AI & SOCIETY
https://doi.org/10.1007/s00146-024-01955-7

OPEN FORUM

Artificial intelligence and human autonomy: the case of driving automation

Fabio Fossa[1]

Received: 4 December 2023 / Accepted: 22 April 2024

Abstract
The present paper aims at contributing to the ethical debate on the impacts of artificial intelligence (AI) systems on human autonomy. More specifically, it intends to offer a clearer understanding of the design challenges to the effort of aligning driving automation technologies to this ethical value. After introducing the discussion on the ambiguous impacts that AI systems exert on human autonomy, the analysis zooms in on how the problem has been discussed in the literature on connected and automated vehicles (CAVs). On this basis, it is claimed that the issue has been mainly tackled on a fairly general level, and mostly with reference to the controversial issue of crash-optimization algorithms, so that only limited design insights have been drawn. However, integrating ethical analysis and design practices is critical to pursue the implementation of such an important ethical value into CAV technologies. To this aim, it is argued, a more applied approach targeted at examining the impacts on human autonomy of current CAV functions should also be explored. As an example of the intricacy of this task, the case of automated route planning is discussed in some detail.

Keywords AI ethics · Ethics of connected and automated vehicles · Human autonomy · Automated route planning · Design ethics

1 Introduction

Artificial intelligence (AI) systems are being adopted in a growing set of practical contexts. From industry, healthcare, and households to warfare, finance, and law enforcement—just to name a few—AI technologies are becoming increasingly embedded into the fabric of individual and social existence (Dubber et al. 2020; Crawford 2021). Respectively, the scope of human autonomous decision-making and agency is inevitably affected and morphs into new configurations. The delegation of tasks to AI systems impacts on human autonomy in intricate ways, reshaping its contours, remodulating its characters, and raising thorny philosophical and ethical questions. Both potential enhancements and constraints to its exercise demand thorough evaluation. Accordingly, its respect and promotion lie at the very core of many regulatory frameworks (Jobin et al. 2019; Floridi and Cowls 2019; Fjeld et al. 2020).

The domain of road transport presents important challenges at the intersection of autonomy and automation. In particular, the development of connected and automated vehicles (CAVs) is expected to revolutionize the role of human vehicle occupants in traffic decisions and actions (Michelfelder 2022; Jenkins et al. 2022; Fossa 2023). As the scope of human choice and agency shifts, threats to and opportunity for the exercise of autonomy require to be carefully assessed. From the perspective of engineering ethics, an inquiry into the effects of driving automation technologies on user autonomy is necessary to steer design decisions away from manipulatory or paternalistic outcomes and toward the support of users' autonomous behavior.

The present paper aims at exploring the complex nature of ethical problems arising at the intersection of human autonomy and AI systems in the domain of driving automation. In a nutshell, it claims that the issue has been mainly tackled in the literature on a fairly general level, and mostly with reference to the controversial issue of crash-optimization algorithms. As a result, only limited design insights can be

✉ Fabio Fossa
fabio.fossa@polimi.it

[1] Department of Mechanical Engineering, Politecnico di Milano, via Giuseppe La Masa 1, 20156 Milan, Italy

Published online: 16 May 2024

drawn from its study.[1] However, integrating ethical analysis and design practices is critical to pursue the implementation of such an important ethical value into CAV technologies. To this aim, it is argued, a more applied approach targeted at examining the impacts on human autonomy of current CAV functions should also be explored. As an example of the intricacy of this task, the case of automated route planning is discussed in some detail.

The paper is structured as follows. Section 2 introduces the general debate on the ambiguous effects of AI systems on human autonomy, showing the importance of nuanced analyses and setting the stage for the subsequent discussion. Section 3 tackles the literature on the impacts of CAVs on human autonomy and provides a critical assessment of its significance, suggesting that focusing primarily on current CAV functions might help take a first step towards the elaboration of viable design guidelines. Section 4 provides a preliminary example of a similar examination by discussing the case of automated route planning. Section 5 concludes the paper by offering some final remarks.

2 Human autonomy and AI systems

Before considering the literature dedicated to the prospected impacts of CAV technologies on human autonomous decision-making and agency, let us briefly introduce the debate on how AI systems influence this ethical value worthy of being protected and fostered.

Many commentators have examined and discussed the diverse repercussions of AI systems on human autonomy—the relational, situated, and multi-layered nature of which also requires to be duly factored in (e.g., Mindell 2015; Rubel et al. 2021; Tiribelli 2023).[2] In sum, two main and opposing effects have been noticed. On the one hand, AI systems can be said to contribute to supporting user autonomous decision-making and its translation into practice. In fact, they show the potential of enlarging our choice and action possibilities by both assisting us in navigating through complex decisional processes and freeing our hands from tasks we cannot or would rather not carry out. Moreover, they offer the possibility of improving the efficiency, effectiveness, and safety of many operations, which represents an indirect condition to the enjoyment of autonomy. In doing so, however, AI systems necessarily process information, compute decisions, and sometimes even implement courses of action on our behalf, thus bypassing our own judgment and constraining our agential possibilities. Therefore, the relationship between human autonomy and AI systems does not present a monolithic profile. Rather, it shows a multifaceted nature that calls for nuanced examinations.

Critical reflection on enhancing and constraining effects has accordingly characterized the literature on both AI systems influencing human decision-making and robotic applications acting on our behalf in the physical world.

For what concerns decision-making, considerable attention has been dedicated to recommendation systems (Prunkl 2022; Bonicalzi et al. 2023). Among others (Calvo et al. 2020; Laitinen and Sahlgren 2021; Rubel et al. 2021), a particularly insightful perspective on the multifaceted impacts of algorithmic tools on the scope of human autonomy has been proposed by John Danaher (2016, 2018, 2019). Danaher claims that algorithmic tools pose relatively new challenges to the exercise of user autonomy due to their pervasiveness, centralization, and targeting capacities. At least three dimensions call for accurate analysis in this sense (see also Raz 1986). First, algorithmic tools might impact our capacity to rationally choose the right means to our ends—the rationality condition. Second, they might foster or hinder our capacity to meaningfully access "an adequate range of options" (Danaher 2019: 105)—the optionality condition. Third, and finally, they might either support or frustrate our freedom to counteract unwanted coercion, encroachment, and manipulation—the independence condition.

While the rationality condition does not seem to be in substantial peril, Danaher notes, optionality and independence might indeed be threatened through the use of recommendation systems. By pre-filtering or drawing attention to given options, guiding decision-making through incentive schemes, or taking choices on our behalf, AI systems might negatively affect our autonomy, setting the stage for "algorithmic micro-domination". Hence, the risk of variously nudging users against their will or manipulating them through recommendation systems should not be underplayed (Sunstein 2015; Vallor 2016: 188–207; Ienca 2023), as it has been stressed in the debate surrounding Yeung's (2017, 2019) notion of "hypernudge".

[1] For clarity purposes, I will center the analysis on which *design choices* might help protect, respect, and promote human autonomy through AI systems such as CAVs. This is not meant to claim that design is the only (or the most promising) practical domain through which to influence the impacts of AI technologies on human autonomy. On the contrary, the same objective can be pursued by other means as well—e.g., regulative efforts, policy measures, institutional action, social pressure, user education and awareness, and so on. It is likely that only a concerted effort toward this goal might arrive at tangible results—or so I believe (see Sect. 5).

