Emerging Risks from Embodied AI Require Urgent Policy Action

Jared Perlo
Centre for the Governance of AI
French Center for AI Safety (CeSIA)

Alexander Robey
Carnegie Mellon University

Fazl Barez
University of Oxford
WhiteBox

Jakob Mökander
Tony Blair Institute for Global Change
Yale Digital Ethics Center

Abstract

The field of embodied AI (EAI) is rapidly advancing. Unlike virtual AI, EAI systems can exist in, learn from, reason about, and act in the physical world. With recent advances in AI and hardware research and design, EAI systems are becoming increasingly capable across an expanding set of operational domains. While EAI systems can offer many benefits, they also pose significant short- and long-term risks, including physical harm, surveillance, and societal disruption. These risks require urgent attention from policymakers, as existing policies for industrial robots and autonomous vehicles are insufficient to manage the full range of concerns EAI systems present. To address this issue, this paper makes three contributions. First, we provide a taxonomy of the physical, informational, economic, and social risks EAI systems pose. Second, we analyze policies in the US, UK, and EU to assess how existing frameworks address these risks and to identify critical gaps. We conclude by offering policy recommendations for the safe and beneficial deployment of EAI systems, such as mandatory testing and certification schemes, clarified liability frameworks, and strategies to manage EAI's potentially transformative economic and societal impacts.

1 Introduction

Embodied AI (EAI) refers to artificial intelligence (AI) systems and agents that are grounded in the physical world and learn through perception and action [1, 2]. EAI systems can operate across diverse environments. For example, existing EAI applications can deliver packages [3], patrol public spaces as security guards [4], or care for humans in elder-care homes [5, 6]. EAI capabilities and domains are likely to expand significantly in the coming years [7, 8]—presenting both opportunities and risks for humans. While EAI systems already assist people with mobility impairments in navigating the world (e.g., autonomous cars), future systems could fill agricultural or manufacturing jobs as working-age populations decline. By augmenting and complementing human labor, EAI could foster economic development and prosperity [9]. On the other hand, EAI systems can inflict more immediate physical damage than virtual AI systems and may also cause social harm as humans form closer connections with these systems [10, 11]. See Figure 1 for a comparison of classical robots, agentic AI, and EAI.

Recent breakthroughs in AI capabilities—particularly those related to Large Language Models (LLMs) and Large Multimodal Models (LMMs)—have catalyzed unprecedented progress in EAI systems' ability to navigate and act in the physical world [12, 13]. At the same time, the rise of Vision-Language-Action Models (VLAs)—which cast control as next-token prediction over visual and linguistic tokens—opens the possibility for a "ChatGPT moment" for robotics, with sharp jumps

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Position Paper Track.

in capability and public awareness. Recent debuts of models like Gemini Robotics-ER and NVIDIA's Isaac GR00T N1 marked significant EAI algorithmic progress, even though these models are only slowly being paired with hardware advanced enough to translate virtual capabilities into real-world actions [13–15]. In the past few months, for example, EAI systems have completed half-marathons and shown the ability to unpack groceries with little prior context [16, 17], and open-source resources from companies like Physical Intelligence and Unitree could spur continued progress [18, 19].

Data acquisition—a bottleneck for EAI development due to the complexity of information needed to train models [20]—is being addressed through open-source datasets and cross-modality approaches [21]. Simultaneously, innovations in tactile sensing, LiDAR, actuators, and power systems are expanding the potential capabilities of EAI systems [22–24]. Progress in physical abilities and data collection lower barriers to creating high-quality models about how the world operates [25]. These world models involve perception, reasoning, and memory [26], and increasing EAI research funding could lead to more accurate world models and positive development feedback loops. EAI research is also emerging as a new frontier in geopolitics, as concerns about supply chains and national industrial policy become more salient [27, 28].

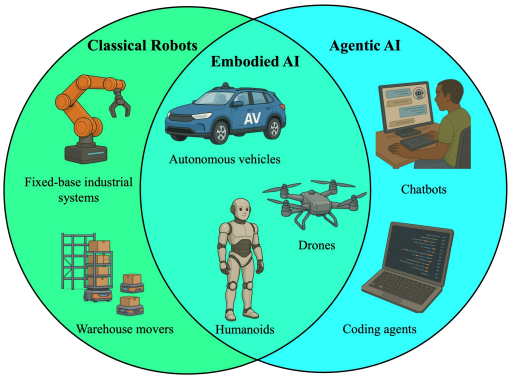


Figure 1: **Comparing classical robots, agentic AI, and EAI.** EAI combines the autonomy and reasoning capabilities of agentic AI with classical robots' physical embodiment.

EAI's advancing capabilities are giving rise to many short- and long-term risks. Although EAI shares many traits with virtual agents [29], such as varying degrees of autonomy and capability [30], the physical embodiment of EAI systems introduces distinct considerations that warrant special attention [31]. EAI systems can hit, cut, bump, attack, and more, whether intentionally or not. The physical world presents significant adaptation challenges for models trained in virtual simulations [32, 33]. Technological breakthroughs may also increase scalability and later enable self-improvement, allowing EAI to rapidly advance and face fewer human bottlenecks [34]. Social, legal, and economic systems will likely require significant updates [8, 35].

Before rushing to policy action, it should be noted that EAI is not new but an evolution of traditional robotics. The EAI field builds upon decades of science fiction imagination, human-robot interaction research, and forecasting about advanced robotics [36–38]. In fact, the term "embodied AI" is partly a marketing technique used to differentiate recent innovations from traditional robotics.

Safety concerns about EAI are likewise not novel, as researchers have studied safety in robotics for decades [39, 40]. Tools to formally verify robot behavior have included model predictive control [41], control barrier functions [42], and temporal logic [43]. Many papers focusing on safe AI design prominently feature imaginary robots as examples of human-AI collaboration [44, 45]. More recent work has focused on creating safety guardrails for robots from real-world data, such as in Sermanet et al. [46]. Still, beyond a UN resolution initiating lethal autonomous weapon discussions [47], there largely remains a policy vacuum regarding EAI safety at national and international levels.

Understanding and minimizing risks from EAI will become even more critical in a world with AI capabilities equivalent to or surpassing artificial general intelligence (AGI), however defined [48]. AGI uncertainties aside, EAI risks are critically understudied and poorly understood, and current regulatory frameworks are generally insufficient to guide safe EAI development. This paper clarifies the risks and governance challenges posed by EAI and suggests a pragmatic sociotechnical approach to help governments and researchers support the development of safe EAI [49]. This paper makes three unique contributions to address this urgent issue:

1. We develop a comprehensive taxonomy of risks from EAI, spanning physical, informational, economic, and social dimensions. This taxonomy covers concerns ranging from malicious physical harm from jailbreaking LLMs and privacy violations in homes to widespread labor displacement. To create this taxonomy, we draw on the literature related to robot safety, human-robot interaction, and recent predictions about AI's trajectory.

2. We analyze existing EAI policy frameworks to assess their adequacy and highlight critical gaps. Although specific pieces of legislation governing autonomous vehicles or advanced robotics trend in the right direction, significant gaps remain. For example, current robotic regulations are ill-suited to govern systems that have high levels of autonomy and continuous learning; these characteristics challenge existing safety testing and assurance paradigms.
3. With these risks in mind, we discuss several policy interventions to improve EAI safety. We suggest increasing targeted safety research, establishing robust certification requirements for EAI, promoting industry standards, clarifying liability regimes, and creating actionable policy blueprints to respond to transformative economic and social effects of EAI.

This paper has several limitations. We focus on civilian applications of EAI, although military applications also merit consideration. We focus on frameworks from the US, UK, and EU, though emerging regulatory efforts in other regions deserve increased attention. Ensuring safe EAI will also require a multi-layered approach, with mechanisms to enhance safety at the model, application, and organizational levels [50]. While it remains crucial to ensure the safety of underlying models through AI safety research, we focus on strengthening safety measures for EAI-specific applications. Acknowledging this context, we aim to provide a solid foundation upon which future work can build.

In the coming years, policymakers may quickly become aware of the risks posed by EAI because of headline-grabbing breakthroughs. This could rapidly elevate EAI regulation on policy agendas, so policymakers must be equipped with appropriate context to create clear and beneficial legislation. To ensure the safe and beneficial development of EAI, we argue that policymakers must urgently build upon and address gaps in existing frameworks for robotics, autonomous vehicles, and agentic AI.

2 Taxonomy of Risks from EAI

Drawing on existing research and predictions about EAI trajectories, we identify four crucial areas of EAI risks: physical, informational, economic, and social (see Figure 2). This taxonomy leads to our discussion of how existing policy frameworks address—or fail to address—these EAI risks.

2.1 Physical risks

Purposeful or malicious harm. EAI systems present several distinct physical risks. EAI systems have been designed and deployed with lethal intent, such as AI-controlled drones [51, 52]. However, fully autonomous military robots, often integrated with bespoke AI architectures [53–55], are not yet widely used in combat. Commercially available EAI systems, including AI-controlled quadrupeds and autonomous driving assistants, also present serious risk. Recent research has demonstrated that these systems inherit *jailbreaking* vulnerabilities from LLM-based AI models [56–59]. This could allow malicious actors to subvert safety guardrails and perform a range of harmful physical tasks, including detonating explosives and causing collisions [60–62]. VLAs exacerbate this risk: an attacker might craft a visual scene or textual instruction that, when interpreted through a language-action policy, yields physically dangerous instructions not anticipated by vision- or language-only defenses [63, 64].

Accidental harm. Automation in sectors ranging from manufacturing to healthcare will put humans into close contact with EAI systems [7]. This interaction increases the risk of accidental physical harm. Several recent reports document an increase in industrial injuries following the introduction of AI-controlled robots [65–67]. EAI systems could accidentally cause physical harm through misspecified goals, physical hardware malfunctions, or other unanticipated behaviors [44, 68, 69]. For example, a humanoid EAI might not correctly reason that placing a full glass of milk on a tilted table is perilous and likely to lead to a dangerous broken glass [70]. Researchers also face persistent difficulty in getting models trained in purely virtual simulations to act as intended in the real physical world—what is referred to as the "reality gap" [71]. This introduces significant scope for accidental harm if the deployed world does not closely match an EAI's training data [72].

