The ironies of autonomy

Ganesh, Maya Indira

The ironies of autonomy

Maya Indira Ganesh [1✉]

Current research on autonomous vehicles tends to focus on making them safer through policies to manage innovation, and integration into existing urban and mobility systems. This article takes social, cultural and philosophical approaches instead, critically appraising how human subjectivity, and human-machine relations, are shifting and changing through the application of big data and algorithmic techniques to the automation of driving. 20th century approaches to safety engineering and automation—be it in an airplane or automobile-have sought to either erase the human because she is error-prone and inefficient; have design compensate for the limits of the human; or at least mould human into machine through an assessment of the complementary competencies of each. The 'irony of automation' is an observation of the tensions emerging therein; for example, that the computationally superior and efficient machine actually needs human operators to ensure that it is working effectively; and that the human is inevitably held accountable for errors, even if the machine is more efficient or accurate. With the emergence of the autonomous vehicle (AV) as simultaneously AI/ 'robot', *and* automobile, *and* distributed, big data infrastructural platform, these beliefs about human and machine are dissolving into what I refer to as the ironies of *autonomy*. For example, recent AV crashes suggest that human operators cannot intervene in the statistical operations underlying automated decision-making in machine learning, but are expected to. And that while AVs promise 'freedom', human time, work, and bodies are threaded into, and surveilled by, data infrastructures, and re-shaped by its information flows. The shift that occurs is that human subjectivity has socio-economic and legal implications and is not about fixed attributes of human and machine fitting into each other. Drawing on Post-phenomenological concepts of embodiment and instrumentation, and excerpts from field-work, this article argues that the emergence of AVs in society prompts a rethinking of the multiple relationalities that constitute humanity through machines.

[1] Leuphana University, 21335 Lüneburg, Germany. ✉email: indira.ganesh@stud.leuphana.de

The crashes

Road safety and driving accuracy are the top research topics in both academic and grey literature about autonomous vehicles (AVs) (Cavoli et al., 2017). The case for autonomous driving is made by showing that humans are error-prone and that software is safer. The US National Highway and Traffic Safety Association proclaims on its website that 'technology can save lives. 94% of crashes due to human error.' AVs are promised to be safer: they follow rules, do not speed nor get drunk and drive, do not check their phones, get sleepy, or distracted. However, four recent crashes involving AVs resulted in human fatalities:

- In Hebei province, China, a driver was killed when his Tesla Model S crashed into a road-sweeping vehicle. The father of the driver claimed that the car was in Auto-pilot mode, however Tesla said that the damage to their vehicle was so severe that it was not possible to retrieve information about how the crash actually occurred (Boudette, 2016)
- In Florida, a test driver was killed when he did not take over control from the Tesla that was in Auto-pilot mode and that drove into the light, white side of an 18 wheel truck, mistaking it for the early morning sky. The driver did not respond to alerts to take over, and eventually had only three seconds to respond before impact (National Transport Safety Board, 2017; Tesla, 2016).
- In Arizona, an Uber test-driver in a Volvo semi-autonomous vehicle did not take control of the car that was in auto-pilot and hit a pedestrian wheeling her bicycle across the road; the pedestrian was not properly identified by the vehicle's computer vision software; the driver was found to be distracted (National Transport Safety Board, 2018).
- In California, a driver was killed when driving a Tesla in Auto-pilot mode that drove into a road works barrier (Shepardson, 2018).

There is a thread in the causes for each of these crashes: faulty handover between the human and the vehicle that was in Auto-pilot[1] mode because of the human driver's distraction, or slow response time in taking over (National Transport Safety Board, 2017, 2018, 2019). With an increase in automation of a task, the human physical and cognitive skills required to complete, monitor and oversee, and eventually step back in to that task become poor, and especially at short notice (Cummings et al., 2013; Cummings and Ryan, 2014). However, it was not just the handover between human and machine that was a problem, but a failure of machine learning in the computer vision systems within the Florida and Arizona test AVs.[2] A gap opened up between the world as it was, and the world as modelled by the computer vision system. John Cheney-Lippold writes with reference to the Florida accident:

> an ontological gap form[ed] between a white truck crossing into one's lane and an algorithmic interpretation of a white truck crossing into one's lane. One is a collection of elements moving through time and space; the other is a probabilistic evaluation of those elements, represented, as best as the algorithm can, as a deviating new world, intelligible as data, where a white truck ceases to be a white truck and becomes a statistical relationship. It is instead a formal acceptance that the statistics underlying Tesla's Autopilot suite are operational precisely because they are not evaluating some mythical, unmediated "real" but rather are processing the world in line with the necessarily objectifying force of statistics. (2019, p. 527)

This article addresses two aspects of the role of the human in the emerging autonomous vehicle. The first is the dominant perception that autonomous driving entails the replacement of the human driver with computation and automation, and thus is sometimes colloquially referred to as 'robot driving'. However, automation does not *replace* the human but *displaces* her to take on different tasks (Sheridan and Parasuraman, 2005). I will show how humans are distributed across the internet as paid and unpaid micro-workers routinely supporting computer vision systems; and as drivers who must oversee the AV in auto-pilot.[3] Aside from online tasks, humans are encouraged to 'empathise' with the emergent machine that struggles to learn how to navigate the world. These are cases of heteromation, also seen across contemporary online platforms and services: a "new economic arrangement in which humans are put on the margins of machines and algorithms, providing labour in unrewarded or minimally rewarded ways" (Ekbia and Nardi, 2018, p. 365). Distinct from automation where "the machine takes centre stage"; or augmentation where "the machine comes to the rescue", heteromation is defined as "the machine calls for help" (Ekbia and Nardi, 2014, n.p.), and in which the human becomes legible as a "computational component" (ibid).

Second, and related, is that the discursive construction of the AV rests on the transition from human to robot driving; precisely because the AV is not *just* a car *or* a robot but is also a distributed data infrastructure running AI technologies, there are subject positions the human may find herself in that she cannot necessarily predict or control given what big data infrastructures are. It is one thing to expect the human to be alert enough to take over; but, we are in different territory with the AV perceiving the world through machine learning and making decisions on this basis. Research finds that object detection by computer vision systems used in AVs perform poorly "when detecting pedestrians with Fitzpatrick skin types between 4 and 6. This behaviour suggests that future errors made by autonomous vehicles may not be evenly distributed across different demographic groups." (Wilson et al., 2019, p. 1) In other words, people with darker skin tones are less likely to be clearly identified by computer vision used in AVs. The problem of darker and female phenotypes being misidentified or not seen by computer vision has a precedent, it is not new (Buolamwini and Gebru, 2018). Thus the recognition of the world through the (not so) 'objectifying force of statistics' that mistakes a truck for the light, almost-white sky, or that may not recognise a darker-skinned person as a human, indicates a scale of computational, automated decision-making that is near impossible to intervene in from the outside. The conditions of optimisation and standardisation of the data in the statistical relationships that underlie computer vision have the power to produce multiple, conflicting subjectivities within the AV: that of an accident victim on a dark night or poorly lit street; in the operator's hot seat and expected to take over at a moment's notice, but without any control over the contingencies set in motion by the computational infrastructures she is embedded in; and as a 'heteromated' worker-cog propping up these material infrastructures, including annotating and labelling visual images for computer vision systems. And yet despite the limited human control in such systems, accountability and liability still fall on the human operator, coupled with surveillance and monitoring systems to discipline the human to remain alert and vigilant in their role as driver-overseer, as I will discuss.

I draw on Postphenomenological approaches to discuss these two aspects of human subjectivities within the AV as 'multiple' (Mol, 2002), that is as AI/robot, as a 'more' automated automobile, and as a big data-infrastructural platform. That heteromated humans are embodied within the informational flows and

decisional capacities of AI-based autonomous systems; and controlled through measurement and surveillance instruments within these systems (Ihde, 1995).

