Autonomy and the social dilemma of online manipulative behavior

Marietjie Botes[1,2]

Abstract
Persuasive online technologies were initially designed and used to gain insights into the online behavior of individuals to personalize advertising campaigns in an effort to influence people and convince them to buy certain products. But recently, these technologies have blurred the lines and morphed into technologies that covertly and gradually manipulate people into attaining a goal that is predetermined by the algorithm and disregards the decision-making rights of the individual. This may lead to people exercising decisions that do not align with their personal values and beliefs, and rob them of their autonomy—an ethical principle, in the absence of which the application of these technologies may be unethical. However, not all technologies that are persuasive are necessarily manipulative which require the careful consideration of a couple of elements to determine whether or not technologies are manipulative and ultimately whether their application is ethical or not. In this article, we analyze the ethical principle of autonomy and unpack the underlying elements of this ethical principle which must be considered to determine whether the application of a technology is ethical or not in the context of it being persuasive or manipulative.

1 Introduction

In a literary twist to the old adage that the road to hell is paved with good intentions, Lewis writes in his novel *The Screwtape Letters* that "the safest road to hell is the gradual one" [1]. Online communication platforms, and more specifically, social media such as Facebook, Twitter, and Instagram (to name a few) had lofty goals when starting out. However, some of these goals quickly took a negative turn when it was discovered that the algorithms used to manage users of these platforms not only posed the risk of addiction, isolation, body dysmorphia, spreading of fake news, undermining democracies, and general erosion of the fabric of society, but were often carefully crafted to manipulate people's online behavior [2–4]. This negative turn of events was slowly but surely brought about by the selling of the

so-called *prediction products* which involve meta data, or behavioral surplus data of social media users to companies who use it to effectively target the marketing of their goods or services to consumers in accordance with their online behavioral patterns [5]. Sayer argued that such economic decisions, behaviors, and institutions "depend on and influence moral/ethical sentiments, norms and behaviors which have ethical implications" [6]. But often, as is the case with social media platforms, ethical or even moral considerations are not reflected in their business models or economic behavior. Ethics in this context is defined by Sayer as "norms (formal and informal), values and dispositions regarding behavior that affects others" [6]. Individual behavior is usually directed by free individual will, which behavior will necessarily affect others with whom such an individual has contact, which will ultimately shape and influence the future of entire societies [7]. In the absence of the ability to exercise individual will, or the manipulation thereof, individuals will not be able to shape their own behaviors and futures, which will lead to the living of a less-fulfilled life. This is the essence of Zuboff's main criticism against the commercialisation of *prediction products,* because it will ultimately deprive humanity of its ability to assert their individual freedom of will to assert their "right to the future tense as

✉ Marietjie Botes
  maria.botes@uni.lu

1 University of Luxembourg, SnT Interdisciplinary Centre for Security, Reliability and Trust, Esch-sur-Alzette, Luxembourg

2 School of Law, University of KwaZulu-Natal, Durban, South Africa

a condition of a fully human life" [8]. This article explores the effects of online manipulation as systematic threat to the ethical principle of autonomy and how autonomy should be protected as a formal human right.

2 Online manipulation

Online information of individuals was originally mined and monitored to gain insights into their online behaviors for purposes of branding and selling products and services [9]. These days, digital technologies are influencing human behavior in ways that were not previously anticipated. Like the Internet, social media networks were not initially built to influence behavior, but to serve as premised on the notion that people should be allowed to invent, shape, and share the content and uses of these platforms [11]. However, this naïve co-creation theory changed dramatically over time when the collection, monitoring and measuring of online behavioral information escalated into the building of algorithms to predict the outcomes of social processes in domains that include economics, public policy, popular culture, and even national security to reveal patterns of market crashes, regime collapses, fads and fashions, and "emergent" social movements which sometimes involve significant segments of society [12]. In addition, algorithms were designed with embedded persuasive technology "tools" for the purpose of changing attitudes or behaviors by means of reduction, tunneling, tailoring, suggestion, surveillance, and conditioning [13]. Automated, pervasive surveillance through the monitoring and logging of online activity, online and offline data, and Internet traffic is also common to digital technologies, and is found in website analytics, tracking cookies, search engines, advertisers, internet service providers, and even governments that mine behavioral patterns of both individuals and groups [15]. Some of these persuasive "tools", that were further refined into "Persuasive Systems Design" models, especially the conditioning and surveillance techniques, were deemed unacceptable by Conti in 2009 from an ethical perspective and even called "weapons of influence" [16]. This criticism was based on (among other) the power of these algorithms to filter and control the information that is made available to people known as the "filter bubble" which narrowly focusses people's interests, preventing them to consider alternative or dissonance points of view [14]. In the social media context, the ultimate goals of these algorithms came to center around how to keep people engaged, grow the network, and tailor advertising opportunities through cleverly crafted "tools" such as ellipses, indicating that someone is busy responding, never ending scrolling, photo tagging, and the like button. These persuasive technologies, which morphed into manipulative technologies over time, are such an integral part of how people interact with the world, that it