[2] It lies outside the scope of this paper to provide a summary of the different philosophical theories and interpretations on human autonomy—a task that has been recently carried out by, e.g., Calvo et al. (2020), Laitinen and Sahlgren (2021), Formosa (2021), Rubel et al. (2021), Bonicalzi et al. (2023). It might be useful to signal here that current research is increasingly criticizing the individualistic, Western-centric character of many discussions on human autonomy and AI, stressing the need to widen the perspective so to include insights from cultures that harbor more relational understandings of what autonomy is and implies (e.g., Mhlambi and Tiribelli 2023, Tiribelli 2023).

Nonetheless, it would be one-sided to conclude that recommendation systems can only be detrimental to human autonomy. Rightly tuned recommendation systems would enable us to delegate uninteresting decision-making tasks or analytical operations on huge amount of data we would feel inadequate to carry out. Given the right conditions, moreover, filtering and ordering options, nudging toward preferred outcomes, and automating choices might support user autonomy, challenged as it is by cognitive limitations, information overload, and continuous decision-making.[3] Besides, opting out of algorithmic tools might prove almost impossible in a world where they constitute an integral part of how we choose and behave. As many scholars further clarify (Danaher 2018; Milano et al. 2020; Calvo et al. 2020), recommendation systems co-shape how individuals access, make sense of, and act on digital information. Accepting their mediating role and learning how to live with it appear much more reasonable and promising than attempting to entirely separate oneself from it.

Given their ubiquity and significance, it is fundamental to ensure that recommendation systems are well-adjusted when released into society. What is critical, then, is to learn how to design, regulate, deploy, and use algorithmic tools so to enhance human rationality, optionality, and independence while mitigating related risks. To do so, challenges to the respect of human autonomy and self-determination should be explicitly and appropriately tackled. Building on similar considerations, Varshney (2020) stresses the need for operationalizing the value of human autonomy. The massive influence that recommendation systems can exert on decision-making processes, it is argued, requires to be experimentally assessed and carefully tamed by-design, so to leave enough space for the expression of user autonomy. As Calvo et al. (2020: 32) claim in their discussion of Youtube recommender system, "designing for autonomy is an ethical imperative to the future design of responsible AI"—but one that requires fine-grained, context-related, and multidimensional analyses to be properly carried out.

Even if to a lesser extent, the complex impacts of AI on human autonomy and the need for tackling related challenges have been studied also in relation to robotic systems acting on our behalf (Formosa 2021). Persuasive approaches based on captology and nudge theory (IJsselstein et al. 2006; Fogg et al. 2008; Siegel et al. 2009) have raised a heated debate on how to protect and support the autonomy of users interacting with robots. Borenstein and Arkin (2016) have discussed the ethical legitimacy of "robotic nudges"—i.e., of programmatically influencing user behavior by-design through robotic technologies. Being embodied and acting in the physical world, the persuasive potential of robots might be of massive help in situations where physical or cognitive limitations work against the exercise of autonomous decision-making and agency. Moreover, specific design choices might also serve wider socio-ethical goals, such as spreading (supposedly) socially beneficial behaviors. However, nudging people towards given decisions and actions poses obvious threats to user autonomy. For instance, design choices aimed at maximizing acceptance or conveying given socio-ethical values have been criticized as possible threats to user autonomy, even though they might support processes of moral growth (Weßel et al. 2021; Fossa 2022a). Illegitimate encroachments on users' autonomous decision-making such as paternalism and manipulation are evident risks that roboticists and regulators have an obligation to minimize, in particular when vulnerable users are involved (Sparrow 2002; Sparrow and Sparrow 2006; Sharkey and Sharkey 2010).

An insightful examination of how robots—more precisely, social robots—could diversely impact on human autonomy has been recently provided by Formosa. As the author suggests, robotic technologies "could enhance and respect, as well as inhibit and disrespect, the autonomy of their users" (Formosa 2021: 596). The dual nature of prospected impacts is particularly evident in Formosa's analysis, where potential benefits to the exercise of user autonomy are systematically associated with opposite inhibiting risks. On the bright side, through interactions with social robots users could set and pursue ends they deem more valuable, improve competencies conducive to autonomy, and have access to more authentic choices. For example, tasks we deem dull and meaningless could be thus offloaded, so to gain time to engage in activities we value the most and that reinforce our sense of autonomy and self-respect. Moreover, robotic support could help us take decisions and act in greater accordance with our own convictions or consider all relevant information and reasons before making up our mind.

Symmetrically, however, human autonomy could also be threatened through social robots. These AI systems might make fewer valuable ends available; obstruct the development, maintenance, and cultivation of autonomy competencies; and push users toward less authentic choices. For instance, social robots might handle by default tasks we deem valuable and would like to carry out ourselves, without our knowledge. As a result, our autonomy competencies might wither, and the authenticity of the ensuing choices might be challenged. Finally, and most seriously, social robots might weaken or altogether disrespect human autonomy. By designing social robots to operate in deceptive,

[3] According to some authors (e.g., Klincewicz 2016), algorithmic tools might also prove useful as *moral advisors*—i.e., as systems assisting human moral decision-making by providing relevant information and stimulating moral reflection, thus "respect(ing) and indeed enhanc(ing) individuals' moral autonomy" (Giubilini and Savulescu 2018: 185–186).

manipulative, and coercive ways, these technologies might incentive dependency and turning users into means to ulterior ends, thus severely threatening human dignity as well. As a result, particular care must be exercised in "design, regulation, and use" (Formosa 2021: 596) so that the ethical value of human autonomy is duly protected and promoted.

To summarize, the ethics literature on human autonomy and AI stresses the necessity to distinguish between technologies designed and deployed in ways that support human autonomous decision-making and agency, and technologies that risk manipulating, coercing, or overly constraining the scope of such critical components of human existence. However, it also acknowledges that separating ethically legitimate from illegitimate effects on user autonomy is rarely straightforward, and that controversial trade-offs between the right to individual autonomy and the pursuit of supposed social goods or values are inevitable. That being said, the literature stresses the central normative role of human autonomy as an ethical value variously connected to other pivotal principles such as responsibility, identity, dignity, and well-being (Laitinen and Sahlgren 2021; Tiribelli 2023). Even though off-loading to automated systems the burden that comes attached to human autonomy might appear alluring, as Chiodo (2023) argues, it would engender an ethically troublesome deterioration of human identity and dignity. As AI technologies increasingly co-shape human decision-making and agency, then, it is critical to protect and promote the exercise of user autonomy.

3 Human autonomy and driving automation

The debate on CAVs and human autonomy has also brought to the surface the inextricably dual nature of prospected impacts. Threats to and opportunities for human autonomy are so numerous and deeply entangled with each other that much philosophical work is needed to clarify how this value is to be specified and upheld in the context of driving automation.

Generally speaking, and much in line with the previous remarks, driving automation has been found to exert an ambiguous effect on human autonomy. On the positive side (e.g., Williams et al. 2020), delegating driving to CAVs is expected to allow for more free time and energies to pursue one's own self-determined interests and goals. Lowering the psychological costs of driving, automation is also expected to support autonomous decision-making on matters that importantly impact on individual well-being—such as, for instance, where to live and where to work. Furthermore, driving automation would offer transport opportunities to cognitively or physically impaired people, elderly people, children, minors, and other social categories that as for today enjoy little or no access to it, thus improving their capacity of implementing autonomous choices (e.g., Goggin 2019).

