Chapter 6
Right to Life, Liberty and Security of Persons

Abstract Artificial intelligence (AI) can support individuals' enjoyment of life, liberty and security, but it can also have adverse effects on them in a variety of ways. This chapter covers three cases affecting human life, liberty and security: one in transportation (self-driving cars), one in the home (smart security systems) and one in healthcare services (adversarial attacks). The chapter discusses ethical questions and three potential solutions to address AI human rights issues related to life, liberty and security of persons: defining and strengthening liability regimes, implementing quality management systems and adversarial robustness. AI developers, deployers and users must respect the sanctity of human life and embed, value and respect this principle in the design, development and use of their products and/or services. Critically, AI systems should *not* be programmed to kill or injure humans.

Keywords Right to life · Safety · Security · Self-driving cars · Smart homes · Adversarial attacks

6.1 Introduction

All humans enjoy the right to life, liberty and security of the person. The right to life is also included as a core right in 77% of the world's constitutions (UN 2018), is the cornerstone of other rights and is enshrined in international human rights instruments (Table 6.1).

State parties who are signatories to the human rights instruments enshrining the right to life have a duty to take necessary measures to ensure individuals are protected from its violation: i.e. its loss, deprivation or removal.

Artificial intelligence (AI) can support an individual's enjoyment of life, liberty and security by, for example, supporting the diagnosis and treatment of medical conditions. Raso et al. (2018) outline how criminal justice risk assessment tools could benefit low-risk individuals through increased pre-trial releases and shorter sentences. Reports suggest that AI tools could help identify and mitigate human security risks and lower crime rates (Deloitte n.d., Muggah 2017).

B. C. Stahl et al., *Ethics of Artificial Intelligence*,
SpringerBriefs in Research and Innovation Governance,
https://doi.org/10.1007/978-3-031-17040-9_6

Table 6.1 Right to life in international human rights instruments

Provision	Human Rights Instrument
Right to life, liberty and security of person	Universal Declaration of Human Rights (UN 1948: art. 3)
Right to life	International Covenant on Civil and Political Rights (UN 1966: art. 6)
Right to life, survival and development	Convention on the Rights of the Child (UN 1989: art. 6)
Right to life	International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families (UN 1990: art. 9)
Right to life	Convention on the Rights of Persons with Disabilities (UN 2006: art. 10)
Right to life	American Convention on Human Rights (Pact of San José) (UN 1969: art. 4)
Right to life	African Charter on Human and Peoples' Rights (Banjul Charter) (ACHPR 1981: art. 4)
Right to life and dignity in old age	Inter-American Convention on Protecting the Human Rights of Older Persons (OAS 2015: art. 6)
Right to life	European Convention on Human Rights (ECHR 1950: art. 2)

AI can have adverse effects on human life, liberty and security in a variety of ways (Vasic and Billard 2013; Leslie 2019), as elaborated in this chapter. Human rights issues around life, liberty and security of persons are particularly serious, and risks from the use of AI need to be weighed up against the risks incurred when not using AI, in comparison with other innovations. AI systems identified as high-risk (European Commission 2021) include those used in critical infrastructure (e.g. transportation) that could put the life and health of people at risk; in educational or vocational training that determine access to education and the professional course of someone's life (e.g. the scoring of exams); in the safety components of products (e.g. AI applications in robot-assisted surgery); in employment, the management of workers and access to self-employment (e.g. CV-sorting software for recruitment procedures); in essential private and public services (e.g. when credit scoring denies citizens the opportunity to obtain a loan); in law enforcement that may interfere with people's fundamental rights (e.g. evaluation of the reliability of evidence); in migration, asylum and border control management (e.g. verification of the authenticity of travel documents); and in the administration of justice and democratic processes (e.g. applying the law to a concrete set of facts). These categories of high-risk AI systems have the potential to impact the right to life, liberty and security (some in more direct ways than others, but nonetheless relevant).

Life-threatening issues have been raised regarding the use of robot-assisted medical procedures and robotics systems in surgery (Alemzadeh et al. 2016), robot accidents and malfunctions in manufacturing, law enforcement (Boyd 2016), retail and entertainment settings (Jiang and Gainer 1987), security vulnerabilities in smart home hubs (Fránik and Čermák 2020), self-driving and autonomous vehicles (AP and Reuters 2021), and lethal attacks by AI-armed drone swarms and autonomous weapons (Safi 2019). We look at three different cases affecting human life, liberty and security, one in the transportation context (self-driving cars), one related to the home (smart home security), and one in the healthcare service setting (adversarial attacks).

