The Power of Algorithms to Redefine Human Autonomy

Alessia Cadelo

Introduction

In recent years, artificial intelligence has developed very rapidly. It is, indeed, a sector that promises great benefits for all of society. It has already been used in medicine, finance, domotics, entertainment, and many other areas. One of these applications is the recommender system, which selects the content being displayed to users in order to recommend the best options. On one hand, it might help users to navigate the online environment; indeed, if suggestions are sufficiently aligned with the users' interests, it could alleviate choice overload and improve self-expression. On the other hand, since recommender systems act like a filter, they shape our perception of available content, information, choice, and, in a way, of the world [1]. They may, therefore, undermine authenticity, defined as being in possession of values and desires considered one's own. Authenticity is a salient element of autonomy, so autonomy could also be influenced. The aim of the present paper is to contribute to the discussion about this topic. Firstly, an investigation into the philosophical concept of autonomy will be carried out, with special regard for two different positions: procedural and relational, respectively. Secondly, it will be shown how recommender systems affect our way of thinking and thus our authenticity and identity.

[1] S. Bonicalzi-M. De Caro-B. Giovanola, *Artificial Intelligence and Autonomy: On the Ethical Dimension of Recommender Systems.*in Topoi XLII, 3, 2023, p. 825.

1. *Autonomy: between procedural and relational views*

In order to understand the relationship between recommender systems and autonomy, a brief philosophical investigation about the meaning of autonomy is necessary. As stated in the introduction, in the current debate there are two main positions, procedural and relational. According to the procedural perspective, autonomy is the ability of self-governance, which consists of determining how to live according to one's beliefs, values and goals [2]. This capacity requires the fulfilment of certain requirements [3]. The first is a minimum degree of rationality; desires or beliefs should not be manifestly inconsistent. In other words, there should be a certain degree of coherence in beliefs and desires [4]; they should respect logic laws. The final ends and purposes must also be harmonised with the rest of the values, preferences, and ideas to which an individual has committed himself. It doesn't count if the beliefs are false; individuals don't lack autonomy simply due to this [5]. It also occurs that a person identifies himself with his projects, values, aims, goals, desires, and so forth. This aspect, named authenticity, is shared among different representatives of this vision of autonomy, such as Joel Feinberg or Gerald Dworkin, but with little differences. For the first, authenticity consists of the capacity to alter his beliefs for reasons of his own. For Dworkin instead, a person has to assimilate the influences that motivate him to himself; they has to recognise them as their own. Otherwise, if a person is alienated from them, he couldn't be defined as autonomous [6]. Moreover, during the process of development of the desires, an autonomous individual should not be under the influence of manipulating factors that inhibit his or her capacity to critically reflect on those desires [7]. In other terms, individuals should have a certain degree of self-control; they should not be influenced by external forces. The procedural perspective owes its name precisely to the fact that it focuses on the process of critical reflection made by the indi-

[2] J. Christman, *Introduction* in The Inner Citadel: Essays on Individual Autonomy, Oxford University Press, New York 1989, pp. 5-6.

[3] Procedural accounts usually focus on factors internal to the psychology of the agent. Nevertheless, Diana Meyers' account is a noticeable exception; she indeed argues that autonomy requires also social and interpersonal skills.

[4] J. Christman, *Autonomy and Personal History*, in Canadian Journal of Philosophy, XXI, 1, 1991, pp. 16 ss.

[5] *Ibidem*.

[6] Id. *Introduction* cit. p. 12.

[7] Id. *Autonomy and Personal History*, cit. p. 19.

ISSN 0039-4130

viduals about their beliefs and values. What matters is the individual's active "participation" in this process, not the content of those beliefs, preferences, and desires. This vision is valuable since it underlines the psychological dimension of self-government. At the same time, it has been widely criticised for various reasons by the feminists; first of all, according to this view, individuals are entities separated from the rest of the world. Individuals are supposed to be atomistic and individualistic. However, human beings are social beings.

«The conviction that persons are socially embedded and that agents' identities are formed within the context of social relationships and shaped by a complex of intersecting social determinants» [8] is indeed the fundamental conviction of all perspectives that are labelled with the umbrella term "relational autonomy".

