AI and Ethics (2022) 2:259–268
https://doi.org/10.1007/s43681-022-00149-5

ORIGINAL RESEARCH

Artificial intelligence and consumer manipulations: from consumer's counter algorithms to firm's self-regulation tools

Nathalie de Marcellis-Warin[1] · Frédéric Marty[2] · Eva Thelisson[3] · Thierry Warin[4]

Received: 6 February 2021 / Accepted: 28 February 2022 / Published online: 29 March 2022

Abstract

The growing use of artificial intelligence (A.I.) algorithms in businesses raises regulators' concerns about consumer protection. While pricing and recommendation algorithms have undeniable consumer-friendly effects, they can also be detrimental to them through, for instance, the implementation of dark patterns. These correspond to algorithms aiming to alter consumers' freedom of choice or manipulate their decisions. While the latter is hardly new, A.I. offers significant possibilities for enhancing them, altering consumers' freedom of choice and manipulating their decisions. Consumer protection comes up against several pitfalls. Sanctioning manipulation is even more difficult because the damage may be diffuse and not easy to detect. Symmetrically, both ex-ante regulation and requirements for algorithmic transparency may be insufficient, if not counterproductive. On the one hand, possible solutions can be found in counter-algorithms that consumers can use. On the other hand, in the development of a compliance logic and, more particularly, in tools that allow companies to self-assess the risks induced by their algorithms. Such an approach echoes the one developed in corporate social and environmental responsibility. This contribution shows how self-regulatory and compliance schemes used in these areas can inspire regulatory schemes for addressing the ethical risks of restricting and manipulating consumer choice.

Keywords Algorithmic manipulation · Deceptive practices · Unfair practices · Compliance · Self-regulation tools

1 Introduction

In digital markets, some companies have gained the power to act as private rule-makers [14]. They shape the informational environment and influence users' experiences and interactions [17]. They can also create unfair conditions for businesses using these platforms and less choice for consumers. For instance, online advertising on these platforms aims to steer consumers' choices towards options they would not have chosen in the absence of the nudge. The idea is to inform about the existence of a recommended option, persuade that a choice is suited to their needs, highlight one option rather than another, or even make the consumer decide that she would not have taken it. Reduction in the space of available choices and manipulation (i.e., pushing consumers to act in a way that is in line with the interests of advertisers and not their own) is not unprecedented. The same is true of the exploitation of cognitive biases and the limited rationality of economic agents, through the implementation of biased choice architectures (dark patterns), whether it is an incentive to make choices that do not conform to the interests of the consumer (nudge) or prevent her from making any preferable choices (sludge[1]). Therefore, advertising creates needs, diverts attention to specific products, and exploits consumer biases. As such, it is subject to regulations. Online advertising multiplies these problems as it is no longer undifferentiated and static. It is specific to

✉ Nathalie de Marcellis-Warin
nathalie.demarcellis-warin@polymtl.ca

1 Polytechnique Montreal, Department of Mathematics and Industrial Engineering, CIRANO and OBVIA, C.P. 6079, succ. Centre-ville, Montreal, QC H3C 3A7, Canada

2 GREDEG - UMR 7321 CNRS Université Côte d'Azur, 250 Rue Albert Einstein, 06560 Valbonne, France

3 A.I. Transparency Institute, Lausanne, Switzerland

4 HEC Montreal, Department of International Affairs, CIRANO and OBVIA, 3000, Chemin de la Côte-Sainte-Catherine, Montreal, QC H3T 2A7, Canada

[1] Following Thaler [47] and Sunstein [44, 46], we could define a sludge as 'a viscous mixture', in the form of excessive or unjustified frictions that make it difficult for consumers, investors, employees, students, patients, clients, small businesses, and many others to get what they want or to do as they wish.

each consumer and can adjust to the instant observation of their decisions, if not their behavior. We are moving from the logic of nudging to hyper-nudging [54].

The risks of algorithmic manipulation are even more significant given that A.I. tools make it possible to predict consumer behavior more precisely in real-time and have diversified, massive data flows, constantly renewed, of good quality, and increasingly from the online and offline worlds. The prohibition of these tools does not make sense insofar as the algorithms used in programmatic advertising and the algorithms for a recommendation, price, etc., generate significant efficiency gains and open the space of choices available to the consumer. In this perspective, the European Commission's proposals for A.I. regulation draw a continuum according to the risks for consumers induced by the algorithms [13]. The algorithms we consider in this article do not involve high-stakes decisions. They do, however, meet the criteria for algorithmic manipulation. This is defined by four necessary and cumulative conditions: the company that develops and implements the algorithm voluntarily and covertly seeks to modify consumers' decisions in its interest by exploiting their vulnerabilities.

A possible avenue would be to empower consumers by providing "counter algorithms" to thwart firms' algorithms or highlight biased choice architectures and algorithmic manipulations [19]. These new tools would allow consumers to assess the quality of the algorithms used by online platforms and their compliance with democratic principles and values, such as the respect of the principle of non-discrimination or the legality of the processing of personal data by the platform in the light of the applicable regulatory framework. However, we should take into consideration informational asymmetries-related concerns. Such supervision tools may be insufficient to detect manipulations. It remains that, at the same time, digital platforms have the moral responsibility to care about the users, not to harm them, and not to put at risk the foundation of societal stability in spreading misinformation polarizing opinions, which may confuse the truth. The European Commission's proposals go toward self-regulation of the effects of algorithms by the companies themselves [13]. This is precisely the purpose of our article, i.e., to investigate the various means available to counteract the risks of harmful manipulation of consumer choices, particularly self-regulation tools. In general terms, self-regulation is defined as the protocols firms put in place to "pre-empt or supplement governmental rules and guidelines that govern their activities" [12]. Damages related to consumer manipulation compromise fundamental market values, such as the freedom of choice. The use of certain market practices cannot be prohibited ex-ante to the extent that they may generate net efficiency gains. However, damages are difficult to detect, characterize as anti-competitive according to the criteria used in competition law, and above all, very difficult to remedy. These damages can be difficult to detect by governments as well as companies. The contribution we want to make is about empowering companies to assess whether they have ethical practices.