6.2 Cases of AI Adversely Affecting the Right to Life, Liberty and Security of Persons

6.2.1 Case 1: Fatal Crash Involving a Self-driving Car

> In May 2016, a Tesla car was the first known self-driving car to be involved in a fatal crash. The 42-year-old passenger/driver died instantly after colliding with a tractor-trailer. The tractor driver was not injured. "According to Tesla's account of the crash, the car's sensor system, against a bright spring sky, failed to distinguish a large white 18-wheel truck and trailer crossing the highway" (Levin and Woolf 2016). An examination by the Florida Highway Patrol concluded that the Tesla driver had not been attentive and had failed to take evasive action. At the same time, the tractor driver had failed, during a left turn, to give right of way, according to the report. (Golson 2017)

In this case, the driver had put his car into Tesla's autopilot mode, which was able to control the car. According to Tesla, its autopilot is "an advanced driver assistance system that enhances safety and convenience behind the wheel" and, "[w]hen used properly" is meant to reduce a driver's "overall workload" (Tesla n.d.). While Tesla clarified that the underlying autonomous software was designed to nudge consumers to keep their hands on the wheels to make sure they were paying attention, that does not seem to have happened in this case and resulted in a fatality. According to Tesla, "the currently enabled Autopilot and Full Self-Driving features require active driver supervision and do not make the vehicle autonomous" (ibid).

In 2018, an Uber test driver in charge of monitoring one of the company's self-driving cars was charged with negligent homicide when it hit and killed a pedestrian. An investigation by the National Transportation Safety Board (NSTB) concluded that the crash had been caused by the Uber test driver being distracted by her phone and implicated Uber's inadequate safety culture (McFarland 2019). The NSTB also found that Uber's system could not correctly classify and predict the path of a pedestrian crossing midblock.

In 2021, two men were killed in Texas after the Tesla vehicle they were in, which was going at a high speed, went off the road and hit a tree. The news report also mentioned that the men been discussing the autopilot feature before they drove off (Pietsch 2021). Evidence is believed to show that no one was driving the vehicle when it crashed.

While drivers seem to expect self-driving cars, as marketed to them, to give them more independence and freedom, self-driving cars are not yet, as stated by Tesla, for example, “autonomous”. The autopilot function and the “Full Self-Driving” capability are intended for use with a fully attentive driver with hands on the wheel and ready to take over at any moment.

While some research (Kalra and Groves 2017; Teoh and Kidd 2017) seems to suggest that self-driving cars may be safer than those driven by the average human driver, the main case and the further examples cited here point to human safety challenges from different angles: the safety of the drivers, passengers and other road users (e.g. cyclists, pedestrians and animals) and objects that encounter self-driving cars.

Other standard issues raised about self-driving cars, as outlined by Jansen et al. (2020), relate to *security* (the potential for their hacking leading to the compromising of personal and sensitive data) and *responsibility*, that is, where does responsibility for harms caused lie: with the manufacturer, the system programmer or software engineer, the driver/passenger, or the insurers? A responsibility gap could also occur, as pointed out by the Council of Europe’s Committee on Legal Affairs and Human Rights, “where the human in the vehicle—the ‘user-in-charge’, even if not actually engaged in driving—cannot be held liable for criminal acts and the vehicle itself was operating according to the manufacturer’s design and applicable regulations.” (Council of Europe 2020). There is also the challenge of shared driving responsibilities between the human driver and the system (BBC News 2020).

The underlying causes that require addressing in these cases include software/system vulnerabilities, inadequate safety risk assessment procedures and oversight of vehicle operators, as well as human error and driver distractions (including a false sense of security) (Clifford Law 2021).

6.2.2 Case 2: Smart Home Hubs Security Vulnerabilities

A smart home hub is a control centre for home automation systems, such as those operating the heating, blinds, lights and internet-enabled electronic appliances. Such systems allow the user to interact remotely with the hub using, for instance, a smartphone. A user who is equipped to activate appliances remotely can arrive at home with the networked gas fire burning and supper ready in the networked oven. However, it is not only the users themselves who can access their smart home hubs, but also external entities, if there are security vulnerabilities, as was the case for three companies operating across Europe. (Fránik and Čermák 2020)

Smart home security vulnerabilities directly affect all aspects of the right to life, liberty and security of the person. E.g., Man-in-the-middle attacks that interrupt or spoof communication between smart home devices and denial-of-service attacks could disrupt or shut devices down and compromise user well-being, safety and security.

Such vulnerabilities and attacks exploiting them can threaten a home, together with the peaceful enjoyment of life and human health within it. Unauthorised access could also result in threats to human life and health. For example, as outlined in a report from the European Union Agency for Cybersecurity (ENISA), safety might be compromised and human life thus endangered by the breach, or loss of control, of a thermostat, a smoke detector, a CO_2 detector or smart locks (Lévy-Bencheton et al. 2015).

When smart home security is exposed to vulnerabilities and threats, these can facilitate criminal actions and intrusions, or could themselves be a form of crime (e.g. physical damage, theft or unauthorised access to smart home assets) (Barnard-Wills et al. 2014).

While there are many other ethical issues that concern smart homes (e.g. access, autonomy, freedom of association, freedom of movement, human touch, informed consent, usability), this case study also further underlines two critical issues connected to the right to life: *security* and *privacy* (Marikyanet al. 2019; Chang et al. 2021). Hackers could spy on people, get access to very personal information and misuse smart-home-connected devices in a harmful manner (Laughlin 2021). Nefarious uses could include the perpetration of identity theft, location tracking, home intrusions and access lock-outs.