The 'irony of autonomy' is a riff on Lisanne Bainbridge's "irony of *automation*": "the automatic control system has been put in because it can do the job better than the operator, but yet the operator is being asked to monitor that it is working effectively." (1983, p. 776). 'Irony' also refers to other contradictions and gaps that emerge from the AV as simultaneously AI/robot, big data infrastructure, and automated automobile. Human-machine relations forged through automation and aviation engineering of the 20th century, are having to be reconsidered as humans become distributed and displaced through the system. Significant legal and accountability loopholes have emerged from these earlier conceptions and require new approaches to protecting human values, as current automation law does not, even though it proposes to (Jones, 2015). What kinds of rights and protections are there for people working within the platforms of autonomous driving? What is the accountability of AV manufacturers in deploying computer vision software that effectively increases the risks of people with darker skin? Such questions about AVs as future technologies that are already rife with errors, lapses, and contradictions, and that are shaping social relations nonetheless, tend *not* to be the focus of research. Between policy and transport research that focuses narrowly on concerns of regulation, safety, and efficiency, and deterministic popular discourses (Bissell et al., 2020), social research about autonomous driving focuses on: 'human factors' in the interaction of humans as drivers of conventional vehicles, pedestrians, cyclists, or distracted operators (Schoettle and Sivak, 2015); challenges for policy and innovation (Stilgoe, 2019); the emergence of sociotechnical imaginaries and the process of technological change (Mladenović et al., 2020; Tennant et al., 2020). This article challenges these trends.

To conclude this section, I synthesise the structure of this article with my main arguments. This article presents interlinked themes that speak to cultural and philosophical approaches to AI through the case of the autonomous vehicle; as such, it is affiliated to Media Studies and Cultural Studies of technology, mobility, and big data. It emerges from a broader study of the material-discursive practices through which automated, machine learning-based decision-making brings instability, doubt, and uncertainty into supposedly stable ontological categories like 'worker', 'driver', 'human', and machine' (Amoore, 2018). Conducted between 2016 and 2019, this study includes primary, empirical data sources in addition to desk research: 20 in-depth, unstructured interviews with academic, policy, and industry experts from Germany, India, and the United States working in the Law, Computer Science, Design, Mapping, Robotics, and Automotive Engineering; two interviews with Tesla owners in North America, and a test drive with one of them; two workshops on ethics and future technologies with engineers and technologists in Germany and the United States; and a two-year long professional association with a futurologist at a leading German auto manufacturer tasked with imagining the future, autonomous vehicle.[4] Excerpts from these research interactions are threaded through this article in support of my arguments. The sites of enquiry in this paper relate to the production of human subjectivities through the unstable ontology of the AV as simultaneously a 20th century automobile emerging from a history of automation and human-machine relations; *and* as a big data infrastructural platform; *and* through the framing of an AI/'robot' car. I address each of these in this article.

The AV, which does not exist yet, is presented to us in terms of the robot trope, a computational brain housed within and directing the car-body to navigate independently, making decisions for its 'self'. I discuss these configurations as a Foucault-ian

apparatus of material knowledge practices, including measurements, scales, standards that influence autonomy in terms of this trope. However this 'brain' is in fact a vast data-infrastructural network spread over multiple commercial, regulatory, legal, and cloud geographies; the material infrastructures within the emergent AV render it a data platform in itself (Alvarez León, 2019). Thus I move away from this trope that locates autonomy and agency as "attributes inherent in entities", and instead towards autonomy as an "effect[s] of discourses and material practices" and as always "enacted within, rather than separable from, particular human-machine configurations" (Suchman and Weber, 2016, p. 2). These human-machine configurations are applied to accident accountability and safety, and have originated in aviation engineering and safety design, like the 'human in the loop'. Yet, the human-in-the-loop sits in contradiction to the 'cascading logics' of automation that now permeate autonomous driving as a big data infrastructure; it becomes almost impossible for an increasingly automated system to be regulated by something that is *not* similarly automated (Andrejevic, 2019). And while 'the loop' suggests continuity between human and machine, there is still a separation, for the human is always understood to be in control and accountable for errors (Elish and Hwang, 2015; Elish, 2019). Within the AV as a big data-infrastructural, machine learning-based platform, I detail examples of how humans are increasingly embodied in the decisional flows of AV, and are controlled by it. I conclude by proposing that future AV technologies must account for the relationalities emerging herein, with attention to design, and implementation contexts, and with concern for humans who find themselves in this apparatus *sans* equity, support, or care.

The measures of robot driving

To ask the seemingly straightforward question, "what is an autonomous vehicle?", is to undertake a mapping of material practices of knowledge-making, metaphors, institutions, and infrastructures that constitute it. By identifying the discursive interplay of language and measurement in constructing autonomy, I want to bring political valence to this term; if not, it remains both opaque and fantastical.

The language of autonomous driving echoes the robot trope, a computational brain inside a vehicular body or automated machinery *sans* humans: driverless car; robot taxi; unmanned vehicle system (includes drones and robots). 'Self'-driving suggests the vehicle might have a sense of self; or that humans see it as having a self because it can navigate itself. The AV is imagined as an artefact that is separated from the human, but is still humanoid in its processing capabilities, referred to as 'intelligence', in the same way that AI/ artificial intelligence is, an 'awesome thinking machine' that will make decisions for itself, automatically or, 'autonomously' (Natale and Ballatore, 2017). AVs in advertising, cinema, literature, TV programs and industry literature are resolutely anthropomorphic (Kröger, 2015). Drive. Ai, a Silicon Valley software company says that they are "building the brain of driverless cars"; and the BMW Sales and Marketing Lead echoes this: "now we're in the 'hands off' and 'eyes off' phase, but only for brief periods. The next phase will be 'brain off'" (2015). Both are directly modelling autonomy on a separation thought to exist between human body and cognition; a fetishised individuality, an atomised independence, and separation (Fisch, 2016). Artificial intelligence is constructed through a fertile and messy exchange of metaphors about human and machine, and measures like 'intelligence'. Metaphors are powerfully entangled with epistemology even when they are not accurate, and are constitutive of theory particularly in young fields of research. Theoretical Psychology, for example, is replete with

analogies between humans and computers; "computer metaphors have an indispensable role in the formulation and articulation of theoretical positions." (Boyd, 1993, p. 487).

I argue that this is more than semantics, it is part of the discursive shaping of the AV within a Foucault-ian 'apparatus', a system of relations and knowledge-making through a "heterogenous ensemble consisting of discourses, institutions, architectural forms, regulatory decisions, laws, administrative measures, scientific statements, philosophical and moral propositions" (Foucault, 1980, pp. 194–195). I argue this apparatus wields power to shape autonomy as a cognitive, data-based state through supposedly-objective metrics, classifications, heuristics, algorithms, models, and quantifications of human affect and bodies. Such specially constructed measures work as "strategies of relations of forces supporting, and supported by, types of knowledge" that are considered scientific and valid ways of speaking about a topic (196-198). A vehicle is autonomous if it navigates the road like a human driver does, that is, on its own and without a human needing to pay attention to it. There is a standard that captures this relationship in a heuristic that has taken on the status of fact. The J3016 standard for automated driving issued by the Society for Automotive Engineering shows Level 5 as "Full automation: the full-time performance by an automated driving system of all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver." (SAE, 2014). This is accompanied by a graphic that has been widely reproduced in popular media, tech writing, legislative and policy documents. In this graphic, autonomous driving is presented as a linear scale from 'no automation' to 'full automation'; at the end of the scale depicting 'no automation' (level 0), a humanoid figure, shaded in solid grey colour is at a steering wheel shaded in a solid blue; as the levels of automation increase upwards to 1, 2, 3, the human and the wheel, both lose their solid shading, become clearer, and are bounded by a dotted line, suggesting a change in their status, like instability or disappearance. Recent updates to the standards do not include the image of the humanoid figure (SAE, 2018). Instead, the text describes increasing automation in behavioural and technical terms only. SAE levels can be misleading and dangerous but also ignore the many layers of automation that already exist in driving at present such as parallel parking, rear mirrors, lane-assist and other features (Stayton and Stilgoe, 2020; Roy, 2018). Liza Dixon proposes the term 'autonowashing' "to describe the gap between the way automation capabilities are described to users and the system's actual technical capabilities" (2019). Similarly, the prefixes 'fully' and 'semi' suggest linear stages leading up to 'full' autonomy as a final destination. The language of handovers and loops that I discuss ahead also fixes autonomy in terms of a dyadic relationship between human and machine.

