Cambridge Forum on AI: Law and Governance (2025), **1**, e28, 1–19
doi:10.1017/cfl.2025.10012

RESEARCH ARTICLE

Exploring fairness in service robotics

Eduard Fosch-Villaronga, Antoni Mut-Piña, Mohammed Raiz Shaffique and Marie Schwed-Shenker

eLaw–Center for Law and Digital Technologies, Leiden University, Leiden, The Netherlands
Corresponding author: Eduard Fosch-Villaronga; Email: e.fosch.villaronga@law.leidenuniv.nl

(Received 1 November 2024; revised 7 May 2025; accepted 12 May 2025)

Abstract
Fairness in service robotics is a complex and multidimensional concept shaped by legal, social and technical considerations. As service robots increasingly operate in personal and professional domains, questions of fairness – ranging from legal certainty and anti-discrimination to user protection and algorithmic transparency – require systematic and interdisciplinary engagement. This paper develops a working definition of fairness tailored to the domain of service robotics based on a doctrinal analysis of how fairness is understood across different fields. It identifies four key dimensions essential to fair service robotics: (i) furthering legal certainty, (ii) preventing bias and discrimination, (iii) protecting users from exploitation and (iv) ensuring transparency and accountability. The paper explores how developers, policymakers and researchers can contribute to these goals. While fairness may resist universal definition, articulating its core components offers a foundation for guiding more equitable and trustworthy human–robot interactions.

Keywords: fairness; service robots; exploitation of users; legal certainty; bias; reflexivity

1. Introduction

Service robots are personal or professional use robots that perform practical tasks for humans, such as handling objects, transporting them from and to different places and providing physical support, guidance or information (ISO 8373:ISO, 2021 Robotics Vocabulary; Fosch-Villaronga & Roig, 2017). For instance, Roomba is a vacuum cleaner robot that autonomously navigates domestic spaces to clean the floors, or delivery robots like those from Starship used in university campuses or office buildings to transport food or packages from one location to another without human intervention. Given their usefulness in both domestic and professional contexts, service robots are expected to be widely deployed for different facets of human life, which range from simple tasks to more complicated ones in care, such as improving children and vulnerable populations' quality of life (Boada, Maestre & Genís, 2021).

Despite the significant contributions service robots bring to daily applications, legal, ethical and societal concerns arise due to the various contexts of application, which range from public spaces to personal and intimate spheres, multiple embodiments and interactions with different populations, including vulnerable users such as the elderly and children (Boada et al., 2021; Salvini, Paez-Granados & Billard, 2021; Wangmo, Lipps, Kressig & Ienca, 2019). These issues involve complex dimensions, including safety, transparency, opacity, trust and accountability, often context-sensitive (Felzmann, Fosch-Villaronga, Lutz & Tamò-Larrieux, 2020; Winfield, 2019). Given that regulations tend to be

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

technology-neutral to ensure applicability over time, as a result, these concepts have multiple meanings and result in a problematic understanding of how they operationalize in concrete embodied robotic systems such as service robots, including socially and physically assistive robots (Giger et al., 2019; Danielescu, 2020; Liu, Nie, Yu, Xie & Song, 2021; Felzmann, Fosch-Villaronga, Lutz & Tamo-Larrieux, 2019a). This regulatory ambiguity becomes especially pressing when service robots are deployed in socially sensitive contexts – such as eldercare settings, where a robot assisting with daily routines might unintentionally reinforce gendered assumptions by defaulting to female voices or caregiver roles – where their decisions and behaviors can reinforce or exacerbate existing societal biases (Groppe & Schweda, 2023).

This brings us to the question of fairness – a foundational but often ambiguously defined concept in both ethical and legal discussions around robotics. Fairness is a contemporary ethical and legal concern that is fundamental for safe robots and mitigates the effects of other concerns (Boada et al., 2021). Fairness is a multifaceted term, enclosing nondiscrimination and equality as the main components (Caton & Haas, 2024). Its prominence in existing frameworks is linked to the risks and power asymmetries of the deployment of artificial intelligence (AI), errors in algorithmic and human decision-making, socioeconomic inequalities, lack of diversity, privacy concerns and inadequate policies in society (Calleja, Drukarch & Fosch-Villaronga, 2022; Leslie et al., 2023; Wangmo et al., 2019). Its importance lies in ensuring safety and preventing unjust outcomes, including discriminatory consequences and social concerns like gender stereotypes (Londoño et al., 2024).

While there has been significant research on fairness and technology in general (Schwartz et al., 2022; Whittaker et al., 2019), there is a gap in comprehensive frameworks that assess fairness at multiple levels within the context of service robots. Indeed, in complex cyber-physical systems such as service robots, fairness issues may not be evident, as it could be attributable to causes like robot designer bias, model prototyping, data, algorithmic or systemic biases in legislation and standards (Winfield, 2019; Fosch-Villaronga & Drukarch, 2023; Leslie et al., 2023). For instance, if the data used to train service robots contain biases – such as gender or sex stereotypes, discriminatory information about certain demographics, or underrepresentation of certain groups – the robots might behave unfairly for different populations. This would explain why speech recognition systems might work better for people with certain accents while struggling with others (Fossa & Sucameli, 2022) (i.e., American-English robots in a low-income Indian context, Singh, Kumar, Fosch-Villaronga, Singh & Shukla, 2023), or why systems might recognize better light and male faces rather than dark-skinned female faces (Addison, Bartneck & Yogeeswaran, 2019). The difference with other AI-based systems is that the physical design or appearance of these robots can also introduce bias. A robot's embodiment may lead to unintended preferences or exclusions. For example, a robot designed to look like a female nurse may reinforce stereotypical gender roles at work (the wrong idea that nurses are reserved jobs for women) (Fosch-Villaronga & Poulsen, 2022). In more specific examples, excluding anatomical differences between genders (i.e., breast and pelvis) in certain robots (exoskeletons, for instance), may also lead to the design of robots that may provide better assistance to men than women (Fosch-Villaronga & Drukarch, 2023). And, on the last note, the safety standards that govern how these robots should be built and used can also be biased. Since many of these standards focus on one aspect (typically safety, which traditionally refers to physical safety, Martinetti, Chemweno, Nizamis & Fosch-Villaronga, 2021) and are created by private companies or small groups of stakeholders, they might not consider the needs and perspectives of diverse groups (e.g., people from different cultures, with disabilities, or various socioeconomic backgrounds) (Haidegger et al., 2013). This could lead to robots that do not serve all people fairly – an especially critical concern given the close proximity these technologies have with end users and the influence they may exert on social norms.

Our contribution addresses this gap by providing a comprehensive framework for understanding fairness within the domain of robotics. As indicated, robots represent a complex intersection of various elements (software and physical embodiment), which implies that multiple factors can influence their performance and potentially lead to unfair outcomes. To that end, we have conducted a study to

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

gain a more comprehensive understanding of the problem of fairness in robotics. Doctrinal research has been used to clarify the various definitions of fairness used within the specific context of service robots. This type of analysis allows us to identify and catalog the underlying themes and perspectives associated with the semantics of the term "fairness." An interdisciplinary framework that integrates insights from law, ethics, social sciences and human–computer interaction (HCI) has been used. In this regard, building on previous research (Leslie et al., 2023), our approach has been to dissect the different elements that make up robots, identifying those that may result in unfair performance.

The structure of the paper is as follows. After explaining the methodological approach followed, we explore the dimensions of the phenomenon of fairness within the domain of service robots. In this section, we examine various understandings of fairness that we consider are mutually complementary and are necessary for achieving fairness. In concrete: fairness (i) as objectivity and legal certainty; (ii) as prevention of bias and discrimination; (iii) as prevention of exploitation of users; and, (iv) finally, as transparency and accountability. We then propose a framework for fairness in service robotics, which builds upon the identified dimensions and provides a working definition. The design of our study is conditioned by several key factors. First, the implementation of service robotics is still in its early stages, as is its intersection with AI. This means that there is a limited amount of data available on unfair behavior exhibited by robots (Cao & Chen, 2024; Hundt, Agnew, Zeng, Kacianka & Gombolay, 2022) or discriminatory practices toward robots (Barfield, 2023b). Additionally, existing literature primarily focuses on the identification of biases and potential solutions, but with a particular emphasis on data-related issues, often overlooking other relevant aspects such as the influence on embodiment in stereotypes or unfair outcomes (Lucas, Poston, Yocum, Carlson & Feil-Seifer, 2016). Finally, we close the paper with our conclusions, reflecting on the implications of the proposed framework and identifying avenues for future research and application.

2. Methods

We conducted multidisciplinary doctrinal research to develop a comprehensive framework for understanding fairness in the domain of service robotics. Doctrinal research, traditionally rooted in legal scholarship, involves the critical analysis, interpretation and synthesis of existing theories and normative principles. It enables the abstraction of concepts and their systematic integration into coherent structures of meaning (Bhat, 2020). This methodological strength makes it particularly well suited for addressing *fairness* – a concept that is both value-laden and context-dependent (Stamenkovic, 2024) – by allowing us to examine its diverse articulations across domains and distill them into a structured, unified framework.

In our case, doctrinal analysis provides the foundation for identifying, comparing and reconciling how fairness is conceptualized in law, ethics, social sciences and technical disciplines. Rather than proposing a new empirical model, we aim to build conceptual clarity and normative depth – outcomes for which doctrinal research is especially well equipped. Thus, this approach directly supports our goal of constructing a framework that not only maps existing understandings of fairness but also offers guidance for its application in the complex sociotechnical environment of service robotics. This capacity is further amplified when doctrinal research is combined with interdisciplinary perspectives, which help surface context-specific dimensions and variations in meaning (Ishwara, 2020).