The responsibilities for ensuring that smart home devices and services do not suffer from vulnerabilities or attacks are manifold, and lie largely with the manufacturers and service providers, and with users. Users of smart-home-connected devices must carry out their due diligence when purchasing smart devices (by buying from reputable companies with good security track records and ensuring that security is up to the task).

6.2.3 Case 3: Adversarial Attacks in Medical Diagnosis

Medical diagnosis, particularly in radiology, often relies on images. Adversarial attacks on medical image analysis systems are a problem (Bortsova et al. 2021) that can put lives at risk. This applies whether the AI system is tasked with the medical diagnosis or whether the task falls to radiologists, as an experiment with mammogram images has shown. Zhou et al. used a generative adversarial network (GAN) model to make intentional modifications to radiology images taken to detect breast cancer (Zhou et al. 2021). The resulting fake images were then analysed by an AI model and by radiologists. The adversarial samples “fool the AI-CAD model to output a wrong diagnosis on 69.1% of the cases that are initially correctly classified by the AI-CAD model. Five breast imaging radiologists visually identify 29–71% of the adversarial samples" (ibid). In both cases, a wrong cancer diagnosis could lead to risks to health and life.

Adversarial attacks are "advanced techniques to subvert otherwise-reliable machine-learning systems" (Finlayson et al. 2019). These techniques, for example by making tiny image manipulations (adversarial noise) to images that might help confirm a diagnosis, guarantee positive trial results or control the rates of medical interventions to the advantage of those carrying out such attacks (Finlayson et al. 2018).

To raise awareness of adversarial attacks, Rahman et al. (2021) tested COVID-19 deep learning applications and found that they were vulnerable to adversarial example attacks. They report that due to the wide availability of COVID-19 data sets, and because some data sets included both COVID-19 patients' public data and their attributes, they could poison data and launch classified inference attacks. They were able to inject fake audio, images and other types of media into the training data set. Based on this, Rahman et al. (2021) call for further research and the use of appropriate defence mechanisms and safeguards.

The case study and examples mentioned in this section expose the problem of machine and deep learning application vulnerabilities in the healthcare setting. They show that a lack of appropriate defence mechanisms, safeguards and controls would cause serious harm by changing results to detrimental effect.

6.3 Ethical Questions

All the case studies raise several ethical issues. Here we discuss some of the core ones.

6.3.1 Human Safety

To safeguard human safety, which has come to the fore in all three case studies, unwanted harms, risks and vulnerabilities to attack need to be addressed, prevented and eliminated throughout the life cycle of the AI product or service (UNESCO 2021). Human safety is rooted in the value of human life and wellbeing. Safety requires that AI systems and applications should not cause harm through misuse, questionable or defective design and unintended negative consequences. Safety, in the context of AI systems, is connected to ensuring their accuracy, reliability, security and robustness (Leslie 2019). *Accuracy* refers to the ability of an AI system to make correct judgements, predictions, recommendations or decisions based on data or models (AI HLEG 2019). Inaccurate AI predictions may result in serious and adverse effects on human life. *Reliability* refers to the ability of a system to work properly using different inputs and in a range of situations, a feature that is deemed critical for both scrutiny and harms prevention (ibid). *Security* calls for protective measures against vulnerabilities, exploitation and attacks at all levels: data, models, hardware and software (ibid). *Robustness* requires that AI systems use a preventative approach to risk. The systems should behave reliably while minimising unintentional and unexpected harm and preventing unacceptable harm, and at the same time ensuring the physical and mental integrity of humans (ibid).

6.3.2 Privacy

As another responsible AI principle, privacy (see Chap. 3) is also particularly implicated in the first and second case studies. Privacy, while an ethical principle and human right in itself, intersects with the right to life, liberty and security, and supports it with protective mechanisms in the technological context. This principle, in the AI context, includes respect for the privacy, quality and integrity of data, and access to data (AI HLEG 2019). Privacy vulnerabilities manifest themselves in data leakages which are often used in attacks (Denko 2017). Encryption by itself is not seen to provide "adequate privacy protection" (Apthorpe et al. 2017). AI systems must have appropriate levels of security to prevent unauthorised or unlawful processing, accidental loss, destruction or damage (ICO 2020). They must also ensure that privacy and data protection are safeguarded throughout the system's lifecycle, and data access protocols must be in place (AI HLEG 2019). Furthermore, the quality and integrity of data are critical, and processes and data sets used require testing at all stages.

6.3.3 Responsibility and Accountability

When anything goes wrong, we look for *who* is responsible for making decisions about liability and accountability. Responsibility is seen in terms of ownership and/or answerability. In the cases examined here, responsibility might lie with different entities, depending on their role and/or culpability in the harms caused. The cases furthermore suggest that the allocation of responsibility may not be simple or straightforward. In the case of an intentional attack on an AI system, it may be possible to identify the individual orchestrating it. However, in the case of the autonomous vehicle or that of the smart home, the combination of many contributions and the dynamic nature of the system may render the attempt to attribute the actions of the system difficult, if not impossible.