Metaphors of a computational brain are materialised by a raft of profitable software companies that are building autonomous driving based on AI technologies (Stewart, 2017). The AV exists in and as a 'formidable' 'intelligent vehicular grid', a big data-infrastructural platform. It is comprised of sensors capturing and processing data about the environment, cameras, radar, Lidar, myriad data processing functions including machine learning, object recognition, tracking, and coordination, mapping and localisation systems, machine-readable road signs, networking and communication architectures including vehicular cloud computing, computer vision, machine-learning based risk and uncertainty assessments, and driving style analysis among others (Gerla et al., 2014; Yurtsever et al., 2020). However, sensors and software only create a notional map of the world around the driverless car; actual driving is about negotiating the real world of pedestrians, cyclists, red lights, uneven curbs, or freak snowstorms that obscure lane markings and signs. In a research

interview,[5] a Human Factors researcher at a US university describes the dissimilarities between human driving and software-led driving. He says that AVs can be slow, ("like a little grandma driving!"), stopping if it doesn't know what to do when faced with new situations that are not covered by the rules in its learning system. Humans, on the other hand, he says, extrapolate from past experience to figure out how to address new challenges. "Autonomous vehicles tend to stop entirely in new or unfamiliar situations, which is not always very helpful" he goes on. A human driver is able to patch together sensing, perception, memory and the body to generate the appropriate response. The same researcher also tells of a research experiment he conducted. In this, he found that human drivers caught in snowstorms that completely obscured lane markings were still able to approximately maintain the required distance in their own lanes; however, AVs were at a loss because they relied on the visual information conveyed by the lane markings, and did not have any data in their databases that might allow them to compute their way out of this problem. It is precisely this inability of the AV to adapt and respond on the fly that the human has to step in to help with. This handover has become a measure called 'disengagements', applied in California, which is the number of miles driven in 'autonomous' mode (with a human required by law to be present) before the human driver has to take over (Hawkins, 2020). Every year, car companies authorised to test their driverless car technology in California must submit disengagement reports: "Manufacturers must track how often their vehicles disengage from autonomous mode, whether that disengagement is the result of technology failure or situations requiring the test driver to take manual control of the vehicle to operate safely." (California Department of Motor Vehicles). However, disengagement reports are contested because they can be misleading; an AV can record a relatively low disengagement number by testing on open, empty highways rather than in more challenging driving environments like a crowded city.

In identifying measures that constitute 'robot' driving, I want to emphasise the relationship of human and machine that positions autonomy as the next evolutionary stage for both car and human. The human has long been considered inimical to safe and efficient driving; it is the union of machine and its infrastructure that captured the imagination of engineers like Norman bel Geddes in the 1960s: "Everything will be designed by engineering, not by legislation, not in piecemeal fashion, but as a complete job. The two, the car and the road, are both essential to the realisation of automatic safety." (in Seiler, 2008). There is an irony unfolding here; the human is the template for driving but also what must be erased, improved on, or displaced for the machine to be transformed. In the next section I continue to examine the human-machine relationship as shaped by 20th century histories of safety engineering in automated systems and its implications for accident accountability.

Automation legacies
The transition from automation to autonomy will occur through machine learning-based decision-making that is subject to the 'cascading logics of automation', meaning that one instance of automation necessitates another; a large scale automated data collection can only be analysed through a similarly large-scale automated process, and not manually (Andrejevic, 2019, p. 8). Automation begets *more* automation; anything other than this is friction that will slow down the process. It is typically this kind of situation that makes it appear that the human is erased. But in fact what is erased is not so much the human, but *how* humans make judgements and decisions; and driving is a case where humans are generally shown to make poor decisions. Driverless

cars are promised to be safer because they can, in theory, receive and process large amounts of data about a complex and changing environment and compute how to act accordingly. Machine decision-making is not just fast, but is also efficient and correct precisely because of the 'god trick' (Haraway, 1988, p. 581) of seeing everything from nowhere, or 'objectively' as big data technologies are thought to. Cascading logics makes it possible to argue that automated and scaled decision-making must be deployed in regulation of such systems a point often made in the argument for lethal 'autonomous' weapon systems that will act on the basis of data, objectively, and automatically in the face of complexity (Asaro, 2019; Suchman and Weber, 2016). Yet, this 'automated management' (Dodge and Kitchin, 2006) breaks and fails; and in the AV context, the history of human-machine decision-making still prevails.

The development of computation and automation through 20th century aviation safety design have been influential in shaping human-machine relations in terms of what exactly machines do better than humans, and vice versa, and what is best pursued collaboratively. These longstanding concerns are now transported to AVs; for example, the SAE's levels of autonomy mentioned earlier emerged from studies of human factors in automation design across industrial contexts (Sheridan, 1992; Jones, 2015, pp. 107–112). The now-discontinued Fitts List aka 'MABA-MABA' (Machines Are Better At-Men Are Better At) is an example of a 1950s heuristic of what humans and machines were thought to perform better than the other; this emerged from the systematisation of national Air Traffic Control in the United States (Cummings, 2014). The Rasmussen skills-rules-knowledge (SRK) model applied in aviator training is a more fine-grained approach to human-machine collaboration that makes a distinction between tasks as skills-based, rules-based, or knowledge-based, each of these being differentiated by how they unfold under conditions of uncertainty in the environment (Cummings, 2017, p. 3). So skills-based tasks such as landing a plane or parallel-parking a car, are well-suited to automation because they entail a routine, specific set of steps. But landing a plane under adverse conditions requires expertise plus intuition and judgement sharpened through a variety of experiences; and thus is notoriously hard to formalise as requirements of an automated system (4–6). The more formalised, specific, and certain an environment and task are, the easier they are to automate. Cummings shows that uncertain environmental conditions may not be addressed by the automation of perception within "brittle" computer vision systems and supervised machine learning,[6] like a pedestrian who wheels her bicycle across a road at a point where she should not. Similarly, Bainbridge presents a detailed discussion of the various conditions under which different kinds of ironies emerge from the automation of tasks (1983. pp. 775–777); one that specifically applies to the AV's auto-pilot is the gradual degradation of human skills through the introduction of automation. As investigations of AV accidents show, the human, notorious for not paying attention, is freed from paying attention by automation that never loses attention; yet she pays a tragic price for inattention when something in the automated system, i.e., computer vision, fails to respond to a sudden change in the environment, and requires her attention to manage.

