PI: How might my work with the Edinburgh University Formula Student's (EUFS) Artificial Intelligence Localisation Team contribute to unintentional harm?

In the computing industry, the rapid innovation rate leads to an ever-increasing number of ethical issues that need to be accounted for. My role in the EUFS society begs the question: what implications will my specific role and the wider subject area carry?

The Edinburgh University Formula Student Society produces a self-driving car for competition every year. I worked on the localisation software which helps the car determine its position. This involves using various algorithms to predict the car's location based on sensor data. Furthermore, this naturally influences both what the car is told to perceive and where it is told to move.

A critical concern that must be taken into consideration is the way in which the localisation system fails. When the car's estimated position does not match the car's real position, failure and then crashes are likely to occur. While these localisation systems are designed to be accurate and take into account as many sources of input as possible, failure will always be a threat. This failure has a cost in its potential for injury and damage to people and infrastructure. Additionally, the mitigation of failure when it does occur introduces harm reduction, a choice between protecting the driver and objects or persons external to the car. Some schools of thought already elect to protect the driver and vehicle above all else in the car's physical design such as Tesla's [1]. This, however, has come under criticism as it could cause a greater net negative impact by causing grievous harm to other parties [4]. In either case, it is essential that the prioritisation of input is thoroughly researched so as to come to an informed decision that agrees with codes of conduct like the Association for Computing Machinery (ACM) Code of Ethics and Professional Conduct [2]. Even the errors due to input that isn't prioritised should only be allowed to occur at a minimal rate. Therefore, it is critical that the failure rate be minimised by ensuring that potential failure is detected as early as possible. The localisation system only has probable estimates of the car's position and a low uncertainty threshold should be in place to allow graceful failure. However, failing gracefully is in itself a challenge as cars suddenly stopping or pulling over is also dangerous. It could be said that once a self-driving car is no more sure than a driver could be, it should stop. Self-driving features like Tesla's elect to leave such decisions in the drivers' hands [3] but there are also calls for those decisions to be made by the car [4, 5]. It isn't enough to say that when it fails, evasion should occur. The system's sheer criticality demands that the failure rate be exceptionally low before deployment in a non-test area.

When attempting to increase the accuracy of localisation, algorithm design and optimisation can only go so far. Increasing the variety and amount of data collected from sensors and images of the car's surroundings enable easier detection and avoidance of civilians. However, this would encompass personal data and reduce the general population's privacy [6] which ACM 1.6 [2] calls to respect. The already fairly large security risk that hacking into the localisation system presents would be accentuated by the possession of private data. GDPR law needs to be followed in any public deployment of the car. Furthermore, the car's movement is largely dependent on output received from the localisation system; should the localisation system be compromised by foreign agents, the car could be led to crash or cause serious damage to civilians and infrastructure with little risk to the hijacking agents. Security is thus paramount in this system due to the physical threat it poses [7].

As a team member, I must ensure that these responsibilities are considered at all development stages. Prioritising human life over all other non-human factors is essential in this system. Ethical codes with regards to harm reduction are still under debate and evolving, so software developers should evolve with them [8]. Furthermore, a test and security-driven approach should be ensured so as to protect both the privacy and livelihood of those in the car's operational zone. Certainty thresholds should not be slackened, as this results in catastrophic consequences as was the case when Boeing's flight aiding tool caused two crashes [14]. Industry-standard security is seen as possibly too frail [9] for even rudimentary web safety and ACM 2.6 [2] stresses outsourcing of security is needed if those worries cannot be remedied. As per ACM 1.1 [2], the responsibility lies not just in relation to the car but also to the wider industry and the way that it will impact the stakeholders of the wider society. Such technology's development will likely lead to the increased use of autonomous robots and vehicles. The weaponisation of such technology has the potential to escalate and create tension between military powers [10] or possibly remove human compassion escalating brutality towards citizens [11] conflicting with ACM 1.2's [2] principle of "avoid harm.". Additionally, without explicit information regarding the capabilities of the developing software, regulators lack the foresight to draft laws regulating industry behaviour. This makes the industry itself responsible for behaving ethically, which has previously been seen to be problematic [12]. This dilemma is set to grow as technology becomes increasingly sophisticated and inaccessible to laypeople. To counter this, individual effort needs to be matched by institutional openness as proposed by ACM 2.7 [2]. Likewise, unregulated release into the open market could cause economic turbulence for industries such as taxis [13] which also carries social and cultural ramifications as job roles are subject to unconstrained, rapid change.

