manuale di Diritto Civile*, 2000, p.360

[36] Directive 2011/83/EU of the European Parliament and of the Council of 25 October 2011 on consumer rights, amending Council Directive 93/13/EEC and Directive 1999/44/EC of the European Parliament and of the Council and repealing Council Directive 85/577/EEC and Directive 97/7/EC of the European Parliament and of the Council Text with EEA relevance.

[37] Cf. European Court of Justice decisions: 12 May 2005, trial C-112/03, Société financière et industrielle du Peloux; 20 January 2005, trialC-464/01, Gruber, point 34; 11 July 2002, trial C-96/00, Gabriel, point 39.

merchant dynamic that is not properly competitive[38]. With this legal regime, Consumer Law bypasses private autonomy and protects it against the individual's will itself. In this regard, it is remarkable how this protection (in some instances) also embraces situations that fall into the "take it or leave it" paradigm[39] in which the weaker party must undergo unfair conditions to purchase a product or have a service. This unfair practice often occurs for essential or necessary goods and services or for those goods supplied by monopolies or players in a dominant position. In this case, it applies the range of protection given by the EU Competition Law[40].

6.2 *Autonomy as informed consent*

As seen, consent is the crystallisation of one's autonomous decision-making process, which allows the power of informational self-determination, unless it also implies the contractual paradigm that so far weakened data subjects' positions in accepting unfair conditions. Nevertheless, for the Law, consent must hold several requirements in order to be valid. From a mere contractual perspective, consent must only be free. In the Italian Civil Code, for instance, no positive norm requires it, but it can be derived - *a contrariis* – from Article 1427[41], which states *"the contractual party, whose consent was given in error, extorted violently or snatched with malice, can request the cancellation of the contract"*. On the other hand, the GDPR requires the consent accorded for the data processing to be freely given, informed, specific and

[38] D. La Rocca, *Eguaglianza e libertà nel diritto europeo*, Giappichelli, Torino 2008, p. 110.

[39] F. J. Zuiderveen Borgesius; S. Kruikemeier; S. C. Boerman; N. Helberger, *Tracking Walls, Take-It-or-Leave-It Choices, the GDPR, and the ePrivacy Regulation*, in «Eur. Data Prot. L. Rev.» 3, 2017; and A. Romero-Medina, M. M. Triossi Verondini, *Take-It-or-Leave-It Contracts in Many-to-Many Matching Markets* (2019). Available at SSRN: https://ssrn.com/abstract=2917189 or http://dx.doi.org/10.2139/ssrn.2917189

[40] Council Regulation (EC) No 1/2003 of 16 December 2002 on the implementation of the rules on competition laid down in Articles 81 and 82 of the Treaty (TFEU).

[41] Titled "Error, violence, malice".

unambiguous[42]. Furthermore, GDPR Article 7 and recital 32 (which is not binding) complete the legal regime for consent. Article 7(2) states that when the data controller[43] asks for consent, the request must be presented in a clear, comprehensible and straightforward manner, in which all the different characteristics of the relationship[44] are clearly distinct. Also, Article 7(4) specifies that *"When assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract"*. In this sense, recital 32 is paradigmatic when it states that *"consent should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data relating to him or her, such as by a written statement, including by electronic means, or an oral statement. […] Silence, pre-ticked boxes or inactivity should not, therefore, constitute consent. […]"*. The concept of free consent is further addressed by the formal interpretation given by the European Data Protection Authorities[45]. The described framework shows how important it is for Data Protection – and the Law in general – the concept of free consent.

From the described normative clarifications, it emerges that a clear subjective understanding of the relationship conditions (data

[42] GDPR article 4(1) 11).

[43] i.e. the subject who performs the data processing, according to GDPR article 4(7): *"'controller' means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law"*.

[44] Which, in the case of data processing, are the different purposes. See GDPR article 4(2).

