

AI4People: Ethical Guidelines for the Automotive Sector – Fundamental Requirements and Practical Recommendations

Christoph Lütge, Technical University of Munich, Munich, Germany

Franziska Poszler, Technical University of Munich, Munich, Germany

Aida Joaquin Acosta, Mobility and the Urban Agenda, Spain & Harvard University, USA

David Danks, Carnegie Mellon University, USA

Gail Gottehrer, Law Office of Gail Gottehrer LLC, USA

Lucian Mihet-Popa, Østfold University College, Halden, Norway

Aisha Naseer, Fujitsu Laboratories of Europe, UK

ABSTRACT

This paper presents the work of the AI4People-Automotive Committee established to advise more concretely on specific ethical issues that arise from autonomous vehicles (AVs). Practical recommendations for the automotive sector are provided across the topic areas: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental wellbeing, as well as accountability. By doing so, this paper distinguishes between policy recommendations that aim to assist policymakers in setting acceptable standards and industry recommendations that formulate guidelines for companies across their value chain. In the future, the automotive sector may rely on these recommendations to determine relevant next steps and to ensure that AVs comply with ethical principles.

KEYWORDS

AI4People, Autonomous Driving, Autonomous Vehicle Ethics, Ethics of AI, Fundamental Rights, Governance, Override, Regulation, Self-Driving Cars, Transparency

AIM AND SCOPE OF THIS PAPER

In the past decade, many policy documents have discussed ethical issues and potential future directions related to new emerging technologies such as artificial intelligence (AI) or autonomous systems. This paper presents the work of the AI4People-Automotive Committee¹ established to advise more concretely on specific ethical issues that arise from autonomous vehicles (AVs). The committee consisted of industry experts and researchers from the fields of ethics, law, philosophy, engineering,

DOI: 10.4018/IJT.20210101.oa2

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

technology and policy. The aim of this paper is to provide the automotive sector, including both companies and public entities such as regulators, with concrete and practical guidelines to comply with ethical principles within the AI systems of AVs. Therefore, this paper could serve as a checklist for policymakers and companies as well as a basis for developing a certification of ethics, an ‘ecosystem of trust’ (European Commission, 2020b) and ultimately a ‘Good AI Society’ (Floridi et al., 2018) in the automotive sector. These guidelines are intended to provide a clearer vision and moral compass on how to proceed and what to consider when developing AVs, rather than additional barriers to innovation. The automotive sector is defined here in the broadest terms possible to encompass a wide range of companies involved in the development of vehicles, including private cars, trucks, busses and shuttles. Sea, air and military-type applications have been excluded due to their functional and ethical specificity. This paper will focus on the ethics of the AI-based tools that are used in automotive technology, rather than on the ethics of vehicles in general.

This paper distinguishes between high-level guidelines for policymakers (‘policy recommendations’) and concrete actionable recommendations for companies (‘industry recommendations’). However, the line between the two cannot always be drawn clearly which also highlights the importance of co-regulation (i.e. the interaction of legal regulation and self-regulation by companies) (Pagallo et al., 2019). The policy recommendations are designed to focus attention on pressing policy issues and assist in setting acceptable standards. Thus, the policy recommendations ultimately influence the industry recommendations. Responsible targets for the execution of the policy recommendations are: policymakers, legislators, ethics standards boards and commissions such as the United Nations Economic Commission for Europe (UNECE). The industry recommendations formulate guidelines for companies across their entire value chain (especially during research & development, production & operations and service). Therefore, original equipment manufacturer (OEM) and car manufacturers are the primary responsible targets for those recommendations.

Before turning to the principles and guidelines, we note three key points of consensus among the authors: (1) a responsible balancing of risks or estimated harm should be permitted at any time for AVs; (2) a large-scale introduction of full-mode AVs (level 4 and higher) onto streets is unlikely in the short run, so we must consider a more incremental, step-by-step approach; and (3) policymakers face significant challenges now, and so there are significant pressures to quickly develop a clear regulatory framework.

THE GUIDELINES

Fundamental Rights Underlying the Guidelines

Particular fundamental rights are the basis for the proposed seven requirements that were originally derived by the High-level Expert Group on Artificial Intelligence (2019) (i.e. human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental wellbeing; accountability) and recommendations in this paper. In addition to general human dignity, key fundamental rights (United Nations, 1948) that policymakers and companies in the automotive sector should recognize are: *Right to self-determination and liberty* which draws attention to human agency in self-driving cars (i.e. importance of override options) (see Guideline 1). *Right to life and security* which entails ensuring technical robustness and safety of operating self-driving vehicles; on a broader level, this includes securing societal and environmental wellbeing (see Guideline 2 and 6). *Right to protection of personal data* drawing attention to data ownership, data governance and privacy of personal data that is generated during the operation of self-driving cars (see Guideline 3). *Right to equality and non-discrimination* requiring the avoidance of unfair bias in operating vehicles as well as the accessibility of benefits for every individual in society (see Guideline 5). *Right to explanation* which, in the field of autonomous driving, demands transparency and communication of the underlying functionality,

23 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/ai4people/269438

Related Content

Perverting Activism: Cyberactivism and Its Potential Failures In Enhancing Democratic Institutions

Tommaso Bertolotti, Emanuele Bardone and Lorenzo Magnani (2011). *International Journal of Technoethics* (pp. 14-29).

www.irma-international.org/article/perverting-activism-cyberactivism-its-potential/54753

Ethics, Decision-Making, and Risk Communication in the Era of Terroredia: The Case of ISIL

Mahmoud Eid (2016). *International Journal of Technoethics* (pp. 91-104).

www.irma-international.org/article/ethics-decision-making-and-risk-communication-in-the-era-of-terroredia/152807

Intellectual Property and Licensing Strategies in Open Collaborative Innovation

Marcel Bogers, Rudi Bekkers and Ove Granstrand (2013). *Digital Rights Management: Concepts, Methodologies, Tools, and Applications* (pp. 1204-1224).

www.irma-international.org/chapter/intellectual-property-licensing-strategies-open/71026

Algorithms vs. Hive Minds: Preserving Democracy's Future in the Age of AI

Rick Searle (2018). *The Changing Scope of Technoethics in Contemporary Society* (pp. 275-288).

www.irma-international.org/chapter/algorithms-vs-hive-minds/202504

Privacy and Property in the Global Datasphere

D. L. Burk (2007). *Information Technology Ethics: Cultural Perspectives* (pp. 94-107).

www.irma-international.org/chapter/privacy-property-global-datasphere/23657