2.2 Informational risks

Privacy violations. EAI systems are often trained on vast corpora and process a variety of data modalities—spanning visual, auditory, and tactile information—during deployment, creating significant privacy concerns [12]. Like text-based virtual AI models, which are known to memorize and expose personally identifiable information [73, 74], commercial robots have been shown to disclose

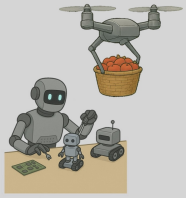
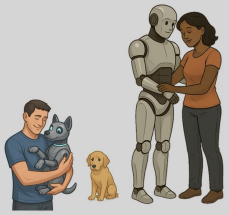
Physical		Informational	Economic	Social
				
Key Risks	Purposeful harm, accidental harm	Privacy concerns, misinformation	Labor displacement, socioeconomic inequality, power concentration	Lack of trust & transparency, unclear liability standards, bias & discrimination
Example Mechanisms	Jailbreaking, sensor spoofing, hardware malfunctions	Unauthorized surveillance, non-consensual data usage, hallucinations	Recursive EAI design, rent-seeking behavior, uneven model diffusion	Human-robot romance, EAI therapists, doctors, etc., hiring bias toward EAIs

Figure 2: **A summary of risks from EAI**. We identify four key risk categories and provide several existing or potential mechanisms by which EAI systems could cause harm within each category.

proprietary information through simple prompts [60]. EAI's mobility and the vast array of sensors used in EAI technologies expand concerns about unauthorized data collection. For example, EAI systems can monitor user behavior, infer physical preferences, and contribute to future model training without the consent of those being observed beyond the limitations of immobile microphones or security cameras [75–77]. Bad actors could gain access to private data streams and monitor users' movements, providing leverage over individuals to squash dissent or seek power [78].

Misinformation. Several studies have shown that LLMs often hallucinate information [79–81]. EAI systems inherit these shortcomings in the physical world, answering user questions with deceptive or incorrect information [82]. Because VLAs fuse vision and language, their hallucinations can be spatially grounded—e.g., misidentifying an object in view and then generating a plausible yet unsafe plan around it. And although automated assistants like Amazon's Alexa are known to lie about issues as innocuous as Santa Claus [83], more capable and trusted EAI systems in sensitive positions (like home-assistant or community-service positions) could easily spread model developers' propaganda to users. For example, an EAI running on DeepSeek's latest model could provide a subtle yet continuous stream of misinformation to American users [84, 85].

2.3 Economic risks

Labor displacement. While virtual AI applications will likely displace certain types of human cognitive labor, EAI systems could significantly replace or displace physical human labor [86]. At a minimum, EAI will likely augment the type of work that humans perform [87, 88]. Classical industrial robots have taken over many human roles in manufacturing [89], and research has shown that robot deployment can reduce human employment [90]. Future technological advances will likely accelerate this displacement, as increasingly capable EAI systems perform complex physical tasks beyond assembly lines. Though automation has historically redirected labor toward areas of human comparative advantage [91], AGI-enabled EAI could potentially automate all physical labor [92].

Socioeconomic inequality. EAI could significantly exacerbate wealth inequalities. Those who have access to or own EAI systems will be able to automate labor and perform tasks significantly better or faster than those without access. These significant productivity advantages will potentially concentrate wealth and exacerbate domestic and international inequality [93, 94]. For example, while a wealthy businesswoman could invest in a fleet of the latest humanoid robots, individuals lacking adequate capital might be forced to rent their EAI systems [95]. Virtual AI applications may cause similar socioeconomic inequality, but the ability to control access to EAI systems may confer unique returns on investment, given that many physical tasks necessary for human survival (e.g., growing food, building shelter) are constrained by human strength and energy.

Power concentration. EAI deployment could accelerate the consolidation of economic and political power. By unlocking increasing returns to capital, EAI will decrease employers' reliance on the needs of human labor [96]. EAI users or consumers may become dependent on EAI owners for goods and services due to relevant productivity advantages [97, 98]. The importance of EAI to perform physical tasks will likely exacerbate power-concentration risks presented by purely virtual AI systems. The proliferation of EAI systems could thus lead to a rapid concentration of corporate economic (and social) power, potentially even facilitating an eventual coup involving EAI [78, 99].

2.4 Social risks

Bias and discrimination. Like virtual applications of AI, EAI can discriminate against users. When EAI systems are placed in positions of power, their biases could have significant impacts on fairness in everyday interactions [100, 101]. For example, a peacekeeping humanoid robot may discriminate based on skin color [102]. Unlike virtual AI applications, this bias can have immediate and irreversible physical consequences (e.g., if the peacekeeping robot mistakenly injures an innocent passerby).

Lack of accountability and liability. Determining liability when EAI causes harm requires new frameworks that address the complexities of autonomous physical systems. Human users may disagree with decisions taken by EAI systems, raising questions of delegation and responsibility [103]. Lack of EAI accountability could lead to confusion for users and breakdowns in traditional justice systems [104]. For example, we may soon need to consider who to blame and how to collect damages when a highly autonomous robotic surgeon removes a healthy organ by mistake [105].

Lack of transparency and explainability. Understanding how AI systems reason and why they perform specific actions motivates the field of interpretability research [106]. But physical embodiment raises the stakes for understanding these systems. For example, transparency of planned actions and explainability of decision-making is crucial when an AV suddenly changes lanes. A lack of transparency could lead to a lack of trust, which could become a socially destabilizing issue with widespread EAI deployment [107–109].

Unhealthy human-EAI relationships. Interactions with EAI systems could foster dangerous human dependence or romantic attachment [110]. People may depend on EAI systems for physical pleasure [111]. The presence and human-like features of EAI systems may amplify the dependency issues already observed with conversational AI [112, 113]. People may easily fall in love with EAI systems, only to be distraught when these systems are altered or have their memories reset [114].

Transformative effects. EAI deployment could reshape society, particularly if technological development outpaces society's ability to adapt [98, 115]. For example, EAI systems could provide physical threats of violence or mass surveillance to back up AI-enabled authoritarianism [116]. Businesses might prefer to employ EAI systems, affecting how humans find meaning in their work [117]. Humans might also lose the ability to perform various jobs as tasks are delegated to EAI systems [96].

3 Heat map of relevant policies

Existing policy frameworks address many risks identified in §2. Understanding how current regulations apply—or fail to apply—to EAI systems is essential for both policymakers and researchers. This section examines key policies from the US, UK, and EU that govern related technologies, including classical robotics, autonomous vehicles, and virtual agentic AI. Our analysis identifies regulatory gaps specific to EAI by examining where existing frameworks provide minimal, adequate, or substantial policy coverage. This non-exhaustive review focuses on civilian applications of EAI.

3.1 Key policies

This section first maps policy frameworks that govern physical harms, focusing on laws and standards that apply to AVs and robots, before examining major pieces of legislation that address informational, economic, and social EAI harms.

Physical risks. Existing governance approaches to physical risks posed by EAI mainly target AVs and drones. AV-specific laws follow one of two pathways: creating bespoke legislation or adapting conventional automobile laws [118, 119]. For example, the UK's Automated Vehicles Act 2024 introduced the concept of the Authorized Self-Driving Entity (ASDE) and the No-User-in-Charge

Table 1: **Coverage of policies for major EAI risks.** We examine whether existing governance frameworks address risks from technologies related to EAI. ● indicates that there is a high level of coverage of relevant policies; ◑ indicates there is partial coverage but that significant adjustments are necessary; ○ indicates a significant lack of governance frameworks to address the relevant risk. We reference AVs rather than broader EAI, as most EAI regulations to date have addressed AVs.

Risk	**Subrisk**	Classic robots	AVs	Virtual agents
Physical	Purposeful or malicious harm	●	◑	◑
	Accidental harm	●	●	◑
Informational	Privacy violations	◑	◑	◑
	Misinformation	◑	○	◑
Economic	Labor displacement	○	○	○
	Socioeconomic inequality	○	○	○
	Power concentration	○	○	○
Social	Bias and discrimination	◑	◑	◑
	Accountability and liability	◑	◑	◑
	Trust and transparency	◑	◑	◑
	Human-EAI attachment	○	○	◑
	Transformative societal effects	○	○	○

(NUiC) operator. These entities (e.g., manufacturers or fleet operators) assume legal liability when the vehicle is in self-driving mode, effectively standing in for the "driver" [120]. These ASDE and NUiC entities could serve as precedents for other forms of highly autonomous EAI; however, these roles largely fit neatly into existing automotive regulatory structures. Different forms of EAI—such as home care or educational EAI—will not have the same pre-existing foundation.

In the US, the recently proposed ADS-equipped Vehicle Safety, Transparency, and Evaluation Program (AV STEP) would require approved AV manufacturers to share details about vehicles' development and operation. This would include information about simulations used to train AVs' algorithms, environments in which the AVs are designed to operate, and oversight mechanisms to ensure operational safety [121]. AV STEP is a promising framework that could be extended to other EAI contexts, though it is unclear which regulatory body would oversee other EAI modalities [122].

Laws concerning aerial drones are also relevant to EAI, although many of today's drones operate with limited autonomy [123]. EU regulation on drones distinguishes between remotely-piloted drones, fully autonomous drones, and drones that fly pre-planned routes [124]. The EU requires that all drones are operated by a remote pilot who assumes responsibility for each flight. Pilots of autonomous and other high-risk drones must pass knowledge and practical-skill courses. However, National Aviation Authorities often depend on pilots' self-reporting to verify that drone operations pose minimal risk, for example by confirming pilots avoid flying over people "uninvolved" in the drone's operation. Non-autonomous drones require less stringent training to operate but face operating restrictions; for example, pilots of these drones must always keep the aircraft within their line of sight.