We propose to detail and discuss possible solutions to prevent such consumer harm. We show that while providing consumers with counter-algorithms may be a partial solution, it may effectively encourage compliance models. Inspired by the experience of corporate social and environmental responsibility policies, these co-regulatory models are based on a well-understood self-interest on the part of firms. Failure to comply exposes them to consumer backlash or reputational damage from stakeholders. However, for this regulation to be effective, it is necessary that the transparency and sincerity of the results of the assessment tools be guaranteed by external norms, standards, or tools. These can help to regulate the (self-) regulator.

Our paper is organized as follows. Section 2 shows how A.I. development could create exacerbated and amplified risks for consumers. Section 3 presents regulatory and non-regulatory measures to tackle these risks. It clarifies why self-regulation tools help manage negative externalities and shows the limits of regulatory and consumer empowerment tools. Section 4 introduces international examples of audit framework of A.I. and specific tools that can help firms self-regulate while comparing themselves with other firms.

2 Algorithmic consumer manipulation's challenges: dark patterns, nudges, and sludges

Following Wagner and Eidenmüller [51], the dark side of A.I. implementation for consumers can be described through three parallel trends. The first one consists of an enhanced capacity to engage price discrimination-based strategies converging to near-from perfect price discrimination. Being able to infer the reservation price of each consumer may allow siphoning its transaction surplus. A second trend also implies the personalization of the offers. The algorithms' proposals may systematically exploit behavioral bias by engaging each consumer in welfare-reducing transactions. The third trend also consists of implementing A.I.-based tools to create micro-targeted ads and recommendations to shape consumers' preferences and steer them in specific consumption patterns.[2] We considered here two types of damage that may occur to the consumer. The first type of damage is the reduction of options available. The second form of damage concerns the manipulation of choices. The range of

[2] A.I.-based recommendation tools may also deprive consumers of access to services, such as the loans market. Such refusals are sometimes based on social bias reflected and aggravated by algorithms bias [11].

available solutions is not artificially closed, but how people decide is biased towards desired solutions. It can result in a risk of unbalanced transaction processing. The ability to predict the characteristics of the consumer (technical expertise, ability to pay, etc.) makes it feasible to construct dedicated offers that lead to the extraction of the entire economic surplus (which would not be possible with uniform prices or imperfectly differentiated prices).

The manipulation of consumers is based on exploiting their cognitive or emotional weaknesses to orient their decisions in a direction that does not correspond to their initial and/or reasoned preferences. Such strategies have long been known in marketing and advertising. The introduction of digital technology and artificial intelligence introduces a change in scale and efficiency, which, as we will see in our following sections, poses a double question: that of the capacity of consumers to detect and respond to these risks and that of external regulations to respond to them. The change induced by the A.I. transforms static and undifferentiated manipulation strategies into "dynamic, interactive, intrusive, and incisively personalizable choices architectures—decision-making contexts that can be specifically designed to adapt and to exploit each individual user's particular vulnerabilities" [43].

The argument stems from the personalization of the action on the consumer's decision and its ability to adjust in real-time from a constant recalculation from intermediate consumer decisions. It is no longer just nudging, as we will see below, but hyper-nudging in that it is specific to each consumer and constantly adjusted independently [54]. An additional characteristic must be considered: it is more a question of manipulation than of coercion in that the consumer cannot be fully aware that her choice is coerced or altered.

The manipulation, which consists of making a decision that would not have been taken spontaneously and which goes in the interests of the manipulator, can take two forms: the first is to play on the choices' space, in other words, to change the options theoretically open to the consumer; the second is to play on her decision-making process, that is, to alter the way she can understand the different options and their consequences. We will successively consider these two strategies in the two sub-sections below.

2.1 Reducing the consumer's choice space

A.I. enables changes that break the usual business paradigms. A.I. can be incorporated to influence consumer behavior via increasingly more targeted recommendations or recommendations to limit consumer choice. The latter may see their range of choices reduced depending on their past consumption or the customer segment that the algorithm links them. A.I. is a predictive analytics tool based on data modeling [2, 3]. In other words, they can be isolated in a “filter bubble”. The consumer’s choice can be more than a little biased. It can be pre-empted through algorithmic predictions and thus, to some extent, "constrained." For instance, in the field of e-commerce, marketplaces may use their algorithms to evolve from a model of shopping-then-shipping to a shipping-then-shopping one [3]. The customer may incur an incidental cost in shipping the item back even if it enjoys the possibility to return the shipped item free. These practices could raise significant ethical concerns. Doesn't the algorithm replace consumer choice? What about the situation of a consumer who has decided to break with his or her past consumption habits, especially if these are addictive and harmful in terms of well-being?

Algorithms can also help manipulate consumer choices through a precise understanding of behavior or an accurate estimate of maximum payment capacity [6]. Indeed, as Ezrachi and Stucke [15] noted, in an age whose economy is data-driven, personal data on user behavior, preferences, weaknesses, and habits are the new currency for advertising and marketing-dependent business models. These capabilities require monitoring massive, diverse and continuously updated data, combined with various analytical tools to help predict other operators' strategies [32, 33].

2.2 Consumer's behavior manipulation

The manipulation is not intended to constrain the consumer's choice but to alter it. Indeed, risks induced by algorithmic manipulation are the more significant that consumers may not be aware of the practices or may not perceive the damage induced. A.I. allows firms to anticipate and control consumers' decisions [29].