[45] Article 29 Data Protection Working Party, "WP29 Opinion 15/2011 on the Definition of Consent (EU 2011) https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2011/wp187_en.pdf; Article 29 Data Protection Working Party, "WP29 Guidelines on Consent under Regulation 2016/679 as Last Revised and Adopted on 10 April 2018" (EU 2017) 17/EN WP259 rev.01 https://ec.europa.eu/newsroom/article29/items/623051.

processing) constitutes in the PDP regime the pivotal element for free consent. Indeed, "information" is the second element of valid consent for the GDPR. It follows that, in order to be free, the consent must be (also) informed, which, therefore, represents a necessary - but not sufficient - element to render the consent free. Although, as seen, the classical Contract law does not consider the "information" as an element of valid consent, it emerges in parallel with the European regime for the protection of the weak party in deceptive commercial practices[46].

On another level, the concept of legal literacy, plays a crucial role for the data subject to be able to technically read and understand the consent form and the information leaflet (and to be capable of detecting any relevant mistake or discrepancy regarding the GDPR requirements). The EU Law addresses the profiles relating to information asymmetries present in commercial practices, in order to establish through the completeness of the information a correct relationship between the contracting parties. The aim is to correct the inadequacies of market information for consumers and to ensure competition between companies. For these reasons, European legislation has intervened through the regulation of commercial advertising and the prohibition of so-called unfair (deceptive) commercial practices[47].

The two further elements required by the GDPR for valid consent are specificity and non-ambiguity (univocity). The former refers to the principle of purpose, which states that the data processing must be

[46] DIRECTIVE 2011/83/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 25 October 2011 on consumer rights, amending Council Directive 93/13/EEC and Directive 1999/44/EC of the European Parliament and of the Council and repealing Council Directive 85/577/EEC and Directive 97/7/EC of the European Parliament and of the Council

[47] Council Regulation (EC) No 1/2003 of 16 December 2002 on the implementation of the rules on competition laid down in Articles 81 and 82 of the Treaty (TFEU). See it in combination with the Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council; ('Unfair Commercial Practices Directive').

performed for specific, legitimate, and disclosed purposes only. The latter affects the concept of clear and affirmative manifestation of the subjective will, represented through an active motion or declaration[48]. Thus, the Regulation requires that the consenting behaviour appears with a visible (objective) manifestation.

From this analysis, it appears that the EU Data Protection regime accords some attention to the subjective sphere of the individual, as the rationale for requiring the elements of clear and straightforward information betrays the legislator's intention to ensure a subjective comprehension. However, this comprehension can be achieved fully only with the proper educational tools. Nevertheless, the Law must adopt the legal fiction of presuming the consent informed if the data controller provides evidence of a signed (accepted) informed-consent form deployed according to the formal requirements of GDPR Article 13 or 14. Therefore, the mechanism of granting an autonomous and free understanding of the contractual conditions to the weak party still passes through the formal compliance-check of the objective legal requirements, which, in turn, represents the external manifestation (effects) of one's activity or behaviour. In this sense, this is not a guarantee at all of the factual comprehension nor, most importantly, of the effective free consent-giving process.

6.3 *Dark patterns (deceptive acts) and vices of the consent*

Indeed, many (somehow deceptive, at least neurologically speaking) techniques in the design of interactive platforms and services have been developed and deployed over time, and they adopt digital marketing strategies, psychological tricks and neuromarketing features, as seen above[49]. This set of techniques has been called

[48] WP29, *WP29 Guidelines on consent under Regulation 2016/679 as Last Revised and Adopted on 10 April 2018* (2017), p.6. https://ec.europa.eu/newsroom/article29/items/623051/en

[49] G. Rupali, J. Singh, *A Review of Neuromarketing Techniques and Emotion Analysis Classifiers for Visual-Emotion Mining*, 9th International Conference System Modeling and Advancement in Research Trends (SMART) (2020); B. MD