Responsibility lies not only at the point of harm but goes to the point of inception of an AI system. As the ethics guidelines of the European Commission's High-Level Expert Group on Artificial Intelligence outline (AI HLEG 2019), companies must identify the impacts of their AI systems and take steps to mitigate adverse impacts. They must also comply with technical requirements and legal obligations. Where a provider (a natural or legal person) puts a high-risk AI system on the market or into service, they bear the responsibility for it, whether or not they designed or developed it (European Commission 2021).

Responsibility faces many challenges in the socio-technical and AI context (Council of Europe 2019). The first, the challenge of "many hands" (Van de Poel et al. 2012) results as the "development and operation of AI systems typically entails contributions from multiple individuals, organisations, machine components, software algorithms and human users, often in complex and dynamic environments". (Council of Europe 2019). A second challenge relates to how humans placed in the loop are made responsible for harms, despite having only partial control of an AI system, in an attempt by other connected entities to shirk responsibility and liability. A third challenge highlighted is the unpredictable nature of interactions between multiple algorithmic systems that generate novel and potentially catastrophic risks which are difficult to understand (Council of Europe 2019).

For now, responsibility for acts and omissions in relation to an AI product or service and system-related harms lies with humans. The Montreal Declaration for a Responsible Development of AI (2018) states that the development and use of AI "must not contribute to lessening the responsibility of human beings when decisions must be made". However, it also provides that "when damage or harm has been inflicted by an AIS [AI system], and the AIS is proven to be reliable and to have been used as intended, it is not reasonable to place blame on the people involved in its development or use".

Accountability, as outlined by the OECD, refers to.

> the expectation that organisations or individuals will ensure the proper functioning, throughout their lifecycle, of the AI systems that they design, develop, operate or deploy, in accordance with their roles and applicable regulatory frameworks, and for demonstrating this through their actions and decision-making process (for example, by providing documentation on key decisions throughout the AI system lifecycle or conducting or allowing auditing where justified). (OECD n.d.)

Accountability, in the AI context, is linked to auditability (assessment of algorithms, data and design processes), minimisation and reporting of negative impacts, addressing trade-offs and conflicts in a rational and methodological manner within the state of the art, and having accessible redress mechanisms (AI HLEG 2019).

But accountability in the AI context is also not without its challenges, as Busuioc (2021) explains. Algorithm use creates deficits that affect accountability: the compounding of informational problems, the absence of adequate explanation or justification of algorithm functioning (limits on questioning this), and ensuing difficulties with diagnosing failure and securing redress. Various regulatory tools have thus become important to boost AI accountability.

6.4 Responses

Given the above issues and concerns, it is important to put considerable effort into preventing AI human rights issues arising around life, liberty and security of persons, for which the following tools will be particularly helpful.

6.4.1 Defining and Strengthening Liability Regimes

An effective liability regime offers incentives that help reduce risks of harm and provide means to compensate the victims of such harms. “Liability” may be defined by contractual requirements, fault or negligence-based liability, or no-fault or strict liability. With regard to self-driving cars, liability might arise from tort for drivers and insurers and from product liability for manufacturers. Different approaches are adopted to reduce risks depending on the type of product or service.

Are current liability regimes adequate for AI? As of 1 April 2022, there were no AI-specific legal liability regimes in the European Union or United States, though there have been some attempts to define and strengthen existing liability regimes to take into account harms from AI (Karner et al. 2021).

The European Parliament’s resolution of 20 October 2020 with recommendations to the European Commission on a civil liability regime for AI (European Parliament 2020) outlined that there was no need for a complete revision of the well-functioning liability regimes in the European Union. However, the capacity for self-learning, the potential autonomy of AI systems and the multitude of actors involved presented a significant challenge to the effectiveness of European Union and national liability framework provisions. The European Parliament recognised that specific and coordinated adjustments to the liability regimes were necessary to compensate persons who suffered harm or property damage, but did not favour giving legal personality to AI systems. It stated that while physical or virtual activities, devices or processes that were driven by AI systems might technically be the direct or indirect cause of harm or damage, this was nearly always the result of someone building, deploying or interfering with the systems (European Parliament 2020). Parliament recognised, though, that the Product Liability Directive (PLD), while applicable to civil liability claims relating to defective AI systems, should be revised (along with an update of the Product Safety Directive) to adapt it to the digital world and address the challenges posed by emerging digital technologies. This would ensure a high level of effective consumer protection and legal certainty for consumers and businesses and minimise high costs and risks for small and medium-sized enterprises and start-ups. The European Commission is taking steps to revise sectoral product legislation (Ragonnaud 2022; Šajn 2022) and undertake initiatives that address liability issues related to new technologies, including AI systems.

A comparative law study on civil liability for artificial intelligence (Karner et al. 2021) questioned whether the liability regimes in European Union Member States provide for an adequate distribution of all risks, and whether victims will be indemnified or remain undercompensated if harmed by the operation of AI technology, even though tort law principles would favour remedying the harm. The study also highlights that there are some strict liabilities in place in all European jurisdictions, but that many AI systems would not fall under such risk-based regimes, leaving victims to pursue compensation via fault liability.