The influence that the algorithm can then play is hidden and deceptive insofar as it leads it to act in a direction contrary to its interests. It is about instilling false beliefs and shaping the consumer's beliefs and expectations favorable to the interests of the manipulator. These strategies, which are based on the cognitive biases of agents, have been known for a long time but have taken on an unprecedented scale with the possibilities of personalization and instant automatic adjustment specific to the digital aspect.

The profit motive ultimately drives markets. The latter plays a crucial role in market efficiency, alerting participants to potential risks. There is abundant literature on the concept "value of information", and authors have looked at the value of information in the context of price strategies on digital markets [52, 53] as well as in regulatory contexts [31]. Using A.I., some firms have access to the aggregated information and customers' value of information, notably recommender systems. With so much valuable information and data, A.I. can be used to partly model consumer behavior and create an incentive for purchase at the right time. For example, it is

possible to implement drip pricing strategies and price partitioning. The customer can be engaged in a purchasing process by attractive features such as low pricing, which leads to a surprise when the total price is discovered. Spending time on subsequent pages will prevent her from remembering the competition's price or prevent her from starting the search process from the beginning [38].

The notion of dark patterns illustrates these practices, which can be aggravated by A.I.-based algorithm implementation [42]. Dark patterns are user interfaces whose designers knowingly confuse users, make it difficult for users to express their actual preferences, or manipulate users into taking specific actions [28]. They cover all the profiling methods, algorithmic proposals, or user interfaces that can restrict the ability to make a free and informed choice on the consumer's part. Dark patterns are also called dark nudges or bad sludges. Therefore, they cover strategies that increase the opacity of consumers' choices, making it more difficult for them to express their preferences freely, or that leads them to make decisions that they would not have made spontaneously (see, for instance, [21, 46]). The dark pattern can be produced by the design of the site or the modes of presentation of the choices [1]. It is not just a matter of using consumers' and users' vulnerabilities to induce them to make choices they could rationally try to curb, but also a matter of eliciting these preferences [34]. Thus, "manipulative by design" devices exist [35].

Manipulation can take on an unprecedented scale with digital technology in so far as the manipulation no longer concerns biases generally present in the population but biases specific to each micro-segment of consumers or even to each consumer. The biases based on an architecture of choice can now be based on a dynamic architecture dependent on the decisions immediately taken by each Internet user. Much like Waze recalculates a route to invite you to join the recommended route after the motorist's recommendation is ignored for the first time, the algorithm can re-invite the consumer to return to a preferable path with the seller.

A.I.'s development could make these strategies more useful by allowing a better understanding of consumer behavior after closely linking it to a given segment based on observed and inferred characteristics. In other words, A.I. can promote the personalization of prices and the personalization of manipulations. Indeed, as the Stigler Committee on Digital Platforms notes ([42], p. 238), the use of sludges will have multiplier effects with A.I.: “Dark patterns are often used to direct users towards outcomes that involve greater data collection and processing. Additionally, the proliferation of data-driven computational methods allows firms to identify vulnerabilities of users and to target specific users with these vulnerabilities” [21].

A.I. can make it possible to determine which stimulus to present to a consumer and when to do so based on an increasingly refined prediction of its characteristics and, therefore, also of its inferred weaknesses. A.I. shifts manipulation from a logic of nudging to a logic of hyper nudging [54]. Influence is no longer undifferentiated: it is micro-targeted. It exploits the vulnerabilities of each consumer, whether based on profiling resulting from their identification in the online world (but also offline through data purchases) or on a constantly updated prediction based on observation of their behavior. Yeung [54] considers that "big data-driven decision-guidance techniques" constitute hyper-nudges in that, unlike traditional undifferentiated and static techniques, they are dynamic, interactive, intrusive, and personalized. Firms' implementation of A.I. tools may also lead to the imposition of unstable contractual conditions. The use of algorithms can lead to an excellent segmentation of customers, allowing them to propose almost personalized prices. The latter's problem is that a discriminatory price makes it possible to confiscate the consumer's total surplus at the level of the consumer's maximum propensity to pay. There is no damage in economic terms to efficiency, but an undue transfer of welfare compared to the distribution that would prevail in perfect competition.

Therefore, the notion of an "augmented dark pattern" can cover several modes. The one we have just seen corresponds to the manipulation of transaction costs. Consumer well-being is degraded by exploitative abuses in the form of a confiscated personalized price (of its surplus), an offer with a degraded price/quality ratio, or in the form of barriers to exit. The notion of a dark pattern also applies to manipulating consumer behavior based on an acceptable identification of their characteristics and, more precisely, their weaknesses.

3 Preventing algorithmic consumer manipulations: from regulations to counter-algorithms

Regulation can be seen as a tool to help prevent such manipulation. However, let us look at the regulatory frameworks that are beginning to be developed, such as the E.U. Commission through its draft on A.I. regulation [13]. The requirements proposed may not be sufficient or not be sufficiently adaptable when considering the speed of the development of algorithms. This is especially true when it is limited to obtaining consumer consent or strengthening information obligations. The latter faces an information overload and presents a certain number of biases that could hinder the actual effectiveness of these ex-ante measures: their myopia regarding the risks associated with a reduction in privacy etc. An avenue would be to empower consumers by providing algorithmic tools to highlight biased choice architectures and algorithmic manipulations or thwart firms’ algorithms.

3.1 Understanding firms' algorithms and highlighting biased choice architectures

In a context where the consumer wants to understand the algorithms and the conditions of using her data, the obligation to provide information about firms' algorithmic practices could be required. However, this may be ineffective to the extent that consumers face information overload [20, 23]. As such, transparency may be insufficient [18]. The obligation of transparency incurs a cost for firms without improving consumers who do not commonly invest their time and cognitive resources in reading and understanding contractual clauses and arrangements [4]. As Yeung [54] points out, consumers do not have the necessary diligence to read the clauses before accepting them: lack of time, understanding, and intense pressure to access the service. Acceptance is a sine qua non of use. Using algorithms here would be useful to mobilize A.I. tools to detect the determinants of the contracts offered to consumers [36].