However, and quite paradoxically, such benefits depend on the full automation of what Michon (1985) defined as operational and tactical driving decisions—i.e., decisions concerning how to handle vehicle controls and how to behave in traffic. In other words, driving systems must be capable of automatedly managing choices concerning, e.g., when to speed up (Smids 2018), when to slow down (Nyholm and Smids 2020), when to let other road users pass (Millard-Ball 2018), when to bend traffic rules for the greater good (Reed et al. 2021), and so on. This delegation evidently entails constraining user autonomous decision-making, sometimes in morally relevant ways.

A great deal of attention has been dedicated in this sense to the hotly debated problem of crash-optimization algorithms (e.g., Nyholm 2018; Dogan et al. 2020; Jenkins et al. 2022). If driving is to be fully automated, so are decisions concerning how to distribute risk among involved parties during unavoidable collisions (Goodall 2016). Whether delegating such decisions would support or limit human moral autonomy is controversial. On the one hand, crash-optimization algorithms would make it possible to ethically deal with situations that used to extend beyond the reach of human moral agency, thus expanding the *general* domain of human moral autonomy. On the other hand, implementing these algorithms would constrain the *more specific* domain of user moral autonomy in ways that might make some legitimate ethical choices impossible by-design and possibly amount to moral paternalism (Millar 2015; Gogoll and Müller 2017; Müller and Gogoll 2020). Therefore, some believe, design solutions should be implemented to allow users to exercise autonomy in adequate ways even when the management of unavoidable collisions is automated and delegated to CAVs (Millar 2016; Contissa et al. 2017).

Full driving automation, however, is not the only way to support human autonomy in road transport. Driving technologies such as Advanced Driving Assistance Systems (ADAS) or partial automation solutions also have a role to play. For instance, these technologies could help drivers better manage physiological and psychological constraints to autonomy by automating emergency functions (e.g., emergency braking) or providing valuable information concerning driving behavior (e.g., lane changing warning and fatigue detection systems). Moreover, offering drivers valuable traffic information, as happens with smart intersections and smart roads, might also be construed as assisting humans in exercising autonomy behind the wheel. However, these forms of assistance clearly presuppose the involvement of a vehicle occupant in the execution of the driving task, thus blocking the enjoyment of the autonomy-enhancing effects discussed in the previous paragraph. At the same time, unclear or cumbersome frameworks of shared control over driving tasks might generate mode confusion or inadequate degrees of user reliance, leading to situations where human autonomous decision-making and agency is impeded (Hancock 2019; Bellet et al. 2019).[4]

Interestingly enough, the same ambiguity can be identified on the regulative side as well. In the case of Europe, the 2020 report *Ethics of Connected and Automated Vehicles. Recommendations on Road Safety, Privacy, Fairness, Explainability and Responsibility* (Horizon 2020) establishes an ethical framework for CAVs and offers twenty recommendations aimed at guiding stakeholders in the effort of aligning driving automation technologies to relevant ethical values. Following the lead of many other frameworks (Floridi et al. 2018; HLEGAI 2019; Jobin et al. 2019), the report grants particular recognition to the value of autonomy as one of the eight overarching ethical principles for driving automation (Santoni De Sio 2021). According to it, human beings are to be conceived as "free moral agents" (Horizon 2020: 22) whose right to self-determination ought to be respected. The importance of autonomy reverberates on several recommendations, ranging from the protection of privacy rights and the promotion of user choice to reducing opacity and enhancing explainability. The principle of personal autonomy, then, demands that CAVs are so designed to "protect and promote human beings' capacity to decide about their movements and, more generally, to set their own standards and ends for accommodating a variety of conceptions of a 'good life'" (Horizon 2020: 22). As argued by Santoni de Sio and Fossa (2023), however, supporting *both* autonomous decision-making about driving decisions *and* the autonomous pursuit of different conceptions of a 'good life' through mobility is hard to achieve, since these two specifications of autonomy point towards seemingly incompatible technological pathways.[5]

In sum, insofar as CAV technologies show the capacity to influence or bypass human decision-making and agency in the traffic context, driving automation too poses complex challenges to the respect and enhancement of user autonomy. On the one hand, human autonomous decision-making and agency conducive to well-being could be enhanced by CAVs, which promises more inclusivity and more meaningful time management at the expenses of the exercise of autonomy with reference to driving decisions. On the other hand, CAV technologies could enhance human autonomous decision-making and agency through the implementation of ADAS and partial automation, which aim at improving driving behavior by limiting the effects of regular drivers' physical and cognitive constraints. However, these solutions presuppose the presence of a driver, which would exclude the enjoyment of the autonomy-enhancing benefits so often associated with driving automation.

These general reflections are useful to understand the impacts of driving automation on human autonomy. However, they provide little practical insights to engineers involved in the development of CAV technologies. Transitioning from the abstract acknowledgment of autonomy to more practical endorsements is critical to ensure that driving automation is pursued in alignment with what the value of autonomy demands. As such, it represents a clear mission of responsible design (Morley et al. 2021, 2023)—and one that has been clearly acknowledged in the field of driving automation (Gerdes and Thornton 2015; Thornton et al. 2017; Gerdes et al. 2019; Millar et al. 2020). However, the generality of the discussion reviewed above, paired with its main application to the controversial issue of crash-optimization algorithms—that many believe too abstract to be relevant (e.g., Davnall 2020; De Freitas et al. 2020)—offers only limited guidance to the task of translating general calls for the respect and promotion of human autonomy into more actionable design guidelines. Next to this discussion, it is suggested, current CAV technologies and their impacts on user autonomy should also be assessed with the aim of raising an interdisciplinary debate centered on design strategies and best practices. Given the importance of tackling this side of the problem as well, the rest of the paper shifts the attention to a more applied discussion intended at showing the relevance and intricacy of design issues surrounding the integration of the value of autonomy into key components of current CAVs.

[4] As some commentators have noted (e.g., Glancy 2012, Schoonmaker 2016, Jannusch et al. 2021), CAVs might also pose threats to user autonomy if privacy rights are not adequately upheld. A common example in this sense is the risk of governmental agencies carrying out surveillance activities by, e.g., accessing the location data of a vehicle and associating them to its owner or user, which would constrain individual autonomy and thwart civil and political rights. Since this group of threats to user autonomy involving the value of privacy requires dedicated considerations to be thoroughly discussed, they are explored only partially in this paper with reference to the problem of location-based targeted advertising in CAVs. I have extensively dealt with this general issue in Fossa 2023: 41–64.

[5] As argued in Fossa (2022b) and Santoni de Sio and Fossa (2023), designing CAVs to promote human autonomous decision-making about *movements* includes allowing users to take and implement real-time road traffic decisions, which is possible only if control is shared between users and driving systems. On the contrary, designing CAVs to promote the autonomous pursuit of the *good life* importantly depends on the possibility of delegating driving as a whole to automated systems. Indeed, full automation would make private road transport accessible to currently excluded categories (disabled people, minors, elderly people, people with no driving license, and so on), thus improving their autonomy in the pursuit of well-being. Moreover, full automation would allow users to spend the time previously occupied by driving the way they see fit. As a result, it is dubious whether the principle of personal autonomy supports the technological paradigm of shared control or that of full automation.

As a first step in this direction, a sharper focus on given CAV functions might help bringing theory and practice closer to each other. A hint in this direction can be drawn from (Horizon 2020: 48), where the authors propose to structure reasoning by first assessing the *ethical relevance* of "CAV applications of algorithm and/or machine learning based operational requirements and decision-making". Considering the effects of specific CAV functions on human autonomy might help anchor the analysis to given design and deployment contexts, thus providing precise starting points for a discussion of technical requirements. Building on similar considerations, a function-based working approach has been recently proposed with the intention of supporting driving automation practitioners in the operationalization of ethical values (Fossa et al. 2022). As a first step, the methodology suggests determining whether the technological function under examination relevantly impacts on (how many of) the eight ethical principles advanced in the European report.[6] If F is the function under assessment and the principle of autonomy is considered, the first questions to ask would then be: "Should F remain under user control for personal autonomy to be respected in high-level automation?" (Fossa et al. 2022: 7).