"dark patterns" by researchers in the field: Brignull coined the term "dark pattern" in 2010, defining it as «*a user interface that has been carefully crafted to trick users into doing things, such as buying insurance with their purchase or signing up for recurring bills*»[50]. However, in terms of legal classification, these activities should be more correctly framed as deceptive acts aimed at undue influence of the volition process. The EU regulatory framework does not have any positive rule to govern this phenomenon (although it started long before the even first draft of the GDPR, back in 2010), and legal solutions must be found via hermeneutical interpretation of the combined provisions and jurisprudence rulings regarding the vices of the will (in private law). Nevertheless, the US legislators are increasingly focusing on the issue. Indeed, under the California Privacy Rights Act (CPRA), which represents the most up-to-date and GDPR-aligned US regulation on Privacy, the use of dark patterns to obtain consent will render consent totally invalid. According to the regulation, a dark pattern is *"a user interface designed or manipulated with the substantial effect of subverting or impairing user autonomy, decision-making, or choice, as further defined by regulation"*[51].

Johnson, *Addiction and Will*, in «Frontiers in Human Neuroscience», 7, 2013; D. Clifford, *Citizen-Consumers in a Personalised Galaxy: Emotion Influenced Decision-Making, a True Path to the Dark Side?*, in «Social Science Research Network», 2017 SSRN Scholarly Paper ID 3037425 <https://papers.ssrn.com/abstract=3037425> accessed 13 June 2021; J. Lauren E. Sherman et al., *What the Brain "Likes": Neural Correlates of Providing Feedback on Social Media*, in «Social Cognitive and Affective Neuroscience», 13, 2017; A. Alter, *Irresistible: Why We Can't Stop Checking, Scrolling, Clicking and Watching*, Random House, 2017; A. Narayanan et al., *Dark Patterns: Past, Present, and Future: The Evolution of Tricky User Interfaces*, in «Queue Pages», 18, 10, 67, 2020; M. Nouwens et al., *Dark Patterns after the GDPR: Scraping Consent Pop-Ups and Demonstrating Their Influence*, Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (Association for Computing Machinery 2020) <http://doi.org/10.1145/3313831.3376321> accessed 9 June 2021.

[50] "Expert Witness in Dark Patterns, Harry Brignull", <https://testimonium.co/> accessed 23 June 2021.

[51] “"Democratic Societies In the Digital Age 2: Drk Patterns and Online Manipulation | European Data Protection Supervisor" <https://edps.europa.eu/press-publications/publications/podcasts/democratic-societies-digital-age-2-dark-patterns-and_en> accessed 23 June 2021.

Although the norm is pretty generic, the abstract feature of the literal provision would strengthen the protective regime considerably for consent if only the EU legislator would consider amending the – already old – GDPR[52], which does not seem the case[53]. More impactful solutions may come from a specific updating in the regulation. An example is the new EU Regulation issued in October 2022 to govern the activities of online intermediates (the so-called Digital Service Act [DSA])[54]. The legislative text at Recital 67 (non-binding) prohibits interfaces to exploit dark patterns practices aimed at misleading users[55]. Precisely, the DSA aims to ban providers of intermediary online services from adopting deceiving or nudging techniques towards recipients of their services, and, therefore, from using dark patterns to distort or impair user autonomous decision-making process.

6.4 The vices of the consent *in light of the Civil and Penal Law regime for conducts implying undue influence on the decision-making process*

[52] See F. Pizzetti, *Intelligenza Artificiale, protezione dei dati personali e regolazione*, Giappichelli ed., 2018. Consider that the GDPR does not address Artificial Intelligence and smart speakers, Internet of Things, wearables, nor blockchain or smart contracts.

[53] On how dark pattern contracts the general regime above described, essentially for the same reasons, see *How Dark Patterns Conflict with GDPR and CCPA* (*Piwik PRO*, 3 December 2020) <https://piwik.pro/blog/how-dark-patterns-conflict-with-gdpr-ccpa/> accessed 23 June 2021.