Another avenue that would be interesting would be to empower consumers by providing algorithmic tools ("counter algorithms") to detect biased choice architectures and algorithmic manipulations. These new tools would allow consumers to assess the quality of the algorithms used by online platforms and their compliance with democratic principles and values, such as the respect of the principle of non-discrimination or the legality of the processing of personal data by the platform in the light of the applicable regulatory framework.

However, it is worth noting that more sophisticated consumers may better perceive the risks and are more likely than average to implement countermeasures. Conversely, the most vulnerable consumers are the most exposed to manipulation. These naive, helpless, and defenseless digital consumers—in terms of these "fragile digital consumers"—will be doomed to find themselves at the mercy of increasingly powerful companies, busy taking advantage of them and manipulating them ([9], p.174).

3.2 Thwarting algorithms

Another option consists of using counter-algorithms to not only make its attributes less decipherable and its online experience less observable by the firms' algorithms,[3] but also in transferring the purchase decision to a dedicated algorithm that will be able to buy at the right moment[4] and make successive requests to exploit the biases of the sellers' algorithms [19]. It is a matter of re-equilibrating the informational balance to benefit the consumers without aggravating their informational overload.

In this perspective, using A.I.-based tools may help to make public regulation more effective and empower consumers [10]. In this perspective, using A.I. based empowering tools may aim at "protecting consumers against unwanted monitoring and data collection, supporting consumers and their organizations in the detection, and contesting of unfair use of A.I. […] enabling consumers and their organizations to control the fairness of commercial practices and so on" ([10], p.5151). A.I. based tools may be oriented toward the different objectives of consumers protection law, e.g., "the weaker party protection principle, privacy protection, the regulated autonomy principle, and the non-discrimination principle" ([10], p.5153). Empowering consumers helps them exert a countervailing power that limits the firm's ability to extract an undue part of the transaction surplus, discriminate among them, and exploit and aggravate behavioral bias to their profits.

To sum up, consumers themselves can use algorithmic tools to detect or counteract possible company manipulation. This can be done through legal requirements, distributed monitoring devices set up by non-profit institutions, or consumer-specific over-the-counter algorithms. They can play with firms' algorithmic strategies or even deceive them by sending false signals.

4 Self-regulation on behalf of the firm itself and its stakeholders

The tools available to consumers may prove insufficient to respond to the risks presented above. Indeed, manipulation by A.I. is challenging for consumers to understand and particularly difficult to detect and control by an external regulator that would be over the loop and could only intervene ex-post. The obligations to certify algorithms, collect and store data on interactions, etc., could be excessively penalizing for firms and congestive for the regulator. Therefore, the regulation should be integrated into one form or another in the code. In other words, it is the firms that can most effectively prevent or stop damage if it is detected as algorithms evolve. However, it would be possible to go beyond the rationale of regulation in a top-down perspective to consider the option

[3] See the cases of Tor or Anonabox, for instance. Sunstein [45] has introduced a distinction between an architecture of control proposed by the algorithm and based on past choices and architecture of serendipity in which the proposals made to the consumers only reflect the average choices of the platform's users.

[4] Algorithmic tools as Shadowbid propose to consumers using marketplaces to "state their personal reservation price [and] then purchases automatically when the price drops below this threshold" ([51], p. 589). The literature on algoactivism can also be investigated to draw parallels between consumers' possible counterstrategies and platforms depending on contractors, such as car drivers, for instance [25].

of self-supervision. Indeed, firms are stakeholders in preventing possible damage to consumers, knowing that the latter induces both reputational and legal risks and may not comply with the ethical commitments of the members of the firm and its funders.

4.1 Motivations and process for self-regulation

The motivations for a firm to self-regulate are numerous, for instance, its reputation, brand recognition, and desire to be consistent with its announced core values. This is where ethics can play a crucial role in a firm's decision to perform a compliance assessment of its data use situation. This history of firms, A.I., and ethics are strongly intertwined. Self-regulation refers to voluntary protocols put in place by firms to preempt or supplement governmental rules. This simple definition requires some nuance. Indeed, self-regulation can apply at the firm and industry levels involving industry associations and sometimes collaborations with governments [30]. Related to this, self-regulation can be implemented by the firm itself or by another non-governmental entity.

Another nuance is called "meta-regulation," referring to "ways outside regulators deliberately, rather than unintentionally, seek to induce targets to develop their own internal, self-regulatory responses to public problems" [12].

In the context of the novelty related to A.I. and platforms, our use of the self-regulation concept relies on definitions, self-regulation, and meta-regulations. It benefits from not being restrictive: we consider self-regulation even if incentivized by implicit threats or signals from outside regulators.

About the efficiency of self-regulation, on the one hand, in the context of the A.I. novelty, firms may have better information to find solutions to public problems. Firms are indeed likely to have greater knowledge about their operations.

On the other hand, critics explain that self-regulations are not a jack-of-all-trades for all problems. The main reasons might be conflicts of interests and/or the lack of incentives [8]. In the context of A.I., for some authors, the race to A.I. is accompanied by the race to A.I. regulations, nationally and even globally [39, 40]. The regulatory playground can take multiple shapes: new regulation, deregulation, re-regulation, soft-law measures, principle-based or rule-based.

In our view, the uncertainty related to A.I. innovation associated with its novelty and the lack of a systematic understanding of its impacts necessitates a mix of self-regulation and government regulation to supervise A.I. developments.

The supervision of corporate governance can be public or private. In many jurisdictions, complaints to the regulator are considered the most effective enforcement mechanism. Public oversight takes time to be implemented for emerging technologies. Therefore, we propose considering private supervision to complement public supervision, considering the firms' responsibility and moral consciousness towards their stakeholders. Considering both legal risks and possible consumers' backlash, firms can be incentivized to use A.I. to maximize their profits and internally monitor their compliance with consumer law [29], p.171, even if they do not use A.I.