In line with this, we examined how fairness is conceptualized not only in legal literature but also across a range of research domains, in order to construct a framework that reflects its multifaceted role in service robotics. In particular, we reviewed literature from five key areas: (1) Robotics and Human–Robot Interaction (HRI), (2) Artificial Intelligence (AI) and Machine Learning (ML), (3) Law and Regulation, (4) Ethics and Philosophy of Technology and finally (5) Social Sciences and Psychology. Although scientific literature constitutes our primary source, we also considered legal texts, ethical guidelines and regulatory sources. Examining these areas of knowledge allowed us

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

to identify the diverse and intersecting factors that shape the concept of fairness, emphasizing its multidimensional nature in the context of robotic systems and HRIs.

From a legal perspective, the principles that emerged from our analysis include legal certainty (Braithwaite, 2002; Shcherbanyuk, Gordieiev & Bzova, 2023), antidiscrimination (Council Directive 2005/29/EC; Rigotti & Fosch-Villaronga, 2024) and consumer protection (Cartwright, 2015; Lim & Letkiewicz, 2023). Ethically, our analysis highlighted fairness theories grounded in equity (Adams, (2015)) and justice (Folger & Cropanzano, 2001; Rawls, 1971) as core normative foundations that consistently inform debates on responsibility, distribution and legitimacy in technological contexts. Similarly, our examination of the sociotechnical literature revealed recurring concerns related to bias and user vulnerability, particularly in robot-mediated environments (Addison et al., 2019; Fosch-Villaronga & Poulsen, 2022). In the psychological and HCI domains, empirical studies highlighted perceptions of fairness, trust and accountability as central themes in HRI (Cao & Chen, 2024; Chang, Pope, Short & Thomaz, 2020; Claure, Candon, Shin & Vázquez, 2024).

3. Identifying the dimensions of fairness in service robotics

Mulligan, Kroll, Kohli and Wong (2019) extensively discussed the different conceptualizations of fairness across disciplines and how this creates confusion owing to the lack of a common vocabulary. For instance, a computer scientist might think about fairness in terms of how an algorithm processes data equally for everyone; a sociologist might focus on whether the system perpetuates inequality in society; and a legal expert might consider fairness from the standpoint of individual rights and whether the system abides by laws. Even though they all use the word "fairness," these different perspectives can lead to confusion because they do not mean the same thing in different contexts. In other words, fairness is a context-specific concept that is universally difficult to define (McFarlane v McFarlane (2006) UKHL 24; Mehrabi, Morstatter, Saxena, Lerman and Galstyan (2021); Londoño et al., 2024) and that has interdisciplinary connotations that can have different meanings based on how one approaches it (Rigotti & Fosch-Villaronga, 2024).

In this sense, fairness has been viewed in terms of justice (Rawls, 1971; Colquitt & Zipay, 2015), proportionate reward from work (Adams, 2015), accountability (Folger & Cropanzano, 2001) and so on, resulting in diverging discussions and understandings, each with its merit. For instance, in robotics literature, fairness has been analyzed in the context of human–robot collaboration, where certain studies examine fairness from the angle of how humans perceive the fairness of task allocation while in human–robot teams (Chang et al., 2020; Chang, Trafton, McCurry & Thomaz, 2021). In Cao and Chen (2024), the authors study how robots' fair behaviors impact the human partner's attitude and their relationship with the robots, highlighting that humans are likely to reward robots that demonstrate cooperative behaviors and penalize those exhibiting self-serving or unfair actions, much like they do with other humans.

Conceptualizing fairness for service robotics is challenging *inter alia* because this field often involves elements of the spheres above, such as law, social sciences, computer science and engineering. Technological advancements have resulted in these robots having sophisticated capabilities (Lee, 2021), and these robots usually operate in an ever-changing environment among humans (Sprenger & Mettler, 2015). As a gamut of laws and standards governs these robots, as they interact with humans physically and socially, and incorporate various hardware and software elements including AI, various interdisciplinary fields are relevant to fairness in service robots. Thus, adopting the fairness frameworks of any of these areas would be insufficient.

Despite the difficulty in identifying fairness for service robotics, it is still the need of the hour. In service robotics, the increasing gap between policy development and technological advancements leads to a regulatory mismatch, which may cause robot developers to overlook crucial safety considerations in their designs, affecting the well-being of many users (Calleja et al., 2022; Fosch-Villaronga et al., 2025). This, in turn, leads to issues such as the legal frameworks overly emphasizing the need

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

for physical safety while not paying due heed to other essential aspects, including cybersecurity, privacy and psychological safety (Martinetti et al., 2021). Additionally, one is confronted with the reality that service robots might not interact with the users in a "fair" manner because of not accounting for culture-specific designs (Mansouri & Taylor, 2024). Thus, four dimensions of fairness in service robotics as distilled from analysis of the literature are presented in this section.

3.1. Fairness as legal certainty

The challenges present with laws being unable to keep pace with the rapid growth of technology adequately termed the Collingridge dilemma (Collingridge, 1980) or the pacing problem (Marchant, 2011), is a pressing issue for service robotics. Even though robotic technology is a critical sphere, given the close cyber-physical interaction robots have with users, the existing regulatory framework does not adequately address all aspects (Goffin and Palmeri, Fosch-Villaronga, 2019; Goffin & Palmieri, 2024; Palmerini et al., 2016). The problem becomes even more pronounced when we consider that the software powering autonomous systems introduces an added layer of complexity, particularly in terms of meeting compliance standards and regulatory requirements. While cloud-based services – such as speech recognition, navigation and AI – can help offload computational tasks and make robots more modular, they also fragment system responsibility across multiple actors. This dispersion complicates compliance, as traditional regulatory frameworks tend to operate in silos and may not adequately address the interconnected nature of these systems. As a result, ensuring regulatory alignment across hardware manufacturers, software providers and cloud service operators becomes an increasingly complex challenge (Fosch-Villaronga & Millard, 2019; Setchi, Dehkordi & Khan, 2020; Wangmo et al., 2019).

EU law does not define robots, and no comprehensive law targets robots (Fosch-Villaronga & Heldeweg, 2018). A number of legislations could apply to service robotics depending on the uses and functionalities of the robots, such as the General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679) (if the robot processes personal data), the Toy Safety Directive (if the robot is a toy) and the Medical Devices Regulation (if the robot is a medical device). While recently enacted legislations such as the Machinery Regulation and the Artificial Intelligence Act (AIA) can also apply to robots, these are still not statutes targeted at robotics and providing a comprehensive regulatory framework (Mahler, 2024). This lack of clarity on the law governing robotics presents dilemmas regarding whether the current legal framework is sufficient or whether a new specific law is needed for robotics (Fosch-Villaronga & Heldeweg, 2018). Further, as mentioned before in this paper, the lack of proper regulation can lead to safety and inclusivity issues in service robotics.

Having legal clarity in this sphere can lead to a better position for all concerned stakeholders so that manufacturers know how to produce legally compliant robots, users know their rights and duties while interacting with robots, and regulators know how to supervise the robots placed in the European Union (EU) market and further the safety for users. Therefore, we argue that the principle of *legal certainty*, as recognized in Europe, can be a crucial plank for further fairness in service robotics. Legal certainty is an essential facet of the rule of law, a fundamental principle of EU law – both for the Council of Europe (2011, 2016) and the EU (European Commission, n.d.-b). It is a legal theory "*aimed at protecting and preserving the fair expectations of the people*" (Shcherbanyuk et al., 2023). Several Court of Justice of the EU judgments have evoked this principle in diverging contexts, and the scope of this principle has expanded over the years (van Meerbeeck, 2016). A vital component of the legal certainty principle is that laws should be *clear and precise* so that natural and legal persons know their rights and duties and can foresee the consequences of their actions (European Commission, n.d.-b; Shcherbanyuk et al., 2023). When applied to the service robotics context, stakeholders should clarify their roles, rights and responsibilities. Of course, we do not mean to suggest that manufacturers ought to be allowed to use the lack of regulatory clarity as a justification for not ensuring the safety of their robots, and they should still be fully accountable for the robots they develop.

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

3.2. Fairness as preventing bias and discrimination

Fairness in ML is an extensively researched field (Pessach & Shmueli, 2022). There is yet to be a consensus on what fairness is in this context, too, given the complexity of the issue and the different dimensions associated with it (Caton & Haas, 2024). In this context, the FAIR Principles – guidelines aimed at ensuring that data is Findable, Accessible, Interoperable and Reusable – specifically highlight the importance of enabling machines to automatically discover and utilize data (Wilkinson et al., 2016). In Rigotti and Fosch-Villaronga (2024), it was highlighted that when it comes to the use of AI in recruitment, different stakeholders, such as job applicants and human resource practitioners, view fairness in markedly different ways. Fairness has been defined in algorithmic decision-making as an absence of prejudice or favoritism based on intrinsic or acquired traits (Mehrabi et al., 2021). Therefore, many studies in this sphere view fairness from the prism of bias and discrimination. Some studies deal with fairness in ML, which explicitly acknowledge that unfairness, bias and discrimination are interchangeable terms (Pessach & Shmueli, 2022).