progressively enables new behaviors to emerge through the silent and gradual chipping away of individuals' autonomy needed to exercise properly considered decisions, while doing so at a scale which is only possible in a digital world.

The concept of "code is law" was established by Lessig when he explained that while the traditional laws regulate the behavior of people in the real world, online behavior is regulated by code which includes both the software and hardware architecture of the Internet [10]. Subsequently, whoever controls the code controls online behavior. Accordingly, the four forces that determine people's behavior, as identified by Lessig, in the context of online manipulative behavior are the following: (1) *digital architecture*, consisting of technology and design that determine how, when, why, by whom, and to what extent people will use social media; (2) *market forces,* including the economic model of the social media platforms, including other models of digital engagement that determines the exposure to and interaction between people and online advertisers, vendors, and other third parties; (3) *norms* or ethics that dictate what is socially acceptable, harmful, or beneficial; and (4) the *law* that prohibits harmful practices and protects the rights of individuals using social media. In this article, we shall discuss how the enforcement of norms or ethics, and law are missing forces from the current ecosystem that determines people's online behavior, which threatens individual autonomy and ultimately democracy.

But not all persuasive technologies are equal, and to effectively shape the future landscape of these technologies to the benefit of humanity, we first need to discern what constitutes persuasion and manipulation and where to draw the ethical line between these actions.

3 Distinguishing definitions

BJ Fogg, pioneer in the field of persuasive technology research, defined *persuasive computing technology* as a "computing system, device, or application intentionally designed to change a person's attitudes or behavior in a predetermined way" [17]. Persuasion from this technical perspective could be interpreted and applied as either good or bad, and must be distinguished from *online manipulative practices* as defined by Susser, Roesler, and Nissembaum as "applications of information technology that impose hidden influences on users, by targeting and exploiting decision-making vulnerabilities … [t]hat means influencing someone's beliefs, desires, emotions, habits, or behaviors without their conscious awareness, or in ways that would thwart their capacity to become consciously aware of it by undermining usually reliable assumptions" [18]. The biggest ethical differentiating factor between these definitions is awareness of the manipulative nature of the technology which influences

an individual's ability to actively consider his or her options which determines his or her own future in accordance with their own free will.

Between persuasion and manipulation lays a vast array of techno-socially engineered technologies aimed at impacting how people think, perceive information, and act upon it [19]. For purposes of this article, we shall limit the differentiation between definitions to that of persuasion, coercion, deception and manipulation.

3.1 Persuasion

Persuasion entails a fairly direct appeal to an individual's decision-making power, but still allows the individual to freely decide after having had the opportunity to understand and consider the information presented to him or her [19]. Ethically speaking, persuasion is acceptable, because it grants an individual the opportunity to think about the presented information, deliberate about available options and, most importantly to consider it against the backdrop of his or her personal beliefs, desires, and commitments before exercising a decision based on his or her own reasons, absent from any unknown or unwelcome outside influence. The persuasive technique is thus clearly visible and acknowledged for what it is—an attempt to influence people what they decide by changing their understanding of presented information. Full understanding of the information on which you base your decision, as well as the consequences of your decision are critical elements of legal and ethical informed consent. While persuasion tries to convince people to make certain decisions, people are still free to consider the information and decide for themselves. Example: lowering dietary guilt with a logical argument by making claims of less fat and less calories in fried potato chips and calling these new and reduced fries "SatisFRIES". An individual is presented with all the facts to consider whether to buy these fries and may either be refused by highly health conscious individuals, or accepted by individuals who may be more prone to comfort eating, which choices as free and compatible with the relevant individual's beliefs and values around diet.