4.2 Audit frameworks for A.I. systems

Many international organizations are engaged in developing audit frameworks for A.I. systems. For instance, the United States published an *A.I. Accountability Framework* to help managers ensure accountability and responsible use of A.I. in government programs and processes [49]. This framework is organized around four complementary principles, which address governance, data, performance, and monitoring. The framework describes critical practices for federal agencies and other entities considering, selecting, and implementing A.I. systems for each principle. Each practice includes a set of questions for entities, auditors, and third-party assessors to consider and procedures for auditors and third-party assessors. This flexible approach plays a vital role because a legally binding instrument takes time to be adopted. A.I. is a transformative technology with applications in medicine, agriculture, manufacturing, transportation, defense, and many other areas. It also holds substantial promise for improving government operations. Federal guidance has focused on ensuring A.I. is responsible, equitable, traceable, reliable, and governable. This approach echoes other initiatives, notably concerning algorithms used by public authorities. As the [5] or the [7], these aim at favoring the implementation of algorithmic impact assessment tools. For instance, the Algorithmic Impact Assessment (AIA) is a mandatory risk assessment tool intended to support the Treasury Board's Directive on Automated Decision-Making [22].

The proposed tools are similar to environmental impact assessments, human rights ones, or data protection ones. In this context, algorithmic impact assessment consists of implementing tools and procedures to evaluate and monitor the effects of an algorithmic system that is likely to negatively affect people throughout its life cycle to limit any adverse effects (see [27], for a comprehensive view).

Various control methods could be envisaged. The first possibility could be regulation aimed at the ex-ante certification of algorithms. The second would be setting up regulatory sandboxes [37]. However, both are matters of supervision implemented to a certain extent ex-ante. The supervision of algorithms must be continuous and can be implemented more effectively by the company that deploys them. The latter must then put in place tools for evaluating the effects of its algorithms. Companies can implement

self-regulation procedures based on a risk-based approach. The more likely the algorithm produces undesirable effects, the more stringent the control measures must be. In other words, as far as we are concerned, it is a question of ensuring that the implementation of A.I. is not likely to harm the ability of the stakeholders affected to make free, independent, and informed decisions regarding their consumption choices. The risks of low effectiveness or false-negative bias that may be a concern in self-regulation schemes can be limited through several mechanisms. These may include transparency of oversight arrangements, involvement of stakeholders in the process, certification by independent bodies, or the use of instruments developed by non-profit institutions. However, if the aim is to measure the effects of algorithms within the framework of a specific regulation, the firm's action may first be part of an ethical and voluntary approach.

Ethics focusing on intentionality, data-driven companies should demonstrate that they have genuinely good intentions to be qualified as ethical [17]. This raises the question of the maleficence or beneficence of business models. In principle, the intentionality of action is necessary for an ethical evaluation [17]. However, Ethics is also a matter of evaluating the consequences of a decision on people affected [17]. Responsibility can therefore be engaged without fault or bad intention. It should be emphasized that the purpose of legal rules is to sanction not so much intentions but harmful effects on the consumer. In other words, ethics goes beyond the management of legal risk and involves a commitment by the company per its values and contribution to society. The commitment, therefore, naturally goes beyond monitoring compliance with legal rules.

4.3 Discussion

The E.U. draft Regulation on A.I. outlines several solutions leading to delegating part of the regulation to the firms concerned.[5] This logic is that of procedural regulation: firms must integrate compliance with the rules of law into their internal procedures, possibly into the code of their algorithms. This logic of compliance and self-regulation (whether direct or transferred to an independent third-party certifier) may not be limited to highly consequential decisions; It participates in the movement of corporate social responsibility and ethical commitments.

Indeed, in terms of self-assessment, firms are starting to show some new alignments with societal values in the ethical context. Although it varies across firms and industries [26], it is interesting to see how firms include environmental, social, and governance (ESG) concerns. Companies have a specific interest in guaranteeing the compliance of their actions to stakeholders indeed, and the use of A.I. falls into the social and governance aspects of the ESG goals, to name a few: managing legal and reputational risk,avoiding a consumer backlash; be accountable to investors who wish to engage in responsible practices. Ethics is not just about compliance or minimizing legal risk in such a context.

It remains that such a self-regulation approach can be put at risk in terms of effectiveness. Providing firms with a wide margin of discretion may run counter to the objectives of regulation, as [39, 40] show about the proposal for a European regulation on A.I. presented by the Commission in April 2021.

The obligations envisaged in the Regulation only relate to high-risk A.I. systems. If the issue is about fundamental rights, we could apply the same analysis grid to manipulations against consumers. We could consider that the absence of manipulation is part of consumers' freedom in the market. The draft regulation obliges the companies concerned to set up a risk management system while leaving them a wide margin of discretion on the system put in place. It is up to the company to decide whether the system is appropriate and suitable. The risk of self-regulation is linked to the combination of vague criteria and the delegation of a wide margin of discretion. Smuha et al. [39, 40] propose an ex-ante certification of algorithms, which we could extend to a certification procedure for self-evaluation tools.

To guarantee the effectiveness of their risk management self-assessment procedure, companies already use certification mechanisms such as the ISO 31000 standard for risk management or the ISO 26000 standard for social responsibility. The idea in the latter case is to help businesses clarify what social responsibility is, help businesses and organizations translate principles into practical actions, and share best practices relating to social responsibility. The stakes regarding algorithmic responsibility related to possible consumer manipulation mean that support for self-assessment tools may prove crucial for the next step towards standardization and possibly regulation.