Answering this question is critical to inform subsequent design choices, but also extremely difficult due to the multilayered analyses it requires. Initially at least, a theoretical examination might be useful to trace the general contours of the discussion. In this spirit, and as a way to show the intricacy of the task at hand, the next section is dedicated to a preliminary theoretical exploration of the possible impacts on human autonomy of an important function within the scope of driving automation: automated route planning. Building on what has been showed in previous sections, the next pages will hopefully contribute to developing a clearer understanding of the challenges that await any attempt to design CAV automated route planning (and possibly many other automated driving tasks) according to the ethical value of autonomy. The identification of pitfalls and difficulties is intended to count not as a dismissal of what can be achieved through design ethics approaches, but rather as an opportunity to kickstart a participated conversation on the issue. Indeed, an accurate representation of the level of difficulty presented by a challenge already marks a step toward its responsible management (Siegel and Pappas 2023).

4 An example: automated route planning

Planning routes from points of origins to destinations is an eminent component of traveling experiences. Tools—e.g., compasses, quadrants, maps—have always been playing an essential part in it. With GPS and digital maps, navigation systems have made it possible to delegate route planning to artificial systems capable of computing various options on our behalf according to pre-set criteria. Automated route planning is of course essential to driving automation too. Understanding whether the automation of route planning is relevant vis-à-vis the ethical value of autonomy is necessary to align CAVs to legitimate moral expectations and, thus, build social trust in the technology.

Interestingly, the automation of route planning through navigation systems—that arguably belong to the class of recommendation systems discussed in Sect. 2—has already raised discussions concerning the impacts on human autonomy. Consider, for example, Nickel et al. (2010). As a way to inquire into the controversial relation between trust and technology, Nickel and colleagues introduce a fictional case study to explore how navigation systems reshape and co-shape human autonomous behavior in the practical context of transportation. The case study presents a situation in which the criteria applied by the planning algorithm to compute the best route fail to reflect the (changing) needs of the driver. Having delegated route planning to the navigation system, however, the driver relies on the route recommendation provided. As a result, she is led to a route she would not have taken otherwise—so, in a sense, against her will—and that turns out to disrupt her plans.[7]

Indeed, misalignment between system settings and user preferences might be opaque to users or generally difficult to realize. As a consequence, the delegation of route planning to navigation systems might turn from supporting

[6] The proposition, advanced in Fossa et al. (2022), of anchoring the analysis to given functions executed by the technological product under study is intended to help practitioners structure the ethical inquiry around well-defined technical aspects. This, in turn, is expected to provide a practical foothold to the generality (and, sometimes, inevitable ambiguity) of high-level ethical guidelines in order to reduce the risk of abstraction and keep theory and practice closer to each other. While addressing the ethical import of a technology as a whole might lose track of relevant fine-grained aspects, proceeding on a function-by-function basis might help realize issues related both to single functions and to their concurrent execution, thus facilitating collaboration between technical and ethical experts.

[7] Rubel et al. (2021: 83) also briefly hint at this problem while discussing how users' (autonomous) practical agency might be impacted by algorithmic systems. According to these authors, similar cases should be understood as "a restriction of practical agency against a baseline of an overall expansion of practical agency", which raises the interesting question whether the moral significance of the specific restriction is to be evaluated by itself or by reference to the overall benefits of using navigation systems. In light of the discussion carried out in Sect. 4—in particular, of the reasons why a given route could be preferred over other alternatives—I believe that the authors' claim according to which route planning systems do not affect "significant facets of a person's life" and "do not impose restrictions on one's practical agency" could be questioned.

user autonomous mobility to overlooking human decision-making in ways that might be perceived as illegitimate. The question of user reliance on the performance of the navigation systems and their ability to adapt to drivers' needs and values is, therefore, key to understand the impact of system recommendations on user autonomy. Even though users are those who eventually determine what route is taken, the ways in which navigation systems are designed and used importantly co-shape this decision-making process, thus redefining the scope of user autonomy.

More recently, Frischmann and Selinger (2018: 81–101) have also offered some noteworthy considerations on the impacts of navigation systems on user autonomy. Navigation systems are here discussed as an example of "mind-extending technologies"—i.e., technologies to which cognitive tasks are delegated. While the liberating and empowering effects on mobility choices and practices are difficult to deny, less evident constraints to human autonomy in terms of intrusive nudges (e.g., to speed up so to beat the prospected time of arrival), manipulative geographically targeted advertising, navigation deskilling, and spatial awareness loss must also be attentively factor in to paint a clear picture of how this technology reshapes human autonomy in navigating the world.

Importantly, Frischmann and Selinger notice that even though it is the user who decides whether to use the technology, it should not be ignored the fact that user autonomy is at least challenged by the intentions, biases, plans, and interests of those who design it. Therefore, the scope of user autonomy can be appropriately measured only by taking into due consideration the mediating role of the technology and the wider context in which its use is inscribed.

The previous observations suggest that there are reasons to consider automated route planning as a relevant function vis-à-vis human autonomy even when it comes in the technological form of navigation systems. Its implementation in CAVs arguably corroborates this claim. Indeed, the impacts on human autonomy are much more tangible in the case of driving automation. Compared to usual navigation systems, CAVs present a further aspect: they *apply* route planning on our behalf, entirely bypassing our judgment concerning how to follow the recommended route. Navigator systems do not exert any direct control on the steering wheel. Even though automation biases might make it difficult to critically assess or reject route recommendations,[8] users do retain the possibility of choosing otherwise. This possibility is considerably reduced with CAVs, where routes are not just computed but, once selected, seamlessly turned into practice. This addition represents a crucial novelty. Route planning through navigation systems only partially automates the transportation task of going from A to B. Route planning through CAVs automates the entire task, reshaping even more substantially the scope of user autonomy.

Some hints suggesting that automated route planning in CAVs should be considered as a relevant function with reference to the value of user autonomy can be found in the literature. For instance, Danaher (2019: 106, 109) briefly refers to route planning in the essay quoted in Sect. 2. Moreover, the authors of (Horizon 2020: 41) recommend to "support user empowerment in (…) choosing routes". However, an analysis dedicated to measure the full scope of how the automation of route planning in CAVs might reshape human autonomy is yet to be carried out. The following lines intend to offer a contribution to this issue by clarifying what is at stake in terms of autonomy when route planning is automated through CAVs.

This case too is characterized by ambiguous impacts. The automation of route planning in CAVs evidently entails a constraint in users' "capacity to decide about their movements" (Horizon 2020: 22). Indeed, their judgement concerning what route to take is mediated by the system, which computes and apply the best route on their behalf. This restriction on autonomous decision-making concerning route planning, however, can concurrently be said to enhance user autonomy in at least two ways. First, it allows vehicle occupants to redirect their energies and attention to what matters to them the most, thus supporting their ability to "set (their) own standards and ends" in the pursuit of their conception of a "good life" (Horizon 2020: 22). By taking care of selecting the best route and driving, CAVs support users' needs and desires on how to best occupy the time they spend en route. In this way, automated route planning removes a constraint to the users' autonomous organization of their own time by giving them the possibility of deciding by themselves how to spend it. Going back to Formosa's framework, all this seems to enhance the pursuit of more valuable ends and more authentic choices on the user part.