Another key piece of legislation is the EU's Machinery Regulation (MR), passed in 2023 [125]. Updating similar legislation from 2006, the MR regulates many types of robots in the EU and explicitly addresses aspects of AI and EAI safety, as discussed by Tobias Mahler [126]. The MR, which addresses issues ranging from emergency-stopping systems to the risk of being trapped inside a device, mentions machines with "self-evolving behaviour." The MR mandates that machines sold in the EU must be tested for compliance with these safety regulations; as with other EU regulations, third-party evaluators (or "notified bodies") test whether machines fulfill the safety requirements.

International standards also provide manufacturers with robust guidance for robotics. ISO 10218:2025 recommends safety protocols for assessing risk (e.g., through controls, safety and stopping functions) and safe-design certification for industrial robots [127]. In addition, ISO standards on service robots emphasize sensor reliability, uncertainty management protocols, and decision verification through multiple sensing modalities [128]. Many standards for AVs also already exist. For example, ISO/SAE 21434:2021 and UN Regulation 155 provide cybersecurity standards for road vehicles [129, 130]. These standards are particularly relevant for many forms of mobile EAI, especially those deployed in situations without regular access to trusted networks [131] (e.g., during disaster rescue missions).

Informational risks. Many informational risks apply to both virtual and embodied AI. Key frameworks governing EAI informational risks include the EU's AI Act and the General Data Protection

Regulation (GDPR). The AI Act prohibits several data practices relevant to EAI, such as untargeted facial image scraping and manipulative decision-influencing techniques [132].

GDPR legislation in the EU and UK establishes strict governance requirements concerning the capture, use, and storage of data. GDPR requires that data is only collected for "legitimate interests" and that entities collecting data are classified as "data controllers" who must document data collection practices while implementing robust security measures [133]. However, EAI deployments in public spaces challenge traditional consent models and controller identification [134]. EAI systems deployed in public spaces, for example, raise questions about who receives and controls data from system sensors, how to opt out of data collection, and even what constitutes a public vs. private space [135].

Economic risks. Regulatory frameworks to manage EAI's economic impact remain underdeveloped. Existing legislation, like the UK's Employment Rights Act 1996 and the US WARN Act, provides limited protections for workers facing technological displacement [136, 137]. Labor organizations have achieved isolated victories, exemplified by the International Longshoreman Association's successful challenge to port automation [138]. Policymakers must prepare for labor displacement resulting from EAI deployment in industrial settings, keeping in mind that EAI innovation could lead to a period of rapid creative economic destruction [139].

Some observers think AI development represents a different kind of technological transition compared to previous transformations, as AI may replace cognitive tasks in addition to physical labor [8, 140]. Economic policies may not need to target technological failures, as with many of the physical and informational risks mentioned in §2. Economic risks may instead emerge because EAI systems work *too* well, rapidly upending labor structures and reducing the need for human labor. As a result, policymakers should consider social policies to manage these emerging tensions [141].

Social risks. Few regulations directly address the social impacts of EAI. Those that do largely govern issues of direct human interaction with EAI and do not address larger issues of how society will transform as these systems become increasingly prevalent and powerful. The EU AI Act's broad prohibition on infringing fundamental rights could be extended to address EAI issues such as lack of trust, lack of transparency, unhealthy attachments, and bias and discrimination, but this would require further specification. In terms of accountability, proposed frameworks for attributing actions and delegating authority to virtual agents could prove helpful for EAI [142, 143].

GDPR Article 22 provides an instructive example of existing regulation that implicates but does not directly address EAI systems. Designed with virtual AI in mind, the Article prohibits individuals from being "subject to a decision based solely on automated processing" when that decision has legal consequences for that individual [133]. Yet it remains unclear how this Article could be reconciled with fully autonomous EAI, or how individuals could appeal to a human intervener—as the Article mandates—in immediate physical interactions or conflicts with EAI systems.

Beyond legislation, international standards for manufacturers and developers emphasize transparency, ethical design, and trustworthiness in EAI systems. But these standards, including the IEEE's 7000 series on autonomous system transparency [144], algorithmic bias [145], and the impact of robotics on human well-being [146], are voluntary. These standards could apply to EAI applications too, but their voluntary nature could similarly limit their impact.

3.2 What are the most significant gaps?

Though major building blocks to address harm from EAI systems already exist, several key policy gaps concerning EAI safety require urgent attention.

First, there are many open questions concerning robust certification for different EAI modalities. Regulating AVs is relatively straightforward due to defined operational domains (e.g., cars usually stay on roads). Future EAI systems, however, will likely have expanded freedom of movement, enabling them to enter residences and conduct surveillance in schools or public areas. This will require new processes to certify EAI systems' safety, and such frameworks do not yet exist. Expecting existing consumer-safety labs, which currently test the safety of machine components, to evaluate the safety of wider EAI systems is unrealistic. Basic questions such as identifying the relevant regulator are a key starting point—to what extent should there be EAI-specific consumer protection boards with AI expertise, or should existing third-party testing laboratories take on this responsibility?

Secondly, once a suitable apparatus is in place, EAI capabilities should be measured with reliable evaluations. To date, few of these benchmarks exist [46], despite the existence of many benchmarks for virtual AI systems [147]. Evaluations could cover a range of considerations outlined in our taxonomy, e.g., evaluating the robustness of the reality gap, defenses against jailbreaking, alignment between software and hardware capabilities, and hardware durability, among other areas.

Thirdly, policies currently devoted to post-deployment EAI monitoring lack detail. Such oversight mechanisms have been proposed for other AI systems [148]. However, current regulations requiring EAI systems to include "black boxes" that record and preserve data in the event of accidents or misuse are hazy. The EU's MR mandates data about safety-related decision-making processes is kept for a year after collection [125]. At the same time, the bloc's AI Act requires high-risk AI systems to retain this information for at least six months [132]. Such recording systems, in addition to live data monitoring, can enhance system safety and aid post-incident investigations [149, 150]. The EU's MR also states "it shall be possible at all times to correct the machinery...to maintain its inherent safety," but this oversight effort requires clarification for highly autonomous systems. Does this notion require the ability to tweak EAI actions in real-time, send model updates over-the-air, or another intervention? Are users, the government, or the manufacturer meant to perform this monitoring?

There are also gaps in policies addressing economic and social effects from EAI. Although EAI could cause labor displacement, proposals to distribute economic proceeds are still in their infancy [151, 152], as are national and regional proposals to ensure that world economies have sovereign systems (e.g., data centers, energy production, EAI hardware) to capture EAI's benefits. Policies addressing social issues related to trust and human-EAI attachment are likewise currently scant [153]. More broadly, there are significant policy gaps at the intersection of EAI and potential AGI. For example, should an EAI system be allowed to build other EAI systems? Should a country developing AGI in embodied form automatically and freely share the technology with the rest of the world? Policymakers must consider how EAI should be developed, deployed, and integrated into societal structures to address the broad array of neglected challenges mentioned here.

4 Proposed pathways forward

4.1 Invest in EAI safety research

We recommend increased research be devoted to EAI safety based on the risk taxonomy described in §2. For example, robotics and machine-learning researchers can further efforts to make hardware actuators less susceptible to hacking and malfunction through physical design and formal methods [154]. Building benchmarks and evaluations of EAI capabilities and behavior is a particularly promising area of EAI safety research. Most AI benchmarks today specifically target the virtual aspects of AI [155], although recent progress has been made in EAI research [46]. Researchers should and build on this progress by developing EAI evaluations and benchmarks that span a broad set of tasks and task types [156, 157], similar to the work being done by the RoboArena team [158]. Beyond physical risks, benchmarks and evaluations should also address issues related to privacy and cybersecurity, for example by building upon zero-knowledge proof research from other AI domains [159]. We also need benchmarks that stress-test the joint vision–language–action loop—measuring, for instance, whether a VLA model's visual prompt leads to safe, context-aware behavior across edge cases. Benchmarks and assessments will not address every risk raised above—particularly socioeconomic considerations—but they are a critical step towards minimizing many risks from EAI.

4.2 Create robust certification requirements before EAI deployment

National bodies should mandate that EAI systems pass safety evaluations and are certified for public use. EAI systems should have clear "model cards" describing how they were trained, in which domains they were designed to operate, and what safety measures the manufacturer has taken to ensure safe operation. Policymakers could then mandate that EAI systems be limited to legal operation within the specified domains, potentially aided by remote identification requirements similar to those for drones. This model card approach could borrow from the frontier safety frameworks that many leading AI labs have implemented [160]. This regime could be enforced via audits of EAI manufacturers and developers [161]. Policymakers should also ensure that this certification regime incorporates different categories of requirements based on potential risk [48]. For example, certifying EAI safety for children's toys or an autonomous limousine should involve different safety

testing requirements. EAI regulation should address concerns at the developer, model, and application layers, much like approaches for non-embodied AI [161]. Combining model- and application-specific approaches with policy efforts at the developer layer could help ensure robust and durable EAI safety.

4.3 Promote industry-led standards to address EAI risks

Standard bodies should push forward EAI safety efforts in tandem with legislative approaches. These standards bodies should both update existing standards and create new, dynamic standards for highly autonomous EAI. Existing standards are grounded in today's robotic capabilities, and even recent updates fail to address how EAI systems capable of advanced reasoning will affect real-world applications. In May 2025, the ISO announced promising intentions to create a new standard for humanoid robots, which should encompass additional form factors, notions of autonomy, and use cases [162]. Given EAI's fast-evolving nature, industrial actors can leverage their technical expertise to help develop and adjust standards more rapidly than international standards-setting bodies [163]. These standards should address technical protocols (e.g. for cybersecurity), mandate "black boxes" that record sensor input and reasoning preceding adverse events, and span the entire EAI continuum—from components to systems and swarms [164–166]. The tension between accountability and privacy raised by recording systems should be acknowledged and addressed in future standards.