Thus there is a need for a considered release of localization technology when paired with the control of vehicles as is the case in the EUFS project. The unintentional harms in isolation are limited to damage to people and infrastructure but it poses wider risks to society, privacy and public safety and all must stay in consideration.

[1] news.com.au. 2021. *New Tesla Cybertruck could put other road users at risk*. [online] Available at: <https://www.news.com.au/technology/innovation/motoring/motoring-news/new-tesla-cybertruck-could-put-other-road-users-at-risk/news-story/da648cad43ef731ae5dbb45632a6ed41> [Accessed 22 October 2021].

[2] Acm.org. 2021. *ACM Code of Ethics and Professional Conduct*. [online] Available at: <https://www.acm.org/code-of-ethics> [Accessed 22 October 2021].

[3] Lin, P., 2021. *Here's How Tesla Solves A Self-Driving Crash Dilemma*. [online] Forbes. Available at: <https://www.forbes.com/sites/patricklin/2017/04/05/heres-how-tesla-solves-a-self-driving-crash-dilemma/?sh=101e1d3b6813> [Accessed 22 October 2021].

[4] MIT Technology Review. 2021. *Why Self-Driving Cars Must Be Programmed to Kill*. [online] Available at: <https://www.technologyreview.com/2015/10/22/165469/why-self-driving-cars-must-be-programmed-to-kill/> [Accessed 22 October 2021].

[5] The Verge. 2021. *Elon Musk, live at the Code Conference*. [online] Available at: <https://www.theverge.com/2021/9/28/22698563/elon-musk-code-conference-liveblog> [Accessed 22 October 2021].

[6] Data Protection Report. 2021. *The Privacy Implications of Autonomous Vehicles - Data Protection Report*. [online] Available at: <https://www.dataprotectionreport.com/2017/07/the-privacy-implications-of-autonomous-vehicles/> [Accessed 26 October 2021].

[7] Gowling WLG. 2021. *Risks and liability - Cyber security and autonomous vehicles*. [online] Available at: <https://gowlingwlg.com/en/insights-resources/articles/2020/cyber-security-and-autonomous-vehicles/> [Accessed 22 October 2021].

[8] Newsweek. 2021. *Ethical Standards for Self-Driving Car Testing Are Still in Their Beta Stage*. [online] Available at: <https://www.newsweek.com/ethical-standards-self-driving-car-testing-are-still-their-beta-stage-1634910> [Accessed 22 October 2021].

[9] Medium. 2021. *What is going on with OAuth 2.0? And why you should not use it for authentication*. [online] Available at: <https://medium.com/securing/what-is-going-on-with-oauth-2-0-and-why-you-should-not-use-it-for-authentication-5f47597b2611> [Accessed 22 October 2021].

[10] Jonny Hallam, C., 2021. *Deadly drone attack on tanker escalates Iran-Israel maritime tensions*. [online] CNN. Available at: <https://edition.cnn.com/2021/07/31/middleeast/iran-israel-tanker-attack-drone-oman-intl/index.html> [Accessed 22 October 2021].

[11] American Civil Liberties Union. 2021. *ACLU News & Commentary*. [online] Available at: <https://www.aclu.org/news/privacy-technology/robot-police-dogs-are-here-should-we-be-worried/> [Accessed 22 October 2021].

[12] BBC News. 2021. *Cambridge Analytica: Warrant sought to inspect company*. [online] Available at: <https://www.bbc.co.uk/news/technology-43465700> [Accessed 22 October 2021].

[13] employees, U., 2021. *Uber loses court battle over whether its drivers are employees*. [online] euronews. Available at: <https://www.euronews.com/2021/09/13/uber-loses-court-battle-over-whether-its-drivers-are-permanent-employees> [Accessed 22 October 2021].

[14] BBC News. 2021. *'Boeing played Russian roulette with people's lives'*. [online] Available at: <https://www.bbc.co.uk/news/extra/jDOe2y9Tbo/boeing-737-max> [Accessed 22 October 2021].