3.2 Coercion

In contrast to persuasion, coercion tightens an individual's autonomy by restricting the available and acceptable options from which he or she can choose, and exploits the weaknesses and vulnerabilities found in an individual's personal beliefs, desires and commitments to steer his or her decision-making ability toward a certain goal as determined by the coercing technique. Although coercion shares some features with persuasion, such as providing an individual with a number of choices, and cognizant of the fact that not all

persuasive technologies necessarily constitute manipulative techniques, coercion is definitely unethical to the extent that it deprives an individual from considering all information to enable a free, well-informed and considered decision [19]. In addition, coercion actively seeks to control the outcome of this decision-making processes by offering "irresistible incentives" which can only be overcome through "heroism, madness, or something similarly extraordinary" [20]. Where persuasion leaves the individual in control of the entire decision-making process, coercion deprives the individual of the capacity to exercise a conscious decision. In this context, it is unethical to treat people as mere resources without respecting their legal autonomy and human rights which manifests in the respecting of their decisions. Example: the repetitive and extended presentation of pop-up sales advertisements on a website to buy or subscribe to certain products or services, specifically targeting the individual's buying preferences or weaknesses, which makes the navigation on the website frustratingly difficult to the extent that such an individual is tempted to act on the pop-up's request simply to get rid of it. This targeted advertising and continuous confrontation method exploits individual weaknesses and severely restricts the availability of choices an individual can choose from.

3.3 Deception

Deception, on the other hand, is not as subtle as persuasion or coercion. Its main goal is to purposefully influence people into making certain decisions by providing them with false information or half-truths, without their knowledge. In this way, people's weaknesses and vulnerabilities are exploited without their knowledge or ability to exercise any form of autonomy over the decision-making process, surrendering their so-called future tense to the deceptive powers, as described by Zuboff [5]. Save for strictly monitored deception that may be permissible in certain research studies, albeit controversial in nature, deceptive practices are clearly not ethical or legal [21]. Example: scams involving the promise to deliver a specific service or product upon receipt of payment, without such a product even existing, or without the intention of ever delivering the service.

3.4 Manipulation

Building on Susser, Roesler, and Nissembaum's above definition of online manipulative practices, the essence of manipulation can be found in its goal to influence or persuade, but *without being detected*. The influence exercise by manipulation on the individual is thus *hidden* or entails the *covert* subversion of another person's decision-making power [18]. Moreover, the success of manipulation lays in the ability of this persuasive technique to target and exploit

people's decision-making vulnerabilities, much like coercive techniques would do. Manipulation keeps vital information from people which deprive them from being able to properly consider their options to exercise a decision that aligns with their personal beliefs and values, and is therefore unethical as it completely disrespects the autonomy of the decision-maker. These manipulative online practices are further not compatible with the requirement of transparency and fairness as contemplated in the GDPR, because collection and any further processing of individuals' behavioral data or *prediction products* are not deemed fair in light of the information imbalance between the social media platform and its user, or the lack of any transparency about data that is collected for purposes of informing manipulative algorithms [22]. By undermining the autonomy of individuals through hidden manipulative techniques, manipulation practices have the ability to change not only how individuals decide, but also their behavior, and ultimately the behavior of the communities that are made up of these individuals. Example: advertising a product, but emphasizing the scarcity of the product, thereby prompting the individual into quickly deciding, based on an assumption of scarcity, in the absence of which the individual may have taken more time to consider his or her options without the pressure of the product's alleged scarcity. Although the individual is prevented from detecting the true number of available products, confirming the hidden nature of manipulation, the individual is still able to consider his options based on the specifications provided in respect of the product, albeit under time pressure, which may be exploiting individual weaknesses.

4 Autonomy and the decision-making process

The ethical principle of autonomy is based on the Kantian claim that one should always respect the special moral status, which later came to be known as dignity, of human beings by treating them as persons instead of mere resources [23]. In practice, this principle manifests by respecting people's rights and choices. However, in an increasing digitized world, the concept of what exactly autonomy entails and to what extent manipulative techniques interferes with people's choices and subsequent control over their own lives is often vague and blurred. In this regard, Kant explains that anyone who violates another's autonomy "intends to use another man merely as a means, without the latter containing the end in himself at the same time. For he whom I want to use for my own purposes … cannot contain the end of this action himself …. instead, persons … must always be esteemed … only as beings who must be able to contain in themselves the end of the very same action" [24]. Accordingly, to determine whether a person is treated as an end, it must be established

whether that person has himself an interest in that end, or can share in the goal of that end. For a person to share in the benefit of the end goal, it means that such a person must be deferred some form of control over the person's ability to choose a certain outcome or path that will advance that person toward the benefits that the goal entails. But to satisfy as a moral or ethical standard, this deference to choose is not sufficient in itself. People's ability to adequately consider their options to exercise an appropriate choice may change over time in accordance with their personal beliefs and values, which may also change in reaction to changing social circumstances. What may be deemed as a good choice today may be considered to be too risky or irresponsible some time in future. Additional information may bring people to understand certain circumstances in a new light, and may subsequently effect changes in their belief systems with resulting changes to the choices they make. Thus, to respect people's autonomy means that people should be respected over time and not only at a single point of contact.