Coming to service robotics, robots are fundamentally designed to serve humans, making it essential to address ethical considerations, among which fairness and, in particular, biases have become critical issues that have recently gained significant attention not only among scholars (Wang et al., 2022) but also among policymakers and regulators.[1] Similar to the concept of fairness, the term "bias" takes on different interpretations depending on the context and the specific academic discipline in which it is applied (FRA, 2022). In AI, for example, *bias* is usually defined as "a systematic error in the decision-making process that results in an unfair outcome" (Ferrara, 2023). To some extent, this definition can be applied to the field of robotics; however, it remains overly simplistic as it fails to consider the significant impact of a robot's physical embodiment and the specific context in which it operates. For example, Fosch-Villaronga and Özcan (2019) highlight design challenges in certain lower-limb exoskeletons, demonstrating the importance of realizing these concepts within the context of these hardware factors. In this regard, robots are inherently complex entities, and their design can be shaped by various sources of bias, for instance, if robots that simulate nurses include breasts, skirts or talk in a female voice tone (Fosch-Villaronga & Poulsen, 2022; Londoño et al., 2024; Tay, Jung & Park, 2014). This complexity arises from the interplay between technological, social and regulatory factors that influence how robots are conceived and function (Boyd & Holton, 2018; Šabanović, 2010). In this paper, we identify two primary sources of bias that impact robot design and, consequently, could lead to unfair performance: (1) design choices and physical embodiment and (2) the data used for training and decision-making processes.

The role of embodied AI or virtual assistant, changes the game's rules. As humans, we associate human-like machines with intelligence or empathy, unknowingly creating deep bonds (Scheutz, 2012). The eye region can decide the level of trustworthiness, and facial color cues and luminance can increase the level of likability, attractiveness and trustworthiness in machines. Young or old appearance controls trustworthiness and explicit facial ethnicity (Song & Luximon, 2020). An embodied agent can produce positive emotions better than a virtual one, especially anthropomorphism, imbuing them with a higher sense of vitality (Yang & Xie, 2024). An example can be found in education, as a humanoid robot promotes … secondary responses conducive to learning" and sets a level of expectations for it to be able to engage socially (Belpaeme, Kennedy, Ramachandran, Scassellati & Tanaka, 2018).

A significant aspect of embodiment can be found in color. In a survey to determine the emotional response to color, 55 percent saw white as "clean," followed by 27 percent who chose blue (Babin, 2013). It is mentioned in the article that many associate black with mourning in Western cultures, and that is why it is absent from medical facilities and equipment (Babin, 2013). Most robots will be

[1]See, for example, the European Parliament resolution of 14 March 2017 on fundamental rights implications of big data: privacy, data protection, non-discrimination, security and law-enforcement (2016/2225(INI)).

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

designed in light colors, to convey cleanliness and familiarity and to promote positive user perceptions. Such design choices may help reduce discomfort by enhancing perceived safety and warmth (Rosenthal-Von Der Pütten& Krämer, 2014).[2] However, such choices are not neutral. The overrepresentation of white-colored robots may reinforce implicit racial biases, as users tend to project human stereotypes – including racial ones – onto robots. In experiments, people were more likely to ascribe negative traits or act aggressively toward darker-skinned robots compared to white ones, even when the robots were functionally identical (Addison et al., 2019; Bartneck et al., 2018).A robot's height also plays a role in emotional response, with studies suggesting that taller robots can elicit intimidation or fear, whereas robots positioned lower than eye level tend to feel less threatening and more approachable (Hiro & Ito, 2016). Bias in color can also be found in the Robot Shooter Bias, where it is more likely to shoot faster at a darker-colored robot than a lighter one (Addison et al., 2019). Moreover, a white-colored robot will receive less discrimination than a black or rainbow-colored one of the same type (Barfield, 2021).

Social bias comes in many forms and may lead to discrimination. Discrimination can be presented in the form of sexism, racism, stereotypes and xenophobia. This can manifest in various ways, such as hospitality and tourism companies that use service robots, where the robot's embodiment and design choices (i.e., gender, color and uniform) can, in turn, awaken negative emotions from employees (Seyitoğlu & Ivanov, 2023). Discrimination can also manifest through large language model (LLM)-driven robots engaging in decision-making processes and discriminatory behavior based on significant, well-known ethical concerns perpetuating existing social injustices. In this regard, Zhou (2024) demonstrates that LLM-driven robots exhibit negative performance when interacting with racial minorities, certain nationalities or individuals with disabilities.

In addition, the rapid progress in ML and the expansion of computing power in recent years have significantly enhanced the learning capabilities of robots (Hitron, Megidish, Todress, Morag & Erel, 2022). These advancements have facilitated the development of more complex robot behavior, enabling robots to perform tasks that require sequential decision-making in dynamic environments. Thus, ML has propelled the growth of robotic domains across the spectrum, from simple automation to complex autonomous systems (Londoño et al., 2024). Most robot learning algorithms follow a data-driven paradigm (Londoño et al., 2024), allowing robots to learn automatically from vast datasets with guidance or supervision (Nwana, 1996; Rani, Liu, Sarkar & Vanman, 2006). The learning process optimizes models to perform specific tasks, such as navigation (Singamaneni et al., 2024). These methods are essential in helping robots generalize their behavior to various contexts and environments, making learning a key component in developing intelligent robotic systems (Soori, Arezoo & Dastres, 2023).

However, with this reliance on data and ML comes the introduction of biases, which can emerge both in the data used to train these models (Shahbazi, Lin, Asudeh & Jagadish, 2023) and the algorithms that process this data – in this sense, algorithmic bias primarily stems from the software used in the ML process (Takan et al., 2023) – . As a result, the ethical implications of these biases, particularly in the domain of robot learning, are becoming increasingly significant (Alarcon et al., 2023). The risks associated with AI bias could be even more severe when applied to robots, as they are often seen as autonomous entities and operate without direct human intervention (Hitron et al., 2022). Data and algorithm biases in the functioning of AI have been widely detected across various fields. In healthcare, biases have been identified in automated diagnosis and treatment systems (Obermeyer, Powers, Vogeli & Mullainathan, 2019). In applicant tracking systems, these biases can influence candidate selection and exclusion (Frissen, Adebayo & Nanda, 2023). In online advertising, algorithms

[2]The uncanny valley was defined by Mori (1970/2005) as "Climbing a mountain is an example of a function that does not increase continuously: a person's altitude y does not always increase as the distance from the summit decreases owing to the intervening hills and valleys. I have noticed that, as robots appear more humanlike, our sense of their familiarity increases until we come to a valley. I call this relation the 'uncanny valley." *See* Mori M. (1970). The uncanny valley. *Energy*, 7, 33–35.

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

may target ads unequally across different groups (Lambrech et al., 2024). Additionally, in image generation tools, biases have been found to perpetuate visual stereotypes (Naik, Gostu & Sharma, 2024). Lastly, predictive policing tools have shown discriminatory tendencies, exacerbating social justice issues (Alikhademi et al., 2022).

3.3. Fairness as preventing exploitation of users

In European consumer law, unfairness primarily addresses imbalances in contractual relationships. Following the EU's framework, such imbalances may stem from the terms of a contract or the commercial practices surrounding it. The concept of "unfair" is central in two primary Directives within the EU's consumer protection framework: Directive 93/13 on unfair terms in consumer contracts and (Directive 2005/29/EC) on unfair business-to-consumer practices in the internal market. These Directives aim to protect consumers from exploitative practices, acknowledging the structural asymmetries in information and bargaining power between consumers and businesses (Willett, 2010). Additionally, in the specific area of consumer law focused on product safety, fairness is also regarded as a central objective of regulation. In this sense, the Recommendation of the Council on Consumer Product Safety of the Convention on the Organisation for Economic Co-operation and Development recognizes that "compliance with product safety requirements by all economic operators can support a safe, fair and competitive consumer product marketplace."

In this context, the term "unfairness" emerges in two specific areas within the sphere of economic relations between consumers and businesses. First, it reflects the informational and economic asymmetries between businesses and consumers (fairness in consumer relations), and second, it serves as a condition for ensuring consumer safety in the marketplace (fairness in markets). The former used to arise through abusive clauses and practices, while the latter arose when minimum safety standards were not guaranteed. Thus, in the specific field of robotics, unfairness, understood as the failure to prevent the exploitation of users, could arise in both contexts.

Although the robot market is still in its early stages, robots may become suitable for widespread consumer adoption in the near future (Randall & Šabanović, 2024). If many consumers could access robotic markets, market asymmetry issues will likely become a significant concern. As demonstrated, informational asymmetries are particularly pronounced in technical contexts such as financial services (Cartwright, 2015; Lim & Letkiewicz, 2023). In robotics, as with other high-tech products, making an optimal and efficient purchasing decision requires a deep understanding of numerous complex technical aspects. In this sense, most consumers could be vulnerable to unfair commercial practices related to the functionality or performance of a particular robot, or they could be pressured into accepting unfair clauses in consumer contracts (Hartzog, 2014).

Furthermore, significant concerns about unfairness, especially regarding the prevention of consumer exploitation, would likely stem from security and safety issues. This issue can also be seen as a problem of informational asymmetry (Choi & Spier, 2014; Edelman, 2009). Without adequate safety regulations, consumers face an adverse selection problem, as they cannot discern which products (robotic devices, in this case) may pose economic or physical risks.

The issue of unfairness in robot performance becomes significantly more concerning when consumer vulnerability is considered. Nowadays, consumer vulnerability is "a dynamic state that varies along a continuum as people experience more or less susceptibility to harm due to varying conditions and circumstances" (Salisbury et al., 2023). In this sense, older adults, children or individuals with disabilities may have limited capacity to recognize or respond to inconsistencies or deficiencies in a robot's functionality (Søraa & Fosch-Villaronga, 2020). For instance, if a robotic device designed to assist elderly users fails to perform accurately, the consequences could range from diminished quality of life to serious physical harm. Furthermore, vulnerable users may be less likely to understand or challenge technical failures, especially if the system lacks transparency features or provides

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

insufficient recourse for complaints. This amplifies the potential for exploitation and places these consumers at a heightened risk, making robust regulatory standards and clear accountability for robot performance essential to protect their well-being. In this sense, fairness could also be understood as protecting consumers from deceptive robots. Leong and Selinger (2019) show how deceptive anthropomorphism can lead users to trust or bond with robots in ways that may not be in their best interest. As a countermeasure, robots should be designed with transparency cues that clearly indicate they are machines – such as maintaining robotic voice patterns or using visual indicators of artificiality – especially in contexts involving vulnerable populations (Felzmann et al., 2019a).