Autonomous driving still requires human intervention as characterised by the language of handovers between the two (Cummings, 2014, p. 7). A dyadic workflow of human operator and machine is not just about efficiency and productivity but about safety and accountability, which aviation engineering has always kept central. Another aviation engineering-import, the 'human in the loop' (Jones, 2015, p. 134; Marra and McNeil, 2012) has shaped legal accountability in robots and autonomous technologies; both an evocative metaphor and a practical guideline, 'the human in the loop' is a safety mechanism. Jones however identifies problems with this conception saying that the human has *always* been part of the loop and cannot be erased or shifted out. She identifies the irony that US automation law builds on this notion of humans and machines as separate and joined by a loop, thus not acknowledging the inherently socio-technical nature of automation; and thus even as it proposes to protect human values, it actually results in less protection because it understands the two as separate (Jones, 2015, p. 81). Jones proposes that the law—and I would say, accountability regimes more broadly—must break the loop and tie a "policy knot" instead with the contexts of design, implementation and social relations. This becomes critical within the big data-infrastructural aspect of the AV; there are many different layers and locations of humans within multiple workflows that the rigid dynamic of the loop does not address. I return to this knot in my concluding statements. For now I want to stay with Jones' emphasis on the sociotechnical nature of automated systems, that they are neither just human nor machine fitted-in to each other, but are a productive imbrication, that cannot be easily disentangled. This emerges quite starkly in terms of embodiment within the AV.

Embodiment

Recent business, AV Engineering, and HCI[7] narratives suggest that the language of human-machine relations is changing, from 'looping' to the affective registers of 'teaming', trust, and empathy (Visser et al., 2018). For example, robotics researchers want to match the personalities of humans to AVs to encourage humans to 'feel connected' to cars; and to encourage uptake as well, no doubt (Zhang et al., 2019). Nissan Labs is proposing 'Human Autonomy Teaming' (HAT),[8] also developed by Human Factors research in aviation in which autonomous agents are not 'tools' but are 'team members' (McNeese et al., 2018). 'Who Wants To Be A Driverless Car' invites people to 'empathise' with AVs in a more physical way; they have to lie down inside the frame of a motorised buggy and wear a headset that replicates the three-dimensional map view that AVs employ so as to 'understand' what they see.[9] Human operators are encouraged to be sensitive to the needs of the emergent AV. When asked what would make future AVs safe, Andrew Ng, Chief Scientist at Baidu, said that pedestrians need to follow the rules and be "lawful and considerate" (Kahn, 2018). He goes on to say that humans have always reshaped their behaviour in relation to driving and cars, and hence what is being asked of humans is just a continuation of history. However, this has been forced by the auto industry in order to promote cars and driving; the creation of jaywalking as a category of criminal offence in the 1920s in the United States is a case in point; it penalised humans for walking across what had always been public, i.e. the street, in order to make way for what was a new invention at the time:cars (Norton, 2011). As I will discuss ahead, the AV's data and AI infrastructure makes demands on the human mind and body to 'lean in', be empathic, and work 'as a service' (Irani and Silberman, 2013). 'Autonomous' driving is less about the promise of freedom *from* the car, and is perhaps *for* the car as 20th century engineers like Bel Geddes imagined.

Auto-pilot in the AV context evokes a granular register of the human body being reshaped, as this test drive from my field research shows. I am just outside Philadelphia with 'Jyoti', a doctor, who is test-driving a Tesla; she is considering buying a car with Auto-pilot so as to free up time on her long commutes.[10] It does not feel like we are driving because I cannot hear or feel the engine. With the car in Auto-pilot Jyoti finds herself not paying attention to the road. She remarks: "it is so easy to forget you are driving. It is so smooth!" The Tesla representative demonstrates

that Auto-pilot is a small switch that is flicked on with a beeping sound to indicate that it is engaged; a different beep indicates when the car is out of Auto-pilot mode. "The best way to think about Auto-pilot is as a Cruise Control function. By stepping on the brakes-". "You disengage Auto-pilot." Jyoti completes the sentence for him. "Exactly. We ask you to keep your hands on the steering wheel at all times…the steering wheel is going to move on its own so rest your hands there. Don't fight the wheel, just let the wheel guide you" says the representative. Jyoti finds that she cannot get the car to stay in Auto-pilot. She is either holding the wheel too tight thus preventing Auto-pilot from engaging because the system reads her grip as control; or, she holds the wheel lightly enough to engage Auto-pilot, but grabs it tighter when she is confused by its decision-making, thus disengaging Auto-pilot. In one instance, she wants to let the car behind overtake her, and as soon as she does, the beep indicates she has taken control from the car that was in Auto-pilot. In another instance, she finds the car in Auto-pilot overtaking another car a little faster and closer than she would have liked: "whoa, that was close" she says, visibly confused by what the car has just done. Auto-pilot is constantly beeping, signalling that it is being repetitively engaged and disengaged. What seems to frustrate her even more is the gentle and persistent instruction from the Tesla representative to "let Auto-pilot do its job". "I can't get this to work, it's like you have to learn to drive all over again!" she exclaims.

Cultural theorists of automobility persuasively show how the automobile is an extension of the human. A "complex hybridisation of the biological body and the machinic body" (Sheller, 2004, p. 232) in which "new forms of kinship are elaborated 'linking animate qualities to the machine'", "not only do we feel the car but we feel through the car and with the car" (228). Anyone who has driven a car with a perfectly-tuned, powerful, and efficient engine on an open highway has experienced something like this. Such phenomenological interactions of humans with technologies constitute a shared lifeworld that shapes knowledge, politics, aesthetics, and normativity, among others. There are no unidirectional, looping, or deterministic relationships here; instead, complicated new agencies emerge through the shifting subject and object positions of human and machine rendered through practices of embodiment, hermeneutics, alterity, and 'background relations' (Ihde, 1995; Rosenberger and Verbeek, 2015). Of these, embodiment is particularly resonant in the case of the AV; it refers to the 'taking in' of a technology device into human bodily experience, and the extension of the human back into the device, such that the technology 'disappears' and becomes notionally transparent (Rosenberger and Verbeek, 2015; Verbeek, 2011, 2017). The embodiment that Jyoti experiences, of being subtly disciplined into a new way of driving, and the language of teaming, trust, and empathy, speak to a kind of a "body that [is itself] fragmented and disciplined to the machine" (Urry, 2004, p. 31). Postphenomenologists typically refer to benign examples of embodiment such as reading glasses and walking sticks that become work by being 'embedded' in the human body, one expanding into the other to work effectively; however in the case of the AV there is a serious edge to the 'hybridisation' of car and driver that goes beyond the body and includes psycho-affective and emotional states as well, for it can spell the difference between life and death as crashes have shown. I discuss another set of relations of the entwining of human bodies and minds with AVs before turning to a discussion of the social implications of embodiment.