Arguably, automated route planning might be said to foster user autonomy also with reference to Danaher's optionality condition. The possibility of utilizing a driving system capable of navigating through spaces with which their users are not familiar might importantly expand their mobility options. Through this function, users might become capable of reaching critical destinations—e.g., hospitals—without having previous knowledge of their location, and without having to worry about taking the wrong turn. More in general, the possibility of delegating route planning to CAVs might be expected to increase user confidence in moving throughout the road network, thus increasing the range of available options.

[8] The match made in hell of navigation systems and automation bias has caused many mishaps in the past years. Some examples can be found in Hansen (2015). Luckily enough, those occurred to me have remained private.

Finally, automated route planning—at least in principle—serves users' autonomy by actualising their intention better than they could. In Danaher's terms, the technology could be said to enhance the users' rationality condition, i.e., their capacity to "plan and execute complex intentions" (Danaher 2019: 105). If we suppose—and there is little reason not to—that CAV users' intention is to get to their destination as quickly and smoothly as possible, avoiding traffic jams and unexpectedly closed streets, then the automated route planning function can count on much more information to do so most effectively. It is safe to hypothesize that delegating route planning sensibly enhances users' skills in avoiding congestion and other time-consuming nuisances. In this sense, alignment between users' main intentions and system performances could be said to foster their autonomy: to provide a powerful means for translating their plans into practice, even if mediately.

However, this last point is hardly generalizable. Indeed, high-level alignment between self-determined user preferences and automated route planning in CAVs might generate misalignment on a finer level of granularity. Many personal reasons, even ethically relevant ones, might influence the roads we decide to take. Bypassing these decisions by delegating route planning to CAVs might have a relevant impact on the exercise of user autonomy and lead to what Formosa terms as less authentic choices. Once a destination is set, automated route planning programs compute multiple strategies and select the most convenient—i.e., the one that maximizes the parameters that programmers have selected to represent various constraints and costs. In the pursuit of high-level transportation goals according to high-level specifications—such as reducing travel time and avoiding traffic jams—more detailed and context-related constraints might be overlooked. For instance, as the author of (Horizon 2020: 42) briefly consider, CAVs could compute and drive along routes that "result in personal data collection that the user could not anticipate from the outset, to which they have not consented, and of which they may never become aware" (Horizon 2020: 42). The fact that automated route planning could expose users to privacy infringements they would have avoided, had they had the chance to do so, seems to point to a possible violation of human autonomy designers should take into account and possibly manage.

Further threats to autonomy might come from considering how automated route planning might be bent to serve the interests of a wider set of stakeholders. Consider, for example, how targeted advertising could be paired with information about planned routes and localization to make potential customers drive by given shops, restaurants, and other commercial activities (Glancy 2012; Hansson et al. 2021; Mulder and Vellinga 2021). If some users might embrace this form of advertising, other might perceive it as intrusive and manipulatory—i.e., as an illegitimate encroachment on a domain that should fall under the purview of their autonomy. Deciding whether to be exposed to commercial advertising, and deciding whether routes should be planned also according to this criterion, fall into the purview of autonomous decision-making concerning road transport. Bypassing users' judgment without their explicit consent would amount to violate Danaher's independence condition and pave the way for instances of algorithmic micro-domination, particularly when it exploits users' psychological vulnerabilities constituting what Rubel et al. (2021: 105–109) define as "affective challenges" to human autonomy (Fossa 2023: 57–59).

Similar reflections offer a starting point for a discussion of design solutions aimed at delivering the benefits of automated route planning while minimizing the related risks. In the cases of both unwanted personal data collection and location-based targeted advertisement, threats to user autonomy mainly stem from misalignment between system settings and user preferences.[9] Indeed, if users were given the possibility of personalizing criteria for route planning, and if reliable information about digitally monitored roads were publicly available, they could autonomously choose whether to include these stretches of road among the ones taken into consideration by the system. Similarly, seeking user consent to location-based targeted advertising on CAV in explicit, fair, and understandable ways through system preference settings might help respect user independence without impeding interested users to enjoy the service.

Perhaps, then, an interface aimed at allowing users to specify route planning preferences would help strike a better "balance between the decision-making power we retain for ourselves and that which we delegate to artificial agents" (Floridi et al. 2018: 698). Indeed, Calvo and colleagues (2020: 45) stress the importance of interface design to empower user control (and, thus, autonomy) by arguing that "design for autonomy-support in this interface sphere is largely about providing meaningful controls that allow users to manipulate content in ways they endorse". Accordingly, Kun et al. (2016: 37) claim that the most relevant challenges for interface design in driving automation have precisely to do with assuring that the user "retains autonomy at the desired level". By designing interfaces that allow users to make sure that automated route planning is carried out in alignment with their own criteria, threats to optionality,

[9] To an extent, issues related to user moral autonomy with reference to crash-optimisation algorithms also stems from a case of misalignment—i.e., misalignment between user moral values and system ethics settings. Even though this problem might be of little significance for engineers due to its technological remoteness, valuable insights could still be drawn from the debate to discuss more concrete situations involving functions that might cause user preferences and system settings to conflict.

independence, and authenticity could be more explicitly brought to the awareness of users and managed by-design.

The idea of fostering optionality and independence by allowing users to set their automated route planning system according to their preferences also exhibits some limitations. For instance, user preferences in terms of route choices might reveal themselves the moment the traffic situation makes them relevant, or shifts depending on context. For example, a user might be willing to accept her routes to be generally computed based also on data concerning her consumer behavior, but not when she is late for work or the day after she went shopping. Similarly, a user might prefer not to be driven through a neighborhood she considers unsafe, but not in the case of an emergency trip to the hospital or if she is late for an important meeting. Were these users behind the wheel, they could exercise their autonomy directly in ways system settings could hardly reproduce. Situations of misalignment, then, are likely to occur even if a wide range of user settings can be implemented in the system.[10] Trade-offs between autonomy-enhancing and constraining aspects of automated route planning seems likely to represent a condition of CAV technologies, rather than a fixable bug.

Finally, but most importantly, these considerations remind that the effort of designing automated route planning systems that respect and promote user autonomy does not occur in a moral vacuum. Other values are relevant to driving automation and calls for adequate consideration. Pursuing user autonomy through design choices and regulative measures is likely to affect how legitimate claims based on other ethical values can be accommodated. The ethical design, development, and use of 'trustworthy' CAVs must be sought with as many relevant ethical values in mind as possible. Otherwise, unexpected side effects conveying value hierarchies that defy rational or social support would likely lead to rejection and negatively affect people's trust.

In this perspective, it might be the case that on some occasions the value of human autonomy should take a back seat. Indeed, there might be strong, ethically relevant reasons to delegate route planning to CAVs, even though some constraints in terms of user autonomy are to be expected. For instance, automating route planning might lead to remarkable collective advantages in terms of traffic efficiency and environmental sustainability. Automating route planning according to shared parameters could contribute to minimizing uncertainties and making vehicle behavior more predictable, which would enable a more optimized and flexible management of traffic fluxes (e.g., Friedrich 2016). In case of need, centralized traffic management could optimize the distribution of CAVs on the road network, ultimately improving traffic efficiency, minimizing congestion, and ensuring optimal use of the available road infrastructure. Moreover, it would allow prioritizing routes characterized by minimum energy consumption and the use of less busy routes where smoother driving could be adopted, which would impact positively on the environment while lowering vehicle wear and tear (e.g., Barth et al. 2014). Fine-grained user control over the system preferences of CAV automated route planning would substantially limit the reach of concerted traffic management and the accomplishment of the prospected social benefits.