[54] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) PE/30/2022/REV/1 - OJ L 277, 27.10.2022

[55] «*Dark patterns on online interfaces of online platforms are practices that materially distort or impair, either on purpose or in effect, the ability of recipients of the service to make autonomous and informed choices or decisions. Those practices can be used to persuade the recipients of the service to engage in unwanted behaviours or into undesired decisions which have negative consequences for them. Providers of online platforms should therefore be prohibited from deceiving or nudging recipients of the service and from distorting or impairing the autonomy, decision-making, or choice of the recipients of the service via the structure, design or functionalities of an online interface or a part thereof. This should include, but not be limited to, exploitative design choices to direct the recipient to actions that benefit the provider of online platforms, but which may not be in the recipients' interests, presenting choices in a non-neutral manner, such as giving more prominence to certain choices through visual, auditory, or other components, when asking the recipient of the service for a decision. […]*»

It is also worth considering the general regime of consent accorded by Contract law[56]. As mentioned before, Contract law does not positively establish specific requirements for consent, which are derived from the analysis of the norm about "errors, violence, and malice". For Contract law, the focus of the relationships is the agreement between parties and how it is formed. The agreement is the conjunction of the two parties' counter-consents as the external manifestation of the will. The Law here does not care about the inner subjective motivations of the parties. Since the consent is formally lawful and there is no objective vice (error, violence, or malice), it is always considered valid. It means that, for the Law, it does not matter if one accepted unfair conditions or behaved against their own interest[57] unless other deceptive elements intervene to undermine the "freedom" of the consent.

Thus, the negotiated agreement requires the willingness of both parties freely manifested through serious and responsible statements. Also, in the Civil Law system, it intervenes the principle of good faith (which automatically applies to every contractual relationship), which aims to protect the very trust that third parties place in the validity of a particular declaration. However, there are sometimes vices of the will that interact in the formation of free consent-giving. Indeed, the contracting party whose consent was given by mistake, extorted with violence, or harvested with malicious intent, may request the cancellation of the contract.

For the sake of this investigation, it is interesting to focus on error and malice[58]. The error is a valid cause for the breach of the contract when it is essential and recognisable by the other party (double condition). It must lead the mistaken party to a false understanding of the reality on essential elements of the contract,

[56] Which is the *"genus"* (gender), while Privacy Law is the *"species"* (specialty). Therefore, the general discipline informs the special one, if not otherwise disposed with a positive norm.
[57] It is said that the Law does not protect the "dumb" person.
[58] Violence is usually considered a mere scholastic example.

meaning only on those elements that the Law establishes as "essential" for a valid agreement. The error can be *de facto* (factual) or *de jure* (legal). The error is considered essential when (i) it refers to the nature or object of the contract; (ii) it refers to the identity of the object of the performance or on a particular quality (e.g. characteristic, legal status) of the performance itself, which must be considered a determinant of the consent in relation to the circumstances; (iii) it refers to the identity or capacity of counter-party if they have been decisive for the consent-giving; (iv) where it was the sole or principal reason for the contract (at the condition that it was a legal error). Outside the provisions of these formal conditions, every error does not affect the validity and efficacy of a contractual agreement.

In relation to malice, the notion must be based on a cross-reference retrieved from the Penal Code[59]. For the Law, the conduct *"is [...] malicious [...] when the harmful or dangerous event, which is the result of the action or omission [...], is foreseen by the agent and intended as a consequence of his action or omission"*. Therefore, for the Penal Law, there must be four concurring dimensions of the will to constitute the malice, i.e. a) the intention, b) the representation, c) the foreseeability of the outcome derived by the conduct (with acceptance of them), and d) the action.

