Discover Artificial Intelligence

Research

Virtuous integrative social robotics for ethical governance

Anshu Saxena Arora[1] · **Arlene Marshall**[1] · **Amit Arora**[1] · **John R. McIntyre**[2]

Received: 1 September 2024 / Accepted: 15 January 2025
Published online: 29 January 2025

Abstract
This research conceptualizes virtuous integrative social robotics (VISR) as a value-driven philosophy for developing and designing social robots and robotic applications. It encompasses shared ethical principles highlighted as autonomy, responsibility, and transparency (ART) for social robotics in the VISR context. Virtue ethics is explored as a means for programming social robots as artificial moral agents, placing human values as the basis of robot design. It is based on the 'non-replacement principle' whereby social robots should behave as a virtuous human would. Finally, this research provides managerial implications and promises to find innovative ways based on ethical decision-making and improving ethical transparency for human–robot interaction (HRI), unification, trust, and collaboration more than ever before.

Keywords Virtuous integrative social robotics (VISR) · Autonomy, responsibility, and transparency (ART) · Artificial intelligence (AI) · Design for values · Virtuous robots · Shared ethics · Human–robot interaction (HRI)

1 Introduction

Artificial intelligence (AI) refers to the capability of machines to perform tasks that typically require human cognition, such as learning, reasoning, problem-solving, and adapting to new situations. AI systems utilize algorithms, data processing, and computational models to simulate intelligent behavior and facilitate decision-making processes. While AI strives to approximate aspects of human intelligence [26], it operates based on predefined rules, training data, and patterns, rather than possessing intrinsic understanding or consciousness. If we can conceive that AI systems may simulate empathy in a manner that appears empathetic to humans, it opens possibilities for exploring how such systems can support human interactions. However, we recognize that while AI can demonstrate behaviors that appear empathetic, the nature and depth of empathy may differ from human emotional experiences in ways we are still exploring and understanding. Virtuous and moral behavior in AI and robotics can be accomplished through the programming of social robots. Additionally, moral behavior may not be broadly required for every AI application. However, it should be required in certain specialized areas of social robotics, where there is human–robot interaction (HRI) with the potential for influencing humans or where it is used to facilitate or improve human–human interaction (HHI). In these special situations, virtuous robots are considered artificial moral agents (AMAs) [18].

Social robotics involves the design, development, and study of robots capable of engaging in meaningful interactions with humans, fostering social behaviors that support collaboration, assistance, and trust [30]. Human–robot interaction (HRI) explores how humans and robots engage with each other in diverse settings, aiming to optimize communication,

✉ Anshu Saxena Arora, anshu.arora@udc.edu; Arlene Marshall, arlene.marshall@udc.edu; Amit Arora, amit.arora@udc.edu; John R. McIntyre, john.mcintyre@scheller.gatech.edu | [1]University of the District of Columbia, Washington, DC, USA. [2]Georgia Institute of Technology, Atlanta, GA, USA.

collaboration, and trust in these interactions [30]. Virtuous robots are designed for Human–Robot Interaction (HRI) "to help humans reach a higher level of moral development" [6], p. 7) through moral reasoning enhancement, ethical reasoning development, and ethical evolution among humans. Cappuccio et al. [6] emphasize that social robots should not merely simulate human behavior but act as agents that inspire humans toward higher levels of virtue and moral reasoning through meaningful interactions. For example, robots used in educational settings, such as teaching children about sharing and cooperation, aim to cultivate positive moral values. However, such applications remain exceptions rather than the norm, highlighting the need for industry-wide adoption of moral programming as a standard in robotics design. Building on this foundation, Shneiderman [36] underscores the necessity of embedding moral programming within the design and development of AI systems. As artificial intelligence (AI) continues to advance, its applications must transcend narrow purposes such as business efficiency or cost optimization to address broader societal implications, including ethical and moral considerations. Shneiderman [36] advocates for a human-centered research, design, and development (RDD) process, where the focus lies on ensuring that AI applications are aligned with societal values, fostering trust, and addressing ethical concerns proactively. This necessitates integrating moral programming through human-centered RDD into social robots that embed ethical values at every stage of development. This approach prioritizes transparency, accountability, and fairness in AI systems, ensuring they serve humanity's broader societal needs rather than merely achieving technical or economic goals [36]. Autonomy in robots, when guided by moral programming, allows them to act as ethical agents in diverse contexts, facilitating trust and accountability in their interactions with humans. At the same time, the impact of AI and robotics on employment underscores the need to balance innovation with ethical governance, ensuring that technological progress aligns with human values and societal needs.

Technological advancements make human lives easier. Few people would give up the convenience of 24-h access to an automated teller machine (ATM) for a return to physically conducting cash transactions with a human teller during regular business hours. Research shows that most jobs require varying degrees of four types of intelligences: mechanical, analytical, intuitive, and empathetic [22, 23]. Conceptually, AI will supplant human intelligence for mechanical and analytical tasks, thereby steadily encroaching on intuitive and empathetic tasks. The loss of human jobs to AI or robotic systems may spark moral debate. However, as in the example of ATMs, businesses cannot ignore innovations like robots and robotic systems that could increase efficiency and effectiveness. While robotics can lead to the displacement of certain jobs, this is not an inherent or unavoidable consequence. The accountability for such outcomes lies with humans, who design, deploy, and regulate robotic systems. By prioritizing collaborative human–robot systems, such as robots that complement human labor rather than replace it, we can create scenarios where robotics enhance quality of life and reduce occupational hazards without significant workforce disruption.