Third-party assessments and audits are essential to achieving these goals if organizations demonstrate independence. Audit assessment can be performed ex-ante before highly automated systems are implemented or ex-post as an audit trail to analyze failures and help assess accountability [16]. They can incorporate a process considering product development, deployment, and acquisition. Ideally,

[5] As mentioned above, the European Commission's proposal draws a continuum between algorithmic systems prohibited insofar as they would exploit vulnerabilities of specific groups, ex-ante obligations for systems involving high stake decisions, and finally, transparency obligations for systems involving less significant risks [50]. The algorithms that concern us fall into this third category. The E.U. Unfair Commercial Practices Directive already imposes conditions on their use, and the General Data Protection Regulation (GDPR) imposes transparency and explainability requirements on companies [24]. Under the A.I. proposal, only high-risk algorithms are obligated to set up a quality system to carry out a lifecycle impact assessment.

this process will be the most effective during conception through product development before building the product. The features can be modified to mitigate the risks if a risk is identified. An impact assessment team will then review the impact of the product before its deployment on the market or before a new product acquisition. If the product is not robust enough based on the company's values, the due diligence process will mitigate the risks and assess the impact of the products for human rights, human autonomy, etc.

A multistakeholder consultation should validate that the new product is aligned with the values promoted by the company in its Ethics Statement or Code of Conduct. If this process is spread in the supply chain with a contractual agreement, this can positively impact business models at scale. Therefore, the pursuit of corporate social responsibility and business ethics policies can be a competitive advantage for organizations [41]. In each company, a red team could identify a group of people being ethically responsible for dealing with an ethical dilemma on a case-by-case basis. This team could monitor its risks and products and services with a dashboard based on a risk assessment framework. This will contribute to managing social, environmental, and safety risks while building social acceptance and public trust in highly automated systems. Such a dashboard would offer companies information to intervene responsibly to shape a culture that is not harmful to its users.

These self-regulation tools optimize the management of the risks induced by highly automated systems. They enhance traceability, which will result in greater transparency and explainability. In the literature, some scholars propose to develop an audit trail for highly automated system operations based on A.I. [16]. The main argument is that A.I. could be used to "identify risks using real-time monitoring and analysis or provide post-event insight into the context surrounding the accident". In this context, data should be stored in a "publicly available data repository" to foster transparency. This proposal is based on an analogy with the aviation industry, which uses audit trails for their systems using black box flight data recorders (FDRs). Used as an ex-ante safety mechanism, it can decrease safety risks like was the case for the aviation industry, considering the complexity of their systems and processes [16]. In this industry, the audit trail is based on a black box that collects evidence of systems actions and the surrounding context for analysis after near misses and failures. It will be essential to adapt the audit trail to different contexts and in the event of an accident to complement it with a human process for the ex-post investigation. Financial trading systems have already implemented recording systems for each trade for basic business needs, but more information may be needed about how machine-aided decision-making was performed.

In the absence of adequate formal legal frameworks, self-regulation methods can be valuable for improving digital readiness. Thelisson, Morin, and Rochel [48] argue that "a Digital Responsibility Index" can play a central role in restoring trust in a data-driven economy and creating a virtuous circle, contributing to sustainable growth. Such indexes consider the interests of all stakeholders, including employees, customers, and the broader community and investors. They encourage companies to behave diligently. Such indexes are accessible to shareholders and can significantly influence the shareholder value of companies by rewarding transparent and diligent behavior. They will show how companies position themselves vis-à-vis their competitors. They will encourage the adoption of similar behavior by economic actors in the same market and improve their rating, thus in their responsible and transparent behavior towards consumers and the community. Making visible what economic actors are doing encourages companies to adopt socially responsible behavior.

These trust indices can impact the behavior of the company and that of the consumer. On this basis, the company can identify its shortcomings and invest its resources in essential issues to both its shareholder value and the common good. It will gain efficiency by improving its internal processes and demonstrating responsible management to all stakeholders. By including the scores in its annual report and on its website, the company increases confidence in its corporate governance.

5 Conclusion

The manipulation of consumers' online choices by A.I. constitutes a significant risk that must be subject to ex-ante regulation. The tools for sanctioning practices contrary to competition and consumer protection rules do not provide a satisfactory solution because the damage is difficult to characterize and because it appears necessary to prevent its occurrence. It is not only a question of protecting the consumer himself but also of protecting fair competition. Indeed, misleading commercial communications lead both to consumer deception and competition distortions.[6]

However, while guaranteeing fairness on the market presupposes the definition of ex-ante rules, the very nature of the technological dynamic presupposes a combination of flexibility in the framework and the gradation of obligations according to the risks involved. Among the proposals for the regulation of A.I., the one formulated by the Commission in April 2021 offers interesting avenues for reflection. The nature of the algorithms used in the sector of interest does not lead us to link them to those that should be prohibited or even those involving high-stakes decisions. These algorithms belong to a third category

[6] https://www.ftc.gov/system/files/documents/public_statements/1582914/final_commissioner_chopra_dissenting_statement_on_zoom.pdf.

within the meaning of the European proposal, which has the advantage, in our view, of combining a requirement to carry out algorithmic impact studies—not only before the study but also on an ongoing basis—and a transfer of this assessment to the company implementing the algorithm.

This co-regulation is a particularly relevant solution for algorithmic tools for which the ex-ante assessment of the effects on consumer choices is difficult to evaluate. Pure compliance solutions, such as the transparency or explicability of algorithmic results, may not be sufficient. Other solutions could be envisaged, such as the use of regulatory sandboxes, which would allow regulation to be adjusted to observe the effects of algorithms in an experimental framework. However, the solutions of co-regulation based on the self-regulation by the firms themselves, in a logic comparable to those implemented in the fields of fundamental rights or societal and environmental responsibility, seem to be particularly interesting.

This advantage could appear in a logic of economic analysis of law: the firm that develops and implements algorithms enjoys an informational advantage over other stakeholders. It is both the actor who can avoid the materialization of the risk (or modify its decisions as quickly as possible to reduce its severity) and the one who can do so at the lowest cost in collective terms. However, the logic of self-regulation echoes a societal commitment of companies.