Contract law instead focuses on the objective manifestation by referring generically to "frauds"[60]. Thus, the malicious intent is a cause for cancellation of the contract when the deceptions used by one of the parties have been such that, without them, the other party would not have agreed. Furthermore, when a third party has

[59] Royal Decree 19 October 1930 n. 1398 "Italian Penal Code". (n 10). Article 43. Given the lack of Civil or Penal codifications in EU Law, it has been performed a regulatory reference upon the Italian codification, which, however, must be confronted with other national law. Also, consider that the continental tradition somehow converges in the same kind if regime and, thus, although the wording of the regulation may differ, the norms preceptive contents are usually aligned. Nevertheless, consider proceeding with a granular micro-comparison in this sense that though exceed the scopes of this study.

[60] Royal Decree 16 March 1942 n. 242 "Italian Civil Code" (n 9). Articles 1439 and 1440.

used deceptions, the contract can be annulled if these were known to the contracting party who benefited from them. Moreover, if the deceptions were not essentials in determining the consent, the contract is still valid, although, without them, it would have been concluded under different conditions. In this case, the party in bad faith is liable for damages.

6.5 *AI, dark patterns and autonomy: legal paradigms for deceptive influential activities*

With this general regime in mind, the concept of dark pattern and AI influence can now be addressed from a wider legal perspective. As mentioned above, the kind of influence that an Artificial Intelligence system can perform via dark pattern schemes ranges from misleading information and content to nudging and neuromarketing stimuli, as well as from content design (echo chambers, false scarcity) to coercion (dominant position, take it or leave it). From a Contract law perspective, it is hard to quantify and address specifically all these nuances and therefore, they must be reconducted to the legal paradigm already existent in the positive codification (error and malice). Indeed, as seen above, it is hard for the Law to consider something that cannot be adequately quantified and – therefore – proved. Thus, the phenomenon must be addressed for its factual effects. Nevertheless, Penal Law offers a broader range of understanding for subjective behaviours, which can be considered for disentangling the differences among the various type of influences. For instance, the Penal Code addresses the concept of "instigation" (provocation, incitation) in many provisions[61].

Another example is Article 640, which punishes fraud when someone *"uses artifices or deceiving* [tricks a.n.] *to make unfair gains for himself or others, causing damage to other people"*. The elements of

[61] For instance, in Article 414 "Instigation to commit a crime", among the others.

the crime must concur all together, and so there must be a profit (and it has to be unfair). Besides, this must be linked to third-party's damage. Then, this must be a direct consequence of the "artifices or deceiving". Thus, in order to prove that a particular practice – such as, for instance, discriminatory pricing based on profiling – can be classifiable as an offence, it must be proven both that the party (e.g. the service provider) gained from it, and that the data subject/consumer suffered damage (having the good for an unequal price determined according to the economic and behavioural weaknesses of the consumer himself, based on its profile). Afterwards, it must be proven that the (alleged) malicious practice can be considered a swindle for consumers. This would render the gain unfair, as obtained by tricking the commercial bargaining and by misleading the consumer behaviour. The prove might be offered even considering the neuromarketing and psychological techniques embedded into the dark pattern design of the service features and AI algorithmic activities. Furthermore, these combined elements might be relevant concerning the consumer's good faith[62], which would be breached.

Moreover, design elements embedded or exploited by service provided through active AI interaction with the subject, such as non-disclosed neuromarketing techniques, non-disclosed discriminatory pricing, deceptive content, dark pattern design or, for instance, fake recensions exploitation, as well as content filtering, may be relevant for another particular crime considered by the Italian Penal Law. It is the so-called "abuse of popular credulity"[63] and states that *"Anyone who, publicly, tries with any imposture, even free of charge, to abuse popular credulity is subject, if the fact may result in a disturbance of public order, to the pecuniary administrative sanction from 5,000 to 15,000 euros."*. It is

[62] Which, for the Italian penal system is considered as an aggravating circumstance when breached, and is mentioned in two different crimes, i.e. "theft" (Italian Penal Code, Article 625) and "malicious damage" (Article 635 n. 3) when the good is exposed to the "public faith".
[63] Italian Penal Code, Article 661.

interesting to consider the case in which deceptive or influencing techniques are preventively disclosed and whether this would render them lawful or not, under the scheme of the consent provided from the subject entitled to the right, which presumes the availability of the right[64].