Technology and innovations like artificial intelligence (AI), AI powered chatbots, social robots, etc. are important to our society and businesses. However, the negative impact of technology on jobs is well documented. For example, since the 1950s, the number of employees in the steel industry has reduced from 500,000 to approximately 143,000 by 2021[1] [5, 9, 33]. From the steel industry to the automotive industry and other areas of manufacturing, machines are doing human jobs [9]. In a human–machine interaction, we can determine how and when AI (and robots) replace human jobs. The consequences of job displacement are context-dependent. For instance, while the automation of steel manufacturing reduced jobs in the U.S. steel industry, it simultaneously mitigated the dangers and physical burdens associated with this work. Similarly, robots used to clean sewer pipes or manage hazardous environments like Fukushima remove humans from life-threatening conditions, presenting clear ethical and practical benefits. This paper argues that the deployment of robotics should focus on aligning technological advancements with societal values, ensuring that labor displacement leads to positive outcomes, such as safer workplaces and improved quality of life. Having achieved most of the mechanical and analytical job skills through the use of robotics, AI has only partially succeeded with intuitive and empathetic skills. Until then, humans should gain or sharpen intuitive and empathetic skills to compete. Table 1 shows the four intelligences and how humans can learn to combat AI with human-based, machine-enhanced solutions.

As society navigates the era of Industry 4.0, opposing research fuels the debate on whether AI will have a negative or a positive impact. AI and robots could eradicate millions of jobs to new technologies in diverse sectors [4]. Alternatively, AI will have a net positive impact on jobs for servicing technologies or creative roles [37]. Businesses and governments seeking to run effectively and efficiently cannot ignore the advantages of AI. Therefore, whether for or against AI, it is coming. How do we, humans prepare ourselves for such staggering changes—that should be the human-centered, AI-focused debate question, as highlighted in Table 1.

[1] https://www.statista.com/statistics/1243935/employment-in-the-us-iron-and-steel-industry/.

Table 1 Four intelligences and humans vs. AI

Four intelligences explained	Humans vs. AI: human-centered, machine-enhanced solutions to combat AI
• Mechanical intelligence is required for tasks that can be done automatically and repetitively • Analytical intelligence is the ability to process information for problem-solving and learn from it—in other words, information processing • Intuitive intelligence is the ability to creatively think and effectively adjust to novel situations—in other words, to use wisdom based on experience • Empathetic intelligence is the ability to recognize and understand other peoples' emotions, respond appropriately emotionally, and influence others' emotions	• Dual service (unification)—whereby some customers may prefer to pay handsomely for human service, especially for mechanical intelligence • Machine serving humans (collaboration)—whereby humans pick the tasks they prefer to perform and increase their quality of life (e.g., humans can pursue creative and artistic tasks while AI performs those tasks that humans prefer not to, for example AI can perform mundane tasks using mechanical and analytical intelligences while humans focus on intuitive and empathetic intelligences) • Human–machine division of labor (trust)—whereby AI is viewed as a collaborative tool that helps humans perform better, especially dealing with intuitive and empathetic intelligences • Machine-enhanced humans (unification, trust, and collaboration)—whereby AI becomes a biological, technological, and even moral extension of humans, venturing into the realm of intuitive and empathetic intelligences

Adapted from Huang and Rust [23] and Gibert [18]

AI applications communicate with humans via electronic output devices, including robots [26]. Humanoid robots resemble humans and can perform mechanical, routine tasks on command. However, as human innovation advances autonomous AI, robots may become ubiquitous. HRI will be even more relevant as we navigate toward a future where robots are as common as cell phones. In this research, we conceptualize "Virtuous Integrative Social Robots (VISR)". Within VISR, elements of design for values methodology, integrated social robotics, HRI, and virtues ethics combine to prepare the populace for increased HRI, eliminate the AI black box, and infuse morality throughout every step of the design process and supply chain of AI and robots, while protecting intellectual property.

This research seeks to address questions related to VISR and find a balance between opposing viewpoints and realistic solutions to this fast-approaching moral and ethical dilemma. Through VISR, every person and/or organization within the supply chain for AI/robotics must be identified and accounted for. The upcoming sections explore VISR in detail along with virtue ethics, ART, and intellectual property.

2 Integrative social robotics

Research has defined "Integrative Social Robotics (ISR)" as a new method or approach to designing and developing social robots and robotic applications [7, 20, 34, 35]. Seibt's work in social robotics emphasizes the Five Principles of Integrative Social Robotics (ISR), which advocate for embedding societal, ethical, and psychological considerations into robotic systems through interdisciplinary collaboration. Seibt highlights the importance of designing robots that align with human values and the necessity of bridging gaps between technical design and broader societal impacts. The discipline of social robotics supports an approach to social robots that is methodical and systematic to move past the "gridlock of description, evaluation, and regulation that are kept far away from the research, design, and development (RDD) process" [34] (p. 29). RDD often stresses what AI can do, with developers and engineers racing to be the first to reach a new AI plateau. From the start, equal (if not more) attention and emphasis should be given to what AI is designed for and thereby allowed to do. ISR asserts that social robots aim to improve human society through applications that enhance or preserve human values, requiring short and long-term research documentation of social robots' ethical, social, anthropological, and psychological impact on individuals and communities [34]. Greater autonomy in social robots requires greater human involvement with humans in the loop. "We should reject the idea that autonomous machines can exceed or replace any meaningful notion of human intelligence, creativity, and responsibility" [36], p. 56). AI and social robots are tools to improve human existence, not replace it. It is important that not only do humans understand ethical nuisances, but social robots need to "understand" them too. Specifically, social robots operating as AMAs should be programmed