With particular respect to self-driving vehicles, existing legal liability frameworks are being reviewed and new measures have been or are being proposed (e.g. Automated and Electric Vehicles Act 2018; Dentons 2021). These will need to deal with issues that arise from the shifts of control from humans to automated driver assistance systems, and to address conflicts of interest, responsibility gaps (who is responsible and in what conditions, i.e. the human driver/passengers, system operator, insurer or manufacturer) and the remedies applicable.

A mixture of approaches is required to address harms by AI, as different liability approaches serve different purposes: these could include fault- or negligence-based liability, strict liability and contractual liability. The strengthening of provisions for strict liability (liability that arises irrespective of fault or of a defect, malperformance or non-compliance with the law) is highly recommended for high-risk AI products and services (New Technologies Formation 2019), especially where such products and services may cause serious and/or significant and frequent harms, e.g. death, personal injury, financial loss or social unrest (Wendehorst 2020).

6.4.2 *Quality Management for AI Systems*

Given the risks shown in the case studies presented, it is critical that AI system providers have a good quality management system in place. As outlined in detail in the proposal for the Artificial Intelligence Act (European Commission 2021), this should cover the following aspects:

1. a strategy for regulatory compliance …
2. techniques, procedures and systematic actions to be used for the design, design control and design verification of the high-risk AI system;
3. techniques, procedures and systematic actions to be used for the development, quality control and quality assurance of the high-risk AI system;
4. examination, test and validation procedures to be carried out before, during and after the development of the high-risk AI system, and the frequency with which they have to be carried out,
5. technical specifications, including standards, to be applied and, where the relevant harmonised standards are not applied in full, the means to be used to ensure that the high-risk AI system complies with the requirements set out [in this law];
6. systems and procedures for data management, including data collection, data analysis, data labelling, data storage, data filtration, data mining, data aggregation, data retention and any other operation regarding the data that is performed

before and for the purposes of the placing on the market or putting into service of high-risk AI systems;

7. the risk management system …
8. the setting-up, implementation and maintenance of a post-market monitoring system …
9. procedures related to the reporting of serious incidents and of malfunctioning …
10. the handling of communication with national competent authorities, competent authorities, including sectoral ones, providing or supporting the access to data, notified bodies, other operators, customers or other interested parties;
11. systems and procedures for record keeping of all relevant documentation and information,
12. resource management, including security of supply related measures,
13. an accountability framework setting out the responsibilities of the management and other staff …

6.4.3 Adversarial Robustness

Case 3 demonstrates the need to make AI models more robust to adversarial attacks. As an IBM researcher puts it, "Adversarial robustness refers to a model's ability to resist being fooled" (Chen 2021). This calls for the adoption of various measures, such as the simulation and mitigation of new attacks, via, for example, reverse engineering to recover private data, adversarial training (Tramèr et al. 2018; Bai et al 2021, University of Pittsburgh 2021), using pre-generating adversarial images and teaching the model that these images are manipulated, and designing robust models and algorithms (Dhawale et al. 2022). The onus is clearly on developers to prepare for and anticipate AI model vulnerabilities and threats.

Examples abound of efforts to increase adversarial robustness (Gorsline et al. 2021). Li et al. (2021) have proposed an enhanced defence technique called Attention and Adversarial Logit Pairing (AT + ALP), which, when applied to clean examples and their adversarial counterparts, would help improve accuracy on adversarial examples over adversarial training. Tian et al. (2021) have proposed what they call "detect and suppress the potential outliers" (DSPO), a defence against data poisoning attacks in federated learning scenarios.

6.5 Key Insights

The right to life is the baseline of all rights: the first among other human rights. It is closely related to other human rights, including some that are discussed elsewhere in this book, such as privacy (see Chap. 3) or dignity (see Chap. 7).

In the AI context, this right requires AI developers, deployers and users to respect the sanctity of human life and embed, value and respect this principle in the design, development and use of their products and/or services. Critically, AI systems should *not* be programmed to kill or injure humans.

Where there is a high likelihood of harms being caused, even if accidental, additional precautions must be taken and safeguards set up to avoid them, for example the use of standards, safety-based design, adequate monitoring of the AI system (Anderson 2020), training, and improved accident investigation and reporting (Alemzadeh et al. 2016).

While the technology may have exceeded human expectations, AI must support human life, *not* undermine it. The sanctity of human life must be preserved. What is furthermore required is *sensitivity* to the value of human life, liberty and security. It is insensitivity to harms and impacts that leads to change-resistant problematic actions. Sensitivity requires the ability to understand what is needed and the taking of helpful actions to fulfil that need. It also means remembering that AI can influence, change and damage human life in many ways. This sensitivity is required at all levels: development, deployment and use. It requires continuous learning on the adverse impacts that an AI system may have on human life, liberty and security and avoiding and/or mitigating such impacts to the fullest extent possible.