However, another crime that once was considered by the Penal Code[65] and that might have been touched by some particular cases of dependency and behavioural addiction for internet platforms or Apps, is mental manipulation. Article 603[66] once provided that *"anyone who subjects a person to his or her own power, so as to reduce him or her to a state of total subjection, shall be punished by imprisonment of from five to fifteen years"*. Nevertheless, as it was not possible to determine precisely in the norm when and how a brainwash could occur, the crime was declared unconstitutional for indeterminacy. Nowadays, if the conduct is proven, it can be reconducted to the crime of "private violence"[67] which punishes *"anyone who, by violence or threat, forces others to do, tolerate or omit anything"*. In this case, the violent element would be the psychological pressure and/or the eliciting to the behavioural addiction.

All these kinds of influences are relevant for Penal Law but do not affect Privacy and Data Protection or Contract law, unless they are reconducted to the paradigms of legal error or malice (or the crime is proved and punished with a ruling in a legal situation that involves PDP or contractual relationships). Notwithstanding, from this analysis, it can be drawn a line to categorise the type of influence that may occur and their legal relevance concerning the

[64] See P. Moro, *I Diritti Indisponibili. Presupposti Moderni e Fondamento Classico Nella Legislazione e Nella Giurisprudenza*, Giappichelli Ed., 2004.

[65] Nowadays it does not exist anymore as it was declared unconstitutional for indeterminacy by the Italian Constitutional Court with the decision n.96/1981.

[66] Titled "plagiarism" (which means brainwashing) and comes from Latin *"plagium"* (subterfuge), which in Roman law indicated the sale of a man who was known to be free as a slave, or the removal through persuasion or corruption of a slave of another. Therefore, for the Civil Law tradition it evolved into a crime against the individual freedom and its self-determination.

[67] Italian Penal Code, Article 610.

validity of consent, and individual's self-determinative autonomy with it. Different types of influence can be described according to their intensity and effects:

1) Nudge, i.e. the activity of suggesting something[68].

2) Incentive, i.e. the range of activities that intervene to align parties' benefits or interests[69].

3) Influence, i.e. the activity that ranges from instigation to persuasion[70].

4) Manipulation, i.e. the all those activities which imply changing, or however conditioning, one's decision through particular techniques[71].

5) Mental manipulation, i.e. those activities that exploit addictions, neurological tricks, awe and brainwashing[72].

It is also essential to consider these paradigms according to their appearance, i.e. the ability of the individual to know that the data processing system is exploiting some of these features. Informing the data subjects appears to be a crucial aspect for the validity of the consent in the Data Protection regime, and this requirement could be extended to all the relationships that ground the potential

[68] C. R. Sunstein, *Nudging: A Very Short Guide*, in A. de Groot, B. van der Sloot *Handbook of Privacy Studies: an interdisciplinary introduction*, Amsterdam University Press, 2018, pp. 173-180.

[69] E. Fehr, A. Falk, *Psychological Foundations of Incentives, in* «European Economic Review», 46, 687, 2002.

[70] R. W. Benson, J. B. Kessler, *Legalese v. Plain English: An Empirical Study of Persuasion and Credibility in Appellate Brief Writing*, in «Loyola of Los Angeles Law Review», 20, 1986; H. Kim, D. R Fesenmaier, *Persuasive Design of Destination Web Sites: An Analysis of First Impression*, in «Journal of Travel Research», 47, 2008; H. A. Alijda Spelt et al., *Psychophysiological Reactions to Persuasive Messages Deploying Persuasion Principles*, in IEEE Transactions on Affective Computing, 1, 2019.