Thus, the effects of social background on individuals' sense of themselves must be acknowledged and consequently autonomy has to be reconceptualised [9]. Secondly, with this normative primacy on individual independence, authors who support a procedural account of autonomy fail to recognise the value of relations of dependency and interconnection. Since such relations of care have been historically associated with femininity, it is argued that this concept of autonomy is masculinist [10]. However, the main criticism of the procedural conception remains the fact that these theorists underestimated the socialisation, which aspects promote human autonomy and those that undermine it [11]. Autonomy should be indeed understood as a socio-relational phenomenon. Oshana particularly insists on the fact that autonomy is compatible only with social conditions that endorse dignity rather than deny it [12]. This doesn't imply that individuals should not reflect critically on their beliefs and desires and recognize them as their own, but is not sufficient; it is necessary to take into account also the social context in which individual ideas and preferences develop. As a consequence, to achieve autonomy, individuals within a community must establish relationships with others that facilitate the pursuit of their objectives in an environment of social and psychological security [13]. In addition, two

[8] C. Mackenzie-N. Stoljar, *Introduction* in Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self, Oxford University Press, 2000, p. 4.

[9] *Ibidem*.

[10] *Ibidem.*

[11] *Ibidem.*

[12] M. Oshana, *Personal Autonomy and Society*, in Journal of Social Philosophy XXIX, 1998, p.81.

[13] *Ibidem.*

ISSN 0039-4130

other conditions must be satisfied: access to a relevant set of options and procedural independence. The latter refers to the possibility for individuals to make their own decisions without being influenced or restricted by others in autonomy-constraining ways, in accordance with the procedural view [14]. In conclusion, to be autonomous, individuals must be capable of critically evaluating their preferences and ideas and acknowledging them as their own, but this ability, however, flourishes properly only when there is a wide and meaningful range of options.

2. *Recommender systems as digital nudging.*

As stated in the introduction, recommender systems are built to suggest the best options to the users. Most of them employ three different methods: content-based recommendations, collaborative recommendations, and knowledge recommendations. The first filters options by considering past user behaviour. The criterion of the second is to present the user items other users with similar user preferences have liked in the past, while the third bases his recommendation on the users' preferences and constraints. There are also hybrid systems, which combine these various techniques [15]. Whatever techniques are used, these systems modify the users' digital architecture; they select and order contents, personalise information, and recommend alternatives. In this sense, they could be considered as a form of digital nudging. This concept, as a matter of fact, refers to «the use of user-interface design elements to guide people's behaviour in digital choice environments» [16]. This is the first definition, but in a broader sense, digital nudging refers to the usage of digital technology (not only user interfaces) to predictably change people's choices and behaviour in both digital and physical choice environments [17]. As with the traditional nudging introduced by Thaler and Sunstein, the aim is to guide individuals' choices. Despite this, digital nudging has some distinctive features, which may have an important impact on human autonomy. First of all, nudging is originally conceived to promote the best interest of the individual according

[14] *Ibidem*

[15] S.Tiribelli - D. Calvaresi, *Rethinking Health Recommender Systems for Active Aging: An Autonomy-Based Ethical Analysis* in Science and Engineering Ethics, XXX, 22, 2024, p.22.

[16] M. Weinmann - C. Schneider - J. Brocke. 2016, *Digital Nudging*, in *Business & Information Systems Engineering*, LVIII, 6, 2016, p.433.

[17] M. Ienca - E. Vayena, *Digital Nudging: Exploring the Ethical Boundaries*, in *Oxford Handbook of Digital Ethics*, Oxford University Press, 2023, p. 361.

ISSN 0039-4130

to the choice architect (e.g health), which usually are public actors. The digital environment on the contrary is largely managed by private companies, who may be motivated by self-interest. Therefore, there is a misalignment between commercial and individual objectives. In other words, the users could be guided towards a certain option not because it is the best for them but rather because this better responds to the interests of those who offer this service. The purpose of these companies is actually to increase profit, which in turn is mostly derived from advertising. Because of this, they have a competitive interest in indefinitely increasing the amount of time that its users spend on the platform, even though it could be detrimental for them. As a consequence, they nudge users towards compulsive and persistent engagement on the platform through different strategies, such as infinite scrolling and autoplay [18]. However, this massive involvement is only one of the ways to increase profit; being able to grasp individual tastes and preferences is also crucial to offering the right content. To this end, recommender system algorithms collect large amounts of diverse data and build a profile of the individual, including preferences, ideas, personality traits, lifestyles and so on. As a result, they can then offer more precise, user-tailored nudges. These nudged 'choices', being the only ones available, will in turn become the basis for new hypotheses about the users. The principle behind this circle is therefore «what has been is what will be» [19]. As a consequence, everything that is not part of individual preferences is not shown. This mechanism is known as personalisation, and it is the second distinctive mark of digital nudging. Hence, these algorithms may affect the agent's ideas and preferences. Due to this influence, individual ideas, values and preferences could not be one's own anymore, so authenticity might be undermined. In other words, what could be damaged is the psychological dimension of autonomy. In line with this hypothesis, in 2012 the Adomavicious' research group investigated precisely the conditioning operated by recommender systems on individual tastes. The researchers analysed three different conditions, in each of which the participants were asked to view an item and give their assessment [20]. In the first and second case studies, the stimulus to be evaluated was an episode of a television

[18] *Ibidem.*

[19] S. Grafanaki, *Autonomy Challenges in the Age of Big Data*, in *Fordham Intellectual Property, Media and Entertainment Law Journal*, XXVII, 2014, p. 834.