Heteromation

Embodiment-as-disappearance into technology manifests in micro-work humans perform to support big data infrastructures that constitute the AV. I refer to this as 'heteromation' (Ekbia and Nardi, 2014), the value extracted from human micro-work across large and small online systems to support them, and that go unrewarded or are minimally rewarded. Similarly, "humanly extended automation" is being implemented in highly digitised work environments like Amazon's 'Fulfilment Centres' where humans and robotic technologies shift tasks between each other (Delfanti and Frey, 2020) Heteromation turns humans into "computational components", just as 'humanly extended automation' relies on "living labor [to] mak[e] up for machine shortages" (21) . In the AV, this takes on a kind of embodiment because humans are seeing for machines, almost literally. Computer vision in AVs is not advanced enough for driving and has emerged as a weak link in all fatal crashes so far. It is not that the AV, fitted with multiple sensors, cameras, Lidar and radar to document the environment, cannot visually sense, but that it cannot *make sense* of what it senses. Humans must annotate images so that computer vision algorithms can learn to distinguish one object from another, and then apply this when encountering new and unfamiliar images. This leads us to believe we are witnessing a "bravura performance of autonomy enabled by machine learning" (Stilgoe, 2017, p. 3). A slew of companies hire workers in low-income countries to do this work (Lee, 2018). This annotation micro-work also happens through reCAPTCHAs: internet users tag images as depicting trees, storefronts or chimneys in order to complete various online transactions (O'Malley, 2018). It is also not unusual for off-the-shelf, already-annotated datasets to be installed wholesale. However, it was found that there are substantial errors in these datasets being used to train driverless car software; correcting and updating these databases require more human work (Dwyer, 2020). But, the world is not static and off-the-shelf databases are not always current, even if they are correct. But there is a workaround; in a research interview, a transport researcher at a US university, tells me, "with 5G, this should not be a problem… if there are doubts or errors, we can just patch in crowd workers from Pakistan or wherever to respond [make sense of the image] to whatever situation arises that the car cannot deal with."[11] (He does not stop to address the assumption that there will indeed be frictionless 5G connectivity between a car anywhere in the world and Pakistan.) The European AV industry contracts specialist online micro-work platforms, often branding themselves as 'AI companies' and distinguishing themselves from the 'legacy generalist' platforms like Amazon Mechanical Turk or Crowdflower (Schmidt, 2019, pp. 4–5). And, in a curious return of 'the human in the loop', all these specialist micro-work platforms brand their work as having some kind of human machine teaming, collaboration, or human oversight when there is algorithmic annotation of images, and particularly in edge cases that are difficult for computer vision to parse. Human workers in such firms are low-wage, low-income workers.

The departure that heteromation makes from automation is not just the shift from the transfer of control from the human to the machine in automation, to the machine handing back to the human in heteromation. It is that 'heteromated' humans are generating significant value for software and 'AI companies'. The notion of fixed roles of humans and machines fitting into each other also dissolves. Moreover, big data infrastructures like computer vision are changing *how* decisions are made and what their contents are. An online micro-worker is making a judgement about every specific image or situation on the road, each one different from the other, and with edge cases inevitably cropping up; this proceeds alongside automated computer vision. Large-scale automated work implies not just the automation of the *form* of a task, but its *content*, the close relationship between form and content notwithstanding (Ekbia and Nardi, 2018, p. 361, emphasis

added). The decision about what a thing is—a pedestrian, a road divider, or the sky—is being made by a machine, or a human, or a human overseeing a machine, which becomes a statistical relationship between data points that make up the world around the autonomous vehicle. In domains such as online content moderation, specific guidelines are drawn up for how human moderators must adjudicate on content, not just because it is a matter of speech, but because of how misrecognition or decontextualised annotation can change history itself. In the AV-as-data infrastructure, heteromated humans are now part of its "cognitive assemblage", its 'just in time' logistics, and "the flow of information through a system and the choices and decisions that create, modify, and interpret the flow". (Hayles, 2017, p. 116). Human and machine decision-making for AVs to navigate the world is also re-making it.

Control through measurement

While the human becomes part of the AV's computational system, she also remains, well, human. In the Arizona crash, the test driver was found to be distracted from her task of overseeing the Uber/Volvo AV-in-testing, although Uber's lack of safe testing protocols and a generally weak testing policy environment were also found to be issues (Levin, 2020). It was possible to identify that the test driver spent 34 percent of her time looking at her phone streaming a TV show; that in the three minutes before the crash, she glanced at her phone 23 times; and that she looked back at the road one second before impact (National Transport Safety Board, 2019). All the crashes discussed earlier, including Jyoti's test drive experience, are evidence of the human operator/driver in the difficult role of having to be simultaneously vigilant and relaxed so as to take over at a moment's notice; and particularly in the context of the auto-pilot, the technology that makes autonomous driving appear 'real' in the sense of the car being *self-driving*. Thus, surveillance and monitoring of human drivers has become a part of the AV-driving experience. AV testing requires that a driver-facing camera be fitted to record and monitor driver behaviour, physiological states, and affect. This is affective computing in action, a booming interdisciplinary field that analyses individual human facial expressions, gait, and stance to map out emotional states through machine learning techniques. Despite research that find these claims incomplete at best, and invalid at worst (Barrett et al., 2019), affective computing is being used to monitor and manage drivers' and passengers' states and moods like road rage and driver fatigue, and particularly to ensure that drivers remain attentive to the road.[12] No doubt this surveillance data will protect car and ride-sharing companies against future liability if drivers are found to be distracted. This monitoring is literal bodily control because it is used to make determinations about people. Trustworthy, productive, and efficient long distance truckers, for example, are identified by surveillance and measurement; data analytics feeds into "governance strategies like measurement, classification, and ranking" as a "means to discipline and control employees" (Levy, 2015, p. 161). Similar kinds of quantification exist within ubiquitous, networked technologies and we use it to monitor and optimise our own health, wellbeing, and personal success (Esmonde and Jette, 2020). Measurement of human activity, bodies, and affect in work contexts becomes the basis for sorting, classification and analysis resulting in the production of social categories that have far reaching consequences; for example, categories such as criminality and creditworthiness are now determined algorithmically based on individual data profiles and run through analytics; these control large swathes of already-disadvantaged communities (Amoore, 2020; Eubanks, 2018).

There is a substantial historical precedent to the control and manipulation of people through the measurement of data about and from their bodies for the extraction of value, and the shaping of knowledge. Information technologies were born as measuring devices in 18th-19th century contexts of colonialism and slavery; these produced unique categories such as race, mental illness, and criminality among others, which eventually served to discipline and control entire populations. These distinctions were also important in identifying who was fit and capable of working in a rapidly industrialising and modernising world (Sekula, 1986). At roughly the same time in the American South and British Caribbean colonies, slave-owners were honing 'scientific' data practices for extracting maximum value from the land, labour and capital; enslaved people were both labour and capital (Rosenthal, 2019). Slaveholders were using "sophisticated" 'data-based agriculture' techniques to compare the productivity of different kinds of bodies-pregnant women, older people, young people, men and women-as the basis for calculating bonuses, which incentives made each type work harder, "and of course, punishment. …They excelled in determining the most labor their slaves could perform and pushing them to attain that maximum." (86)

The operations of automated data science to classify and rate communities of people is what Postphenomenology refers to as 'instrumentation': Measurement practices that create transformations in human experience, and knowledge of the world (Ihde, 1995). Such practices mirror the two meanings of the word 'apparatus': as Foucault-ian ensemble of institutions and discursive practices that shape knowledge, and as a literal measuring device. Apparatuses as measuring devices are neither inert, objective, nor universal, they are productive of phenomena they purportedly measure, and betray their origins if we study them (Barad, 2007, p. 146). In other words, apparatuses do not sit apart from the world to passively observe and record it but are as large or small or expansive as the determinate local conditions of its assembly. Thus the absence of dark-skinned people or light-coloured trucks from computer vision datasets tells us about how and where these systems are architected. A device that measures up to 100 on a scale does not allow for a value of 101 to exist in the universe circumscribed by that device. And a phenomenon like 'autonomy' can be measured by a 'device' like a disengagement report; and the 'ethics' of autonomous driving can be based on crowdsourced values held by people playing an online game about how an imaginary AV should react in the case of an unexpected accident (Awad et al., 2018). Thus measuring devices do not just observe and record, but actively create categories and realities like 'trustworthy', 'efficient' or 'autonomous'. However, the big data technologies underlying these devices are not 'objective', and only replay and amplify pre-existing racial, gendered, and socio-economic biases and disadvantages (Buolamwini and Gebru, 2018; Noble, 2017). Thus we cannot be certain that all humans will assessed in quite the same way despite the presumed 'objectivity' of measurement.