The latest remarks raise another objection to the idea of providing users with the possibility of exercising fine-grained control on automated route planning through system settings. So far, benefits and threats to autonomy have been mostly discussed by reference to individual users and the related independence, optionality, and rationality conditions. However, collective benefits that might ensue from a centralized management of traffic—for instance, in terms of environmental sustainability—shed a different light on the importance of user autonomy as an ethical value vis-à-vis other noteworthy moral objectives. That being said, the effects on human autonomy of centralized traffic management also require to be thoroughly evaluated. Indeed, this form of traffic control could be opposed by pointing to potential threats to user autonomy in terms of privacy infringements and surveillance risks. Moreover, cybersecurity risks involved in centralizing traffic management would require to be attentively evaluated. Balancing legitimate claims and striking acceptable trade-offs become unavoidable when the intersection between ethics and technology is acknowledged in its full complexity. The value of autonomy cannot be pursued in isolation from the wider ethical framework of driving automation. A full-fledged ethical analysis of automated route planning, then, must determine how to support user autonomy while also pursuing other ethical objectives relevant to the context of driving automation.

5 Conclusion

To conclude our discussion, there is little doubt that automated route planning implemented in CAVs will reshape and co-shape user autonomy in the context of road transport. On the one hand, opportunities to respecting, protecting, and empowering autonomous behavior are easily detectable. On the other hand, pursuing these mobility benefits might lead to situations where roads are planned according to criteria that are not aligned with those of the users, which could be perceived as an illegitimate encroachment on their autonomy. Coping with these issues

[10] Here, I am presupposing that users own their own CAV. In a scenario where CAVs were mainly available as shared mobility solutions, problems concerning personal autonomy would arguably change and require specific discussion.

by supporting user choice in route planning through dedicated interfaces and settings can only go so far, offering only partial assurance that situations of misalignment will not arise. Moreover, the consideration of other potential ethical benefits that might ensue from automating route planning calls for the establishment of a clear value hierarchy for trustworthy driving automation to be practically pursued.

As a result, the analysis has confirmed that automated route planning is a CAV function that should be designed with an eye to the respect, protection, and promotion of human autonomy. However, the ways in which this automated function could reshape our autonomy leave many doubts on how to properly answer this obligation. In analogy with many other AI applications, this case too has exhibited an ambiguous net of opportunities and threats that are extremely difficult to disentangle. Even though the problem remains open, the contours of the challenges to be faced are now clearer. The road to ethically adequate, trustworthy AI technologies is paved with such difficult, multifaceted, and nuanced issues. Their interdisciplinary exploration and discussion is critical to accomplish on the field what has been theoretically acknowledged as a fundamental ethical objective.

Finally, the outcome of the analysis shows that design practices and solutions only go so far in managing ethical problems raised by AI systems such as CAVs. On the contrary, design teams can play their part in the effort of realizing trustworthy technologies only if the same objective is consistently pursued along with the other stakeholders of driving automation. Identifying relevant individual and social values, proposing and debating value hierarchies, translating them into design requirements, enforcing their respect, validating and auditing technological products accordingly, regulating their deployment and use, and so on, are all necessary ingredients of a mission that extends far beyond design practices to involve the whole sociotechnical system of driving automation. As Stilgoe (2018, 2020) and Santoni de Sio (2021) suggest, resisting the simplification of technological solutionism and remaining aware of the social complexity of the task at hand is critical not to underestimate the actual size of the challenge.

Acknowledgements I would like to express my gratitude to Teresa Scantamburlo, whose kind support and suggestions have been essential to refine the research presented in this paper; and to Mario Verdicchio for his insightful notes on a previous version of the article. Also, I'm grateful to my colleagues at the Department of Mechanical Engineering, who helped me bring many technical aspects of the problems discussed in Sections 3 and 4 into sharper focus.

Funding Open access funding provided by Politecnico di Milano within the CRUI-CARE Agreement. This paper is supported by "Sustainable Mobility Center (Centro Nazionale per la Mobilità Sostenibile—CNMS)", project CN00000023, funded by the European Union NextGenerationEU program within the National Recovery and Resilience Plan (PNRR), Mission 4 Component 2 Investment 1.4 "Potenziamento strutture di ricerca e creazione di 'campioni nazionali di R&S' su alcune Key Enabling Technologies".

Data availability Data sharing is not applicable: We do not analyze or generate any datasets, because our work proceeds within a theoretical approach.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Barth M, Boriboonsomsin K, Wu G (2014) Vehicle automation and its potential impacts on energy and emissions. In: Meyer G, Beiker S (eds) Road vehicle automation lecture notes in mobility. Springer, Berlin, pp 103–112. https://doi.org/10.1007/978-3-319-05990-7_10

Bellet T, Cunneen M, Mullins M, Murphy F, Pütz F, Spickermann F, Braendle C, Baumann MF (2019) From semi to fully autonomous vehicles: new emerging risks and ethico-legal challenges for human-machine interactions. Transp Res F Traffic Psychol Behav 63:153–164. https://doi.org/10.1016/j.trf.2019.04.004

Bonicalzi S, De Caro M, Giovanola B (2023) Artificial intelligence and autonomy: on the ethical dimension of recommender systems. Topoi 42:819–832. https://doi.org/10.1007/s11245-023-09922-5

Borenstein J, Arkin R (2016) Robotic nudges: the ethics of engineering a more socially just human being. Sci Eng Ethics 22:31–46. https://doi.org/10.1007/s11948-015-9636-2

Calvo RA, Peters D, Vold K, Ryan RM (2020) Supporting human autonomy in AI systems: a framework for ethical enquiry. In: Burr C, Floridi L (eds) Ethics of digital well-being philosophical studies series, vol 140. Springer, Cham. https://doi.org/10.1007/978-3-030-50585-1_2

Chiodo S (2023) Technology and the overturning of human autonomy. Springer, Cham

Contissa G, Lagioia F, Sartor G (2017) The Ethical Knob: ethically-customisable automated vehicles and the law. Artif Intell Law 25:365–378. https://doi.org/10.1007/s10506-017-9211-z

Crawford K (2021) Atlas of AI. Power, politics, and the planetary costs of artificial intelligence. Yale University Press, New Haven

Danaher J (2016) The threat of algocracy: reality, resistance and accomodation. Philos Technol 29:245–268. https://doi.org/10.1007/s13347-015-0211-1

Danaher J (2018) Toward an ethics of AI assistants: an initial framework. Philos Technol 31:629–653. https://doi.org/10.1007/s13347-018-0317-3

Danaher J (2019) The ethics of algorithmic outsourcing in everyday life. In: Yeung K, Lodge M (eds) Algorithmic regulation. Oxford University Press, Oxford, pp 98–117. https://doi.org/10.1093/oso/9780198838494.003.0005

Davnall R (2020) Solving the single-vehicle self-driving car trolley problem using risk theory and vehicle dynamics. Sci Eng Ethics 26:431–449. https://doi.org/10.1007/s11948-019-00102-6

De Freitas J, Anthony SE, Censi A, Alvarez GA (2020) Doubting driverless dilemmas. Perspect Psychol Sci 15(5):1284–1288. https://doi.org/10.1177/1745691620922201