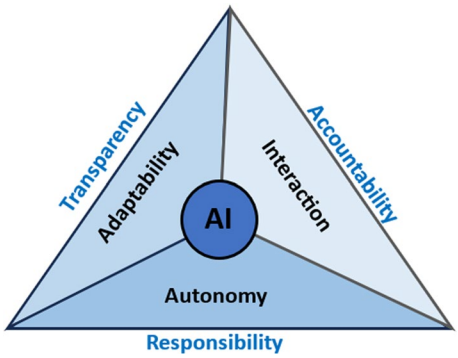

Fig. 1 The ART (accountability, responsibility, transparency) principle. Adapted by Dignum et al. [14]

to respond to the ethical and societal environment in which they are utilized by applying the social practices, people in that environment or field are used to.

The principles of autonomy, responsibility, and transparency (ART) for social robots can help explain why designing applications for values is important. According to Digmum et al. [13], autonomy is exemplified through interactivity, accountability and adaptability as the main characteristics of social robots (p. 44). To have trusted interaction, there must be accountability. Accountability is giving an account of your actions or choices [38]. Just as humans in their respective fields explain decisions, social robots must report the data and algorithms involved in arriving at an action or decision. This is currently lacking within the current algorithms. Not only is human decision-making guided by human experiences (that can be replicated by big data) but also by human morality or integrity—the origins of which the data does not reveal. While big data can model patterns of human behavior and decision-making, human experiences encompass more than data points, they involve subjective, emotional, and situational contexts that big data can approximate but not fully replicate. This distinction is important in understanding the limitations of AI systems designed to mimic human interaction.

Robots, as artificial agents, lack the capacity for accountability or moral responsibility. Instead, the responsibility for their design, deployment, and use rests with humans. By embedding ethical principles such as accountability, responsibility, and transparency (ART) into their design, we can ensure that robots serve humanity's interests without undermining societal values or displacing human workers unnecessarily. To have autonomy, we must be able to allocate responsibility. As a tool, robots cannot take responsibility. Ultimately, humans or human organizations are responsible for the purpose and use of the robot tool. Depending on the situation, different stakeholders could be responsible for social robots, including but not limited to the software developer, manufacturer, legislator, owner, and/or user. AI adaptability (which refers to the ability to learn and adjust) requires transparency (openness) to the program, data, and other mechanisms that enable it. Figure 1 illustrates the ART principles that lend themselves perfectly to the Design for Values methodology, which seeks to bring value and ethics into research, development, and design (RDD) [13]. According to Gibert [19], the first step to building a virtuous robot is establishing a group of virtuous people. The Design for Value approach's most significant step is determining which values the robots will sustain [13].

Embedding values into robots involves integrating ethical principles throughout their design and development lifecycle. For example, programming empathy into robots requires algorithms that detect and respond to emotional cues in human behavior, such as tone of voice, facial expressions, or body language. Machine learning models trained on diverse, ethically sourced datasets can simulate empathic responses, ensuring interactions are contextually appropriate and culturally sensitive. To sustain values like accountability, transparency must be integrated into robot decision-making processes. This could include mechanisms for logging and reporting actions, enabling users to trace decisions to their algorithmic origins. These approaches align with the ART principles of autonomy, responsibility, and transparency, ensuring that robots act as tools for ethical engagement rather than merely executing predefined tasks.

3 Conceptualizing virtuous integrative social robots (VISR)

In this research, we conceptualize Virtuous Integrative Social Robots (VISR) in the field of robotics and AI. VISR integrates elements of design for values methodology, integrated social robotics, HRI, and virtues ethics. In March 2023, calls for a moratorium on AI by tech giants Elon Musk (Tesla CEO), Steve Wozniak (Apple co-founder), and many more have

Fig. 2 Virtuous integrative social robotics (VISR) framework

highlighted the need for concern that artificial intelligence is causing a threat to society.[2] The safety constraints of self-policing by industry standards have become optional in the race to be the next big thing in AI [8]. Figure 2 shows our conceptualization of Virtuous Integrative Social Robots (VISR).

Virtuous integrative social robotics (VISR) is a comprehensive, value-driven framework for the design, development, governance, and implementation of social robotics and AI applications. At its core, VISR integrates ethical ART (autonomy, responsibility, and transparency) principles, governance mechanisms, regulatory considerations, intellectual property protections, and user-centered design to ensure that social robotics serves humanity ethically and sustainably. Unlike methodologies such as "design for values" or "value-sensitive design," VISR embraces an interdisciplinary approach, recognizing that the ethical and societal implications of robotics extend beyond design into regulatory and societal domains.

The breadth of VISR reflects the multifaceted nature of social robotics, which intersects with diverse disciplines and societal layers including design and ethics (focusing on empathy and accountability in robot behavior during the development phase), governance and regulation (focusing on establishing legal frameworks to ensure accountability, protect intellectual property, and enhance public trust in AI technologies), user behavior (shaping responsible usage patterns and minimizing biases through education and engagement), and systemic integration (bridging gaps between the design and deployment of social robots to address long-term societal challenges like unemployment, data privacy, and equitable access). This integrative approach is necessary to address existing gaps in AI governance and the ethical deployment of robotics. Alternatives, such as design-for-values methodologies, while valuable, focus predominantly on design stages and lack comprehensive strategies for addressing post-deployment challenges.