3.4. Fairness as transparency and accountability

At its core, transparency entails rendering AI systems' decision-making processes and underlying algorithms understandable and accessible to users and stakeholders (Felzmann et al., 2020; Larsson & Heinz, 2020). This is particularly needed given the complexity and opacity often inherent in AI systems and the significant risks they can pose to fundamental rights and public interests (Eschenbach, 2021). Transparency helps users understand how decisions are made, which is essential for building trust and ensuring accountability (Felzmann, Villaronga, Lutz & Tamò-Larrieux, 2019b).

In HRI, research on transparency has mainly focused on the explainability of systems, looking at either how easily the robot's actions can be understood (intelligibility) or how well users grasp the robot's behavior (understandability) (Fischer, 2018). The results are mixed: some studies report no significant findings, while others point out the drawbacks of transparency (Felzmann et al., 2019b). These include disrupting smooth interactions and leading to misunderstandings or inaccurate assumptions about the robot's abilities in specific contexts if it is too transparent about its course of action. Other research emphasizes technical limitations, showing that transparency can create misleading distinctions, prioritize visibility over proper comprehension, and sometimes even cause harm (Ananny & Crawford, 2018).

In the context of robotics, transparency also can refer to users knowing how robots collect, process and use data (Felzmann et al., 2019a). For example, robots that assist vulnerable populations, like children or older adults, must be clear about how decisions are made, particularly if those decisions impact users' well-being. Unlike privacy regulations, transparency is not just about disclosing information but ensuring that this information is meaningful and understandable to users and stakeholders. This is especially important because, as research shows, transparency can improve accountability by providing insights into decision-making processes (Lepri, Oliver, Letouzé, Pentland & Vinck, 2018). Accountability requires that those responsible for the robot – manufacturers, developers or deployers – are answerable for the robot's decisions. Although part of the community has reasoned about whether there is a responsibility gap if the robot learns as it operates (Johnson, 2015), it is the contemporary understanding, also as reflected from the EU regulations (the AIA in concrete) that transparency plays a key role by enabling different stakeholders, including authorities to inspect and users to challenge decisions that are significant for them. This is particularly important when robots are used in sensitive areas like healthcare, where biased or inaccurate decisions could have serious consequences (Cirillo et al., 2020). However, as noted in the literature, accountability is not automatic, even with transparency. Systems can be transparent but still avoid accountability if there are no mechanisms to process and act on the disclosed information (Felzmann et al., 2020).

4. Furthering fairness in service robotics

Against this background, in this paper, we understand fairness as the broad and all-encompassing concept that it is without attempting to define it exhaustively. As mentioned, while much discussion on fairness is centered around bias and discrimination, and these are, in fact, essential elements, this

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

paper does not address fairness solely from this lens. Other elements, such as the weaker party protection rationale inherent in consumer protection law, are also relevant for fairness in service robotics because manufacturers could be better financially and information-wise than the users (Kotrotsios, 2021). Therefore, in line with the four identified dimensions, the definition of fairness in service robotics proposed in this paper, which is primarily centered on a user safety perspective, is as follows:

Fairness in service robotics refers to the responsible and context-aware approach to the design, deployment, and use of robotic systems, with a focus on preventing harm, reducing systemic and individual inequalities, and upholding individual rights.

This broad approach is essential to preserve fairness's expansive and evolving nature, which remains a fundamental regulatory and ethical goal. To give practical shape to this definition, we propose a framework based on four dimensions: legal certainty, prevention of bias and discrimination, protection against user exploitation and the promotion of transparency and accountability. The remainder of this section outlines specific measures aimed at furthering these dimensions, as depicted in Fig. 1:

4.1. Legal certainty: clarity and precision as a valuable tool

Legal certainty is a complex regulatory goal. While overly detailed rules might prove ineffective, broader guiding principles might provide more adaptable and meaningful direction (Braithwaite, 2002). Principles alone, however, are not sufficient; legal certainty depends on their integration into a well-defined legal regime that enables actors to anticipate legal outcomes. In the context of service

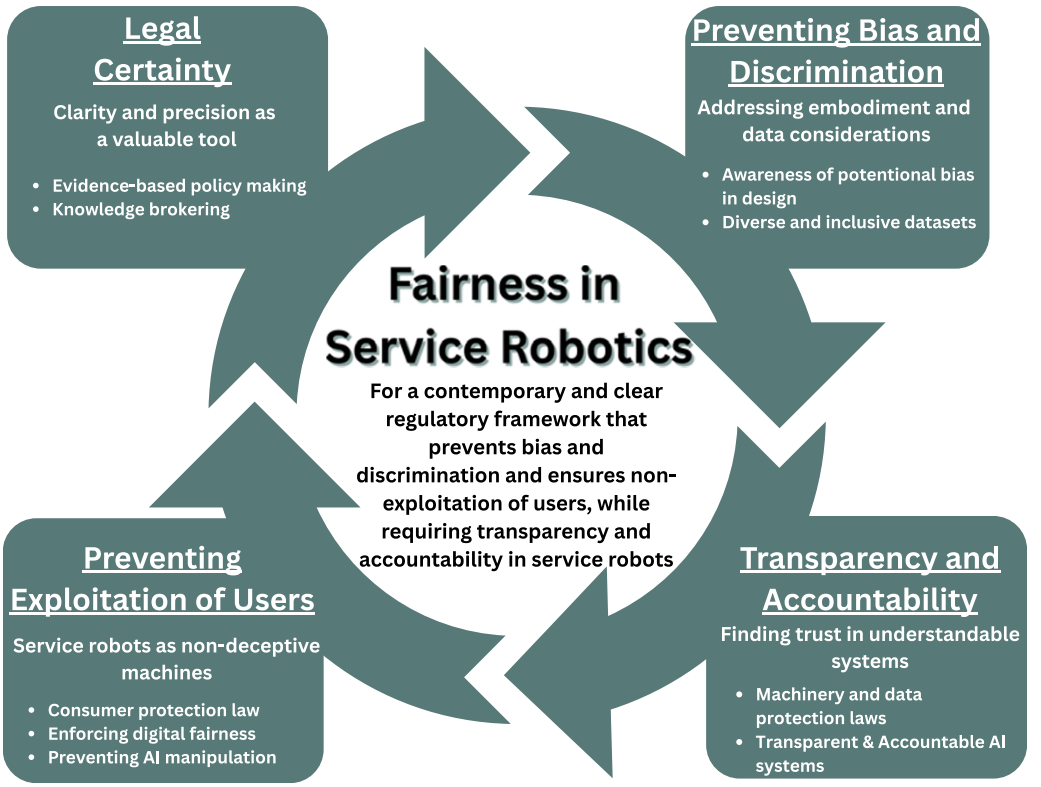


Figure 1. Fairness in service robotics.

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

robotics, this requires coordinated efforts from academia, industry, and policymakers and can be furthered in two ways.

First, the evidence-based policymaking efforts of the EU, as outlined in the "Better Regulation agenda" (European Commission, n.d.-c) and the "Knowledge 4 Policy" platform (European Commission, n.d.-e), offer promising foundations for robotics regulation. Evidence-based policymaking seeks to ground legal decisions in scientific data and analysis (Pflücke, 2024), helping ensure that regulatory choices are guided by objective and rational criteria rather than other considerations such as political ideologies (Princen, 2022). In the context of robotics, this would mean using cogent scientific evidence to promote legal certainty and fairness (Calleja et al., 2022).

Second, knowledge brokering – the practice of linking research and policy through science communication and multi-stakeholder engagement – can also contribute to robotics regulation design (Bielak, Campbell, Pope, Schaefer & Shaxson, 2008; Turnhout, 2013). *Knowledge brokers* act as intermediaries between researchers and policymakers, helping translate scientific insights into actionable guidance while retaining trust, transparency and legitimacy (Gluckman, Bardsley & Kaiser, 2021). Several EU-funded projects have already contributed to this goal, including RoboLaw (RoboLaw, 2014), INBOTS (European Commission, n.d.-f), SIENNA (European Commission, n.d.-g), H2020 COVR LIAISON (Leiden University, n.d.-a) and H2020 Eurobench PROPELLING (Leiden University, n.d.-b). More in particular, advancing fairness and achieving regulatory certainty in service robotics using scientific methods for policy formation is the core of the (omitted for reviewing purposes) project (Leiden University, n.d.-c).

4.2. Bias and discrimination: addressing embodiment and data considerations

Bias in AI systems has long been recognized as a major fairness challenge, particularly when algorithms are trained on datasets that underrepresent or misrepresent specific demographic groups (Verhoef & Fosch-Villaronga, 2023). In such cases, models may reproduce and even amplify existing social inequalities (Howard & Borenstein, 2018; Yapo & Weiss, 2018). While expanding datasets and increasing diversity may help, algorithms optimized to identify patterns in historical data are still prone to reflecting the structural imbalances embedded within that data.

To detect these biases, scholars in ML rely on fairness metrics such as equal opportunity difference, odds difference, statistical parity difference, disparate impact and the Theil index (González-Sendino, Serrano, Bajo & Novais, 2023). These mathematical tools offer structured ways of identifying disparities in how models perform across groups. However, such metrics often fall short of capturing the broader societal and contextual harms that biased systems can produce – particularly when fairness is reduced to statistical parity alone (Carey & Wu, 2022, 2023).