Deference must be extended to choices made after reflection and consideration of information that enables considered choices by means of the so-called informed consent, which is made possible by complete and easy to understand information. In this regard, encumbrances caused by manipulative technologies that restrict or impede people's consideration of online information may result in defective choices that do not reflect an individual's free will and directly impacts his or her ability to control their right to their so-called future tense [5].

But what hinders a person's capability to make proper decisions in certain circumstances may not be hindering them in other circumstances. For example: some algorithms are specifically designed to target consumer vulnerabilities and actively draw their attention to certain goals, as opposed to allowing people to make decisions when they are calm, clear headed and at their best ability to decide [25]. Encumbrances can be exogenous such as online manipulative techniques, including dark patterns, disguised ads, and blackhat copyrighting, such as testiphonials, false scarcity, and damning admissions [40], that specifically targets vulnerabilities and weaknesses; or endogenous involving individual emotional stress at the time or other mental vulnerabilities that may impact an individual's decision-making abilities [23]. It is thus not encumbrances per se that prevent people from considering information and exercising a choice or give informed consent; rather, it is the combination of exogenous and endogenous encumbrances that creates unethical obstacles in the decision-making process. The elements for the perfect unethical storm in this regard are: (1) lack of adequate information to allow an individual to consider his or her options; (2) lack of full capacity to exercise a decision due to the individual being targeted at a time when he or she

may be at their most vulnerable; and (3) undue influence in the form of online manipulation that specifically targets those weaknesses and vulnerabilities. In essence, the principle of respect for persons in this context is summed up by Hodson as "all actions [that] must be consistent with recognition of the supreme moral importance of each person having control over his or her own life in accordance with his or her own unencumbered choices" [23].

We agree with Klenk and Hancock that not all online manipulative techniques may necessarily result in autonomy loss [26]. However, it seems to be the covert or hidden character of manipulative techniques that aims to influence or persuade people, dissonant to their inherent values and beliefs, that makes the manipulation-autonomy connection explicit. Hidden online manipulative techniques are unethical to the extent that they undermine or prevent an individual's ability to control his or her own life in accordance with his or her own unencumbered decisions [23]. In this context, it is not really the exploitation of cognitive biases and vulnerabilities that poses the most alarming problem that confronts society today, but the fact that many of these exploitative practices are specifically designed to be executed covertly, and by doing so causes people to surrender control over their future tenses to the coders. This form of manipulation facilitated by these technologies translates into what we came to know as digital or rather informational manipulation.

In addition, as discussed above, carefully manipulated information may change the way in which people understand certain circumstances, leading to changes in their belief systems, resulting in changes to the choices they make. If this process happens without people being aware of it, people are being instrumentalised by these manipulative technologies and controlled as means to the ends of the intentions of these technologies, instead of being treated and respected as autonomous human beings [27]. The ease with which technologies integrate with society adds to the perfect storm in which manipulative can erode people from their autonomy and control over their decision-making process. Habituation, or the gradual and undetected manipulation of people's online behavior deductively means that the influences that these technologies facilitate are predominantly hidden and accordingly potentially manipulative. Because there is no co-ownership or democracy of the relationship between social media users and providers, the structural power relationship is perfectly positioned to shape habituation, and the more habituated people become to these platforms, the less attention they pay to them [28]. Research about the exploration and identification of consumer responses to multimedia content that focusses on the reduction of habituation is, however, ongoing [29].

5 The impact of autonomy loss on society and democracy

In February 2019, the Council of Europe expressed its fears that online manipulation may not only weaken people's exercise and enjoyment of their human rights, but may also lead to "the corrosion of the very foundation of the Council of Europe" [30]. The central pillars of human rights, democracy, and the rule of law are grounded on the fundamental belief in the equality and dignity of all humans as independent moral agents, and online manipulation, as discussed above, may compromise these rights by compromising people's autonomy.