Conclusions

At the time of writing, it was reported that the test driver in the Arizona crash, Rafaela Vasquez, was found guilty of 'negligent homicide' resulting in the death of Elaine Hertzberg. Arizona's easing of testing norms in a bid to "lure" AV companies, and the Uber/Volvo test vehicle's failed technologies, were also found to be at fault (Levin, 2020). However neither the companies nor the state were eventually liable, only Vasquez ended up in a "moral crumple zone": humans are ultimately responsible for failures of more advanced software that are supposed to replace them (Elish, 2019). This irony might be compounded by the Autonomy-Safety paradox: "as the level of robot autonomy grows, the risk of accidents will increase, and it will become more and more difficult to identify who is responsible for any damage." (Matsuzaki

and Lindemann, 2016, p. 502). It is possible that this will recur, thus there is a real urgency to research, development, and regulation going forward. In conclusion I want to argue that we need to acknowledge the gaps arising here and how we might think differently about the future.

This article has threaded 20th century histories of automation and automobility with current material-discursive practices that are shaping 'autonomy' as a measurable transition to 'robot' driving; my intention has been to bring multiple empirical and philosophical histories of technology to bear on the shifts taking place in human-machine relations. Far from being erased as 'autonomy' suggests, the human is in fact tightly woven into every aspect of the AV, from its computational and decisional flows to its accident accountability mechanisms. The legacies of high end aviation engineering are particularly influential in locating the human as part of a loop with the AV, 'the loop' suggesting bidirectional, shared opportunities to communicate and intervene in an automated process. Bainbridge's evocative 'ironies of automation' indicates the frictions and recursions in this process; these are amplified in the case of the autonomous vehicle that exists as multiple: as robot, and car, and as data-infrastructure. Yet contemporary approaches imagine the AV as an automobile or an airplane. As advanced as an airplane is, it is still only operated by skilled and specially trained people who are constantly accountable for the lives of fellow humans. The AV as it emerges now is not quite the same thing.

Scholars specifically call on philosophers of technology to attempt a "mobility turn" through "a hermeneutic circle between deeper conceptualisation and rigorous empirical investigations", "moving away from an "object-centred perspective.. to acknowledge existing and emerging social practices" towards a greater understanding of our relationships to technology (Mladenović et al., 2019, pp. 160–161). In that vein, I reprise Meg Leta Jones' critique of the human-in-the-loop paradigm; she argues that the task for the law is to bring the treatment of automation in line with responsible design and practical implementation; she calls for tying a "policy knot" across these different domains (2015, p. 102) This means being mindful of how emerging computing "practices and design impact, and are impacted by, structures and processes in the realm of policy" (117), which is critical if we are to protect human values. At minimum, we might begin by recognising the displacements humans inhabit as workers, managers, overseers, drivers, consumers and other publics. Peeling back the layers of the practices that validate 'autonomy', as I have attempted here, and identifying the role of the human in it, is a key part of this. The history of science and technology is replete with examples of disadvantaged people being even further marginalised; so in our breathless enthusiasm to roll out a new technology, we must acknowledge that inequities in human society will play into this emergence too.

Quite urgently, there must be new kinds of social protections and insurance for test drivers, and future drivers or owners who have only limited intervention in the AV's operations. Responsibility lies with car companies to address the practices of the computer vision industry in developing, benchmarking, and rolling out their products. There are automated systems that humans cannot intervene in and cannot be held accountable for, and to, like the logics of computer vision. In contexts where testing takes place unregulated and in public, how are local communities assured that they will not be mis-recognised, or altogether erased by, machine vision? What might it mean to have solidarity with movements of scholars and activists resisting being subjected to algorithmic classification? Further, if AVs are indeed more than just cars, and are commercial data platforms, then questions of labour, data protection, and data use, must become central as well. Just as gig and platform workers have organised,

as have other tech workers in Silicon Valley, what kinds of protections exist for people engaged in developing AV capabilities? The autonomous vehicle community of practice and research has not seriously addressed these concerns, and as such this reaches across different domains of automobility regulation, data protection, and AI governance. Innovation and policy research and advocacy could become more attentive to how multiple new publics and stakeholders are emerging in the shaping of this technology, in addition to traditional institutional actors and investors. All these networks and connections matter and must muddy the discursive construction and emergence of the AV. The irony of autonomy must be emphasised: that autonomy is not about separation or isolation, but is a matter of consistent connection and relations of mutual influence.

Data availability

Notes

1. 'Auto-pilot' is the Tesla-trademarked name for its auto-pilot technology and hence is capitalised; without the capitalisation it refers to the technology that was originally developed in aviation and now is seen in different kinds of AVs.
2. Here I refer chiefly to the cases in the United States, in Florida and in Arizona, for which there are accessible accident forensics and reports
3. While critical research into platform and 'gig' economies have opened our eyes to conditions of exploitative and inequitable human employment and labour here (for example Graham et al., 2017 and Gray and Suri, 2019), this article is only inspired by this work and does not contribute to nor draw on these works.
4. This was an unpaid association that unfolded over time as a series of regular discussions between professionals working on similar topics but different industries. There was one event with journalists that I was invited to speak at and was remunerated for by his employer, the car company.
5. Interview with 'Jason' (name changed), Ann Arbor, Michigan, 27 October 2017.
6. Supervised machine learning is distinguished from unsupervised learning in that the former entails algorithms learning from a labelled data set and reproducing that; in an unsupervised process, algorithms must extract features and patterns in an unlabelled data set.
7. The field of Human Computer Interaction (HCI) has been central to discussions of automation and engineering design, however the focus of this paper prevents a more thorough inclusion of it here. It is relevant to note that while earlier 'paradigms' of HCI were in step with the comparative approach to human and machine as entities that needed to be 'fit' into each other, or that were similar kinds of information processing devices there have been advances. The limitations of a comparative approach have given way to the embodied, affective, and sociotechnical and political aspects of human-machine relations.
8. https://twitter.com/sladner/status/1193576799772520448?s=03
9. Move Lab: https://www.move-lab.com/
10. Test drive with 'Jyoti' (name changed) Philadelphia, PA, 2 March 2018.
11. Interview with 'Dr. Jacob' (name changed), Virginia, United States, 29 September 2019.
12. The artist Paolo Cirio's database, Sociality Today, tracks patents related to affective and social manipulation. In this, 40 "affect patents" can be identified in response to the search term 'driving'. Affectiva, a Boston-based software company, is a frontrunner in developing this for autonomous driving: https://www.affectiva.com/.

References
Alvarez León LF (2019) Eyes on the road: surveillance logics in the autonomous vehicle economy. Surveil Soc 17(1/2):198–204

Amoore L (2018) Doubt and the algorithm: on the partial accounts of machine learning. Theory Culture Soc 0(0):1–23. https://doi.org/10.1177/0263276419851846

Amoore L (2020) Cloud ethics: algorithms and the attributes of ourselves and others. Duke University Press

Andrejevic M (2019) Automating surveillance. Surveil Soc 17(1/2):7–13

Asaro PM (2019) AI ethics in predictive policing: from models of threat to an ethics of care. IEEE Technology and Society Magazine 38(2):40–53

Awad E, D'Souza S, Kim R, Schulz J, Henrich J, Shariff A, Bonnefon J-F, Rahwan I (2018) The moral machine experiment. Nature 563(7729):59–64. https://doi.org/10.1038/s41586-018-0637-6

Bainbridge L (1983) The ironies of automation. Automatica 6:775–779

Barad K (2007) Meeting the universe halfway: quantum physics and the entanglement of matter and meaning. Duke University, Durham and London

Bardzell S (2018) Utopias of participation: feminism, design, and the futures. ACM Trans Comput Hum Interaction 25(1):1–24. https://doi.org/10.1145/3127359