These limitations become even more pronounced in service robotics, where bias is not confined to data or decision outputs but also emerges through physical design and human interaction. Unlike purely virtual AI, service robots are embodied agents – they move through physical space, engage with people face-to-face and signal meaning through voice, gesture and form. Their hardware and interface design can unintentionally convey stereotypes or reinforce normative assumptions about gender, race, ability or cultural identity.

Howard and Borenstein (2018) highlight such risks through examples including a robot peacekeeper, a self-driving car and a medical assistant – each demonstrating how bias in design or behavior can have tangible, unequal impacts depending on the context of deployment.

Given this complexity, mitigating bias in service robotics must begin at the conceptualization and design stage. Manufacturers need to account for how robots are perceived and how different users might experience them. Robots such as social robots and exoskeletons should be developed with attention to diversity and inclusion, especially regarding sex, gender and cultural representation (Barfield, 2023a; Fosch-Villaronga & Drukarch, 2023; Søraa & Fosch-Villaronga, 2020).

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

Reembodiment – adapting a robot's physical or behavioral presentation based on the people it interacts with – has been proposed as one way to foster cultural sensitivity and reduce stereotyping (Reig et al., 2021).

This area is evolving rapidly, and several promising technical and policy responses continue to emerge. The Regulation (EU) 2024/1689 (AIA) takes an important step in this direction, explicitly addressing fairness through the lens of bias and non-discrimination. Recital 27 emphasizes that AI systems must be developed inclusively to avoid "discriminatory impacts and unfair biases that are prohibited by Union or national law."

4.3. Preventing exploitation of users: service robots as nondeceptive machines

Preventing the exploitation of users is a key component of fairness, particularly in contexts where individuals may be in a weaker informational or bargaining position compared to manufacturers. This concern is especially relevant in service robotics, where automated systems can subtly influence user behavior, exploit cognitive biases or obscure information through complex interfaces or persuasive design.

The existing and upcoming legal initiatives at the EU level provide a strong foundation for addressing these risks. Consumer protection law looks at fairness as a principle designed to prevent exploitation of consumers who have a weaker bargaining position and less information than businesses, whether such exploitation is by way of *mala fide* contractual clauses (Council Directive 93/13/EEC), or material distortion of consumer behavior (Directive 2005/29/EC), or concluding online contracts from a distance without providing adequate information (Directive (EU) 2023/2673).

Recently, the European Commission (EC) published its report on the fitness check related to "digital fairness" for consumers, which culminated in the observation that further measures, including legal certainty and simplification of rules, were required to protect consumers online (European Commission, n.d.-a). These measures can go a long way in service robotics, where users may not fully understand how decisions are made, or how their behavior is shaped by the system (Felzmann et al., 2019a).

The AIA also provides a regulatory mechanism to address manipulation and deception risks. Article 5(1)(a) prohibits the use of AI systems that exploit vulnerabilities of specific user groups in ways that cause harm. As service robots increasingly integrate AI capabilities, the effective implementation and enforcement of these provisions will be essential to ensuring that such systems do not mislead, manipulate or otherwise take unfair advantage of users (e.g. Article 5(1)(a), AIA).

4.4. Transparency and accountability: finding trust in understandable systems

Effective enforcement of EU legislations applicable to service robotics are foundational for ensuring transparency and accountability in these systems, particularly since recent legal initiatives have sought to address different elements of these robots. For instance, transparency and accountability are core elements highlighted in the recently adopted AIA, which can address these aspects in service robots that have AI systems. All embodied service robots made available in the EU market are also likely to fall under the scope of the Machinery Regulation 2023 (Regulation (EU) 2023/1230), (similar to the Machinery Directive 2006), which prescribes obligations for the economic operators (manufacturers, importers and distributors) of these robots. These economic operators are *inter alia* required to ensure that certain essential health and safety requirements are met by the robots (e.g., Article 10(1), Machinery Regulation 2023). Further, Article 5(1)(a) of the GDPR specifies that personal data should be processed in a "fair" manner. The meaning of fairness itself is unclear and not further elaborated (Clifford & Ausloos, 2018). However, the European Data Protection Board (EDPB) recently stated regarding transparency under the GDPR that transparency is fundamentally linked to fairness and not being transparent about the processing of personal data is likely to be unlawful

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

(EDPB, 2025). The overall framework provided by GDPR also ensures accountability of entities when it comes to the processing of personal data by service robots.

5. Conclusions

This paper has explored what fairness means in the context of service robotics, aiming to capture its normative, legal, technical and social dimensions. Through doctrinal research, supported by a multidisciplinary review, we proposed a working definition of fairness that considers both software and physical embodiment in robotics. Our analysis focused on four key dimensions of fairness that can advance service robotics in a fair manner: (i) Furthering legal certainty, (ii) Preventing bias and discrimination, (iii) Preventing exploitation of users and (iv) Ensuring transparency and accountability. These dimensions serve not as a rigid checklist, but as guiding principles for fostering fairer practices in the design, deployment and governance of service robots.

At the regulatory level, service robotics presents specific challenges that often fall between the cracks of broadly defined, technology-neutral legal frameworks. While technology neutrality in regulation ensures broad applicability, it can also obscure the specific risks posed by service robots. The principle holds that regulations should avoid privilege or disadvantage particular technologies, and should instead focus on the functions they perform or the outcomes they produce (Greenberg, 2015). However, service robots – operating in both personal and professional environments – where fairness concerns are deeply entangled with social vulnerability, human–machine interaction and embedded power asymmetries. In these environments, fairness becomes a core component of both safety and legitimacy, raising questions around bias, exclusion, privacy and the adequacy of current policy frameworks.

Central to this approach is a recognition that fairness cannot be addressed solely through legal or technical mechanisms. It also requires reflexivity – a continuous awareness of how researchers' own values, assumptions and social positions shape the knowledge they produce (Palaganas, Sanchez, Molintas & Caricativo, 2017). The reflection process involves a continuous dialogue (through introspection) between the assumptions and values that researchers bring into their field of study and the social ecosystems in which they are embedded. This, in turn, enriches the process and results (Palaganas et al., 2017). Reflexive researchers engage in the production of evidence-based scientific findings with an awareness that there is no true objectivity, only aspirations. It is therefore necessary for authors to critically examine their appeal for "pure" legal certainty, acknowledging the influence of their own worldview throughout the research process.

This need for reflexivity extends to *knowledge brokers* as well, who mediate between academic evidence and policy decisions (Fosch-Villaronga et al., 2025). It is also imperative that policymakers engage with this knowledge in a reflective manner, free from political bias, recognizing that their decisions carry tangible consequences in the real world. Complementing reflexivity, efforts undertaken by researchers to minimize biases and to embark on long-term projects can also be helpful to achieve fairer outcomes. In this regard, reviewing each other's work as peers can help academics minimize their biases in their work, consider how their worldview is presented in their articles, and, ultimately, achieve a greater and fairer model for the next generation of scientists in quantitative and qualitative methods. Using slow science and embarking on long-term projects can, in turn, foster collaborations and can also help assess the quality of research and entrust academics to reflect on their work and biases (Frith, 2020). In short, while complete objectivity is unattainable, acknowledging one's positionality can support more transparent, inclusive and responsible policymaking.

Ultimately, while perfect fairness may remain aspirational, critically engaging with its complexities – conceptually, legally and ethically – can help create more equitable and trustworthy HRIs.

Funding statement. This work was funded by the ERC StG Safe and Sound project, a project that has received funding from the European Union's Horizon-ERC program, Grant Agreement No. 101076929.

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

Competing interests. The authors declare no competing interests.

References

Adams, J. S. (2015). Equity theory. In J. B. Miner. *(2015). Organizational behavior 1: Essential theories of motivation and leadership*, Routledge, 134–158.

Addison, A., Bartneck, C., & Yogeeswaran, K. (2019, January). Robots can be more than black and white: Examining racial bias towards robots. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 493–498).

Alarcon, G. M., Capiola, A., Hamdan, I. A., Lee, M. A., & Jessup, S. A. (2023). Differential biases in human-human versus human-robot interactions. *Applied Ergonomics*, *106*, 103858.

Alikhademi, K., Drobina, E., Prioleau, D., Richardson, B., Purves, D., & Gilbert, J. E. (2022). A review of predictive policing from the perspective of fairness. *Artificial Intelligence and Law*, *30*(1), 1–17.

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989.

Babin, S. E. (2013). Color theory: The effects of color in medical environments. Honors Theses. 115. *The Aquila Digital Community*. The University of Southern Mississippi. Retrieved from https://aquila.usm.edu/honors_theses/115.

Barfield, J. (2021, June). Discrimination and stereotypical responses to robots as a function of robot colorization. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, 109–114.

Barfield, J. (2023a). Designing social robots to accommodate diversity, equity, and inclusion in human-robot interaction. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*, 463–466.

Barfield, J. K. (2023b). Discrimination against robots: Discussing the ethics of social interactions and who is harmed. *Paladyn, Journal of Behavioral Robotics*, *14*(1), 20220113.

Bartneck, C., Yogeeswaran, K., Ser, Q. M., Woodward, G., Sparrow, R., Wang, S., & Eyssel, F. (2018, February). Robots and racism. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*, 196–204.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, *3*(21).