[71] EDPS, "EDPS Opinion No. 3/2018 on Online Manipulation and Personal Data" (European Data Protection Supervisor 2018) Opinion 3/2018 <https://edps.europa.eu/sites/edp/files/publication/18-03-19_online_manipulation_en.pdf>; R. Calo, *Digital Market Manipulation*, in «George Washington Law Review», 82, 2013; D. Susser, B. Roessler, H. Nissenbaum, *Online Manipulation: Hidden Influences in a Digital World*, in (2019) 4 «Georgetown Law Technology Review», 4, 2019.

[72] B. M. D. Johnson, *Addiction and will*, in «Frontiers in Human Neuroscience», 7, 2013; C. Montag et al., *Addictive Features of Social Media/Messenger Platforms and Freemium Games against the Background of Psychological and Economic Theories*, in «International Journal of Environmental Research and Public Health», 16, 2612, 2019.

influence over the individual's autonomy on some particular kind of personal data processing (e.g. for very dangerous processing or processing affecting very sensitive data), but it would require a positive regulatory intervention. However, consent accorded for establishing these legal data processing relationships should not be considered entirely free because it is determined by needs and decisions by default[73]. Indeed, this consent is often related to products or services that can be considered necessary, or to some extent, essentials. Finally, the types of influence must also be considered in light of the quality of the relationship and the powers of the counterparty. Thus, it is relevant to properly frame the type of influence as above categorised in connection with:

(i) Service providers' systems and their personal information knowledge (unbalanced position);

(ii) Service provider and their factual, economic, contractual and informational dominant position;

(iii) Contract by adhesion, in which data subjects/consumers cannot negotiate the conditions;

(iv) Level of transparency and information provided about the type of influence performed.

It must be noted that a basic level of both legal and technological education would prevent or, to some extent, attenuate these deceptive influences to take place. Nevertheless, as addressing these situations would require a regulatory intervention from the legislator, which is not likely to happen in a short, reasonable time, other practical solutions must be explored to protect data subjects – or, however, weaker parties, from AI systems exploiting dark patterns techniques and their deceptive influential powers.

[73] C. Sunstein, *Deciding by Default*, in «University of Pennsylvania Law Review», 162, 2013.

7. *Gaps and solutions*

It is important to consider that the current norms and legal paradigms concerning the consent have been developed in a time in which the offline relationships were the only option. Accordingly, the regulation considered – and still considers – the influence only in relation to the offline capabilities, i.e. a one to one relationship in which the elements to leverage the influence were constituted by the contractual power/dominant position and the awe one party could play on the other. Even if nowadays our society and individuals' socio-legal relationships belong to the "onlife" paradigm, in which weak parties deal with algorithms instead of human beings, the regulation still adopts the old offline paradigm. Therefore, the range of potential influence, or even manipulation, that comes from profiling activities, AI algorithmic decisions and dark patterns remains uncovered. Indeed, when today one accepts online terms and conditions and privacy terms[74], they are not only in a weaker patrimonial and negotiating position but also, and *a fortiori*, in a position in which the counterparty holds a "multi-monopoly"[75] on necessary goods and a wide range of invasive personal information related to the consumer. This information is even unknown by the consumer itself and ranges from psychological profiles, predictions, behavioral patterns, and individual preferences to economic analyses of one's personality type and inclination to accept unfair prices[76]. This power creates unreasonable informational and negotiating asymmetries and provides commercial players that exploit AI for managing personal

[74] A.M. McDonald, L. Faith Cranor, *The Cost of Reading Privacy Policies*, in «I/S: A Journal of Law and Policy for the Information Society 2008 Privacy Year», 4, 2008, pp. 543-568.

[75] G.M. Riva, M. Barry, *Net Neutrality matters: Privacy antibodies to face information monopolies and mass profiling*, in «Publicum», 5, 2019.