Barrett LF, Adolphs R, Marsella S, Martinez AM, Pollak SD (2019) Emotional expressions reconsidered: challenges to inferring emotion from human facial movements [published correction appears in Psychol Sci Public Interest. 2019 Dec;20(3):165–166]. Psychol Sci Public Interest 20(1):1–68. https://doi.org/10.1177/1529100619832930

Bissell D, Birtchnell T, Elliott A, Hsu EL (2020) Autonomous automobilities: the social impacts of driverless vehicles. Curr Sociol 68(1):116–134. https://doi.org/10.1177/0011392118816743

Boudette NE (2016) Autopilot cited in death of Chinese Tesla driver. The New York Times https://www.nytimes.com/2016/09/15/business/fatal-tesla-crash-in-china-involved-autopilot-government-tv-says.html. Retrieved 18 Sep 2020

Boyd R (1993) Metaphor and theory change: What is 'metaphor' a metaphor for? In: Ortony A (ed.) Metaphor and thought, 2nd edn. Cambridge University, Cambridge, p 481–532

Buolamwini J, Gebru T (2018) Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency. Vol. 81. PMLR, pp. 77–91

California Department of Motor Vehicles. Website. https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/testing-autonomous-vehicles-with-a-driver/

Cavoli C, Phillips B, Cohen T, Jones P (2017) Social and behavioral questions associated with automated vehicles: a literature review. UCL Transport Institute, London. https://www.ucl.ac.uk/transport-institute/pdfs/social-and-behavioural-literature-review.pdf. Accessed 12 July 2017

Cheney-Lippold J (2019) Accidents happen. Soc Res Int Quart 86:513–535

Cummings ML (2014) Man vs. machine or man + machine? IEEE Intell Syst 29 (5):62–69

Cummings M (2017) Informing autonomous system design through the lens of skill-, rule-, and knowledge-based behaviors. J Cogn Eng Decision Making 12 (1):58–61. https://doi.org/10.1177/1555343417736461

Cummings ML, Mastracchio C, Thornburg KM, Mkrtchyan A (2013) Boredom and distraction in multiple unmanned vehicle supervisory control. Interact Comput 25(1):34–47

Cummings ML, Ryan JC (2014) Shared Authority Concerns in Automated Driving Applications. HAL Reports, DSpace@MIT Retrieved https://dspace.mit.edu/handle/1721.1/86937. Accessed 12 Nov 2020

Delfanti A, Frey B (2020) Humanly extended automation or the future of work seen through Amazon patents. Sci Technol Human Values 1–28. https://doi.org/10.1177/0162243920943665

Dixon L (2019) Autonowashing: the greenwashing of vehicle automation. https://doi.org/10.13140/RG.2.2.19836.69761

Dodge M, Kitchin R (2006) Code, vehicles and governmentality: the automatic production of driving spaces (NIRSA) Working Paper Series. No. 29 [Monograph]. NIRSA–National Institute for Regional and Spatial Analysis. http://www.nuim.ie/nirsa/research/documents/WPS29.pdf. Accessed 15 August 2020

Dwyer B (2020) Self-driving car dataset missing labels for hundreds of pedestrians. Roboflow Blog. February 11, 2020. https://blog.roboflow.ai/self-driving-car-dataset-missing-pedestrians/. Accessed 7 July 2020

Ekbia H, Nardi B (2014) Heteromation and its (dis)contents: the invisible division of labor between humans and machines. First Monday 19.6. Retrieved http://firstmonday.org/ojs/index.php/fm/article/view/5331. Accessed 15 Sep 2018

Ekbia H, Nardi B (2018) From form to content. Cultur Anthropol 33(3):360–367. ISSN 0886-7356

Elish MC, Hwang T (2015) Praise the machine! Punish the human! The contradictory history of accountability in automated aviation. Comparative Studies in Intelligent Systems–Working Paper #1 Intelligence and Autonomy Initiative1. Data & Society. http://www.datasociety.net/pubs/ia/Elish-Hwang_AccountabilityAutomatedAviation.pdf. Accessed 23 Sep 2018

Elish MC (2019) Moral crumple zones: cautionary tales in human-robot interaction. Engaging Sci Technol Soc 5:40–60. 2019 Mar 23

Esmonde K, Jette S (2020) Assembling the 'Fitbit subject': a Foucauldian-sociomaterialist examination of social class, gender and self-surveillance on Fitbit community message boards. Health 24(3):299–314. https://doi.org/10.1177/1363459318800166

Eubanks V (2018) Automating inequality: how high-tech tools profile, police, and punish the poor. St. Martin's Press

Fisch M (2016) Remediating infrastructure: Tokyo's commuter train network and the new autonomy. In: Penny H, Casper BJ, Atsuro M (eds) Infrastructure

and social complexity: a Routledge companion. pp. 115–127. Routledge, London and New York

Foucault M (1980) Power/knowledge: selected interviews and other writings, 1972–1977. In: Gordon C (ed) Harvester Press, New York

Gerla M, Lee E-K, Pau G, Lee U (2014) Internet of vehicles: from intelligent grid to autonomous cars and vehicular clouds. IEEE World Forum on Internet of Things. IEEE 2014:241–246

Graham M, Hjorth I, Lehdonvirta V (2017) Digital labour and development: impacts of global digital labour platforms and the gig economy on worker livelihoods. Eur Rev Labour Res 23(2):135–162. https://doi.org/10.1177/1024258916687250

Gray ML, Suri S (2019) Ghost work: how to stop Silicon Valley from building a new global underclass. Houghton Mifflin Harcourt, Boston and New York

Haraway D (1988) Situated knowledges: the science question in feminism and the privilege of partial perspective. Femin Stud 14:3. Fall 1988

Harrison S, Sengers P, Tatar D (2007) Three paradigms of HCI. Conference on Human Factors in Computing Systems CHI 2007, April 28–May 3, 2007, San Jose, USA. https://people.cs.vt.edu/~srh/Downloads/TheThreeParadigmsofHCI.pdf Accessed 14 November 2017

Hawkins, AJ (2020, February 26). Everyone hates California's self-driving car reports. The Verge. https://www.theverge.com/2020/2/26/21142685/california-dmv-self-driving-car-disengagement-report-data

Hayles NK (2017) Unthought: The power of the cognitive nonconscious. Chicago, University of Chicago

Ihde D (1995). Postphenomenology: Essays in the postmodern context. Northwestern University Press

Irani LC, Silberman MS (2013) Turkopticon: Interrupting worker invisibility in Amazon Mechanical Turk. In Proceedings of Conference of Human Factors in Computing Systems. Apr 28-May 2, 2013. CHI, 2013

Jones ML (2015) The ironies of automation law: tying policy knots with fair automation practices principles. Vanderbilt J Entertain Technol Law 18:77

Kahn J (2018) To get ready for robot driving, some want to reprogram pedestrians. Bloomberg. https://www.bloomberg.com/news/articles/2018-08-16/to-get-ready-for-robot-driving-some-want-to-reprogram-pedestrians. Accessed 13 May 2019

Kröger F (2015) Automated driving in its social, historical and cultural contexts. In: Maurer MJ, Gerdes C, Lenz B, Winner H (eds) Autonomous driving: technical, legal and social aspects. Springer, Heidelberg and Berlin, pp 41–68

Lee D (2018) Why Big Tech pays poor Kenyans to teach self-driving cars. BBC News Internet. https://www.bbc.com/news/technology-46055595

Levin S (2020) Safety driver charged in 2018 Incident where self-driving Uber car killed a woman. The Guardian. https://www.theguardian.com/us-news/2020/sep/16/uber-self-driving-car-death-safety-driver-charged. Accessed 18 Sep 2020