Bhat, P I (2020). Doctrinal legal research as a means of synthesizing facts, thoughts, and legal principles. *Idea and Methods of Legal Research*, 143–168. https://doi.org/10.1093/oso/9780199493098.003.0005

Bielak, A. T., Campbell, A., Pope, S., Schaefer, K., & Shaxson, L. (2008). *From science communication to knowledge brokering: The shift from 'science push' to 'policy pull.'* In D. Cheng, M. Claessens, T. Gascoigne, *et al.*

Boada, J. P., Maestre, B. R., & Genís, C. T. (2021). The ethical issues of social assistive robotics: A critical literature review. *Technology in Society*, *67*, 101726.

Boyd, R., & Holton, R. J. (2018). Technology, innovation, employment and power: Does robotics and artificial intelligence really mean social transformation? *Journal of Sociology*, *54*(3), 331–345.

Braithwaite, J. (2002). Rules and principles: A theory of legal certainty. *Australasian Journal of Legal Philosophy*, *27*(2002), 47–82.

Calleja, C., Drukarch, H., & Fosch-Villaronga, E. (2022). *Harnessing robot experimentation to optimize the regulatory framing of emerging robot technologies. Data & Policy*, Cambridge University Press, 1–15.

Cao, J., & Chen, N. (2024). The influence of robots' fairness on humans' reward-punishment behaviors and trust in human-robot cooperative teams. *Human Factors*, *66*(4), 1103–1117.

Carey, A. N., & Wu, X. (2022). The causal fairness field guide: Perspectives from social and formal sciences. *Frontiers in Big Data*, *5*, 892837.

Carey, A. N., & Wu, X. (2023). The statistical fairness field guide: Perspectives from social and formal sciences. *AI and Ethics*, *3*(1), 1–23.

Cartwright, P. (2015). Understanding and protecting vulnerable financial consumers. *Journal of Consumer Policy*, *38*(2), 119–138.

Caton, S., & Haas, C. (2024). Fairness in machine learning: A survey. *ACM Computing Surveys*, *56*(7), 1–38.

Chang, M. L., Pope, Z., Short, E. S., & Thomaz, A. L. (2020, August). Defining fairness in human-robot teams. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (pp. 1251–1258). IEEE.

Chang, M. L., Trafton, G., McCurry, J. M., & Thomaz, A. L. (2021, August). Unfair! perceptions of fairness in human-robot teams. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, 905–912.

Choi, A. H., & Spier, K. E. (2014). Should consumers be permitted to waive products liability? Product safety, private contracts, and adverse selection. *The Journal of Law, Economics, & Organization*, *30*(4), 734–766.

Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., … Mavridis, N. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *NPJ Digital Medicine*, *3*(1), 1–11.

Claure, H., Candon, K., Shin, I., & Vázquez, M. (2024). Dynamic Fairness Perceptions in Human-Robot Interaction. *arXiv preprint arXiv:2409.07560*, https://arxiv.org/pdf/2409.07560.

Clifford, D., & Ausloos, J. (2018). Data protection and the role of fairness. *Yearbook of European Law*, *37*, 130–187.

Collingridge, D. (1980). The dilemma of control. *The Social Control of Technology*, 13–22.

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

Colquitt, J. A., & Zipay, K. P. (2015). Justice, fairness, and employee reactions. *Annual Review of Organizational Psychology and Organizational Behavior*, 2(1), 75–99.

Council Directive 93/13/EEC of 5 April 1993 on unfair terms in consumer contracts, OJ L 95, 21.4.1993, p. 29–34.

Council of Europe. 4 April 2011. *EUROPEAN COMMISSION FOR DEMOCRACY THROUGH LAW (VENICE COMMISSION) | REPORT ON THE RULE OF LAW*. https://www.venice.coe.int/webforms/documents/?pdf=CDL-AD(2011)003rev-e.

Council of Europe. 18 March 2016. *EUROPEAN COMMISSION FOR DEMOCRACY THROUGH LAW (VENICE COMMISSION) | RULE OF LAW CHECKLIST*. https://venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-AD(2016)007-e.

Danielescu, A. (2020). Eschewing gender stereotypes in voice assistants to promote inclusion. In *Proceedings of the 2nd conference on conversational user interfaces* (pp. 1–3).

Directive (EU) 2023/2673 of the European Parliament and of the Council of 22 November 2023 amending Directive 2011/83/EU as regards financial services contracts concluded at a distance and repealing Directive 2002/65/EC, OJ L, 2023/2673, 28.11.2023, pp. 1–21.

Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive'), OJ L 149, 11.6.2005, pp. 22–39.

Directive 2009/48/EC of the European Parliament and of the Council of 18 June 2009 on the safety of toys, OJ L 170, 30.6.2009, pp. 1–37

Edelman, B. (2009). Adverse selection in online "trust" certifications. Proceedings of the 11th International Conference on Electronic Commerce. https://doi.org/10.1145/1593254.1593286

Eschenbach, W. J. V. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy and Technology*, *34*(4), 1607–1622.

European Commission. (n.d.-a). *Digital fairness – Fitness check on EU consumer law*. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13413-Digital-fairness-fitness-check-on-EU-consumer-law_en.

European Commission. (n.d.-b). *What is the rule of law?*. Retrieved October 25, 2024, from https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/upholding-rule-law/rule-law/what-rule-law_en.

European Commission. (n.d.-c). *Better regulation: Guidelines and toolbox*. Retrieved October 25, 2024, from https://commission.europa.eu/law/law-making-process/planning-and-proposing-law/better-regulation/better-regulation-guidelines-and-toolbox_en.

European Commission. (n.d.-d.) *Better regulation: Why and how*. Retrieved October 25, 2024, from https://commission.europa.eu/law/law-making-process/planning-and-proposing-law/better-regulation_en.

European Commission. (n.d.-e). *Knowledge for policy*. Retrieved October 25, 2024, from https://knowledge4policy.ec.europa.eu/home_en.

European Commission. (n.d.-f). *Inclusive Robotics for a better Society (INBOTS)*. Retrieved 29 October 2024, from https://cordis.europa.eu/project/id/780073.

European Commission. (n.d.-g). *Stakeholder-informed ethics for new technologies with high socio-economic and human rights impact*. Retrieved 29 October 2024, from https://cordis.europa.eu/project/id/741716.

European Data Protection Board. (2025). *Statement 1/2025 on Age Assurance*. Retrieved 29 April 2025, from https://www.edpb.europa.eu/our-work-tools/our-documents/other-guidance/statement-12025-age-assurance_en.

European Union Agency for Fundamental Rights (FRA, 2022). *Bias in algorithms: Artificial intelligence and discrimination*. Publications Office of the European Union. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf.

Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamo-Larrieux, A. (2019a). Robots and transparency: The multiple dimensions of transparency in the context of robot technologies. *IEEE Robotics and Automation Magazine*, *26*(2), 71–78.

Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019b). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data and Society*, *6*(1), 2053951719860542.

Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards transparency by design for artificial intelligence. *Science and Engineering Ethics*, *26*(6), 3333–3361.

Ferrara, E. (2023). Should chatgpt be biased? Challenges and risks of bias in large language models. *arXiv preprint arXiv:2304.03738*, https://arxiv.org/pdf/2304.03738.

Fischer, K. (2018, March). When transparent does not mean explainable. In *Explainable Robotic Systems—Workshop in conjunction with HRI 2018*. Retrieved from https://explainableroboticsystems.wordpress.com/wp-content/uploads/2018/03/1-fischer-being-honest-about-transparency-final.pdf.

Folger, R., & Cropanzano, R. (2001). Fairness theory: Justice as accountability. *Advances in Organizational Justice*, *1*(1–55), 12.

Fosch-Villaronga, E. (2019). *Robots, Healthcare and the Law: Regulating Automation in Personal Care*. Routledge.

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

Fosch-Villaronga, E., & Drukarch, H. (2023). Accounting for diversity in robot design, testbeds, and safety standardization. *International Journal of Social Robotics*, *15*(11), 1871–1889.

Fosch-Villaronga, E., & Heldeweg, M. (2018). "Regulation, I presume?" said the robot–Towards an iterative regulatory process for robot governance. *Computer Law & Security Review*, *34*(6), 1258–1277.

Fosch-Villaronga, E., & Millard, C. (2019). Cloud robotics law and regulation: Challenges in the governance of complex and dynamic cyber–physical ecosystems. *Robotics and Autonomous Systems*, *119*, 77–91.

Fosch-Villaronga, E., & Poulsen, A. (2022). Diversity and inclusion in artificial intelligence. B. Custers & E. Fosch-Villaronga (Eds.), *(2022). Law and artificial intelligence: Regulating AI and applying AI in legal practice*. (Springer Nature) vol 35. 109–134.

Fosch-Villaronga, E., & Roig, A. (2017). European regulatory framework for person carrier robots. *Computer Law & Security Review*, *33*(4), 502–520.

Fosch-Villaronga, E., Shaffique, M. R., Schwed-Shenker, M., Mut-Piña, A., van der Hof, S., & Custers, B. (2025). Science for Robot Policy: Advancing robotics policy through the EU science for policy approach. *Technological Forecasting and Social Change*, *218*, 124202.

Fossa, F., & Sucameli, I. (2022). Gender bias and conversational agents: An ethical perspective on social robotics. *Science and Engineering Ethics*, *28*(3), 23.

Frissen, R., Adebayo, K. J., & Nanda, R. (2023). A machine learning approach to recognize bias and discrimination in job advertisements. *AI & Soc*, *38*(2), 1025–1038. https://doi.org/10.1007/s00146-022-01574-0

Frith, U. (2020). Fast lane to slow science. *Trends in Cognitive Sciences*, *24*(1), 1–2.