3.1 VISR and ethical challenges in social robotics

To further elucidate the role of VISR, this section explores the specific ethical challenges it seeks to address and the rationale for the ART principles at its core. The development and deployment of AI-enabled social robotics present numerous ethical challenges that require immediate and comprehensive attention. Virtuous Integrative Social Robotics (VISR) addresses these challenges by focusing on issues such as opaque decision-making, accountability, bias, trust, and ethical agency in robots. Many AI systems operate as "black boxes," making their decision-making processes difficult to understand and evaluate, which undermines trust and accountability [2]. This lack of transparency becomes particularly problematic in sensitive domains such as healthcare, education, and law enforcement, where the stakes are high. Furthermore, the question of responsibility for the actions of autonomous systems remains unresolved, creating a gap that risks ethical violations and legal ambiguities [38]. Another significant challenge is the potential for bias and discrimination, as the absence of transparency in AI design and training data can perpetuate systemic inequities [18]. These concerns, coupled with the erosion of public trust in AI systems due to unclear guidelines and insufficient accountability, limit the societal acceptance and effectiveness of social robotics. Additionally, social robots often lack mechanisms to understand and apply moral considerations, which can result in unintended harm or ethical violations in human–robot interactions [15].

[2] https://www.nytimes.com/2023/03/29/technology/ai-artificial-intelligence-musk-risks.html.

VISR adopts the ART principles as foundational elements for addressing ethics related challenges. Transparency ensures that decision-making processes in AI are explainable and accountable, enabling the identification and correction of biases and fostering trust among stakeholders [13]. Responsibility emphasizes the need for clearly defined accountability across all phases of the AI lifecycle, ensuring that the actions and decisions of social robots align with societal values and ethical norms [36]. Autonomy, when guided by ethical principles and balanced with human oversight, allows robots to operate effectively and responsibly while adapting to diverse environments and ethical contexts [6]. Together, these principles form the core of VISR and provide the foundation for achieving broader ethical goals such as fairness, justice, and minimizing harm. Transparency, for instance, is essential for detecting and addressing biases, which supports fairness and prevents discrimination. Responsibility ensures that ethical norms are upheld, promoting justice in human–robot interactions, while autonomy minimizes harm by ensuring that robot behavior is ethically guided and contextually appropriate.

Respectful treatment of robots, particularly those designed with human-like attributes, serves a dual purpose. It fosters positive and virtuous interactions in humans, training them in empathy, respect, and other moral qualities. This co-developmental aspect of HRI is particularly significant in educational and caregiving contexts, where interactions with social robots can reinforce human ethical behavior. While robots themselves are not moral agents, their design and use can encourage moral development in human users. Dignum et al. [13] emphasize that values must be intentionally embedded into robots to guide their interactions, ensuring they align with the ethical goals of promoting human growth and social harmony.

In practice, VISR integrates these ART principles into the design, development, and governance of social robotics. Transparency and explainability are embedded into the design process to ensure that the functioning of social robots is comprehensible [2]. Responsibility frameworks are established to define clear accountability among stakeholders, from developers to users, ensuring adherence to ethical principles [38]. Autonomy is balanced with robust human oversight to prevent ethical lapses and ensure that robots act as virtuous agents within their operational contexts [15]. By addressing these specific challenges and integrating the ART principles, VISR provides a robust framework for designing and implementing social robotics in a manner that aligns with societal values and ethical considerations.

3.2 Positioning VISR within the AI ethics landscape

Virtuous integrative social robotics (VISR) emerges as a response to the ethical and societal challenges posed by the increasing use of social robotics. While many frameworks in AI ethics have sought to address similar issues, VISR distinguishes itself through its integrative and transdisciplinary approach. To contextualize VISR within the broader literature, it is necessary to compare it with established methodologies, such as Responsible AI (R-AI), Value-Sensitive Design (VSD), and Trustworthy AI, and to articulate its unique contributions.

Responsible AI (R-AI) emphasizes accountability, fairness, and inclusivity in the development and deployment of AI systems [16]. While these principles are crucial, R-AI often focuses on the ethical dimensions of specific use cases or domains without providing a comprehensive framework for addressing systemic challenges in social robotics. VISR builds on R-AI by embedding these principles into its broader framework, ensuring that accountability and fairness are upheld across the entire lifecycle of social robots, from design to governance.

Value-Sensitive Design (VSD) prioritizes the integration of human values into the technical design of AI systems, ensuring that these values are reflected in their functionality [17]. VISR expands on VSD by incorporating governance and regulatory considerations, recognizing that the ethical impact of social robotics extends beyond the design phase. By addressing the societal and legal dimensions of AI implementation, VISR offers a more holistic approach to embedding values in robotics.

Trustworthy AI focuses on building public trust by ensuring transparency, accountability, and fairness in AI systems [24]. While Trustworthy AI aligns with VISR's emphasis on trust and transparency, VISR goes further by explicitly integrating these principles with autonomy and responsibility. This integration allows VISR to address not only trust but also the ethical programming and societal integration of social robots.

VISR's unique contributions lie in its comprehensive scope, its emphasis on interdisciplinary collaboration, and its focus on phased implementation. Unlike other frameworks, VISR seeks to address systemic challenges by bridging gaps between technical innovation, ethical principles, regulatory mechanisms, and societal engagement. Its reliance on the principles of autonomy, responsibility, and transparency (ART) ensures that it is both ethically robust and practical for implementation [13]. Furthermore, VISR recognizes the importance of gradual adoption, beginning with pilot programs and stakeholder engagement to build trust and refine the framework.