The specter of self-regulation serving only the interests of the regulated entity can also be highlighted through two tools. The first is the transparency of the self-regulation of algorithmic risks, for example, through stakeholders' involvement. The second, outlined in this contribution, is external assessment tools that can be subject to certification procedures. These logics seem to reconcile the gains linked to the use of A.I. and the prevention of the risks that may be linked to it, not only for the other market players but also for the company that implements the algorithms. Therefore, they can combine efficiency gains related to innovation, and effectiveness of regulation, and trust in A.I.

Funding Not applicable.

Code availability Not applicable.

Declarations

Conflicts of interest No conflict of interest for this manuscript.

References

1. Acquisti, A., Brandimarte, L., Loewenstein, G.: Secrets and likes: The drive for privacy and the difficulty of achieving it in the digital age. J. Consum. Psychol. **30**(4), 736–758 (2020)
2. Athey, S.: Beyond prediction: Using big data for policy problems. Science **355**(6324), 483–485 (2017)
3. Agrawal, A., Gans, J., Goldfarb, A.: How AI Will Change Strategy: A Thought Experiment. Harvard Business Review. https://hbr.org/2017/10/how-ai-will-change-strategy-a-thought-experiment (2017).
4. Bakos, Y., Marotta-Wurgler, F., Trossen, D.R.: Does anyone read the fine print? Consumer attention to standard-form contracts. J. Legal Stud. **43**, 1 (2014)
5. British Office for Artificial Intelligence : Understanding artificial Ethics and Safety- Understand how to use artificial intelligence ethically and Safely https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety (2019)
6. Calo, M.R.: Digital market manipulation. George Washington Law Rev. **82**(4), 995–1051 (2014)
7. Canadian Treasury Board Secretariat. Directive on Automated Decision-Making. Policy on Service and Digital. https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592 (2020)
8. Coglianese, C., Mendelson, E.: Meta-regulation and self-regulation. In: Cave, M., Baldwin, R., Lodge, M. (eds.) The Oxford Handbook on Regulation, pp. 146–168. Oxford University Press, Oxford (2010)
9. Colangelo, G., Maggiolino, M.: From fragile to smart consumers: Shifting paradigm for the digital era. Comput. Law Secur. Rev. **35**(2), 173–181 (2019)
10. Contissa, G., Lagioia, F., Lippi, M., Micklitz, H-W, Pałka, P., Sartor, G., Torroni, P.: Towards Consumer-Empowering Artificial Intelligence. Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI-18), pp.5150–5157 (2018)
11. Citron, D.K., Pasquale, F.: The scored society: Due process for automated predictions. Washington Law Rev. **89**, 1 (2014)
12. Cusumano, M.A., Gawer, A., Yoffie, D.B.: Can self-regulation save digital platforms? Ind. Corpor. Change (2021). https://doi.org/10.1093/icc/dtab052
13. E.U. Commission: Proposal for a Regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) (2021)
14. E.U. Commission: The Digital Services Act Package, https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package (2020)
15. Ezrachi, A., Stucke, M.E.: Digitalisation and its impact on innovation. Working Paper 2020/07, October R&I Paper Series, European Commission https://wbc-rti.info/object/document/20829/attach/KIBD20003ENN_en.pdf (2020)
16. Falco, G., Shneiderman, B., Badger, J., Carrier, R., Dahbura, A., Danks, D., Eling, M., Goodloe, A., Gupta, J., Hart, C., Jirotka, M., Johnson, H., LaPointe, C., Llorens, A., Mackworth, A., Maple, C., Pálsson, S., Pasquale, F., Winfield A., Yeong. Z.: "Governing A.I. safety through independent audits" Independent audit of A.I. systems serves as a pragmatic approach to an otherwise burdensome and unenforceable assurance challenge", Nature – Machine Intelligence, VOL 3, July, pp 566–571, https://t.co/ksb7ZYozHi (2021)
17. Floridi, L.: Faultless responsibility: on the nature and allocation of moral responsibility for distributed moral actions. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. (2016). https://doi.org/10.1098/RSTA.2016.0112
18. Gal, M.: Algorithmic challenges to autonomous choice. Michigan Telecommun. Technol. Law Rev. **25**(1), 59–104 (2018)