Individual autonomy plays a critical role in both social and political arenas. Individuals exercise their autonomy in daily decisions taken in their homes, marketplaces, and the political sphere. Democratic institutions are thus fated to reflect the political decisions made by autonomous individuals. Concerns relating to the undermining of individual autonomy can thus extend beyond ethical considerations into the social and political arena. In this regard, Killmister proffers the theory that autonomy is based on four dimensions which include self-definition, self-realization, self-unification, and self-constitution that collectively inform a wide range of socio-political decisions and are holistically understood as self-governance [31]. The assumption that individuals are capable of this type of self-governance informs the idea that individuals can also collectively and democratically govern themselves. However, if manipulative technologies covertly, gradually, and persistently effect changes to individuals' personal beliefs and values, it will lead to changes in the way in which individuals think, evaluate their choices, form intentions about them, and act on the basis of those intentions. This may impact collective decision-making processes and ultimately affect democracies. The impact that manipulative technologies may have on a collective scale and its possible influence on democracy were illustrated by the Cambridge Analytica data scandal in which the data of approximately 87 million Facebook users was collected for purposes of creating detailed profiles with the intent to psycho-graphically tailor advertisements to influence people's voting preferences during the 2016 US presidential election [32]. This covert profiling was specifically targeted at decision-making vulnerabilities in people that undermined the autonomy of people by exploiting their decision-making vulnerabilities and preventing people from considering information at their best decision-making ability to enable them to act authentically on their own, and not for reasons actually endorsed by them. The employment of techniques in this way threatens people's ability to self-govern and to act in their own best interest, ultimately also threatening democracy.

Political scientist Colin Bennett, who calls this online political micro-targeting, explains how four trends he observed may explain its rise: (1) the move from voter management databases to integrated voter management platforms; (2) the shift from mass-messaging to micro-targeting using personal data; (3) the analysis of social media; (4) and the decentralization of data to local campaigns through mobile applications [41]. These trends do not only affect political decision-making, but has a much broader impact on societies in general if the prevalence of news consumption via social media is considered. The subjective feeling or emotions of individuals consuming news via social media can be manipulated by means of so-called affective news or emotional social media memes that may push the covert agenda of certain groups [42]. Shareability, which is one of the most valued characteristics of social media in general, but more so when it entails breaking news events [43] has been found to be highly connected to emotional responses or passionate online discussion of readers [44]. Consequently, the way in which breaking news are framed for readers who access news via social media platforms is increasingly aimed at sensational news tweets that are emotionally appealing and engaging to attract more readers and ultimately a viral transmission. This exploitation of individual emotional vulnerabilities may have critical consequences for collective decision-making processes in the spreading of false news, creating collective hype which may lead to unfounded protests, or manipulating social agendas regarding public health issues such as vaccination policies.

6 Emerging fields of research

Persuasion profiling, used to influence online users, is based on both *explicit* measures, such as people's tendencies to react in certain ways to distinct persuasive strategies, as well as *implicit* or behavioral methods that involve previous individual experiences that relate and influence decision-making and behavioral tendencies [33]. Kaptein explains that implicit influence principles are developed and refined based on interactions with the user without the user being aware of the profiling and resulting adaptations [33]. Such profiling for purposes of personalizing persuasion technologies brings a sleuth of ethical challenges as discussed above. The ethical challenges in this regard could potentially further be complicated by emerging fields of research and applications using affective computing methods which aims at creating software that recognizes and processes human emotions [34]. In addition to detecting different emotions in people, affective computing also uses sentiment analysis or sentiment mining in which the software uses "natural language processing, text analyses, computational linguistics, and biometrics to systematically identify, extract, quantify

and study affective states and subjective information" [34]. Using these methods, the psychological vulnerabilities of people, such as agreeableness, neuroticism, or risk-aversion, can be determined with increased accuracy. Considering the dawn of manipulative techniques, discussed above, this highly sensitive and personal information may make people even more vulnerable to psychological exploitation and being used as resources, gradually surrendering their autonomy to algorithmic pressures. Any behavioral engineering through which psychological weaknesses or emotions are exploited without revealing it will seriously impact an individual's ability to consider his or her choices, thereby increasing the risk of being manipulated. We must be aware that manipulation, facilitated by artificial intelligence, does not become standard commercial practice where people's behavior is purposefully engineered to realize predetermined commercial goals through skillful coding.