[76] *Ibid.* See Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council; ('Unfair Commercial Practices Directive').

data processing and transaction with the capability to influence individual's consent thanks to nudging, necessary coercion ("take it or leave it" practices), deceptive practices (such as discriminatory pricing and social credit scoring) and neuromarketing techniques]. Unfortunately, the Law is still either too slow and too fragmented to face the issue in proper timing and with a global attitude. Consider for instance, that the discussion about a European Civil Code (which is far from being developed) traces back to the early 2000s[77]. No effective legal remedies are in place to tackle the substantive question of online influence and manipulation. On the one hand, it is auspicial that the issue was appropriately regulated on a high and shared level, such as at the European regulatory one. On the other hand, the society is already on delay to address the issue with the regulation, and considering that legislative processes may involve years of discussions and compromises, there is an urgent need to address this gap from a different point of view. Indeed, waiting for the regulation to be updated appropriately, it is crucial to find some practical solution that could temper the AI unbalanced power of influence. Ethics and Privacy by design can play the role of self-regulative solutions for eliciting the transparency of deceptive commercial practices. In this sense, potential solutions should deploy the programming capabilities to notify the users when their interaction with AI or whatsoever algorithm-based ICT system the deal with, implies some influential behavior. For instance, service providers could be nudged reversely (by proper economic or legal incentive, if not by the ethical urge derived by consumers' unions) to implement such services. Flag notifications can help to achieve this factual goal and render users aware of the type of interaction they undergo. This solution, for instance, could be developed as a browser extension or service-app. Another potential

[77] Italian Penal Code, article 646, in G. Alpa, E.N. Buccico, R. Danovi, *Il codice civile europeo*, Milano, 2001.

self-regulative approach can point in the direction of external Ethics certification (as it happens for secure payment certificates) for such platforms that adopt forms of algorithmic influence.

Finally, the crucial point to highlight is that, in the absence of an immediate regulative response, those forms of ethical, social and economic incentives represent the only way to spur (or force) service providers to elicit these kinds of solutions.

8. *Conclusions*

In this article we addressed the concept of autonomy from both an ethical and legal perspective, with a specific focus on those AI systems and algorithmic techniques that as currently operating undermine individuals' decision-making. The contribution approaches the phenomenon with an interdisciplinary approach, which combines two different hermeneutical perspectives, placing and exploring the issue of autonomy and AI's consent-giving influence at the intersection between Moral Philosophy and Privacy Law. From a philosophical perspective, we have pursued an ethical inquiry rooted in Moral Philosophy on how human autonomy can be influenced by novel algorithmic techniques such as ML and DL. Such ethical inquiry helped us to theorize a double-level impact and constraint raised by AI on human autonomy and specifically on at least two dimensions of autonomy, the epistemological dimension and the moral dimension *stricto sensu*. Thanks to this ethical inquiry, we have showed that also the deepest moral dimension of human autonomy, grounded in the capacity of individuals' reflective endorsement (intrinsic consent to what steer their choices and actions), can be negatively affected and even suspended by algorithmic techniques. From the legal perspective, we have showed then that the concept of autonomy deals with categories and paradigms. For this reason, in continuum with the ethical inquiry, the contribution has addressed the different legal regimes (specifically, Private and Penal Law, in light of Privacy Law informed consent) to investigate how the AI's consent influence activity can be reconducted to key legal concepts and further analyzed via them. Finally, the study has underlined the limits of the regulation and provided a few insights to pave the way for alternative self-regulative solutions, which connects Ethics and Privacy by Design.

The contribution aims to constitute an example of multidisciplinary humanistic inquiry and to prompt the discussion on autonomy and related issues, as related to AI and new technologies. Further contributions could investigate in-depth the single phenomena addressed in the paper and elicit the discussion in the direction of interactive design focused on safeguarding and enhancing of human autonomy and centrality.

GIANLUIGI RIVA è Assegnista di Ricerca in Privacy and Digital Health e Research Fellow presso l'Università degli Studi di Milano

gianluigi.riva@unimi.it

SIMONA TIRIBELLI è Ricercatrice in Filosofia morale presso l'Università di Macerata

simona.tiribelli@unimc.it