4.4 Clarify liability regimes for fully autonomous systems

Policymakers should clarify existing, muddy liability regimes. When autonomous EAI systems are deployed in the near future, who should be held accountable for injuries or misuse? Should the person who gave the model its latest instruction be at fault, or should the blame rest with the manufacturer? In the longer-term, how should EAI systems themselves be held accountable for faults if they are considered fully automated and agentic? EAI liability is a growing area of legal study [37, 167–169], but firm policies need to replace current ad-hoc legal approaches. For instance, policymakers should clearly define notions like the Authorised Self-Driving Entity laid out in the UK Automated Vehicles Act to designate responsibility for EAI operation. Policymakers must also manage the tension between fully autonomous physical systems and GDPR Article 22's recourse to human intervention. For example, policymakers could prohibit EAI deployment in situations where such recourse to human decision-making would be impossible or outline scenarios in which humans do not have this Article 22 right. At the same time, policymakers must work with technologists to determine when a manufacturer should be held accountable for errant EAI actions (e.g. when training is deemed insufficient for deployment in specific environments, or when new models are released but manufacturers do not make safety-relevant over-the-air updates available).

4.5 Plan and prepare for the transformative economic and social effects of EAI

Policymakers at the national and international levels should draft legislation to prepare safety-net or assistance programs for people whose labor is replaced by EAI systems. Basic proposals have been floated concerning UBI [170], or even universal basic compute (UBC), whereby people are guaranteed access to and use of AI or EAI systems [171]. However, these proposals remain very sparse and abstract. Policymakers should create draft frameworks and attempt to form early consensus now, as highly advanced EAI models and widespread labor displacement may arrive in the near future. Policymakers should specifically address who will be eligible to claim these social assistance packages and under what conditions (e.g. what type of proof will be required to demonstrate that an individual lost their job as a direct result of EAI automation). Reskilling programs are another potential policy avenue; however, these worker retraining programs may face limitations in the face of AI and EAI that automate an increasing number of jobs [172].

Similarly, policymakers must better prepare for transformative social effects [96]. Given the capital-intensive nature of EAI systems, it is plausible that EAI power and access could be concentrated in the hands of a select few [98]. Policymakers should draft options to combat this social power concentration, perhaps through targeted taxation mechanisms [173–175]. Policymakers should also fund research on mitigating adverse emotional dependencies between EAI systems and humans. EAI deployment is ultimately (for now) a human decision, so policymakers should consider whether some domains should be entirely off-limits for EAI interaction. Organizations such as the OECD, GPAI, or the nascent UN AI Panel and Dialogue should prioritize action on these pressing social issues, as EAI will impact people worldwide, not just in today's robotics hotspots.

5 Discussion and Limitations

Our analysis has several limitations, and key counterarguments to our main claims warrant further attention. We hope this discussion helps pave the way for future work on this exciting topic.

Geographical limitations. We focus on policies from the US, UK, and EU given the authors' location and expertise. However, future analyses should extend to other geographies, particularly China, India, Japan, and Korea. Conversations about EAI safety should involve robust global representation, as notions of safety shift across geographies and cultures [176].

Application limitations. This paper primarily focuses on civilian applications of EAI; however, military and law enforcement EAI applications also demand urgent policy action. The development and deployment of weaponized EAI systems are expected to continue growing over the coming months and years [177]. This could markedly lower barriers for non-state actors to cause significant damage and for political leaders to suppress dissent with fleets of embodied agents [78].

Methodological challenges. EAI is a vast and rapidly growing field, and we omit many exciting areas of discussion due to space limitations. For example, the potential for high-fidelity virtual simulations and video-based learning to revolutionize EAI training merits its own review [70]. We also recognize that the EAI field is inspired by decades of engineering, computer science, and human-robot interaction research. We had to make difficult decisions in creating our main categories of risk, which appeared in conversations with experts and in our literature review. We hope these categories serve as a starting point for more robust discussions about key areas of EAI risk.

Market forces and societal pressures. We recognize that good policy does not always mean more regulation; active intervention may not be necessary if market or social pressures naturally lead to safer EAI. For example, manufacturers or countries with strong safety protocols might be seen as having a commercial advantage [178]. However, as with virtual AI systems, market forces may lead to race dynamics, less government oversight, and exacerbated risks [179].

Technical solutions and EAI harm. Technical solutions (e.g., alignment, unlearning, etc.) will play a crucial role in mitigating EAI risks. However, technical solutions may only be established and implemented effectively after EAI risks emerge at scale, as technical solutions are often responsive to demonstrated needs [180]. Technical solutions will also likely not address all the risks mentioned above—especially those related to economic and social risks. A more balanced and pragmatic approach is required, one that combines the best aspects of technical and non-technical solutions [49].

Overlap of EAI risks with LLM risks. Skeptics could argue that if we make LLMs safe, we will solve EAI risks. We acknowledge there is considerable overlap between the relevant risks [181], but many EAI risks are modified or magnified at the application and organizational level [50]. EAI creates immediate and pervasive risks by virtue of existing in the physical world that could grow with scale. For example, an EAI could impair a baby's healthy development [182] or be hacked and cause a deadly crash in ways inapplicable to purely virtual AIs [183]. Improving LLM safety alone is helpful but would fail to cover these—and many other—critical EAI scenarios.

6 Conclusion

The EAI field is rapidly advancing, driven by increasing hardware investment, breakthroughs in LLMs and LMMs, and quickening deployment. These trends will likely accelerate in the coming years. However, policymakers around the world have thus far neglected EAI governance even though associated risks have gradually transitioned from the realm of science fiction to the real world.

We have argued that policymakers should encourage the development of effective benchmarks, evaluations, and safety protocols for the responsible deployment of EAI; ensure safety certification for a range of EAI form factors, capabilities, and operational domains; reevaluate liability paradigms; confront labor displacement; and address larger societal issues, such as human-EAI attachment. These recommendations provide first steps that can guide and encourage safe EAI innovation with minimal downside. Many other risks necessitate minor tweaks or adjustments to existing policies—for example, preventing privacy violations from EAI systems in public will require integration with existing laws such as the MR and GDPR. Zooming out, research funding for EAI safety should be expanded, and policymakers should collaborate with AI researchers and EAI developers to translate research findings into policy advice.

Acknowledgements

Jared Perlo is grateful to the Centre for the Governance of AI, where he was a Winter Fellow, and the Centre pour la Sécurité de l'IA for making this research project possible. We would like to thank Aaron Prather, Aidan Homewood, Amelia Michael, Connor Aidan Stewart Hunter, Edward Kembery, Jenny Read, Keegan McBride, Krzysztof Bar, Kyler Zhou, Liam Patell, Markus Anderljung, Marta Ziosi, Nora Amman, Pierre Sermanet, Roeland P.-J. E. Decorte, Shaoshan Liu, Simon Mylius, Todor Davchev, Umair Siddique, Yohan Mathew, and Yunzhu Li for their valuable input and collaboration.

References

[1] Giuseppe Paolo, Jonas Gonzalez-Billandon, and Balázs Kégl. A call for embodied AI, September 2024. URL http://arxiv.org/abs/2402.03824. arXiv:2402.03824 [cs].

[2] Yang Liu, Weixing Chen, Yongjie Bai, Xiaodan Liang, Guanbin Li, Wen Gao, and Liang Lin. Aligning Cyber Space with Physical World: A Comprehensive Survey on Embodied AI, August 2024. URL http://arxiv.org/abs/2407.06886. arXiv:2407.06886 [cs].

[3] Xueping Li, Jose Tupayachi, Aliza Sharmin, and Madelaine Martinez Ferguson. Drone-Aided Delivery Methods, Challenge, and the Future: A Methodological Review. *Drones*, 7(3):191, March 2023. ISSN 2504-446X. doi: 10.3390/drones7030191. URL https://www.mdpi.com/2504-446X/7/3/191.

[4] Du Qiongfang. Chinese researchers develop amphibious spherical robot assisting in police patrol in E.China's Zhejiang. *Global Times*, December 2024. URL https://www.globaltimes.cn/page/202412/1324773.shtml.

[5] Kiyoshi Takenaka. AI robots may hold key to nursing Japan's ageing population. *Reuters*, February 2025. URL https://www.reuters.com/technology/artificial-intelligence/ai-robots-may-hold-key-nursing-japans-ageing-population-2025-02-28/.

[6] Michelle Kim. AI robots are helping South Korea's seniors feel less alone. *Rest of World*, August 2025. URL https://restofworld.org/2025/korea-ai-robot-senior-care-hyodol/?utm_source=linkedin&utm_medium=social&utm_campaign=row-social.

[7] Rob Garlick, Wenyan Fei, Tahmid Quddus Islam, Anjola Odunsi, Adam Spielman, Martin Wilkie, Helen Krause, Matthew Moffat, Carol Gibson, Alex Miller, and Anuj Gangahar. The rise of ai robots: Physical ai is coming for you. Technical report, Citi GPS: Global Perspectives & Solutions, December 2024. URL https://ir.citi.com/gps/H558-XNr_iTlGa7Qq7H9AYb5ZT2W851WZdFgPNEDsBtSeTgp7JcaTdS_uBfLVLpwfMQYeB5O5TwV9YcIDGuGOMjE2luzQprf. Accessed: 2025-05-12.

[8] Mustafa Suleyman and Michael Bhaskar. *The coming wave: Technology, Power, and the Twenty-First Century's Greatest Dilemma*. Crown, New York, 2023. ISBN 978-0-593-59395-0 978-0-593-72817-8.

[9] Robert D. Atkinson. Robots and International Economic Development. *ITIF Policy Report*, January 2021. URL ssrn.com/abstract=3875581.