Giger, J. C., Piçarra, N., Pochwatko, G., Almeida, N., & Almeida, A. S. (2025). Intention to Work with Social Robots: The Role of Perceived Robot Use Self-Efficacy, Attitudes Towards Robots, and Beliefs in Human Nature Uniqueness. *Multimodal Technologies and Interaction*, *9*(2), 9, 1–13.

Gluckman, P. D., Bardsley, A., & Kaiser, M. (2021). Brokerage at the science–policy interface: From conceptual framework to practical guidance. *Humanities and Social Sciences Communications*, *8*(1), 1–10. https://doi.org/10.1057/s41599-021-00756-3.

Goffin, T., & Palmieri, S. (2024). Regulating smart healthcare robots: The european approach. *Research Handbook on Health, AI and the Law*, 75–91. https://doi.org/10.4337/9781802205657.ch05

González-Sendino, R., Serrano, E., Bajo, J., & Novais, P. (2023). A review of bias and fairness in artificial intelligence.

Greenberg, B. A. (2015). Rethinking technology neutrality. *Minnesota Law Review*, *100*, 1495.

Groppe, N., & Schweda, M. (2023). Gender stereotyping of robotic systems in eldercare: An exploratory analysis of ethical problems and possible solutions. *International Journal of Social Robotics*, *15*(11), 1963–1976.

Haidegger, T., Barreto, M., Gonçalves, P., Habib, M. K., Ragavan, S. K. V., Li, H., Prestes, E. (2013). Applied ontologies and standards for service robots. *Robotics and Autonomous Systems*, *61*(11), 1215–1223.

Hartzog, W. (2014). Unfair and deceptive robots. *Maryland Law Review*, *74*, 785.

Hiroi, Y., & Ito, A. (2016). Influence of the height of a robot on comfortableness of verbal interaction. *IAENG International Journal of Computer Science*, *43*(4), 447–455.

Hitron, T., Megidish, B., Todress, E., Morag, N., & Erel, H. (2022, August). Ai bias in human-robot interaction: An evaluation of the risk in gender biased robots. In 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) (pp. 1598-1605). IEEE.

Howard, A., & Borenstein, J. (2018). The ugly truth about ourselves and our robot creations: The problem of bias and social inequity. *Science and Engineering Ethics*, *24*(5), 1521–1536.

Hundt, A., Agnew, W., Zeng, V., Kacianka, S., & Gombolay, M. (2022, June). Robots enact malignant stereotypes. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 743–756.

ISO. (2014). *ISO 13482:2014_Robots and robotic devices—Safety requirements for personal care robots*. https://www.iso.org/standard/53820.html.

ISO. (2021). *ISO 8373:2021 Robotics—Vocabulary*. https://www.iso.org/standard/75539.html.

Johnson, D. G. (2015). Technology with no human responsibility? *Journal of Business Ethics*, *127*(4), 707–715.

Judgment of the House of Lords, United Kingdom of 24 May 2006, *McFarlane v McFarlane* [2006] UKHL 24. Retrieved from https://publications.parliament.uk/pa/ld200506/ldjudgmt/jd060524/mill.pdf.

Kotrotsios, G. (2021). *Data, New Technologies, and Global Imbalances: Beyond the Obvious*. Cambridge Scholars Publishing.

Lambrecht, A., & Tucker, C. (2024). Apparent algorithmic discrimination and real-time algorithmic learning in digital search advertising. *Quantitative Marketing and Economics*, *22*(4), 357–387. https://doi.org/10.1007/s11129-024-09286-z

Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Review*, *9*(2). https://doi.org/10.14763/2020.2.1469

Lee, I. (2021). Service robots: A systematic literature review. *Electronics*, *10*(21), 2658.

Leiden University. (n.d.-a). *LIAISON*. Retrieved 25 October 2024, from https://www.universiteitleiden.nl/en/research/research-projects/law/liaison.

Leiden University. (n.d.-b). *PROPELLING*. Retrieved 25 October 2024, from https://www.universiteitleiden.nl/en/research/research-projects/law/propelling.

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

Leiden University. (n.d.-c). *SAFE and SOUND: Towards Evidence-based Policies for Safe and Sound Robots*. Retrieved 25 October 2024, from https://www.universiteitleiden.nl/en/research/research-projects/law/safe–sound-towards-evidence-based-policies-for-safe-and-sound-robots.

Leong, B., & Selinger, E. (2019, January). Robot eyes wide shut: Understanding dishonest anthropomorphism. In *Proceedings of the conference on fairness, accountability, and transparency*. (pp. 299–308).

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes: The premise, the proposed solutions, and the open challenges. *Philosophy and Technology*, *31*(4), 611–627.

Leslie, D., Rincón, C., Briggs, M., Perini, A., Jayadeva, S., Borda, A., … Kherroubi Garcia, I. (2023). AI Fairness in Practice. The Alan Turing Institute. Retrieved from *arXiv preprint arXiv:2403.14636*, https://arxiv.org/pdf/2403.14636.

Lim, H., & Letkiewicz, J. C. (2023). Consumer experience of mistreatment and fraud in financial services: Implications from an integrative consumer vulnerability framework. *Journal of Consumer Policy*, *46*(2), 109–135. https://doi.org/10.1007/s10603-023-09535-w

Liu, X. J., Nie, Z., Yu, J., Xie, F., & Song, R. (Eds.). (2021). *Intelligent Robotics and Applications: 14th International Conference, ICIRA 2021, Yantai, China, October 22–25, 2021, Proceedings, Part III* (Vol. 13015). Springer Nature.

Londoño, L., Hurtado, J. V., Hertz, N., Kellmeyer, P., Voeneky, S., & Valada, A. (2024). Fairness and bias in robot learning. *Proceedings of the IEEE*, *112*(4), 305–330.

Lucas, H., Poston, J., Yocum, N., Carlson, Z., & Feil-Seifer, D. (2016, August). Too big to be mistreated? Examining the role of robot size on perceptions of mistreatment. In *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*, 1071–1076.

Mahler, T. (2024). Smart robotics in the EU legal framework: The role of the machinery regulation. *Oslo Law Review, DOI, 10*.

Mansouri, M., & Taylor, H. (2024). Does cultural robotics need culture? Conceptual fragmentation and the problems of merging culture with robot design. *International Journal of Social Robotics*, *16*(2), 385–401.

Marchant, G. E. (2011). Addressing the pacing problem. *The growing gap between emerging technologies and legal-ethical oversight: The pacing problem*, 199–205.

Martinetti, A., Chemweno, P., Nizamis, K., & Fosch-Villaronga, E. (2021). Redefining safety in light of human-robot interaction: A critical review of current standards and regulations. *Frontiers in Chemical Engineering*, *3*(666237), 1–12.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR) 54*(6), 1–35

Mulligan, D. K., Kroll, J. A., Kohli, N., & Wong, R. Y. (2019). This thing called fairness: Disciplinary confusion realizing a value in technology. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–36.

Naik, T., Gostu, H., & Sharma, R. (2024, March). Navigating Ethics of AI-Powered Creativity in Midjourney. In *2024 3rd International Conference for Innovation in Technology (INOCON)* (1–6). IEEE.

Nwana, H. S. (1996). Software agents: An overview. *The Knowledge Engineering Review*, *11*(3), 205–244. https://doi.org/10.1017/S026988890000789X

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Palaganas, E. C., Sanchez, M. C., Molintas, M. V. P., & Caricativo, R. D. (2017). Reflexivity in qualitative research.

Palmerini, E., Bertolini, A., Battaglia, F., Koops, B. J., Carnevale, A., & Salvini, P. (2016). RoboLaw: Towards a European framework for robotics regulation. *Robotics and Autonomous Systems*, *86*, 78–85.

Pessach, D., & Shmueli, E. (2022). A review on fairness in machine learning. *ACM Computing Surveys (CSUR) 55*(3), 1–44

Pflücke, F. (2024). Evidence-based Consumer Law and Policy. In F. Pflücke (Ed.), *Compliance with European Consumer Law: The Case of E-Commerce* (p. 8). Oxford University Press. https://doi.org/10.1093/9780198906414.003.0002

Princen, S. (2022). The use of evidence in evidence-based legislation: A reflection. *European Journal of Law Reform*, *24*(1), 147–160. https://doi.org/10.5553/EJLR/138723702022024001010

Randall, N., & Šabanović, S. (2024). Designing robots for marketplace success: A case study with technology for behavior and habit change. *International Journal of Social Robotics*, *16*(3), 461–487. https://doi.org/10.1007/s12369-023-01093-y

Rani, P. A. J., Liu, C., Sarkar, N., & Vanman, E. J. (2006). An empirical study of machine learning techniques for affect recognition in human–robot interaction. *Pattern Analysis and Applications*, *9*(1), 58–69. https://doi.org/10.1007/s10044-006-0025-y

Rawls, J. (1971). *Atheory of justice*. Cambridge (Mass.).

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation. OJ L 119, 27.4.2016, pp. 1–88.

Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. OJ L 117, 5.5.2017, pp. 1–175.

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

Regulation (EU) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on machinery and repealing Directive 2006/42/EC of the European Parliament and of the Council and Council Directive 73/361/EEC. OJ L 165, 29.6.2023, pp. 1–102.

REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). OJ L 2024/1689, 12.7.2024, pp. 1–144.

Reig, S., Luria, M., Forberger, E., Won, I., Steinfeld, A., Forlizzi, J., & Zimmerman, J. (2021, June). Social robots in service contexts: Exploring the rewards and risks of personalization and re-embodiment. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (pp. 1390–1402).

Rigotti, C., & Fosch-Villaronga, E. (2024). Fairness, AI & recruitment. *Computer Law & Security Review*, *53*, 105966.