7 Ethical guidelines

To avert some of the ethical fears expressed above, the High-Level Expert Group on Artificial Intelligence of the European Commission published 'Ethical Guidelines for a Trustworthy AI' on 8 April 2019 in which they expressly stipulate that individuals must be treated "as moral [autonomous] subjects, rather than merely as objects to be sifted, sorted, scored, herded, conditioned or manipulated" [35]. They also list the adherence to ethical principles as one of three components necessary to establish trustworthy AI, and advocates for a "human-centric approach" when designing and implementing AI, in which the moral status of humans is the prime enabler for decision-making in civil, political, economic, and social fields in which individual freedom and respect for human dignity is made technologically possible and meaningful, as opposed to over emphasizing an over individualistic account of the human. In view of the above-discussed emerging research fields, these ethical guidelines specifically state that not only the physical integrity of humans must be protected, but also their mental integrity, which include an individual's personal and cultural sense of identity [36]. This means that persuasive technologies must refrain from manipulating mental autonomy, unjustified surveillance, deception, and unfair manipulation, but instead enable people to better control their lives through the decisions they make. By doing so, AI systems, the guidelines conclude, should be able to enhance democratic processes if they respect the autonomy of individuals [35]. Floridi correctly argues that adherence to ethical principles goes beyond formal compliance with existing laws, and it is only through ethical reflection that we are able to understand how the development, deployment and use of AI systems will impact human rights and their underlying values to guide

us toward what we *should* do with technology, instead of what we *can* do [37].

It is thus clear that humans must retain their capability for self-determination when interacting with AI systems to prevent the manipulation and control, not only of their economic choices, but also their social and political decision-making behaviors. In recognition hereof the Council of Europe adopted the so-called declaration on "the manipulative capabilities of algorithmic processes" in February 2019 to mitigate the capability of AI systems to generate "(in) direct illegitimate coercion, threats to mental autonomy and mental health, unjustified surveillance, deception and unfair manipulation" [38]. The main risk envisaged in this declaration entails the manipulative capabilities of algorithmic processes on individuals' cognitive autonomy, and their right to form opinions and take independent decisions. The main aim of the declaration is subsequently to protect individuals' right of choice and self-determination. However, acknowledging that the ethical acceptability of online persuasion is not easy to determine, this declaration also urges member States to engage in public debates to obtain insights into what forms of persuasion seem to be permissible and what constitutes unacceptable manipulation with the goal of providing appropriate protective measures. Public engagement in this regard will provide critical insights into the social values and beliefs that inform individuals' decision-making.

In line with these guidelines, the ethical principle of autonomy is similarly echoed in the ACM Code of Ethics whose self-regulating principles for future persuasive-software design include the consideration whether the intended outcome of, or motivation behind any persuasive technology will be deemed unethical if such persuasion were undertaken without the technology, and the contention that creators of persuasive technologies must take responsibility for all reasonably predictable outcomes of its use [39].

8 Conclusion and recommendations

Some technologies that started out with the aim to persuade people into better lives and futures unfortunately morphed into manipulative technologies that are used to satisfy the intentions of its coders and other relevant stakeholders, which end goals may not be aligned with the values and beliefs of the individuals these technologies are targeting, whilst doing so without being noticed. However, the distinction between persuasion and manipulation is often blurred and uncertain. Despite the existence of sufficient ethical principles to safeguard people against manipulative technologies, it is a thorough analysis of the elements to the ethical principles that is lacking, which is needed by scientists to guide their technical design and employment of technologies on a practical level toward ethically sound technologies. In this article we analyzed the ethical principle of autonomy, which is the ethical principle that is impacted the most by these technologies and plays a determining role in deciding whether a technology is ethical or not. Subsequently, we propose the consideration of the following elements in deciding whether a technology is persuasive and ethically acceptable, or manipulative and unethical:

1. *Intention disclosure*
   Is the intention, end goal, or aim of the algorithm clear to the user of online social media services or is the user consciously aware of the influence that the algorithm may have on him or her?
2. *Option consideration*
   Does the user have the opportunity to consciously and actively consider, deliberate, and think about his or her options against the backdrop of his or her values, beliefs, desires and commitments?
3. *Exploitation*
   Is the algorithm exploiting some form of psychological, emotional or behavioral vulnerability or weakness of the user for its subjective purpose, which purpose may not be aligned with the values and beliefs of the user?
4. *Resource v person*
   Is the algorithm using the user as a resource toward attaining its predetermined goal, or is the algorithm respecting the individual autonomy of the user to allow the user to self-determine the outcome?
5. *Control*

Does the user retain control over his or her decision-making ability, and can the user exercise a free decision?