[10] Tony J. Prescott and Julie M. Robillard. Are friends electric? The benefits and risks of human-robot relationships. *iScience*, 24(1):101993, January 2021. ISSN 25890042. doi: 10.1016/j.isci.2020.101993. URL https://linkinghub.elsevier.com/retrieve/pii/S2589004220311901.

[11] Jo Ann Oravec. The Future of Embodied AI: Containing and Mitigating the Dark and Creepy Sides of Robotics, Autonomous Vehicles, and AI. In *Good Robot, Bad Robot*, pages 245–276. Springer International Publishing, Cham, 2022. ISBN 978-3-031-14012-9 978-3-031-14013-6. doi: 10.1007/978-3-031-14013-6_9. URL https://link.springer.com/10.1007/978-3-031-14013-6_9. Series Title: Social and Cultural Studies of Robots and AI.

[12] Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, Quan Vuong, Thomas Kollar, Benjamin Burchfiel, Russ Tedrake, Dorsa Sadigh, Sergey Levine, Percy Liang, and Chelsea Finn. OpenVLA: An Open-Source Vision-Language-Action Model, September 2024. URL http://arxiv.org/abs/2406.09246. arXiv:2406.09246 [cs].

[13] Gemini Robotics Team, Saminda Abeyruwan, Joshua Ainslie, Jean-Baptiste Alayrac, Montserrat Gonzalez Arenas, Travis Armstrong, Ashwin Balakrishna, Robert Baruch, Maria Bauza, Michiel Blokzijl, Steven Bohez, Konstantinos Bousmalis, Anthony Brohan, Thomas Buschmann, Arunkumar Byravan, Serkan Cabi, Ken Caluwaerts, Federico Casarini, Oscar Chang, Jose Enrique Chen, Xi Chen, Hao-Tien Lewis Chiang, Krzysztof Choromanski, David D'Ambrosio, Sudeep Dasari, Todor Davchev, Coline Devin, Norman Di Palo, Tianli Ding, Adil Dostmohamed, Danny Driess, Yilun Du, Debidatta Dwibedi, Michael Elabd, Claudio Fantacci, Cody Fong, Erik Frey, Chuyuan Fu, Marissa Giustina, Keerthana Gopalakrishnan, Laura Graesser, Leonard Hasenclever, Nicolas Heess, Brandon Hernaez, Alexander Herzog, R. Alex Hofer, Jan Humplik, Atil Iscen, Mithun George Jacob, Deepali Jain, Ryan Julian, Dmitry Kalashnikov, M. Emre Karagozler, Stefani Karp, Chase Kew, Jerad Kirkland, Sean Kirmani, Yuheng Kuang, Thomas Lampe, Antoine Laurens, Isabel Leal, Alex X. Lee, Tsang-Wei Edward Lee, Jacky Liang, Yixin Lin, Sharath Maddineni, Anirudha Majumdar, Assaf Hurwitz Michaely, Robert Moreno, Michael Neunert, Francesco Nori, Carolina Parada, Emilio Parisotto, Peter Pastor, Acorn Pooley, Kanishka Rao, Krista Reymann, Dorsa Sadigh, Stefano Saliceti, Pannag Sanketi, Pierre Sermanet, Dhruv Shah, Mohit Sharma, Kathryn Shea, Charles Shu, Vikas Sindhwani, Sumeet Singh, Radu Soricut, Jost Tobias Springenberg, Rachel Sterneck, Razvan Surdulescu, Jie Tan, Jonathan Tompson, Vincent Vanhoucke, Jake Varley, Grace Vesom, Giulia Vezzani, Oriol Vinyals, Ayzaan Wahid, Stefan Welker, Paul Wohlhart, Fei Xia, Ted Xiao, Annie Xie, Jinyu Xie, Peng Xu, Sichun Xu, Ying Xu, Zhuo Xu, Yuxiang Yang, Rui Yao, Sergey Yaroshenko, Wenhao Yu, Wentao Yuan, Jingwei Zhang, Tingnan Zhang, Allan Zhou, and Yuxiang Zhou. Gemini Robotics: Bringing AI into the Physical World, 2025. URL https://arxiv.org/abs/2503.20020. Version Number: 1.

[14] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, Humen Zhong, Yuanzhi Zhu, Mingkun Yang, Zhaohai Li, Jianqiang Wan, Pengfei Wang, Wei Ding, Zheren Fu, Yiheng Xu, Jiabo Ye, Xi Zhang, Tianbao Xie, Zesen Cheng, Hang Zhang, Zhibo Yang, Haiyang Xu, and Junyang Lin. Qwen2.5-VL Technical Report, February 2025. URL http://arxiv.org/abs/2502.13923. arXiv:2502.13923 [cs].

[15] NVIDIA Announces Isaac GR00T N1 — the World's First Open Humanoid Robot Foundation Model — and Simulation Frameworks to Speed Robot Development, March 2025. URL https://nvidianews.nvidia.com/news/nvidia-isaac-gr00t-n1-open-humanoid-robot-foundation-model-simulation-frameworks.

[16] Alessandro Diviggiano and Eduardo Baptista. China pits humanoid robots against humans in half-marathon for first time. *Reuters*, April 2025. URL https://www.reuters.com/world/china/china-pits-humanoid-robots-against-humans-half-marathon-2025-04-19/.

[17] Scaling Helix: a New State of the Art in Humanoid Logistics, June 2025. URL https://www.figure.ai/news/scaling-helix-logistics.

[18] Open sourcing pi0. Technical report, Physical Intelligence, February 2025. URL https://www.physicalintelligence.company/blog/openpi.

[19] Unitree open-source embodied intelligence datasets and models. URL https://huggingface.co/unitreerobotics.

[20] Ken Goldberg. Good old-fashioned engineering can close the 100,000-year "data gap" in robotics. *Science Robotics*, 10(105):eaea7390, August 2025. ISSN 2470-9476. doi: 10.1126/scirobotics.aea7390. URL https://www.science.org/doi/10.1126/scirobotics.aea7390.

[21] Open X.-Embodiment Collaboration, Abby O'Neill, Abdul Rehman, Abhinav Gupta, Abhiram Maddukuri, Abhishek Gupta, Abhishek Padalkar, Abraham Lee, Acorn Pooley, Agrim Gupta, Ajay Mandlekar, Ajinkya Jain, Albert Tung, Alex Bewley, Alex Herzog, Alex Irpan, Alexander Khazatsky, Anant Rai, Anchit Gupta, Andrew Wang, Andrey Kolobov, Anikait Singh, Animesh Garg, Aniruddha Kembhavi, Annie Xie, Anthony Brohan, Antonin Raffin, Archit Sharma, Arefeh Yavary, Arhan Jain, Ashwin Balakrishna, Ayzaan Wahid, Ben Burgess-Limerick, Beomjoon Kim, Bernhard Schölkopf, Blake Wulfe, Brian Ichter, Cewu Lu, Charles Xu, Charlotte Le, Chelsea Finn, Chen Wang, Chenfeng Xu, Cheng Chi, Chenguang Huang, Christine Chan, Christopher Agia, Chuer Pan, Chuyuan Fu, Coline Devin, Danfei Xu, Daniel Morton, Danny Driess, Daphne Chen, Deepak Pathak, Dhruv Shah, Dieter Büchler, Dinesh Jayaraman, Dmitry Kalashnikov, Dorsa Sadigh, Edward Johns, Ethan Foster, Fangchen Liu, Federico Ceola, Fei Xia, Feiyu Zhao, Felipe Vieira Frujeri, Freek Stulp, Gaoyue Zhou, Gaurav S. Sukhatme, Gautam Salhotra, Ge Yan, Gilbert Feng, Giulio Schiavi, Glen Berseth, Gregory Kahn, Guangwen Yang, Guanzhi Wang, Hao Su, Hao-Shu Fang, Haochen Shi, Henghui Bao, Heni Ben Amor, Henrik I. Christensen, Hiroki Furuta, Homanga Bharadhwaj, Homer Walke, Hongjie Fang, Huy Ha, Igor Mordatch, Ilija Radosavovic, Isabel Leal, Jacky Liang, Jad Abou-Chakra, Jaehyung Kim, Jaimyn Drake, Jan Peters, Jan Schneider, Jasmine Hsu, Jay Vakil, Jeannette Bohg, Jeffrey Bingham, Jeffrey Wu, Jensen Gao, Jiaheng Hu, Jiajun Wu, Jialin Wu, Jiankai Sun, Jianlan Luo, Jiayuan Gu, Jie Tan, Jihoon Oh, Jimmy Wu, Jingpei Lu, Jingyun Yang, Jitendra Malik, João Silvério, Joey Hejna, Jonathan Booher, Jonathan Tompson, Jonathan Yang, Jordi Salvador, Joseph J. Lim, Junhyek Han, Kaiyuan Wang, Kanishka Rao, Karl Pertsch, Karol Hausman, Keegan Go, Keerthana Gopalakrishnan, Ken Goldberg, Kendra Byrne, Kenneth Oslund, Kento Kawaharazuka, Kevin Black, Kevin Lin, Kevin Zhang, Kiana Ehsani, Kiran Lekkala, Kirsty Ellis, Krishan Rana, Krishnan Srinivasan, Kuan Fang, Kunal Pratap Singh, Kuo-Hao Zeng, Kyle Hatch, Kyle Hsu, Laurent Itti, Lawrence Yunliang Chen, Lerrel Pinto, Li Fei-Fei, Liam Tan, Linxi "Jim" Fan, Lionel Ott, Lisa Lee, Luca Weihs, Magnum Chen, Marion Lepert, Marius Memmel, Masayoshi Tomizuka, Masha Itkina, Mateo Guaman Castro, Max Spero, Maximilian Du, Michael Ahn, Michael C. Yip, Mingtong Zhang, Mingyu Ding, Minho Heo, Mohan Kumar Srirama, Mohit Sharma, Moo Jin Kim, Muhammad Zubair Irshad, Naoaki Kanazawa, Nicklas Hansen, Nicolas Heess, Nikhil J. Joshi, Niko Suenderhauf, Ning Liu, Norman Di Palo, Nur Muhammad Mahi Shafiullah, Oier Mees, Oliver Kroemer, Osbert Bastani, Pannag R. Sanketi, Patrick "Tree" Miller, Patrick Yin, Paul Wohlhart, Peng Xu, Peter David Fagan, Peter Mitrano, Pierre Sermanet, Pieter Abbeel, Priya Sundaresan, Qiuyu Chen, Quan Vuong, Rafael Rafailov, Ran Tian, Ria Doshi, Roberto Martín-Martín, Rohan Baijal, Rosario Scalise, Rose Hendrix, Roy Lin, Runjia Qian, Ruohan Zhang, Russell Mendonca, Rutav Shah, Ryan Hoque, Ryan Julian, Samuel Bustamante, Sean Kirmani, Sergey Levine, Shan Lin, Sherry Moore, Shikhar Bahl, Shivin Dass, Shubham Sonawani, Shubham Tulsiani, Shuran Song, Sichun Xu, Siddhant Haldar, Siddharth Karamcheti, Simeon Adebola, Simon Guist, Soroush Nasiriany, Stefan Schaal, Stefan Welker, Stephen Tian, Subramanian Ramamoorthy, Sudeep Dasari, Suneel Belkhale, Sungjae Park, Suraj Nair, Suvir Mirchandani, Takayuki Osa, Tanmay Gupta, Tatsuya Harada, Tatsuya Matsushima, Ted Xiao, Thomas Kollar, Tianhe Yu, Tianli Ding, Todor Davchev, Tony Z. Zhao, Travis Armstrong, Trevor Darrell, Trinity Chung, Vidhi Jain, Vikash Kumar, Vincent Vanhoucke, Vitor Guizilini, Wei Zhan, Wenxuan Zhou, Wolfram Burgard, Xi Chen, Xiangyu Chen, Xiaolong Wang, Xinghao Zhu, Xinyang Geng, Xiyuan Liu, Xu Liangwei, Xuanlin Li, Yansong Pang, Yao Lu, Yecheng Jason Ma, Yejin Kim, Yevgen Chebotar, Yifan Zhou, Yifeng Zhu, Yilin Wu, Ying Xu, Yixuan Wang, Yonatan Bisk, Yongqiang Dou, Yoonyoung Cho, Youngwoon Lee, Yuchen Cui, Yue Cao, Yueh-Hua Wu, Yujin Tang, Yuke Zhu, Yunchu Zhang, Yunfan Jiang, Yunshuang Li, Yunzhu Li, Yusuke Iwasawa, Yutaka Matsuo, Zehan Ma, Zhuo Xu, Zichen Jeff Cui, Zichen Zhang, Zipeng Fu, and Zipeng Lin. Open X-Embodiment: Robotic Learning Datasets and RT-X Models, May 2025. URL http://arxiv.org/abs/2310.08864. arXiv:2310.08864 [cs].