19. Gal, M., Elkin-Koren, N.: Algorithmic consumers. Harvard J. Law Technol. **30**, 309 (2017)
20. Grafanaki, S.: Drowning in big data: Abundance of choice, scarcity of attention and the personalization trap, a case for regulation. Richmond J. Law Technol. **24**(1), 1–66 (2017)
21. Gray, C.M., Kou, Y., Battles, B., Hoggatt, J., Toombs, A.: The Dark (Patterns) Side of UX Design. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, New York, NY, USA: Association for Computing Machinery, 1–14. https://doi.org/10.1145/3173574.3174108 (2018)
22. Government of Canada: Algorithmic Impact Assessment, https://open.canada.ca/data/en/dataset/5423054a-093c-4239-85be-fa0b36ae0b2e/resource/7381144a-8e88-4d74-b83b-746c13de4093 (2020)
23. Jin, G.Z. Wagman, L.: Big data at the crossroads of antitrust and consumer protection. Information Economics and Policy, https://lwagman.org/BigDataAntitrustCP.pdf (2020)
24. Kaminski, M., Malgieri G.: Algorithmic impact assessments under the GDPR: Producing multi-layered explanations, international data privacy law, https://doi.org/10.1093/idpl/ipaa020 (2020)
25. Kellog, K.C., Valentine, M.A., Christin, A.: Algorithms at work: the new contested terrain of control. Acad. Manag. Ann. **14**(1), 366–410 (2020)
26. Kouloukoui, D., De Marcellis-Warin, N., Armellini, F., Warin, T., Andrade, T.E.: Factors influencing the perception of exposure to climate risks: evidence from the world's largest carbon-intensive industries. J. Clean. Prod. (2021). https://doi.org/10.1016/j.jclepro.2021.127160
27. Lehagre, E.: L'évaluation d'impact algorithmique : un outil qui doit encore faire ses preuves, Rapport Etalab, https://www.etalab.gouv.fr/wp-content/uploads/2021/07/Rapport_EIA_ETALAB-.pdf (2021)
28. Luguri, J., Strahilevitz, L.: Shining a Light on Dark Patterns, Journal of Legal Analysis 43, University of Chicago Coase-Sandor Institute for Law & Economics Research Paper No. 879, U of Chicago, Public Law Working Paper No. 719, Available at SSRN: https://ssrn.com/abstract=3431205 or https://doi.org/10.2139/ssrn.3431205 (2021)
29. Lippi, M.C., Lagioia, G., Micklitz, F., Pałka, P., Sartor, G., Torroni, P.: The force awakens artificial intelligence for consumer law. J. Artif. Intell. Res. **67**, 169–187 (2020)
30. Maitland, I.: The limits of self-regulation. Calif. Manage. Rev. **27**(3), 135 (1985)
31. Marciano, A., Nicita, A., Ramello., G., B.: Big data and big techs: understanding the value of information in platform capitalism. Eur. J. Law Econ. (2020). https://doi.org/10.1007/s10657-020-09675-1
32. Marty, F., Warin, T.: Innovation in Digital Ecosystems: Challenges and Questions for Competition Policy. CIRANO Working Paper Series. https://cirano.qc.ca/fr/sommaires/2020s-10 (2020a)
33. Marty, F., Warin, T.: Keystone Players and Complementors: An Innovation Perspective. CIRANO Working Paper Series 2020s-61. https://cirano.qc.ca/fr/sommaires/2020s-61 (2020b)
34. Mulligan, D.K., Regan, P., King, J.: The fertile dark matter of privacy takes on the dark patterns of surveillance. J. Consum. Psychol. **30**(4), 767–773 (2020)
35. Obar, J., Oeldorf-Hirsch, A.: The Clickwrap: A Political Economic Mechanism for Manufacturing Consent on Social Media. Social Media Society **4**(3), 2056305118784770 (2018)
36. Pałka, P., Lippi, M.: Big data analytics, online terms of service, and privacy policies. in Vogl R., ed., Research Handbook on Big Data Law, Edward Elgar (2020)
37. Ranchordas, S.: Experimental Regulations for A.I.: Sandboxes for Morals and Mores, University of Groningen Faculty of Law Research Paper No. 7/2021, Available at SSRN: https://ssrn.com/abstract=3839744 or https://doi.org/10.2139/ssrn.3839744 (2021)
38. Rasch, A., Thöne, M., Wenzel, T.: Drip pricing and its regulation: Experimental evidence. J. Econ. Behav. Organ. **176**, 353–370 (2020)
39. Smuha, N.: From a "A race to A.I." to a "race to A.I. regulation" regulatory competition for artificial intelligence. Law Technol. **13**, 1 (2021)
40. Smuha, N., Ahmed-Rengers, E., Harkens, A., Li, W., MacLaren, J., Piselli R., Yeung, K., How the E.U. can achieve legally trustworthy A.I.: a response to the European Commission's proposal for an artificial intelligence act, LEADS working paper, University of Birmingham, august (2021)
41. Stephenson, A.: The pursuit of CSR and business ethics policies: Is it a source of competitive advantage for organizations? J. Am. Acad. Bus. **14**(2), 251–262 (2009)
42. Stigler Center: Stigler committee on digital platforms final report. The University of Chicago, Chicago (2019)
43. Susser, D., Roessler, B., Nissenbaum, H.: Online manipulation: Hidden influences in a digital world. Georgetown Law Technol. Rev. **4**(1), 2–45 (2020)
44. Sunstein, C.: Sludge and ordeal. Duke Law J. **68**(8), 1843–1883 (2019)
45. Sunstein, C.: *The Ethics of Nudging*, 32 Yale J. on Reg. Available at: https://digitalcommons.law.yale.edu/yjreg/vol32/iss2/6 (2015)
46. Sunstein, C.: Sludge Audits. Behavioural Public Policy: 1-20. (2020)
47. Thaler, R.H.: Nudge, not sludge. Science **361**(6401), 431 (2018)
48. Thelisson, E., Morin, J.H., Rochel, J.: A.I. governance: Digital responsibility as a building block—towards an index of digital responsibility. Delphi Interdiscip. Rev. Emerg. Technol. **2**(4), 167–178 (2019)
49. U.S. Government Accountability Office, Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities, GAO-21–519SP. (2021)
50. Veale, M., Zuiderveen Borgesius F.: "Demystifying the Draft E.U. Artificial Intelligence Act." SocArXiv. July 6. doi:https://doi.org/10.31235/osf.io/38p5f. (2021)
51. Wagner, G., Eidenmüller, H.: Down by algorithms? Siphoning rents, exploiting biases, and shaping preferences: regulating the dark side of personalized transaction. Univ. Chicago Law Rev. **86**, 581–609 (2019)
52. Warin, T., Leiter, D.: Homogenous goods markets: An empirical study of price dispersion on the internet. Int. J. Econ. Bus. Res. **4**(5), 514–529 (2012). https://www.inderscienceonline.com/doi/abs/10.1504/IJEBR.2012.048776
53. Warin, T., Troadec, A.: Price strategies in a big data world. Encyclopedia of E-commerce development, implementation, and management. IGI Global, Chapter 46, pp 625–638, March (2016)
54. Yeung, K.: Hypernudge: Big data as a mode of regulation by design. Inf. Commun. Soc. **20**(1), 118–136 (2017)