The VISR framework aligns with and extends ongoing discourse in social robotics, particularly the Design for Values approach [13]. While Hafner [21] focuses on biologically inspired robotics and Prescott et al. [31] explore adaptive robot behavior in complex environments, VISR emphasizes the integration of ethical principles into robot design and governance. By combining ART principles with interdisciplinary collaboration, VISR seeks to address not only technical challenges but also the societal and moral implications of social robotics. VISR does not aim to replace existing frameworks but rather complements and advances them by providing an integrative approach that addresses the multifaceted challenges of social robotics. By situating VISR within the broader AI ethics landscape, we demonstrate its relevance and potential to contribute meaningfully to the ongoing dialogue on responsible and ethical AI development.

3.3 VISR and HRI

The 'Computers Are Social Actors' (CASA) paradigm suggests that humans interact with computers, including AI systems, as if they were social beings, applying similar social rules as in human-to-human interactions [27, 28] (p. 9). While this idea remains controversial and debated, it provides a useful lens for understanding how people attribute social characteristics to technology, which is critical in studying human–robot interaction. Increasingly, humans will use AI as a standard everyday tool. As a result, our dependency on AI, as well as AI robots, will deepen. VISR determines that the best mechanism for this task is the federal government that is charged with the *virtuous* duty of ensuring domestic tranquility, providing for the common defense, and promoting the general welfare; and thus should be responsible for finding a balance that prepares/protects society while also encouraging, not stifling, human ingenuity. Some predict that robots will become more common by 2050 [41]. There should be a noticeable increase in HRI observations, and some examples are as follows.

- Federal funding could support the states' use of robots in schools. Such programs could provide empirical data on how students interact with robots.
- Universities, trade schools, and colleges can provide training and research on robotics and robotic applications (including AI).
- Grants are available to small to medium enterprises that utilize robots to support and assist staff in front-line positions (again providing HRI data).
- The robots are increasingly used to support front-line services within federal and local government.

The HRI data collected from these and other AI- enabled initiatives can help determine how well society is adjusting and at what pace advancements in autonomous AI technology should be introduced.

3.4 VISR and virtue ethics

A central challenge in conceptualizing virtuous robots lies in the distinction between simulating virtue and embodying it. AI systems are fundamentally imitative, replicating patterns of human behavior and decision-making without intrinsic understanding or moral agency. This raises ethical questions about whether robots that 'appear' virtuous through programming can genuinely promote human moral development or risk reinforcing deceptive interactions under the guise of ethical behavior.

Who determines what is moral or ethical in VISR? Since all parties involved in AI will be licensed appropriately based on their level of involvement in accordance with VISR, we can better address ethical decision-making challenges. For example, social robotics for early childhood education may include ethical decision-makers in teaching, psychology, counseling, and education. VISR requires the participants to be screened and vetted, including their behavioral assessments and references. However, AI for general use should include properly screened ethical decision-makers where the diversity of the populace is represented. These ethical cohorts would serve in that role for a pre-determined term. The cohort's role is to test AI products for ethical concerns and approve the same before it goes to market, even at every new iteration of deep learning. Anderson and Anderson [3] proposed utilizing Ross [32]'s theory of *prima facie* duty in social robots to incorporate virtuous and moral principles of benevolence, autonomy, non-malevolence, or justice.

What values do VISR propose to endorse? This will also vary to match the intent and purpose of the AI applications. For example, AI used for social robotics in early childhood education may focus on values like sharing, manners, honesty, respect, and empathy. Whereas AI used for social robots in adolescent education may include gratitude, authenticity, leadership, and discipline. The cohort would ascertain if the AI application accurately reflected the societal and cultural norms approved by school administrators. This does pose the concern that the cohort could create unknown biases in

the AI application. AI is a tool used by humans and not a human replacement. However, biases may be unavoidable. Establishing diverse cohorts with term limits could help minimize biases. For example, robotic designers may develop deontological robots in open environments [18, 40] whereby the robots are programmed in advance to comprehend rules and rule-sets depending on varying situations. Since there is no complete repository of all moral rules / values; the challenges of utilizing VISR and virtue ethics are evident. Thus, we have the following research proposition.

Research proposition 1 For AI/robotics to be moral and ethical, society (We, the people) may determine how to utilize AI/robots, and which values should be prioritized in VISR using the ART (accountability, responsibility, and transparency) principles.

3.5 VISR and ART

By utilizing the ART principles within VISR, 'accountability' allows for trusted interactions, 'responsibility' underpins autonomy, and the data involved in how AI adapts showcases 'transparency' [13]. Multiple concerns are addressed, and there are several advantages of incorporating ART into virtuous robots [15, 19]. Some examples include:

- A mandatory government-issued license is required from every party within the AI supply chain, including the end user/customer.
- Society can trust that the person or organization involved in creating the AI application has met the appropriate guidelines and will be held accountable for violations.
- Industries can trust that industry standards are not being ignored.
- Autonomous does not mean autocratic.
- The deep-learning algorithm is being collected from licensed users' data.
- Any learned adaptation (iterative advancement) will not be entirely hidden.
- The government holds the spare key to all black boxes for use in the event of ethical concerns that require legal action.