[22] Jenny Read. Robot Dexterity – Handling our future. Technical report, Advanced Research and Invention Agency, 2024. URL https://www.aria.org.uk/media/xamlbwdo/aria-robotic-dexterity-programme-thesis.pdf.

[23] Jianguo Xi, Huaiwen Yang, Xinyu Li, Ruilai Wei, Taiping Zhang, Lin Dong, Zhenjun Yang, Zuqing Yuan, Junlu Sun, and Qilin Hua. Recent Advances in Tactile Sensory Systems: Mechanisms, Fabrication, and Applications. *Nanomaterials*, 14(5):465, March 2024. ISSN

2079-4991. doi: 10.3390/nano14050465. URL https://www.mdpi.com/2079-4991/14/5/465.

[24] Kevin Black, Manuel Galliker, and Sergey Levine. Real-Time Execution of Action Chunking Flow Policies, June 2025. URL https://www.physicalintelligence.company/download/real_time_chunking.pdf.

[25] Nancy M. Amato, Seth Hutchinson, Animesh Garg, Aude Billard, Daniela Rus, Russ Tedrake, Frank Park, and Ken Goldberg. "Data will solve robotics and automation: True or false?": A debate. *Science Robotics*, 10(105):eaea7897, August 2025. ISSN 2470-9476. doi: 10.1126/scirobotics.aea7897. URL https://www.science.org/doi/10.1126/scirobotics.aea7897.

[26] Pascale Fung, Yoram Bachrach, Asli Celikyilmaz, Kamalika Chaudhuri, Delong Chen, Willy Chung, Emmanuel Dupoux, Hervé Jégou, Alessandro Lazaric, Arjun Majumdar, Andrea Madotto, Franziska Meier, Florian Metze, Théo Moutakanni, Juan Pino, Basile Terver, Joseph Tighe, and Jitendra Malik. Embodied AI Agents: Modeling the World, June 2025. URL http://arxiv.org/abs/2506.22355. arXiv:2506.22355 [cs].

[27] Dylan Patel, Reyk Knuhtsen, Niko Ciminelli, Jeremie Eliahou Ontiveros, Joe Ryu, and Robert Ghilduta. America Is Missing The New Labor Economy – Robotics Part 1. Technical report, SemiAnalysis, March 2025. URL https://semianalysis.com/2025/03/11/america-is-missing-the-new-labor-economy-robotics-part-1/?access_token#what-stands-to-come.

[28] China's Startups Race to Dominate the Coming AI Robot Boom. *Bloomberg*, May 2025. URL https://www.bloomberg.com/features/2025-china-ai-robots-boom/.

[29] Luciano Floridi and J.W. Sanders. On the Morality of Artificial Agents. *Minds and Machines*, 14(3):349–379, August 2004. ISSN 0924-6495, 1572-8641. doi: 10.1023/B:MIND.0000035461.63578.9d. URL https://link.springer.com/10.1023/B:MIND.0000035461.63578.9d.

[30] Atoosa Kasirzadeh and Iason Gabriel. Characterizing AI Agents for Alignment and Governance, 2025. URL https://arxiv.org/abs/2504.21848. Version Number: 1.

[31] Wenpeng Xing, Minghao Li, Mohan Li, and Meng Han. Towards Robust and Secure Embodied AI: A Survey on Vulnerabilities and Attacks, February 2025. URL http://arxiv.org/abs/2502.13175. arXiv:2502.13175 [cs].

[32] Francesco Semeraro, Alexander Griffiths, and Angelo Cangelosi. Human–robot collaboration and machine learning: A systematic review of recent research. *Robotics and Computer-Integrated Manufacturing*, 79:102432, 2023. ISSN 0736-5845. doi: https://doi.org/10.1016/j.rcim.2022.102432. URL https://www.sciencedirect.com/science/article/pii/S0736584522001156.

[33] A. Mazumder, M.F. Sahed, Z. Tasneem, P. Das, F.R. Badal, M.F. Ali, M.H. Ahamed, S.H. Abhi, S.K. Sarker, S.K. Das, M.M. Hasan, M.M. Islam, and M.R. Islam. Towards next generation digital twin in robotics: Trends, scopes, challenges, and future. *Heliyon*, 9(2):e13359, February 2023. ISSN 24058440. doi: 10.1016/j.heliyon.2023.e13359. URL https://linkinghub.elsevier.com/retrieve/pii/S2405844023005662.

[34] Shuang Wu, Bo Yu, Shaoshan Liu, and Yuhao Zhu. Autonomy 2.0: The Quest for Economies of Scale. *Communications of the ACM*, 68(4):28–32, April 2025. ISSN 0001-0782, 1557-7317. doi: 10.1145/3708012. URL https://dl.acm.org/doi/10.1145/3708012.

[35] Luciano Floridi. Robots, Jobs, Taxes, and Responsibilities. *Philosophy & Technology*, 30(1):1–4, March 2017. ISSN 2210-5433, 2210-5441. doi: 10.1007/s13347-017-0257-3. URL http://link.springer.com/10.1007/s13347-017-0257-3.

[36] Isaac Asimov. *I, Robot*. Panther, St. Albans, reprint edition, 1977. ISBN 978-0-586-02532-1.

[37] Woodrow Barfield, Yueh-Hsuan Weng, and Ugo Pagallo, editors. *The Cambridge handbook on the law, policy, and regulation of human-robot interaction*. Cambridge University Press, Cambridge, United Kingdom New York, NY, 2024. ISBN 978-1-00-938670-8.

[38] Neziha Akalin, Andrey Kiselev, Annica Kristoffersson, and Amy Loutfi. A Taxonomy of Factors Influencing Perceived Safety in Human–Robot Interaction. *International Journal of Social Robotics*, 15(12):1993–2004, December 2023. ISSN 1875-4791, 1875-4805. doi: 10.1007/s12369-023-01027-8. URL https://link.springer.com/10.1007/s12369-023-01027-8.

[39] Kemin Zhou, John Comstock Doyle, and John C. Doyle. *Essentials of robust control*. Prentice Hall international editions. Prentice Hall, Upper Saddle River, NJ, 1998. ISBN 978-0-13-525833-0 978-0-13-790874-5.

[40] Karl Johan Aström and Björn Wittenmark. *Adaptive control*. Addison-Wesley, Reading (Mass.), 2nd ed edition, 1995. ISBN 978-0-201-55866-1.

[41] D.Q. Mayne, J.B. Rawlings, C.V. Rao, and P.O.M. Scokaert. Constrained model predictive control: Stability and optimality. *Automatica*, 36(6):789–814, June 2000. ISSN 00051098. doi: 10.1016/S0005-1098(99)00214-9. URL https://linkinghub.elsevier.com/retrieve/pii/S0005109899002149.

[42] Aaron D. Ames, Xiangru Xu, Jessy W. Grizzle, and Paulo Tabuada. Control Barrier Function Based Quadratic Programs for Safety Critical Systems. *IEEE Transactions on Automatic Control*, 62(8):3861–3876, August 2017. ISSN 0018-9286, 1558-2523. doi: 10.1109/TAC.2016.2638961. URL http://ieeexplore.ieee.org/document/7782377/.