Declarations

References

1. Lewis, C.S.: The screwtape leters. Geoffrey Bless, London (1942)
2. Hajli, N., Shanmugam, M., Powell, P., Love, P.E.D.: A study on the continuance participation in on-line communities with social commerce perspective. Technol. Forecast. Soc. Chang. **96**, 232–241 (2015)
3. Bradshaw, S., Howard, P.N.: The global organisation of social media disinformation campaigns. J. Int. Aff. **71**, 23–32 (2018)
4. Ferrara, E.: "Manipulation and abuse on social media" by Emilio Ferrara with Ching-man Au Yeung as coordinator. ACM Digit. Libr. (2015). https://doi.org/10.1145/2749279.2749283
5. Zuboff, S.: The age of surveillance capitalism, p. 96. Profile Books, London (2019)
6. Sayer, A.: Moral economy. Department of Sociology, Lancaster University. p. 2–4, www.lancs.ac.uk/fass/sociology/papers/sayer-moral-economy.pdf (2004) Accessed 11 November 2021.
7. Arendt, H.: The human condition, p. 244. University of Chicago Press, Chicago (1998)
8. Zuboff, S.: The age of surveillance capitalism, p. 331. Profile Books, London (2019)
9. Salampasis, M., Paltoglou, G., Giachanou, A.: Using social media for continuous monitoring and mining of consumer behavior. Int. J. Electron. Bus. pp. 85–96, https://www.inderscienceonline.com/doi/abs/10.1504/IJEB.2014.057905 (2013), Accessed 25 November 2021.
10. Lessig, L.: Code and other laws of cyberspace, p. 86. Basic Books, United States, New York (1999)
11. Zittrain, J.: The future of the internet and how to stop it, p. 43. Allen Lane, United Kingdom, London (2008)
12. Colbaugh, R., Glass, K., Ormerod, P.: Predictability and prediction for an experimental cultural market. In: Chai, S.K., Salerno, J.J., Mabry, P.L. (eds.) Advances in social computing, Third International Conference on Social Computing, Behavioral Modeling, and Prediction, SBP, pp. 79–81 Bethesda, MD, USA (2010).
13. Fogg, B.J.: Persuasive technology: using computers to change what we think and do, pp. 32–53. Morgan Kaufmann, United States, California, San Francisco (2003)
14. Zuboff, S.: The Age of Surveillance Capitalism, p. 399. Profile Books, London (2019)
15. Lockton, D.: Attitudes, meaning, emotion and motivation in design for behaviour change. SSRN: https://ssrn.com/abstract=2123495 or https://doi.org/10.2139/ssrn.2123495 (2012). Accessed 26 November 2021.
16. Conti, G.: Googling security: how much does google know about you?, p. 310. Addison-Wesley, United States, New Jersey, Upper Saddle River (2009)
17. Lockton, D.: Cognitive biases, heuristics and decision-making in design for behaviour change. SSRN: https://ssrn.com/abstract=2124557 or https://doi.org/10.2139/ssrn.2124557 (2012). Accessed 26 November 2021.
18. Fogg, B.J.: Persuasive technologies. Commun. ACM **42**(5), 26–29 (1999)
19. Susser, D., Roessler, B., Nissenbaum, H.: Online manipulation : hidden influences in a digital world. Georgetown Law Technol. Rev. **4**(1), 1–45 (2019)
20. Frischmann, B., Selinger, E.: Re-engineering humanity, p. 4. Cambridge University Press, United Kingdom, Cambridge (2018)
21. Rudinow, J.: Manipulation. Ethics **88**(4), 341 (1978)
22. Boynton, M.H., Portnoy, D.B., Johnson, B.T.: Exploring the ethics and psychological impact of deception in psychological research. IRB **35**(2), 7–13 (2013)
23. General Data Protection Regulation (EU) 2016/679 (GDPR), Recital 39.
24. Hodson, J.D.: The ethics of legal coercion, p. 1. D Reidel Publishing Company, The Netherlands, Dordrecht (1983)
25. Kant, I.: Groundwork of the metaphysics of morals. In: Gregor, M., Kant, M. (eds.) Practical philosophy, p. 429. Cambridge University Press, United Kingdom, Cambridge (1996)
26. Adams, A.: Algorithmic decisions and their human consequences. The Regulatory Review. https://www.theregreview.org/2021/11/11/adams-algorithmic-decisions-human-consequences/ (2021). Accessed 29 November 2021.
27. Klenk, M., Hancock, J.: Autonomy and online manipulation. Internet Policy Review. https://policyreview.info/articles/news/autonomy-and-online-manipulation/1431 (2019). Accessed 29 November 2021.
28. Coeckelbergh, M. Defamiliarizing technology, habituation, and the need for a structuralist approach. Foundations of Science. https://doi-org.proxy.bnl.lu/10.1007/s10699-020-09739-0 (2021). Accessed 30 November 2021.
29. Lewandowska A., Jankowski J., Sałabun W., Wątróbski J.: Multicriteria selection of online advertising content for the habituation effect reduction. In: Nguyen N., Gaol F., Hong TP., Trawiński B. (eds). Intelligent information and database systems. Asian conference on intelligent information and database systems (ACIIDS). https://doi-org.proxy.bnl.lu/10.1007/978-3-030-14802-7_43
30. Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes (Adopted by the Committee of Ministers on 13 February 2019 at the 1337th meeting of the Ministers' Deputies) Decl (13/02/2019)1.
31. Killmister, S.: Taking the measure of autonomy: a four-dimensional theory of self-governance. Routledge, Unites States, New York (2017)
32. Hinds, J., Williams, E.J., Joinson, A.N.: "It wouldn't happen to me": privacy concerns and perspectives following the Cambridge Analytica scandal. Int. J. Hum Comput Stud. (2020). https://doi.org/10.1016/j.ijhcs.2020.102498
33. Kaptein, M., de Markopoulos, P., Ruyter, B., Aarts, E.: Personalizing persuasive technologies: explicit and implicit personalization using persuasion profiles. Int. J. Hum-Comput. Stud. **77**, 38–51 (2015)
34. Cambria, E.: Affective computing and sentiment analysis. IEEE Intell. Syst. **1**(2), 102–107 (2016)
35. European Commission, high-level expert group on artificial intelligence: ethics guidelines for trustworthy AI. (2019) https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1#:~:text=Ethics%20guidelines%20for%20trustworthy%20AI.%20The%20aim%20of,be%20ethical%2C%20ensuring%20adherence%20to%20ethical%20principles%20. Accessed 3 December 2021.
36. Hilgendorf, E.: Problem areas in the dignity debate and the ensemble theory of human dignity. In: Grimm, D., Kemmerer, A., Möllers, C. (eds.) Human dignity in context. Explorations of a contested concept, p. 325. Nomos Verlagsgesellschaft mbH & Co. KG, Baden-Baden (2018)
37. Floridi, L.: Soft ethics and the governance of the digital. Philos. Technoly. **31**(1), 1–8 (2018)
38. Council of Europe, Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes. Declaration (13/02/2019) (adopted by the Committee of Ministers on 13 February 2019 at the 1337th meeting of the Ministers' Deputies) (2019) https://search.coe.int/cm/pages/result_detai