Dogan E, Costantini F, Le Boennec R (2020) Chapter Nine—Ethical issues concerning automated vehicles and their implications for transport. In: Milakis D, Thomopoulos N, van Wee B (eds) Advances in transport policy and planning, vol 5. Academic Press, Cambridge, pp 215–233. https://doi.org/10.1016/bs.atpp.2020.05.003

Dubber MD, Pasquale F, Das S (eds) (2020) The Oxford handbook of ethics of AI. Oxford University Press, Oxford

Fjeld J, Achten N, Hilligoss H, Nagy A, Srikumar M (2020) Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI SSRN Scholarly Paper, 3518482. https://papers.ssrn.com/abstract=3518482

Floridi L, Cowls J (2019) A unified framework of five principles for AI in society. Harv Data Sci Rev. https://doi.org/10.1162/99608f92.8cd550d1

Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, Luetge C, Madelin R, Pagallo U, Rossi F, Schafer B, Valcke P, Vayena E (2018) AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Mind Mach 28:689–707. https://doi.org/10.1007/s11023-018-9482-5

Fogg BJ, Cuellar G, Danielson D (2008) Motivating, influencing, and persuading users: an introduction to captology. In: Sears A, Jacko JA (eds) The human–computer interaction handbook fundamentals, evolving technologies, and emerging applications, 2nd edn. Lawrence Erlbaum Associates, Hillsdale, pp 133–146

Formosa P (2021) Robot autonomy vs. human autonomy: social robots, artificial intelligence (AI), and the nature of autonomy. Mind Mach 31:595–616. https://doi.org/10.1007/s11023-021-09579-2

Fossa F (2022a) Social robotics as moral education? Fighting discrimination through the design of social robots. In: Hakli R, Mäkelä P, Seibt J (eds) Social robots in social institutions. Proceedings of Robophilosophy'22. Frontiers in artificial intelligence and applications, vol 366. IOS Press, Amsterdam, pp 184–193. https://doi.org/10.3233/FAIA220617

Fossa F (2022b) Autonomy and automation. The case of connected and automated vehicles. In: Kommers P, Macedo M (eds) Proceedings of the international conferences on ICT, society and human beings 2022, web based communities and social media 2022, and e-health 2022. IADIS Press, pp 244–248

Fossa F (2023) Ethics of driving automation. Artificial agency and human values. Springer, Cham

Fossa F, Arrigoni S, Caruso G, Cholakkal HH, Dahal P, Matteucci M, Cheli F (2022) Operationalizing the ethics of connected and automated vehicles: an engineering perspective. Int J Technoethics 13(1):1–20. https://doi.org/10.4018/IJT.291553

Friedrich B (2016) The effect of autonomous vehicles on traffic. In: Maurer M, ChristianGerdes J, Lenz B, Winner H (eds) Autonomous driving. Springer, Berlin, pp 317–334. https://doi.org/10.1007/978-3-662-48847-8_16

Frischmann B, Selinger E (2018) Re-engineering humanity. Cambridge University Press, Cambridge

Gerdes J, Thornton S (2015) Implementable ethics for autonomous vehicles. In: Maurer M, Gerdes J, Lenz B, Winner H (eds) Autonomes Fahren. Springer, Berlin-Heidelberg, pp 87–102. https://doi.org/10.1007/978-3-662-45854-9_5

Gerdes J, Thornton SM, Millar J (2019) Designing automated vehicles around human values. In: Meyer G, Beiker S (eds) Road vehicle automation 6. AVS 2018. Lecture notes in mobility. Springer, Berlin, pp 39–48. https://doi.org/10.1007/978-3-030-22933-7_5

Giubilini A, Savulescu J (2018) The artificial moral advisor. The "Ideal Observer" meets artificial intelligence. Philos Technol 31(2):169–188. https://doi.org/10.1007/s13347-017-0285-z

Glancy DJ (2012) Privacy in autonomous vehicles. Santa Clara Law Rev 52(4):3, 1171–1239. https://digitalcommons.law.scu.edu/cgi/viewcontent.cgi?article=2728&context=lawreview&httpsredir=1&referer=

Goggin G (2019) Disability, connected cars, and communication. Int J Commun 13:2748–2773. https://ijoc.org/index.php/ijoc/article/view/9021

Gogoll J, Müller JF (2017) Autonomous cars: in favor of a mandatory ethics setting. Sci Eng Ethics 23:681–700. https://doi.org/10.1007/s11948-016-9806-x

Goodall NJ (2016) Away from trolley problems and toward risk management. Appl Artif Intell 30(8):810–821. https://doi.org/10.1080/08839514.2016.1229922

Hancock PA (2019) Some pitfalls in the promises of automated and autonomous vehicles. Ergonomics 62(4):479–495. https://doi.org/10.1080/00140139.2018.1498136

Hansen L (2015) 8 drivers who blindly followed their GPS into disaster. The Week. https://theweek.com/articles/464674/8-drivers-who-blindly-followed-gps-into-disaster

Hansson SO, Belin M-Å, Lundgren B (2021) Self-driving vehicles—an ethical overview. Philos Technol 34:1383–1408. https://doi.org/10.1007/s13347-021-00464-5

HLEGAI-High Level Expert Group on Artificial Intelligence (2019) Ethics guidelines for trustworthy artificial intelligence. Publication Office of the European Union: Luxembourg. https://op.europa.eu/s/xLWx

Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (E03659) (2020) Ethics of connected and automated vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility. Publication Office of the European Union, Luxembourg. https://op.europa.eu/en/publication-detail/-/publication/89624e2cf98c-11ea-b44f-01aa75ed71a1/language-en

Ienca M (2023) On artificial intelligence and manipulation. Topoi 42:833–842. https://doi.org/10.1007/s11245-023-09940-3

Ijsselsteijn W, de Kort Y, Midden C, Eggen B, van den Hoven E (2006) Persuasive technology for human well-being: setting the scne. In: Ijsselsteijn WA, de Kort YAW, Midden C, Eggen B, van den Hoven E (eds) Persuasive technology. PERSUASIVE 2006. Lecture notes in computer science, vol 3962. Springer, Berlin, pp 1–5. https://doi.org/10.1007/11755494_1

Jannusch T, David-Spickermann F, Shannon D, Ressel J, Völler M, Murphy F, Furxhi I, Cunneen M, Mullins M (2021) Surveillance and privacy—beyond the panopticon. An exploration of 720-degree observation in level 3 and 4 vehicle automation. Technol Soc 66:101667. https://doi.org/10.1016/j.techsoc.2021.101667

Jenkins R, Cerny D, Hribek T (2022) Autonomous vehicles ethics. Oxford University Press, Oxford

Jobin A, Ienca M, Vayena E (2019) The global landscape of AI ethics guidelines. Nat Mach Intell 1:389–399. https://doi.org/10.1038/s42256-019-0088-2

Klincewicz M (2016) Artificial intelligence as a means to moral enhancement. Stud Logic Grammar Rhetor 48(61):171–187. https://doi.org/10.1515/slgr-2016-0061

Kun AL, Boll S, Schmidt A (2016) Shifting gears: user interfaces in the age of autonomous driving. IEEE Pervasive Comput 15(1):32–38. https://doi.org/10.1109/MPRV.2016.14

Laitinen A, Sahlgren O (2021) AI systems and respect for human autonomy. Front Artif Intell 4:705164. https://doi.org/10.3389/frai.2021.705164

Mhlambi S, Tiribelli S (2023) Decolonizing AI ethics: relational autonomy as a means to counter AI harms. Topoi 42:867–880. https://doi.org/10.1007/s11245-022-09874-2