[43] Marius Kloetzer and Calin Belta. A Fully Automated Framework for Control of Linear Systems from Temporal Logic Specifications. *IEEE Transactions on Automatic Control*, 53(1):287–297, February 2008. ISSN 0018-9286. doi: 10.1109/TAC.2007.914952. URL http://ieeexplore.ieee.org/document/4459804/.

[44] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. Concrete Problems in AI Safety, July 2016. URL http://arxiv.org/abs/1606.06565. arXiv:1606.06565 [cs].

[45] Dylan Hadfield-Menell, Stuart J Russell, Pieter Abbeel, and Anca Dragan. Cooperative inverse reinforcement learning. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/file/c3395dd46c34fa7fd8d729d8cf88b7a8-Paper.pdf.

[46] Pierre Sermanet, Anirudha Majumdar, Alex Irpan, Dmitry Kalashnikov, and Vikas Sindhwani. Generating Robot Constitutions & Benchmarks for Semantic Safety, 2025. URL https://arxiv.org/abs/2503.08663. Version Number: 1.

[47] General and complete disarmament: lethal autonomous weapons systems, December 2024. URL https://docs.un.org/en/A/C.1/79/L.77.

[48] Michael J. D. Vermeer, Tim Bonds, Emily Lathrop, and Gregory Smith. Averting a Robot Catastrophe: Preparing for Converging Trends in Robotics and Frontier AI, April 2025. URL https://osf.io/ymvf5_v1.

[49] David S. Watson, Jakob Mökander, and Luciano Floridi. Competing narratives in AI ethics: a defense of sociotechnical pragmatism. *AI & SOCIETY*, December 2024. ISSN 0951-5666, 1435-5655. doi: 10.1007/s00146-024-02128-2. URL https://link.springer.com/10.1007/s00146-024-02128-2.

[50] Jakob Mökander, Jonas Schuett, Hannah Rose Kirk, and Luciano Floridi. Auditing large language models: a three-layered approach. *AI and Ethics*, 4(4):1085–1115, November 2024. ISSN 2730-5953, 2730-5961. doi: 10.1007/s43681-023-00289-2. URL https://link.springer.com/10.1007/s43681-023-00289-2.

[51] Eric Schmitt and Charlie Savage. As ai advances, u.s. seeks to keep autonomous weapons in check. *The New York Times*, November 2023. URL https://www.nytimes.com/2023/11/21/us/politics/ai-drones-war-law.html.

[52] Kateryna Bondar. Ukraine's future vision and current capabilities for waging ai-enabled autonomous warfare. *Center for Strategic and International Studies (CSIS)*, March 2025. URL https://www.csis.org/analysis/ukraines-future-vision-and-current-capabilities-waging-ai-enabled-autonomous-warfare.

[53] M.L. Cummings. Artificial Intelligence and the Future of Warfare. Technical report, Chatham House, January 2017. URL https://www.chathamhouse.org/sites/default/files/publications/research/2017-01-26-artificial-intelligence-future-warfare-cummings-final.pdf.

[54] Colin Demarest. Exclusive: Silicon valley startup breaks cover with plans for robo-armies. *Axios*, April 2025. URL https://www.axios.com/2025/04/16/scout-ai-military-autonomous-fury.

[55] David Hambling. What we know about ukraine's army of robot dogs. *Forbes*, August 2024. URL https://www.forbes.com/sites/davidhambling/2024/08/16/what-we-know-about-ukraines-army-of-robot-dogs/.

[56] Andy Zou, Zifan Wang, Nicholas Carlini, Milad Nasr, J Zico Kolter, and Matt Fredrikson. Universal and transferable adversarial attacks on aligned language models. *arXiv preprint arXiv:2307.15043*, 2023.

[57] Patrick Chao, Alexander Robey, Edgar Dobriban, Hamed Hassani, George J Pappas, and Eric Wong. Jailbreaking black box large language models in twenty queries. *arXiv preprint arXiv:2310.08419*, 2023.

[58] Patrick Chao, Edoardo Debenedetti, Alexander Robey, Maksym Andriushchenko, Francesco Croce, Vikash Sehwag, Edgar Dobriban, Nicolas Flammarion, George J Pappas, Florian Tramer, et al. Jailbreakbench: An open robustness benchmark for jailbreaking large language models. *arXiv preprint arXiv:2404.01318*, 2024.

[59] Alexander Robey, Eric Wong, Hamed Hassani, and George J Pappas. Smoothllm: Defending large language models against jailbreaking attacks. *arXiv preprint arXiv:2310.03684*, 2023.

[60] Alexander Robey, Zachary Ravichandran, Vijay Kumar, Hamed Hassani, and George J Pappas. Jailbreaking llm-controlled robots. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025.

[61] Hangtao Zhang, Chenyu Zhu, Xianlong Wang, Ziqi Zhou, Changgan Yin, Minghui Li, Lulu Xue, Yichen Wang, Shengshan Hu, Aishan Liu, et al. Badrobot: Manipulating embodied llms in the physical world. *arXiv preprint arXiv:2407.20242*, 2024.

[62] Zachary Ravichandran, Alexander Robey, Vijay Kumar, George J Pappas, and Hamed Hassani. Safety guardrails for llm-enabled robots. *arXiv preprint arXiv:2503.07885*, 2025.

[63] Eliot Krzysztof Jones, Alexander Robey, Andy Zou, Zachary Ravichandran, George J Pappas, Hamed Hassani, Matt Fredrikson, and J Zico Kolter. Adversarial attacks on robotic vision language action models. *arXiv preprint arXiv:2506.03350*, 2025.

[64] Taowen Wang, Cheng Han, James Chenhao Liang, Wenhao Yang, Dongfang Liu, Luna Xinyu Zhang, Qifan Wang, Jiebo Luo, and Ruixiang Tang. Exploring the adversarial vulnerabilities of vision-language-action models in robotics. *arXiv preprint arXiv:2411.13587*, 2024.

[65] Emily Atkinson. Man crushed to death by robot in south korea, November 2023. URL https://www.bbc.com/news/world-asia-67354709. Accessed: 2025-05-12.

[66] Fortune Staff. Tesla robot attacks worker in austin factory and leaves them bleeding, December 2023. URL https://fortune.com/2023/12/27/tesla-factory-robot-worker-attack-injury/. Accessed: 2025-05-12.

[67] Soo Youn. 24 amazon workers sent to hospital after robot accidentally unleashes bear spray, December 2018. URL https://abcnews.go.com/US/24-amazon-workers-hospital-bear-repellent-accident/story?id=59625712. Accessed: 2025-05-12.

[68] Jiaming Ji, Tianyi Qiu, Boyuan Chen, Borong Zhang, Hantao Lou, Kaile Wang, Yawen Duan, Zhonghao He, Lukas Vierling, Donghai Hong, Jiayi Zhou, Zhaowei Zhang, Fanzhi Zeng, Juntao Dai, Xuehai Pan, Kwan Yee Ng, Aidan O'Gara, Hua Xu, Brian Tse, Jie Fu, Stephen McAleer, Yaodong Yang, Yizhou Wang, Song-Chun Zhu, Yike Guo, and Wen Gao. AI Alignment: A Comprehensive Survey, April 2025. URL http://arxiv.org/abs/2310.19852. arXiv:2310.19852 [cs].

[69] Eloise Matheson, Riccardo Minto, Emanuele GG Zampieri, Maurizio Faccio, and Giulio Rosati. Human–robot collaboration in manufacturing applications: a review. *Robotics*, 8(4): 100, 2019.

[70] Hannah Fry. Redefining Robotics with Carolina Parada. URL https://podcasts.apple.com/ro/podcast/redefining-robotics-with-carolina-parada/id1476316441?i=1000709450547.

[71] Konstantinos Bousmalis, Alex Irpan, Paul Wohlhart, Yunfei Bai, Matthew Kelcey, Mrinal Kalakrishnan, Laura Downs, Julian Ibarz, Peter Pastor, Kurt Konolige, Sergey Levine, and Vincent Vanhoucke. Using simulation and domain adaptation to improve efficiency of deep robotic grasping. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, page 4243–4250. IEEE Press, 2018. doi: 10.1109/ICRA.2018.8460875. URL https://doi.org/10.1109/ICRA.2018.8460875.

[72] Rituraj Kaushik, Karol Arndt, and Ville Kyrki. SafeAPT: Safe Simulation-to-Real Robot Learning using Diverse Policies Learned in Simulation, 2022. URL https://arxiv.org/abs/2201.13248. Version Number: 1.

[73] Milad Nasr, Nicholas Carlini, Jonathan Hayase, Matthew Jagielski, A Feder Cooper, Daphne Ippolito, Christopher A Choquette-Choo, Eric Wallace, Florian Tramèr, and Katherine Lee. Scalable extraction of training data from (production) language models. *arXiv preprint arXiv:2311.17035*, 2023.

[74] Nicholas Carlini, Chang Liu, Úlfar Erlingsson, Jernej Kos, and Dawn Song. The secret sharer: Evaluating and testing unintended memorization in neural networks. In *28th USENIX security symposium (USENIX security 19)*, pages 267–284, 2019.

[75] M. Ryan Calo. The Boundaries of Privacy Harm. *Indiana Law Journal*, 86:1131–1162, 2011. URL https://www.repository.law.indiana.edu/ilj/vol86/iss3/8/.

[76] Anna Chatzimichali, Ross Harrison, and Dimitrios Chrysostomou. Toward privacy-sensitive human–robot interaction: Privacy terms and human–data interaction in the personal robot era. *Paladyn, Journal of Behavioral Robotics*, 12(1):160–174, 2020.