VISR requires all parties to conduct themselves with honesty and integrity since humans treat AI, robotics, and AI-related emerging technologies with the same social rules as interacting with other human beings using the CASA paradigm. The linchpin that holds VISR together is government oversight. The threat that government can access black box content if deemed necessary is an additional incentive to self-govern and reign in impulses that negatively impact society. Deep fakes, misinformation, and questionably fraudulent actions would be deemed crimes under VISR. VISR is an emerging transdisciplinary field that merges governments, industries, and consumers in a legally binding commitment to deal ethically with each other. VISR can be extended to include all AI technologies, as well as virtual reality, augmented reality, and social media. We propose that any computer or internet application that captures, stores, or uses individual data for purposes of monetary consideration (buying or selling) products that can impact the general public incorporate this virtuously integrated structure. Thus, we have the following research proposition.

Research proposition 2 As humans become more reliant on AI technologies, robots are more likely to follow social rules using the CASA paradigm, and VISR needs to be seamlessly integrated for governments, industries, and consumers.

3.6 VISR and intellectual property

The first step in VISR is government regulation. However, before this occurs, there must be an infrastructure to support it. The answer to this lies within the current intellectual property laws (IP) regulations and processes. As previously explained, as AI becomes more advanced, it requires more oversight (human involvement, value ethics, regulations). Additionally, IP should be broadened to include protection for users. Undoubtedly, buried in the fine print authorizations, every user permits the use of their personal data in exchange for using the product or service. However, this individualized decision is detrimental to society.

How does VISR protect IP rights and society? The purpose of government is to protect society. It is unfortunate that many would not describe the government as virtuous. However, the purpose of government (to protect society) is to be virtuous [39]. The patent office could also issue license numbers to individual VISR users. This license number would be required whenever an individual authorizes the use of their personal data. Thereafter, the data's origins would be available if legal actions require opening an organization or company's AI black box. Additionally, the relatively easy free

individual licensing process could include lessons on personal data safety, such as recognizing scams and user rights. The fee-based licensing process for professionals, manufacturers, developers, and businesses would be much more detailed and specific to their part of the supply chain. The party reveals how they are funded within the application, as well as connections, partnerships, purpose, etc. Likewise, the professional license number will be required within any contractual agreements and available as needed. At this professional level, the licensee will be warned that the license can be revoked if determined that their actions were knowingly or willfully fraudulent or malicious. The professional license will require the licensee to attest to understanding the rules that govern AI and/or robotics development. There will be penalties for misuse, and fines can be substantial. Especially egregious violations could result in jail time. Whether individual or professional, licenses should be renewed periodically to ensure the validity of data origins as well as to consent to/acknowledge changing rules/standards and continuing education [1]. Future of virtuous integrative social robotics (VISR) is incumbent upon how humans design and develop robots and robotic systems with possibilities of making mistakes and to be able to differentiate (and reason) between good and bad, or virtuous over vicious actions [10]. However, the bigger question is: are humans prepared to interact with vicious robotic AI systems in the future? Thus, we have the following research proposition.

Research proposition 3 Individual VISR user agreements (terms and conditions) used to compile big data grafted into AI black boxes of organizations are likely to be detrimental to the good of society since these agreements violate the ART principle.

Research proposition 4 Future of VISR depends on how humans design and interact with AI, robotics, and robotic systems using the ART and CASA paradigms.

In this manuscript, we have provided four research propositions to denote specific theoretical and practical areas of inquiry that emerge from the VISR framework. These propositions are intended to inspire further academic research into the ethical, societal, and governance challenges of social robotics. While they may also carry policy or governance implications, their primary purpose is to highlight critical issues for scholarly exploration and empirical validation.

4 General discussion: VISR and ethical governance

Virtuous Integrative Social Robotics (VISR) is our proposed value-driven philosophy for developing and designing AI-enabled robots and robotic applications, and integrating elements of design for values, integrated social robotics, HRI, and virtues ethics. In this research, we conceptualize VISR and develop three research propositions in the VISR context. The United States Constitution, in Article I, Section 8, Clause 1, emphasizes the responsibility of governance to promote the general welfare of society [29]. While this does not constitute a direct legal mandate for moral robot development, it provides a philosophical basis for proposing that government should play a leading role in ensuring that emerging technologies, including social robotics, align with societal values and ethical principles. This alignment involves prioritizing initiatives that enhance quality of life and foster moral development, a goal that requires both ethical oversight and societal investment. Ensuring the well-being of humans remains in the foreground is critical and can be accomplished through the following VISR elements:

- *Utilizing CASA paradigm for robots* will bring awareness that as society becomes more reliant on AI/robotics, there is an inherent vulnerability that moves beyond the elderly, sick, or disabled.
- *Expanding HRI monitoring.* Currently, HRI is primarily designated for robot interactions with the elderly and children with learning disabilities. HRI monitoring and data collection should be expanded to provide insight into how to prepare for when robots will become more common in the not-too-distant future.
- *Gaining consensus on virtue ethics*. Using a diverse interdisciplinary cohort to determine which values should be endorsed when addressing ethical concerns will drastically reduce programming biases.
- *Infusing the ART principles into RDD.* Research, development, and design require accountability, responsibility, and transparency. The federal government will be the secure repository for spare black box keys.
- *Recognizing that intellectual property rights and protections apply to individuals and businesses*. Educating the public on how their data is collected and utilized for AI models and how AI operates within robots. This will help in

allowing legal action by individuals when their intellectual property (data) is misused. Establishing penalties for misuse will help build public trust.