Michelfelder D (ed) (2022) Test-driving the future. Autonomous vehicles and the ethics of technological change. Rowman & Littlefield, London-New York

Michon JA (1985) A critical review of driver behavior models: what do we know, what should we do. In: Evans L, Schwing RC (eds) Human behavior and traffic safety. Springer, Boston, pp 485–524

Milano S, Taddeo M, Floridi L (2020) Recommender systems and their ethical challenges. AI & Soc 35:957–967. https://doi.org/10.1007/s00146-020-00950-y

Millar J (2015) Technology as moral proxy. Autonomy and paternalism by design. IEEE Technol Soc Mag. https://doi.org/10.1109/MTS.2015.2425612

Millar J (2016) An ethics evaluation tool for automating ethical decision-making in robots and self-driving cars. Appl Artif Intell 30(8):787–809. https://doi.org/10.1080/08839514.2016.1229919

Millar J, Paz D, Thornton S, Parisi C, Gerdes J (2020) A framework for addressing ethical considerations in the engineering of automated vehicles (and other technologies). In: Proceedings of the Design Society: DESIGN conference 1, pp 1485–1494. https://doi.org/10.1017/dsd.2020.78

Millard-Ball A (2018) Pedestrians, autonomous vehicles, and cities. J Plan Educ Res 38(1):6–12. https://doi.org/10.1177/0739456X16675674

Mindell DA (2015) Our robots, ourselves: robotics and the myths of autonomy. Viking, New York

Morley J, Elhalal A, Garcia F, Kinsey L, Mökander J, Floridi L (2021) Ethics as a service: a pragmatic operationalisation of AI ethics. Mind Mach 31:239–256. https://doi.org/10.1007/s11023-021-09563-w

Morley J, Kinsey L, Elhalal A, Garcia F, Ziosi M, Floridi L (2023) Operationalising AI ethics: barriers, enablers and next steps. AI & Soc 38:411–423. https://doi.org/10.1007/s00146-021-01308-8

Mulder T, Vellinga NE (2021) Exploring data protection challenges of automated driving. Comput Law Secur Rev 40:105530. https://doi.org/10.1016/j.clsr.2021.105530

Müller JF, Gogoll J (2020) Should manual driving be (eventually) outlawed? Sci Eng Ethics 26:1549–1567. https://doi.org/10.1007/s11948-020-00190-9

Nickel PJ, Franssen M, Kroes P (2010) Can we make sense of the notion of trustworthy technology? Knowl Technol Policy 23:429–444. https://doi.org/10.1007/s12130-010-9124-6

Nyholm S (2018) The ethics of crashes with self-driving cars: a roadmap I. PhilosCompass 13(7):e12507. https://doi.org/10.1111/phc3.12507

Nyholm S, Smids J (2020) Automated cars meet human drivers: responsible human-robot coordination and the ethics of mixed traffic. Ethics Inf Technol 22:335–344. https://doi.org/10.1007/s10676-018-9445-9

Prunkl C (2022) Human autonomy in the age of artificial intelligence. Nat Mach Intell 4:99–101. https://doi.org/10.1038/s42256-022-00449-9

Raz J (1986) The morality of freedom. Oxford University Press, Oxford

Reed N, Leiman T, Palade P, Martens M, Kester L (2021) Ethics of automated vehicles: breaking traffic rules for road safety. Ethics Inf Technol 23:777–789. https://doi.org/10.1007/s10676-021-09614-x

Rubel A, Castro C, Pham A (2021) Algorithms and autonomy. The ethics of automated decision systems. Cambridge University Press, Cambridge. https://doi.org/10.1017/9781108895057

Santoni de Sio F (2021) The European Commission report on ethics of connected and automated vehicles and the future of ethics of transportation. Ethics Inf Technol 23:713–726. https://doi.org/10.1007/s10676-021-09609-8

Santoni de Sio F, Fossa F (2023) Designing driving automation for human autonomy: self-determination, the good life, and social deliberation. In: Fossa F, Cheli F (eds) Connected and automated vehicles: integrating engineering and ethics. Springer, Cham, pp 19–36. https://doi.org/10.1007/978-3-031-39991-6_2

Schoonmaker J (2016) Proactive privacy for a driverless age. Inf Commun Technol Law 25(2):96–128. https://doi.org/10.1080/13600834.2016.1184456

Sharkey N, Sharkey A (2010) The crying shame of robot nannies: an ethical appraisal. Interact Studi Soc Behav Commun Biol Artif Syst 11(2):161–190. https://doi.org/10.1075/is.11.2.01sha

Siegel J, Pappas G (2023) Morals, ethics, and the technology capabilities and limitations of automated and self-driving vehicles. AI & Soc 38:213–226. https://doi.org/10.1007/s00146-021-01277-y

Siegel M, Breazeal C, Norton MI (2009) Persuasive robotics: the influence of robot gender on human behavior. In: 2009 IEEE/RSJ international conference on intelligent robots and systems, St. Louis, MO, USA, 2009, pp 2563–2568. https://doi.org/10.1109/IROS.2009.5354116

Smids J (2018) The moral case for intelligent speed adaptation. J Appl Philos 35(2):205–211. https://doi.org/10.1111/japp.12168

Sparrow R (2002) The March of the robot dogs. Ethics Inf Technol 4:305–318. https://doi.org/10.1023/A:1021386708994

Sparrow R, Sparrow L (2006) In the hands of machines? The future of aged care. Mind Mach 16:141–161. https://doi.org/10.1007/s11023-006-9030-6

Stilgoe J (2018) Machine learning, social learning and the governance of self-driving cars. Soc Stud Sci 48(1):25–56. https://doi.org/10.1177/0306312717741687

Stilgoe J (2020) Who's driving innovation? New technologies and the Collaborative State. Palgrave Macmillan, Cham

Sunstein C (2015) The ethics of nudging. Yale J Regul 32(2):413–450. http://digitalcommons.law.yale.edu/yjreg/vol32/iss2/6

Thornton SM, Pan S, Erlien SM, Gerdes JC (2017) Incorporating ethical considerations into automated vehicle control. IEEE Trans Intell Transp Syst 18(6):1429–1439. https://doi.org/10.1109/TITS.2016.2609339

Tiribelli S (2023) Moral freedom in the age of artificial intelligence. Mimesis International, Sesto San Giovanni

Vallor S (2016) Technology and the virtues: a philosophical guide to a future worth wanting. Oxford University Press, Oxford

Varshney LR (2020) Respect for human autonomy in recommender systems. http://arxiv.org/abs/2009.02603. https://doi.org/10.48550/arXiv.2009.02603

Weßel M, Ellerich-Groppe N, Schweda M (2021) Gender stereotyping of robotic systems in eldercare: an exploratory analysis of ethical problems and possible solutions. Int J Soc Robot. https://doi.org/10.1007/s12369-021-00854-x

Williams E, Das V, Fisher A (2020) Assessing the sustainability implications of autonomous vehicles: recommendations for research community practice. Sustainability 12:1902. https://doi.org/10.3390/su12051902

Yeung K (2017) 'Hypernudge': big data as a mode of regulation by design. Inf Commun Soc 20(1):118–136. https://doi.org/10.1080/1369118X.2016.1186713

Yeung K (2019) Why worry about decision-making by machine? In: Yeung K, Lodge M (eds) Algorithmic regulation. Oxford University Press, Oxford, pp 21–48. https://doi.org/10.1093/oso/9780198838494.003.0002