[20] G. Adomavicious et al., *Do recommender systems manipulate consumer preferences? A study of anchoring effects*, in *Information Systems Research*, XXIV, 2013, pp. 962 e ss.

ISSN 0039-4130

show, with one substantial difference: in order to verify the possible ascendancy of algorithmic suggestions, in the first, the subjects were given artificial ratings, which had not been produced by any recommender system. In the second, individual preferences were also taken into account, which were then used by a well-known algorithm to give personalised recommendations. In the third, the same method as the second was used, but this time jokes were evaluated. In all conditions, it was observed that individual preferences were significantly influenced by the recommendations received [21]. Because recommender systems might have an effect on ideas and preferences, they also indirectly act on the global identity of the individual. The mechanism of personalisation, as we just said, could strengthen our values and consequently reinforce personal identity. Nevertheless, it may restrict access to different, relevant options, which, as mentioned earlier, is one of the requirements of autonomy. In regard to this closure, Eli Pariser speaks significantly of the filter bubble, describing it as a lens that controls what we see and what we do not see [22]. Secondly, the activity of profilation may poorly reflect categories that are perceived as central by the agent. As a result, the corresponding suggestions could become irrelevant [23]. At the same time, these algorithms indeed tell people who they are, but the representation of identity provided by the algorithm is crystallised at the present moment, i.e., the moment in which we use the platforms or search the web. The reason lies in the fact that, when we click on a given piece of content, it is always our present self that does so [24], so recommender systems are unlikely to reflect our future aspirations. However, the purpose of the algorithm is not only to describe the individuals in the present but also to provide an estimate of future identities and behaviours. This estimate is consequently partial and probably inaccurate. This representation will be used in turn to nudge the users in the future, so the recommender systems might reshape the subjective experience of one's own identity. Moreover, in this context, Smith suggests that individual values could be replaced by those that can be economically exploited. In this regard, he recalls that Google/Alphabet paid for the well-known game Pokèmon Gò through

[21] *Ibidem.*

[22] E. Pariser, *The filter bubble: What The Internet Is Hiding From You*, Penguin, Londra, 2011, pp.48-49.

[23] S. Bonicalzi - M. De Caro - B. Giovanola, *Artificial Intelligence and Autonomy: On the Ethical Dimension of Recommender Systems*, cit. p. 827.

[24] E. Pariser, *The filter bubble: What The Internet Is Hiding From You,* cit. p.66.

ISSN 0039-4130

the sale of virtual lands in real locations [25]. Thus, Starbucks paid for the game's monsters to reside near their cafés so as to gather many people and increase sales. Of course, no one had any idea of the motive behind the distribution of monsters. In this case, the desire to play video games was channelled in a distorted way towards the sale of cafés and similar [26]. In sum, recommender systems, by profiling individuals, could help users to save time and attention in the research but they could also be detrimental both for identity and autonomy.

Conclusion

In conclusion, although recommender systems facilitate users navigate the online environment, at the same time, they may restrict their freedom of choice and heavily influence their opinions and ideas. In this way, they could undermine autonomy, both in a procedural and relational sense. These algorithms determine the content to be shown to users based on their profile. On the one hand, they can reinforce individual tastes and preferences and thus their identity. On the other hand, the risk is that the representation is static, i.e. it does not adequately take future aspirations into account. The suggestions made by the algorithm may therefore not be relevant or inaccurate. Nevertheless, they might still influence users and consequently they could reshape their identity. Thus, in order to protect autonomy and freedom to form our identity, it would be necessary to reconsider the humans' relational dimension in AI ethics and in the algorithms' design.

[25] C. H. Smith, *Corporatised Identities ≠ Digital Identities: Algorithmic Filtering on Social Media and the Commercialisation of Presentations of Self*, in Ethics of Digital Well-Being, SpringerLink, 2020, pp. 58-59.

[26] *Ibidem*.

ISSN 0039-4130