The governance of AI/robotics and its economic impact has been a subject of interest for policymakers and researchers for many years. The European Union (EU) has been actively involved in developing and supporting strategic initiatives to improve the competitive situation and foster technological and economic growth. For instance, the European Robotics Research Network (EURON) was established in 2000 to promote research, education, and technology transfer in the field of robotics [14]. Additionally, the European Robotics Platform (EUROP) was founded in 2005 as an industry-driven initiative to strengthen the EU's competitive position in robotics research and development [12]. In terms of economic and societal impact, the increasing adoption of robots has been a focus of EU (macro) economic policy making. Previous studies have emphasized the potential effects of robots on employment and the positive impact of robotics on economic growth [25]. The concept of "robotic governance" involves establishing a regulatory framework to address issues related to intelligent and autonomous machines. It encompasses research and development activities as well as the handling of these machines. Efforts have been made to include robot ethics in robotics policy, drawing from Isaac Asimov's Three Laws of Robotics and aiming to develop a taxonomy of potential ethical issues. The Robotic Governance Foundation, an international non-profit organization, has been involved in realizing the impact of robotics, automation technology, and artificial intelligence on society from a holistic, global perspective [11].

Ethical governance is considered essential to building public trust in robotics and artificial intelligence systems. Thus, our conceptualized VISR is crucial to address the potential risks associated with AI and robotics through a multifaceted approach to governance. Private firms and governments worldwide are increasingly recognizing the need for structured AI governance to ensure responsible and ethical development and deployment of these technologies. VISR requires establishment of regulatory frameworks, addressing ethical considerations, and anticipating the economic and societal impact of increased robotization. VISR aims to ensure responsible and ethical development, and deployment of robotics and artificial intelligence technologies.

While Virtuous Integrative Social Robotics (VISR) offers an ambitious framework for addressing the ethical and societal challenges of AI and robotics, its success depends on its ability to integrate existing methodologies, ethical principles, and regulatory practices into a cohesive strategy. VISR's potential for success lies in its focus on foundational principles—autonomy, responsibility, and transparency—and its recognition of the interconnected nature of ethical, legal, and societal issues in social robotics. The framework builds on established efforts in AI ethics and governance, providing a structured approach to address gaps that siloed methodologies have failed to resolve. For instance, its emphasis on transparency enables stakeholders to identify and address biases, while responsibility mechanisms ensure clear accountability for ethical lapses. By integrating these principles into a single framework, VISR offers a pragmatic pathway for addressing systemic issues in social robotics.

However, we recognize that some of the proposed regulatory, legal, and commercial reforms will require significant effort and collaboration. VISR does not advocate for abrupt radical changes but rather envisions an incremental implementation strategy. This approach begins with pilot programs in targeted industries or regions, generating empirical evidence to refine the framework and demonstrate its feasibility. International efforts, such as the European Union's AI Act, provide a model for the gradual adoption of comprehensive regulations through stakeholder engagement and phased implementation. Practical limitations to VISR include technological biases, legal ambiguity, and societal resistance. For example, current machine learning models often encode biases present in training data [18]. VISR mitigates these issues by requiring bias audits and validation through diverse stakeholder panels. Legal challenges, such as the lack of clear liability frameworks for autonomous systems, are addressed by proposing updates to intellectual property laws and mandating licensing for all actors in the robotics supply chain. These measures ensure accountability while fostering public trust.

It is important to emphasize that VISR is not a panacea for all challenges in AI and robotics. Instead, it provides a structured framework to address key ethical and societal concerns, serving as a catalyst for creating actionable solutions through collaboration among governments, industries, and civil society. The VISR framework offers a transformative vision for social robotics by aligning technological innovation with societal values. For instance, robots designed for eldercare can reduce caregiver burnout while fostering trust and companionship. In education, robots programmed to teach empathy can help children develop social and emotional skills. These applications demonstrate the potential of VISR to address pressing societal challenges while promoting ethical engagement with technology. By integrating ART principles, Design for Values methodology, fostering dialogue, building trust, and ensuring adaptability, VISR provides

a roadmap for responsible and impactful robotics development and offers a viable pathway for navigating the complex landscape of social robotics.

Acknowledgements The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study received funding from the National Science Foundation (NSF). Data Availability Statement: Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Author contributions All authors contributed equally and reviewed the manuscript. A.M.: Conceptualization, Writing–original draft, Writing–review and editing. A.S.A: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing–original draft, Writing–review and editing. A.A.: Investigation, Supervision, Validation, Writing–original draft, Writing–review and editing. J.M.: Funding acquisition, Investigation, Resources, Writing–review and editing.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