ls.aspx?ObjectId=090000168092dd4b. Accessed 3 December 2021.

39. Berdichevsky, D., Neuenschwander, E.: Towards an Ethics of Persuasive Technology. Commun. ACM **42**(5), 51–58 (1999)

40. Birkett, A.: Online manipulation: all the ways you're currently being deceived. https://cxl.com/blog/online-manipulation-all-the-ways-youre-currently-being-deceived/. (2020). Accessed 6 January 2022.

41. Bennett, C.: Trends in voter surveillance in western societies: privacy intrusions and democratic implications. Surveill. Soc. **13**(3/4), 370–389 (2015)

42. Papacharissi, Z., de Oliveira, M.F.: Affective news and networked publics: the rhythms of news storytelling on #Egypt. J. Commun. **62**(2), 266–282 (2012). https://doi.org/10.1111/j.1460-2466.2012.01630.x

43. Hurcombe, E., Burgess, J., Harrington, S.: What's newsworthy about 'Social News'? Characteristics and potential of an emerging genre. Journalism (2019). https://doi.org/10.1177/1464884918793933

44. Harcup, T., O'Neill, D.: What is news? News values revisited (again). Journal. Stud. **18**(12), 1470–1488 (2017). https://doi.org/10.1080/1461670X.2016.1150193

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.