References

ACHPR (1981) African (Banjul) Charter on Human and Peoples' Rights. Adopted 27 June. African Commission on Human and Peoples' Rights, Banjul. https://www.achpr.org/public/Document/file/English/banjul_charter.pdf. Accessed 24 May 2022

AI HLEG (2019) Ethics guidelines for trustworthy AI. High-Level Expert Group on Artificial Intelligence, European Commission, Brussels. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419. Accessed 25 Sept 2020

Alemzadeh H, Raman J, Leveson N et al (2016) Adverse events in robotic surgery: a retrospective study of 14 years of FDA data. PLoS ONE 11(4):e0151470. https://doi.org/10.1371/journal.pone.0151470

Anderson B (2020) Tesla autopilot blamed on Fatal Japanese Model X crash. Carscoops, 30 April. https://www.carscoops.com/2020/04/tesla-autopilot-blamed-on-fatal-japanese-model-x-crash/. Accessed 24 May 2022

AP, Reuters (2021) US regulators probe deadly Tesla crash in Texas. DW, 19 April. https://p.dw.com/p/3sFbD. Accessed 22 May 2022

Apthorpe NJ, Reisman D, Feamster N (2017) A smart home is no castle: privacy vulnerabilities of encrypted IoT traffic. ArXiv, abs/1705.06805. https://doi.org/10.48550/arXiv.1705.06805

Automated and Electric Vehicles Act (2018) c18. HMSO, London. https://www.legislation.gov.uk/ukpga/2018/18/contents. Accessed 24 May 2022

Bai T, Luo J, Zhao J et al (2021) Recent advances in adversarial training for adversarial robustness. In: Zhou Z-H (ed) Proceedings of the thirtieth international joint conference on artificial intelligence (IJCAI-21), International Joint Conferences on Artificial Intelligence, pp 4312–4321. https://doi.org/10.24963/ijcai.2021/591

Barnard-Wills D, Marinos L, Portesi S (2014). Threat landscape and good practice guide for smart home and converged media. European Union Agency for Network and Information Security (ENISA). https://www.enisa.europa.eu/publications/threat-landscape-for-smart-home-and-media-convergence. Accessed 25 May 2022

BBC News (2020) Uber's self-driving operator charged over fatal crash. 16 September. https://www.bbc.com/news/technology-54175359. Accessed 23 May 2022

Bortsova G, González-Gonzalo C, Wetstein SC et al (2021) Adversarial attack vulnerability of medical image analysis systems: unexplored factors. Med Image Anal 73:102141. https://doi.org/10.1016/j.media.2021.102141

Boyd EB (2016) Is police use of force about to get worse—with robots? POLITICO Magazine, 22 September. https://www.politico.com/magazine/story/2016/09/police-robots-ethics-debate-214273/. Accessed 22 May 2022

Busuioc M (2021) Accountable artificial intelligence: holding algorithms to account. Public Adm Rev 81(5):825–836. https://doi.org/10.1111/puar.13293

Chang V, Wang Z, Xu QA et al (2021). Smart home based on internet of things and ethical issues. In: Proceedings of the 3rd international conference on finance, economics, management and IT business (FEMIB), pp 57–64. https://doi.org/10.5220/0010178100570064

Chen P-Y (2021) Securing AI systems with adversarial robustness. IBM Research. https://research.ibm.com/blog/securing-ai-workflows-with-adversarial-robustness. Accessed 15 May 2022

Clifford Law (2021) The dangers of driverless cars. The National Law Review, 5 May. https://www.natlawreview.com/article/dangers-driverless-cars. Accessed 23 May 2022

Council of Europe (2019) Responsibility and AI: a study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework. Prepared by the Expert Committee on human rights dimensions of automated data processing and different forms of artificial intelligence (MSI-AUT). https://rm.coe.int/responsability-and-ai-en/168097d9c5. Accessed 25 May 2022

Council of Europe (2020) Legal aspects of "autonomous" vehicles. Report Committee on Legal Affairs and Human Rights, Parliamentary Assembly, Council of Europe. https://assembly.coe.int/LifeRay/JUR/Pdf/DocsAndDecs/2020/AS-JUR-2020-20-EN.pdf. Accessed 25 May 2022

Deloitte (n.d.) Urban future with a purpose: 12 trends shaping the future of cities by 2030. https://www2.deloitte.com/global/en/pages/public-sector/articles/urban-future-with-a-purpose.html.

Denko MW (2017) A privacy vulnerability in smart home IoT devices. Dissertation, University of Michigan-Dearborn. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/139706/49698122_ECE_699_Masters_Thesis_Denko_Michael.pdf. Accessed 25 May 2022

Dentons (2021) Global guide to autonomous vehicles 2021. http://www.thedriverlesscommute.com/wp-content/uploads/2021/02/Global-Guide-to-Autonomous-Vehicles-2021.pdf. Accessed 24 May 2022

Dhawale K, Gupta P, Kumar Jain T (2022) AI approach for autonomous vehicles to defend from adversarial attacks. In: Agarwal B, Rahman A, Patnaik S et al (eds) Proceedings of international conference on intelligent cyber-physical systems. Springer Nature, Singapore, pp 207–221. https://doi.org/10.1007/978-981-16-7136-4_17

ECHR (1950) European Convention on Human Rights. 5 November. European Court of Human Rights, Strasbourg. https://www.echr.coe.int/documents/convention_eng.pdf. Accessed 25 May 2022

European Commission (2021) Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts. European Commission, Brussels. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206. Accessed 1 May 2022