1. Abbott R, editor. Research handbook on intellectual property and artificial intelligence. Cheltenham: Edward Elgar Publishing; 2022.
2. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access. 2018;6:52138–60.
3. Anderson M, Anderson SL. Machine ethics: creating an ethical intelligent agent. AI Mag. 2007;28(4):15–15.
4. Awogbemi O, Von Kallon DV, Kumar KS. Contributions of artificial intelligence and digitization in achieving clean and affordable energy. Intell Syst Appl. 2024;22:200389.
5. Bessen J. AI and Jobs: The role of demand (No. w24235). National Bureau of Economic Research. 2018.
6. Cappuccio ML, Sandoval EB, Mubin O, Obaid M, Velonaki M. Can robots make us better humans? Virtuous robotics and the good life with artificial agents. Int J Soc Robot. 2021;13:7–22.
7. Capasso M. Responsible social robotics and the dilemma of control. Int J Soc Robot. 2023;15(12):1981–91.
8. Clarke L. Call for AI pause highlights potential dangers. Science (New York, NY). 2023;380(6641):120–1.
9. Chui M, Issler M, Roberts R, Yee L. Technology Trends Outlook 2023. 2023.
10. Constantinescu M, Crisp R. Can robotic AI systems be virtuous and why does this matter? Int J Soc Robot. 2022;14(6):1547–57.
11. de Pagter J. From EU Robotics and AI governance to HRI Research: implementing the Ethics Narrative. Int J Soc Robot. 2023;1–15.
12. DG INFSO. EUROP—the European Robotics Platform, 2006, https://op.europa.eu/.
13. Dignum V, Dignum F, Vázquez-Salceda J, Clodic A, Gentile M, Mascarenhas S, Augello A. Design for values for social robot architectures. In: Robophilosophy/TRANSOR. 2018; pp. 43–52.
14. euRobotics. About euRobotics. https://www.eu-robotics.net/eurobotics/about/about-eurobotics/about-eurobotics.html. 2020; Accessed 15 Nov 2023.
15. Floridi L, Sanders JW. On the morality of artificial agents. Mind Mach. 2004;14:349–79.
16. Floridi L, Cowls J. A unified framework of five principles for AI in society. Harvard Data Sci Rev. 2019;1(1).
17. Friedman B, Kahn PH, Borning A. Value sensitive design and information systems. In: Himma KE, Tavani HT, editors. The handbook of information and computer ethics. Wiley; 2008. p. 69–101.
18. Gibert M. The case for virtuous robots. AI Ethics. 2023;3(1):135–44.
19. Gigerenzer G. Moral satisficing: rethinking moral behavior as bounded rationality. Top Cogn Sci. 2010;2(3):528–54.
20. Gualtieri L, Fraboni F, Brendel H, Pietrantoni L, Vidoni R, Dallasega P. Updating design guidelines for cognitive ergonomics in human-centred collaborative robotics applications: an expert survey. Appl Ergon. 2024;117: 104246.
21. Hafner VV. Cognitive maps in rats and robots. Adapt Behav. 2005;13(2):87–96.
22. Huang X, Yang F, Zheng J, Feng C, Zhang L. Personalized human resource management via HR analytics and artificial intelligence: theory and implications. Asia Pac Manage Rev. 2023;28:598.
23. Huang H, Rust RT. Artificial intelligence in service. J Serv Res. 2018. https://doi.org/10.1177/1094670517752459.
24. High-Level Expert Group on AI. Ethics Guidelines for Trustworthy AI. European Commission. 2016. https://ec.europa.eu/futurium/en/ai-alliance-consultation.

25. Jäger A, Moll C, Lerch C. Analysis of the impact of robotic systems on employment in the European Union-2012 data update. Update of Final Report. 2016.
26. Jiang Y, Li X, Luo H, Yin S, Kaynak O. Quo vadis artificial intelligence? Discov Artif Intell. 2022;2(1):4.
27. Liao S, Lin L, Chen Q. Research on the acceptance of collaborative robots for the industry 5.0 era–the mediating effect of perceived competence and the moderating effect of robot use self-efficacy. Int J Industr Ergon. 2023;95:103455.
28. Lee JER, Nass CI. Trust in computers: The computers-are-social-actors (CASA) paradigm and trustworthiness perception in human-computer communication. In: Trust and technology in a ubiquitous modern environment: Theoretical and methodological perspectives (pp. 1–15). IGI Global. 2010.
29. Leibowitz AH. Defining status: a comprehensive analysis of United States territorial relations. BRILL. 2023.
30. Nocentini O, Fiorini L, Acerbi G, Sorrentino A, Mancioppi G, Cavallo F. A survey of behavioral models for social robots. Robotics. 2019;8(3):54.
31. Prescott TJ, Vogeley K, Wykowska A. Understanding the sense of self through robotics. Sci Robot. 2024;9(95): eadn2733.
32. Ross WD. The right and the good. Oxford: Clarendon Press; 1930.
33. Rowthorn R, Ramaswamy R. Growth, trade, and deindustrialization. IMF Staff Pap. 1999;46(1):18–41.
34. Seibt J, Damholdt MF, Vestergaard C. Five Principles of integrative social robotics. In: Robophilosophy/TRANSOR; 2018. pp. 28–42.
35. Seibt J, Nørskov M, Andersen SS. Editors. What social robots can and should do: proceedings of robophilosophy 2016/TRANSOR 2016; Vol. 290, IOS Press. 2016.
36. Shneiderman B. Human-centered AI. Oxford University Press; 2022.
37. Smith A, Anderson J. AI, robotics, and the future of jobs. Pew Res Center. 2014;6:51.
38. Smith H. Clinical AI: opacity, accountability, responsibility and liability. AI Soc. 2021;36(2):535–45.
39. Taplin R, Editor. Artificial intelligence, intellectual property, cyber risk and robotics: a new digital age. Taylor & Francis. 2022.
40. Tonkens R. A challenge for machine ethics. Mind Mach. 2009;19:421–38.
41. West DM. Brookings survey finds 52 percent believe robots will perform most human activities in 30 years. Brookings.edu. 2018, https://www.brookings.edu/articles/brookings-survey-finds-52-percent-believe-robots-will-perform-most-human-activities-in-30-years/.