Levy KEC (2015) The contexts of control: information, power, and truck-driving work. Inform Soc 31(2):160–174. https://doi.org/10.1080/01972243.2015.998105

Marra WC, McNeil SK (2012) Understanding "the loop": regulating the next generation of war machines. 36:3 Harvard J Law Public Policy 1139

Matsuzaki H, Lindemann G (2016) The autonomy-safety-paradox of service robotics in Europe and Japan: a comparative analysis. AI Soc 31(4):501–517. https://doi.org/10.1007/s00146-015-0630-7

McNeese NJ, Demir M, Cooke NJ, Myers C (2018) Teaming with a synthetic teammate: insights into human-autonomy teaming. Hum Fact 60:262–273. https://doi.org/10.1177/0018720817743223

Mladenović MN, Lehtinen S, Soh E, Martens K (2019) Emerging urban mobility technologies through the lens of everyday urban aesthetics: case of self-driving vehicle. Essay Philos 20(2):146–170

Mladenović MN, Stead D, Milakis D, Pangbourne K, Givoni M (2020) Governance cultures and sociotechnical imaginaries of self-driving vehicle technology: Comparative analysis of Finland, UK and Germany. Policy Implications of Autonomous Vehicles. Elsevier, Amsterdam, The Netherlands, 2020 Jan 27. pp. 235–262

Mol A (2002) The body multiple. Durham and London, Duke University

Natale S, Ballatore A (2017) Imagining the thinking machine: technological myths and the rise of artificial intelligence. Convergence: Int J Res New Media Technol 1–16. https://doi.org/10.1177/1354856517715164

National Transportation Safety Board (2017) Collision between a car operating with automated vehicle control systems and a tractor-semitrailer truck near Williston, Florida. May 7, 2016. Accident Report. NTSB/HAR-17/02 PB2017-102600. https://dms.ntsb.gov/pubdms/search/document.cfm?docID=453252&docketID=59989&mkey=93548. Accessed 14 Nov 2017

National Transportation Safety Board (2018) Preliminary report highway: HWY18MH010. https://www.ntsb.gov/investigations/AccidentReports/Pages/HWY18MH010-prelim.aspx. Accessed 7 Oct 2018

National Transportation Safety Board (2019) Public meeting. Collision between vehicle controlled by developmental automated driving system and pedestrian. November 19, 2019. https://www.ntsb.gov/news/press-releases/Pages/NR20191119c.aspx. Accessed 13 Dec 2019

Noble S (2017) Algorithms of oppression: how search engines reinforce racism. NYU, New York

Norton P (2011) Fighting traffic: the dawn of the motor age in the American City. MIT, Cambridge

O'Malley J (2018) Captcha if you can: How you've been training AI for years without realizing it. Tech Radar. https://www.techradar.com/news/captcha-if-you-can-how-youve-been-training-ai-for-years-without-realising-it. Accessed 3 Oct 2018

Rosenberger R, Verbeek P-P (2015) Postphenomenological investigations: essays on human–technology relations. Lexington Books, Lanham MD

Rosenthal C (2019) Accounting for slavery: masters and management. Harvard University, Cambridge MA

Roy A (2018) How the language of self-driving is killing us. The Drive. https://www.thedrive.com/opinion/20495/how-the-language-of-self-driving-iskilling-us Retrieved 15 October 2020

Schmidt FA (2019) Crowdproduktion von Trainingsdaten. Zur Rolle von Online-Arbeit beim Trainieren autonomer Fahrzeuge. Study Nr. 415, Düsseldorf: Hans-Böckler-Stiftung. https://www.boeckler.de/pdf/p_study_hbs_417.pdf. Accessed 19 Jan 2020, ISSN 2509-2359

Schoettle B, Sivak M (2015) Potential impact of self-driving vehicles on household vehicle demand and usage (Report 2015-3). University of Michigan Transportation Research Institute, Ann Arbor

Seiler C (2008) Republic of drivers: a cultural history of automobility in America. University of Chicago, Chicago

Sekula A (1986) The body and the archive. October. 1986 Dec 1;39:3-64.

Sheller M (2004) Automotive emotions: feeling the car. Theory Culture Soc 21(4/5):221–242. https://doi.org/10.1177/0263276404046060

Shepardson D (2018) Tesla says crashed vehicle had been on autopilot prior to accident. Reuters. https://www.reuters.com/article/us-tesla-crash/tesla-says-crashed-vehicle-had-been-on-autopilot-prior-to-accident-idUSKBN1H7023. Accessed 6 Oct 2018

Sheridan TB (1992) Telerobotics, automation, and human supervisory control. Cambridge, MIT Press

Sheridan TB, Parasuraman R (2005) Human-automation interaction. Rev Human Factors Ergon 1(1):89–129. https://doi.org/10.1518/155723405783703082

Society of Automotive Engineers (2014) Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems J3016_20140. https://www.sae.org/standards/content/j3016_201401/. Accessed 5 July 2020

Society of Automotive Engineers (2018) Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles J3016_201806. https://www.sae.org/standards/content/j3016_201806/. Accessed 30 June 2020

Stayton E, Stilgoe J (2020) It's time to rethink levels of automation for self-driving vehicles. SSRN. https://doi.org/10.2139/ssrn.3579386

Stewart J (2017) Mapped: The top 263 companies racing toward autonomous cars. Wired 10 May 2017. https://www.wired.com/2017/05/mapped-top-263-companies-racing-toward-autonomous-cars/. Accessed 14 Nov 2017

Stilgoe J (2017) Seeing like a Tesla. How can we anticipate self-driving worlds? Glocalism: J Culture Polit Innovat 2017(3):1–20. https://doi.org/10.12893/gjcpi.2017.3.2

Stilgoe J (2019) Who's driving innovation? New technologies and the collaborative state. Palgrave MacMillan, Switzerland

Suchman L, Weber J (2016) Human-machine autonomies. In:Bhuta N, Beck S, Geis R, Liu H-Y, Kreis C (eds) Autonomous weapons systems. Cambridge, Cambridge UK, p 75–102

Tennant CJ, Howard S, Stares S (2020/forthcoming). Imaginaries of driverless mobility: a UK case study

Tesla (2016) A Tragic Loss. https://www.teslamotors.com/blog/tragic-loss. Retrieved 15 Mar 2017

Urry J (2004) The 'system' of automobility. Theory Culture Soc 21:25–39. https://doi.org/10.1177/0263276404046059

Verbeek P-P (2011) Moralizing technology: understanding and designing the morality of things. University of Chicago, Chicago/London

Verbeek P-P (2017) Designing the morality of things: The ethics of behaviour-guiding technology. In: van den Hoven J, Miller S, Pogge T (eds) Designing in ethics. Cambridge University, Cambridge, UK, pp. 78–94

de Visser EJ, Pak R, Shaw TH (2018) From 'automation' to 'autonomy': the importance of trust repair in human–machine interaction. Ergonomics 61 (10):1409–1427. https://doi.org/10.1080/00140139.2018.1457725

Wilson B, Hoffman J, Morgenstern J (2019) Predictive inequity in object detection. arXiv:1902.11097v1 [cs.CV] 21 Feb 2019

Yurtsever E, Lambert J, Carballo A, Takeda K (2020) A survey of autonomous driving: common practices and emerging technologies. IEEE Access 8:58443–58469. https://doi.org/10.1109/ACCESS.2020.2983149

Zhang Q, Esterwood C, Yang XJ, Robert Jr LP (2019) An automated vehicle (AV) like me? The Impact of personality similarities and differences between humans and AVs. http://arxiv.org/abs/1909.11766. Accessed 30 June 2020

Competing interests

The author declares no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.I.G.

Reprints and permission information is available at http://www.nature.com/reprints

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