RoboLaw. JUNE, 2014. *RoboLaw: Project Overview*. Retrieved 25 October 2024, from http://www.robolaw.eu/projectdetails.htm.

Rosenthal-Von Der Pütten, A. M., & Krämer, N. C. (2014). How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior*, *36*, 422–439.

Šabanović, S. (2010). Robots in society, society in robots: Mutual shaping of society and technology as a framework for social robot design. *International Journal of Social Robotics*, *2*(4), 439–450.

Salisbury, L. C., Blanchard, S. J., Brown, A. L., Nenkov, G. Y., Hill, R. P., & Martin, K. D. (2023). Beyond income: Dynamic consumer fnancial vulnerability. *Journal of Marketing*, *87*(5), 657–678.

Salvini, P., Paez-Granados, D., & Billard, A. (2021). On the safety of mobile robots serving in public spaces: Identifying gaps in EN ISO 13482: 2014 and calling for a new standard. *ACM Transactions on Human-Robot Interaction (THRI) 10*(3), 1–27

Scheutz, M. (2012). The inherent dangers of unidirectional emotional bonds between humans and social robots. In P. Lin, K. Abney & G. A. Bekey (Eds.), *Robot ethics: The ethical and social implications of robotics* (pp. 205–221). MIT Press.

Schiele, B., & Shi, S. (2008). *Communicating science in social contexts. New models, new practices*, 201–226.

Schwartz, R., Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). *Towards a standard for identifying and managing bias in artificial intelligence* (Vol. 3, p. 00). US Department of Commerce, National Institute of Standards and Technology. Retrieved from: https://doi.org/10.6028/NIST.SP.1270.

Setchi, R., Dehkordi, M. B., & Khan, J. S. (2020). Explainable robotics in human-robot interactions. *Procedia Computer Science*, *176*, 3057–3066.

Seyitoğlu, F., & Ivanov, S. (2023). Service robots and perceived discrimination in tourism and hospitality. *Tourism Management*, *96*, 104710.

Shahbazi, N., Lin, Y., Asudeh, A., & Jagadish, H. V. (2023). Representation bias in data: A survey on identification and resolution techniques. *ACM Computing Surveys*, *55*(13s), 1–39.

Shcherbanyuk, O., Gordieiev, V., & Bzova, L. (2023). Legal nature of the principle of legal certainty as a component element of the rule of law. *Juridical Tribune-Review of Comparative and International Law*, *13*(1), 21–31.

Singamaneni, P. T., Bachiller-Burgos, P., Manso, L. J., Garrell, A., Sanfeliu, A., Spalanzani, A., & Alami, R. (2024). A survey on socially aware robot navigation: Taxonomy and future challenges. *The International Journal of Robotics Research*, 02783649241230562.

Singh, D. K., Kumar, M., Fosch-Villaronga, E., Singh, D., & Shukla, J. (2023). Ethical considerations from child-robot interactions in under-resourced communities. *International Journal of Social Robotics*, *15*(12), 2055–2071.

Song, Y., & Luximon, Y. (2020). Trust in AI agent: A systematic review of facial anthropomorphic trustworthiness for social robot design. *Sensors*, *20*(18), 5087.

Soori, M., Arezoo, B., & Dastres, R. (2023). Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cognitive Robotics*, *3*, 54–70.

Søraa, R. A., & Fosch-Villaronga, E. (2020). Exoskeletons for all: The interplay between exoskeletons, inclusion, gender, and intersectionality. *Paladyn, Journal of Behavioral Robotics*, *11*(1), 217–227.

Sprenger, M., & Mettler, T. (2015). Service robots. *Business and Information Systems Engineering*, *57*(4), 271–274.

Stamenkovic, P. (2024). Straightening the 'value-laden turn': Minimising the influence of extra-scientific values in science. *Synthese*, *203*(1), 20, 1–38.

Takan, S., TAKAN, D. E., Yaman, S. G., & Kılınççeker, O. (2023). Bias in human data: A feedback from social sciences. *WIREs: Data Mining and Knowledge Discovery*, *13*(4). https://doi.org/10.1002/widm.1498.

Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots: The double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior*, *38*, 75–84.

Turnhout, E., Stuiver, M., Klostermann, J., Harms, B., & Leeuwis, C. (2013). New roles of science in society: Different repertoires of knowledge brokering. *Science and Public Policy*, *40*(3), 354–365.

van Meerbeeck, J. (2016). The principle of legal certainty in the case law of the European court of justice: From certainty to trust. *European Law Review*, *41*(2), 275–288.

Verhoef, T., & Fosch-Villaronga, E. (2023, September). Towards affective computing that works for everyone. In *2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII)* IEEE, 1–8.

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press

Verma, S., & Rubin, J. 2018. Fairness definitions explained. In Proceedings of the 2018 IEEE/ACM International Workshop on Software Fairness (FairWare'18). IEEE, Los Alamitos, CA, 1–7.

Waheed, H. (2016). Subjective objectivity and objective subjectivity: the paradox in social science. *Unpublished manuscript. Retrieved from* https://www.academia.edu/18040949/Subjective_Objectivity_and_Objective_Subjectivity.

Wang, G., Phan, T. V., Li, S., Wang, J., Peng, Y., Chen, G., … Liu, L. (2022). Robots as models of evolving systems. *Proceedings of the National Academy of Sciences*, *119*(12), e2120019119.

Wangmo, T., Lipps, M., Kressig, R. W., & Ienca, M. (2019). Ethical concerns with the use of intelligent assistive technology: Findings from a qualitative study with professional stakeholders. *BMC Medical Ethics*, *20*(1), 1–11.

Whittaker, M., Alper, M., Bennett, C. L., Hendren, S., Kaziunas, L., Mills, M., West, S. M. (2019). Disability, bias, and AI. *AI Now Institute*, *8*. Retrieved from https://tinyurl.com/yr7cft3h.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Mons, B. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1). https://doi.org/10.1038/sdata.2016.18.

Willett, C. (2010). Fairness and consumer decision making under the unfair commercial practices directive. *Journal of Consumer Policy*, *33*(3), 247–273.

Winfield, A. (2019). Ethical standards in robotics and AI. *Nature Electronics*, *2*(2), 46–48.

Yang, W., & Xie, Y. (2024). Can robots elicit empathy? The effects of social robots' appearance on emotional contagion. *Computers in Human Behavior: Artificial Humans*, *2*(1), 100049.

Yapo, A., & Weiss, J. W. (2018). Ethical implications of bias in machine learning. Proceedings of the Annual Hawaii International Conference on System Sciences. https://doi.org/10.24251/hicss.2018.668.

Zhou, R. (2024). Risks of discrimination violence and unlawful actions in LLM-driven robots. *Computer Life*, *12*(2), 53–56.

Dr. Eduard Fosch-Villaronga Ph.D. LL.M M.A. is Associate Professor and Director of Research at the eLaw–Center for Law and Digital Technologies at Leiden University (NL). Eduard is an ERC Laureate who investigates the legal and regulatory aspects of robot and AI technologies, focusing on healthcare, governance, diversity, and privacy. Eduard Fosch-Villaronga is the Principal Investigator (PI) of his personal ERC Starting Grant SAFE & SOUND where he works on science for robot policies. Eduard is also the PI of eLaw's contribution to the Horizon Europe BIAS Project: Mitigating Diversity Biases of AI in the Labour Market. In 2023, Eduard received the EU Safety Product Gold Award for his contribution to making robots safer by including diversity considerations. Eduard is part of the Royal Netherlands Standardization Institute (NEN) as an expert and the International Standard Organization (ISO) as a committee member in the ISO/TC 299/WG 2 laying down Safety Requirements for Service Robots (ISO/CD 13,482).

Dr. Antoni Mut Piña is a postdoctoral researcher at the eLaw–Center for Law and Digital Technologies at Leiden University, investigating science for robot policy at the ERC StG SAFE & Sound project. He completed his Ph.D. at the University of Barcelona, focusing on empirical legal analysis and consumer behavior. Antoni's work employs advanced data analysis techniques to generate policy-relevant data. He holds degrees in Law, Business Administration, along with postgraduate studies in Political Analysis and Consumer Contract Law, with experience in research on consumer vulnerability and regulatory policies.

Mohammed Raiz Shaffique is a PhD candidate at the eLaw–Center for Law and Digital Technologies at Leiden University researching law and robotics. Raiz investigates the generation and use of policy-relevant data to improve the regulatory framework governing physical assistant robots and wearable robots, as part of the ERC StG Safe & Sound project. Previously, he worked on age assurance and online child safety under the EU's Better Internet for Kids+ initiative. Raiz earned his Advanced Master's in Law and Digital Technologies at Leiden University with top honors and authored a book on white-collar crime in India. He also has six years of legal experience and expertise in cyber law and dispute resolution.

Marie Schwed-Shenker is a PhD candidate at the eLaw–Center for Law and Digital Technologies at Leiden University researching law and robotics. Her work focuses on science for social assistive robot policy as part of the ERC StG Safe & Sound project. Previously, she studied Practical Criminology at Hebrew University and Sociology and Anthropology at Tel Aviv University. Marie's work has focused on ethical issues in youth volunteering and the role of media in radicalization. In previous research, she examined stereotypes of female prisoners in media and language's role in group radicalization.

Cite this article: Fosch-Villaronga, E., Mut-Piña, A., Shaffique, M.R., & Schwed-Shenker, M. (2025). Exploring fairness in service robotics. *Cambridge Forum on AI: Law and Governance*, *1*, e28, 1–19. https://doi.org/10.1017/cfl.2025.10012

https://doi.org/10.1017/cfl.2025.10012 Published online by Cambridge University Press