European Parliament (2020) Resolution of 20 October 2020 with recommendations to the commission on a civil liability regime for artificial intelligence (2020/2014(INL)). https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.pdf. Accessed 24 May 2022

Finlayson SG, Bowers JD, Ito J et al (2019) Adversarial attacks on medical machine learning. Science 363(6433):1287–1289. https://doi.org/10.1126/science.aaw4399

Finlayson SG, Chung HW, Kohane IS, Beam AL (2018) Adversarial attacks against medical deep learning systems. ArXiv preprint. https://doi.org/10.48550/arXiv.1804.05296

Fránik M, Čermák M (2020) Serious flaws found in multiple smart home hubs: is your device among them? WeLiveSecurity, 22 April. https://www.welivesecurity.com/2020/04/22/serious-flaws-smart-home-hubs-is-your-device-among-them/. Accessed 22 May 2022

Golson J (2017) Read the Florida Highway Patrol's full investigation into the fatal Tesla crash. The Verge, 1 February. https://www.theverge.com/2017/2/1/14458662/tesla-autopilot-crash-accident-florida-fatal-highway-patrol-report. Accessed 23 Msay 2022

Gorsline M, Smith J, Merkel C (2021) On the adversarial robustness of quantized neural networks. In: Proceedings of the 2021 Great Lakes symposium on VLSI (GLSVLSI '21), 22–25 June 2021, virtual event. Association for Computing Machinery, New York, pp 189–194. https://doi.org/10.1145/3453688.3461755

ICO (2020) Guidance on AI and data protection. Information Commissioner's Office, Wilmslow, UK. https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection/. Accessed 25 May 2022

Jansen P, Brey P, Fox A et al (2020). SIENNA D4.4: Ethical analysis of AI and robotics technologies V1.https://doi.org/10.5281/zenodo.4068083

Jiang BC, Gainer CA Jr (1987) A cause-and-effect analysis of robot accidents. J Occup Accid 9(1):27–45. https://doi.org/10.1016/0376-6349(87)90023-X

Kalra N, Groves DG (2017) The enemy of good: estimating the cost of waiting for nearly perfect automated vehicles. Rand Corporation, Santa Monica CA

Karner E, Koch BA, Geistfeld MA (2021) Comparative law study on civil liability for artificial intelligence. Directorate-General for Justice and Consumers, European Commission, Brussels. https://data.europa.eu/doi/10.2838/77360. Accessed 24 May 2022

Laughlin A (2021) How a smart home could be at risk from hackers. Which?, 2 July. https://www.which.co.uk/news/article/how-the-smart-home-could-be-at-risk-from-hackers-akeR18s9eBHU. Accessed 23 May 2022

Leslie D (2019) Understanding artificial intelligence ethics and safety: a guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. https://doi.org/10.5281/zenodo.3240529

Levin S, Woolf N (2016) Tesla driver killed while using autopilot was watching Harry Potter, witness says. The Guardian, 1 July. https://www.theguardian.com/technology/2016/jul/01/tesla-driver-killed-autopilot-self-driving-car-harry-potter. Accessed 23 May 2022

Lévy-Bencheton C, Darra E, Tétu G et al (2015) Security and resilience of smart home environments. good practices and recommendations. European Union Agency for Network and Information Security (ENISA). https://www.enisa.europa.eu/publications/security-resilience-good-practices. Accessed 25 May 2022

Li X Goodman D Liu J et al (2021) Improving adversarial robustness via attention and adversarial logit pairing. Front ArtifIntell 4. https://doi.org/10.3389/frai.2021.752831

Marikyan D, Papagiannidis S, Alamanos E (2019) A systematic review of the smart home literature: a user perspective. Technol Forecast Soc Change 138:139–154. https://doi.org/10.1016/j.techfore.2018.08.015

McFarland M (2019) Feds blame distracted test driver in Uber self-driving car death. CNN Business, 20 November. https://edition.cnn.com/2019/11/19/tech/uber-crash-ntsb/index.html. Accessed 23 May 2022

Montreal Declaration (2018) Montréal declaration for a responsible development of artificial intelligence. Université de Montréal, Montreal. https://www.montrealdeclaration-responsibleai.com/the-declaration. Accessed 21 Sept 2020

Muggah R (2017) What happens when we can predict crimes before they happen? World Economic Forum, 2 February. https://www.weforum.org/agenda/2017/02/what-happens-when-we-can-predict-crimes-before-they-happen/. Accessed 16 May 2022

New Technologies Formation (2019) Liability for artificial intelligence and other emerging digital technologies. Expert Group on Liability and New Technologies, Directorate-General for Justice and Consumers, European Commission, Brussels. https://data.europa.eu/doi/10.2838/573689. Accessed 24 May 2022

OAS (2015) Inter-American Convention on Protecting the Human Rights of Older Persons. Forty-fifth regular session of the OAS General Assembly, 15 June. http://www.oas.org/en/sla/dil/docs/inter_american_treaties_A-70_human_rights_older_persons.pdf. Accessed 25 May 2022

OECD (n.d.) Accountability (Principle 1.5). OECD AI Policy Observatory. https://oecd.ai/en/dashboards/ai-principles/P9. Accessed 23 May 2022

Pietsch B (2021) 2 killed in driverless Tesla car crash, officials say. The New York Times, 18 April. https://www.nytimes.com/2021/04/18/business/tesla-fatal-crash-texas.html. Accessed 23 May 2022

Ragonnaud G (2022) Legislative train schedule: revision of the machinery directive (REFIT). European Parliament. https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-revision-of-the-machinery-directive. Accessed 24 May 2022

Rahman A, Hossain MS, Alrajeh NA, Alsolami F (2021) Adversarial examples: security threats to COVID-19 deep learning systems in medical IoT devices. IEEE Internet Things J 8(12):9603–9610. https://doi.org/10.1109/JIOT.2020.3013710

Raso F, Hilligoss H, Krishnamurthy V et al (2018). Artificial intelligence and human rights: opportunities and risks. Berkman Klein Center for Internet and Society Research, Harvard University, Cambridge MA. http://nrs.harvard.edu/urn-3:HUL.InstRepos:38021439. Accessed 25 May 2022

Safi M (2019) Are drone swarms the future of aerial warfare? The Guardian, 4 December. https://www.theguardian.com/news/2019/dec/04/are-drone-swarms-the-future-of-aerial-warfare. Accessed 22 May 2022

Šajn N (2022) Legislative train schedule: general product safety regulation. European Parliament. https://www.europarl.europa.eu/legislative-train/theme-a-new-push-for-european-democracy/file-revision-of-the-general-product-safety-directive. Accessed 24 May 2022

Teoh ER, Kidd DG (2017) Rage against the machine? Google's self-driving cars versus human drivers. J Safety Rs 63:57–60. https://doi.org/10.1016/j.jsr.2017.08.008

Tesla (n.d.) Support: autopilot and full self-driving capability. https://www.tesla.com/support/autopilot. Accessed 23 May 2022

Tian Y, Zhang W, Simpson A. et al (2021). Defending against data poisoning attacks: from distributed learning to federated learning. Computer J bxab192. https://doi.org/10.1093/comjnl/bxab192

Tramèr F, Kurakin A, Papernot N et al (2018) Ensemble adversarial training: attacks and defenses. Paper presented at 6th international conference on learning representations, Vancouver, 30 April – 3 May. https://doi.org/10.48550/arXiv.1705.07204

UN (1948) Universal Declaration of Human Rights. http://www.un.org/en/universal-declaration-human-rights/. Accessed 4 May 2022

UN (1966) International Covenant on Civil and Political Rights. General Assembly resolution 2200A (XXI), 16 December. https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights. Accessed 24 May 2022

UN (1969) American Convention on Human Rights: "Pact of San José, Costa Rica". Signed at San José, Costa Rica, 22 November. https://treaties.un.org/doc/publication/unts/volume%201144/volume-1144-i-17955-english.pdf. Accessed 24 May 2022

UN (1989) Convention on the Rights of the Child. General Assembly resolution 44/25, 20 November. https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child. Accessed 24 May 2022

UN (1990) International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families. General Assembly resolution 45/158, 18 December. https://www.ohchr.org/en/instruments-mechanisms/instruments/international-convention-protection-rights-all-migrant-workers. Accessed 24 May 2022

UN (2006) Convention on the Rights of Persons with Disabilities. General Assembly resolution A/RES/61/106, 13 December. https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-persons-disabilities. Accessed 24 May 2005

UN (2018) Universal Declaration of Human Rights at 70: 30 articles on 30 articles – article 3. Press release, 12 November. Office of the High Commissioner for Human Rights, United Nations. https://www.ohchr.org/en/press-releases/2018/11/universal-declaration-human-rights-70-30-articles-30-articles-article-3. 24 May 2022

UNESCO (2021) Recommendation on the ethics of artificial intelligence. SHS/BIO/REC-AIETHICS/2021. General Conference, 41st, 23 November. https://unesdoc.unesco.org/ark:/48223/pf0000380455. Accessed 25 May 2022

University of Pittsburgh (2021) Cancer-spotting AI and human experts can be fooled by image-tampering attacks. Science Daily, 14 December. https://www.sciencedaily.com/releases/2021/12/211214084541.htm. Accessed 24 May 2022.

Vasic M, Billard A (2013) Safety issues in human-robot interactions. In: Proceedings of the 2013 IEEE international conference on robotics and automation, Karlsruhe, 6–10 May, pp 197–204. https://doi.org/10.1109/ICRA.2013.6630576

Van de Poel I, Fahlquist JN, Doorn N et al (2012) The problem of many hands: climate change as an example. Sci Eng Ethics 18(1):49–67. https://doi.org/10.1007/s11948-011-9276-0

Wendehorst C (2020) Strict liability for AI and other emerging technologies. JETL 11(2):150–180. https://doi.org/10.1515/jetl-2020-0140

Zhou Q, Zuley M, Guo Y et al (2021) (2021) A machine and human reader study on AI diagnosis model safety under attacks of adversarial images. Nat Commun 12:7281. https://doi.org/10.1038/s41467-021-27577-x