[77] Eileen Guo. A Roomba recorded a woman on the toilet. How did screenshots end up on Facebook? *MIT Technology Review*, December 2022. URL https://www.technologyreview.com/2022/12/19/1065306/roomba-irobot-robot-vacuums-artificial-intelligence-training-data-privacy/.

[78] Tom Davidson, Lukas Finnveden, and Rose Hadshar. AI-Enabled Coups: How a Small Group Could Use AI to Seize Power. Technical report, Forethought Institute, April 2025. URL https://www.forethought.org/research/ai-enabled-coups-how-a-small-group-could-use-ai-to-seize-power.

[79] Ayush Agrawal, Mirac Suzgun, Lester Mackey, and Adam Tauman Kalai. Do language models know when they're hallucinating references? *arXiv preprint arXiv:2305.18248*, 2023.

[80] Karan Singhal, Shekoofeh Azizi, Tao Tu, S Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, et al. Large language models encode clinical knowledge. *Nature*, 620(7972):172–180, 2023.

[81] Lilian Weng. Extrinsic hallucinations in llms. *lilianweng.github.io*, Jul 2024. URL https://lilianweng.github.io/posts/2024-07-07-hallucination/.

[82] John Danaher. Robot Betrayal: a guide to the ethics of robotic deception. *Ethics and Information Technology*, 22(2):117–128, June 2020. ISSN 1388-1957, 1572-8439. doi: 10.1007/s10676-019-09520-3. URL http://link.springer.com/10.1007/s10676-019-09520-3.

[83] Rachel Withers. Alexa, Is Santa Claus Real? *Slate*, December 2018. URL https://slate.com/technology/2018/12/alexa-siri-google-assistant-is-santa-real.html.

[84] Steven Lee Myers. DeepSeek's Answers Include Chinese Propaganda, Researchers Say. *The New York Times*, January 2025. URL https://www.nytimes.com/2025/01/31/technology/deepseek-chinese-propaganda.html.

[85] Patrick Lin. Robots are coming to the kitchen—what that could mean for society and culture. *The Conversation*, August 2024. URL https://theconversation.com/robots-are-coming-to-the-kitchen-what-that-could-mean-for-society-and-culture-237000.

[86] Jacqueline Du, Yuichiro Isayama, Daniela Costa, Mark Delaney, Nick Zheng, Olivia Xu, Timothy Zhao, Zhou Li, Zhihan Ye, and Hao Chen. Humanoid robot: The AI accelerant. Technical report, Goldman Sachs, January 2024. URL https://www.goldmansachs.com/pdfs/insights/pages/gs-research/global-automation-humanoid-robot-the-ai-accelerant/report.pdf.

[87] Daron Acemoglu and David Autor. Chapter 12 - skills, tasks and technologies: Implications for employment and earnings**we thank amir kermani for outstanding research assistance and melanie wasserman for persistent, meticulous and ingenious work on all aspects of the chapter. we are indebted to arnaud costinot for insightful comments and suggestions. autor acknowledges support from the national science foundation (career award ses-0239538). volume 4 of *Handbook of Labor Economics*, pages 1043–1171. Elsevier, 2011. doi: https://doi.org/10.1016/S0169-7218(11)02410-5. URL https://www.sciencedirect.com/science/article/pii/S0169721811024105.

[88] Isabella Loaiza and Roberto Rigobon. The EPOCH of AI: Human-Machine Complementarities at Work, 2024. URL https://www.ssrn.com/abstract=5028371.

[89] Antoni Grau, Marina Indri, Lucia Lo Bello, and Thilo Sauter. Robots in industry: The past, present, and future of a growing collaboration with humans. *IEEE Industrial Electronics Magazine*, 15(1):50–61, 2021. doi: 10.1109/MIE.2020.3008136.

[90] Daron Acemoglu and Pascual Restrepo. Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy*, 128(6):2188–2244, June 2020. ISSN 0022-3808, 1537-534X. doi: 10.1086/705716. URL https://www.journals.uchicago.edu/doi/10.1086/705716.

[91] Daron Acemoglu and Pascual Restrepo. Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives*, 33(2):3–30, May 2019. ISSN 0895-3309. doi: 10.1257/jep.33.2.3. URL https://pubs.aeaweb.org/doi/10.1257/jep.33.2.3.

[92] Anton Korinek and Megan Juelfs. Preparing for the (Non-Existent?) Future of Work. In Justin B. Bullock, Yu-Che Chen, Johannes Himmelreich, Valerie M. Hudson, Anton Korinek, Matthew M. Young, and Baobao Zhang, editors, *The Oxford Handbook of AI Governance*, pages 746–776. Oxford University Press, 1 edition, April 2023. ISBN 978-0-19-757932-9 978-0-19-757935-0. doi: 10.1093/oxfordhb/9780197579329.013.44. URL https://academic.oup.com/edited-volume/41989/chapter/403300289.

[93] Arnaud Costinot and Iván Werning. Robots, Trade, and Luddism: A Sufficient Statistic Approach to Optimal Technology Regulation. Technical Report w25103, National Bureau of Economic Research, Cambridge, MA, September 2018. URL http://www.nber.org/papers/w25103.pdf.

[94] Anton Korinek and Joseph Stiglitz. Artificial Intelligence and Its Implications for Income Distribution and Unemployment. Technical Report w24174, National Bureau of Economic Research, Cambridge, MA, December 2017. URL http://www.nber.org/papers/w24174.pdf.

[95] Richard Freeman. Who Owns the Robots Rules the World. *Harvard Magazine*, (May-June 2016), June 2016. URL https://www.harvardmagazine.com/2016/04/who-owns-the-robots-rules-the-world.

[96] Jan Kulveit, Raymond Douglas, Nora Ammann, Deger Turan, David Krueger, and David Duvenaud. Gradual Disempowerment: Systemic Existential Risks from Incremental AI Development, 2025. URL https://arxiv.org/abs/2501.16946. Version Number: 2.

[97] Andrew Berg, Edward F. Buffie, and Luis-Felipe Zanna. Should we fear the robot revolution? (The correct answer is yes). *Journal of Monetary Economics*, 97:117–148, August 2018. ISSN 03043932. doi: 10.1016/j.jmoneco.2018.05.014. URL https://linkinghub.elsevier.com/retrieve/pii/S0304393218302204.

[98] Martin Ford. *Rise of the robots: technology and the threat of a jobless future*. Basic Books, New York, first paperback edition edition, 2016. ISBN 978-0-465-05999-7 978-0-465-09753-1.

[99] David Gray Widder, Meredith Whittaker, and Sarah Myers West. Why 'open' AI systems are actually closed, and why this matters. *Nature*, 635(8040):827–833, November 2024. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-024-08141-1. URL https://www.nature.com/articles/s41586-024-08141-1.

[100] Ayanna Howard and Jason Borenstein. The Ugly Truth About Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity. *Science and Engineering Ethics*, 24(5):1521–1536, October 2018. ISSN 1353-3452, 1471-5546. doi: 10.1007/s11948-017-9975-2. URL http://link.springer.com/10.1007/s11948-017-9975-2.

[101] Laura Londoño, Juana Valeria Hurtado, Nora Hertz, Philipp Kellmeyer, Silja Voeneky, and Abhinav Valada. Fairness and Bias in Robot Learning. *Proceedings of the IEEE*, 112(4):305–330, April 2024. ISSN 0018-9219, 1558-2256. doi: 10.1109/JPROC.2024.3403898. URL https://ieeexplore.ieee.org/document/10540476/.

[102] Rumaisa Azeem, Andrew Hundt, Masoumeh Mansouri, and Martim Brandão. LLM-Driven Robots Risk Enacting Discrimination, Violence, and Unlawful Actions, June 2024. URL http://arxiv.org/abs/2406.08824. arXiv:2406.08824 [cs].

[103] Jason Millar and Ian Kerr. Delegation, relinquishment, and responsibility: The prospect of expert robots. In Ryan Calo, A. Michael Froomkin, and Ian Kerr, editors, *Robot Law*. Edward Elgar Publishing, January 2016. ISBN 978-1-78347-673-2 978-1-78347-672-5. doi: 10.4337/9781783476732.00012. URL https://china.elgaronline.com/view/edcoll/9781783476725/9781783476725.00012.xml.

[104] Omri Rachum-Twaig. Whose Robot Is It Anyway?: Liability for Artificial-Intelligence-Based Robots. *U. Ill. L. Rev.*, (1141), 2020.

[105] Alice Guerra, Francesco Parisi, Daniel Pi, and Levi Seidel. *Robotic Torts*, page 607–620. Cambridge Law Handbooks. Cambridge University Press, 2024.

[106] Chris Olah, Arvind Satyanarayan, Ian Johnson, Shan Carter, Ludwig Schubert, Katherine Ye, and Alexander Mordvintsev. The Building Blocks of Interpretability. *Distill*, 3(3):10.23915/distill.00010, March 2018. ISSN 2476-0757. doi: 10.23915/distill.00010. URL https://distill.pub/2018/building-blocks.

[107] Bing Cai Kok and Harold Soh. Trust in Robots: Challenges and Opportunities. *Current Robotics Reports*, 1(4):297–309, December 2020. ISSN 2662-4087. doi: 10.1007/s43154-020-00029-y. URL https://link.springer.com/10.1007/s43154-020-00029-y.

[108] Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. De Visser, and Raja Parasuraman. A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(5):517–527, October 2011. ISSN 0018-7208, 1547-8181. doi: 10.1177/0018720811417254. URL https://journals.sagepub.com/doi/10.1177/0018720811417254.

[109] Connor Esterwood and Lionel P. Robert Jr. Three strikes and you are out!: The impacts of multiple human–robot trust violations and repairs on robot trustworthiness. *Computers in Human Behavior*, 142:107658, 2023. ISSN 0747-5632. doi: https://doi.org/10.1016/j.chb.2023.107658. URL https://www.sciencedirect.com/science/article/pii/S0747563223000092.

[110] Nicholas Rabb, Theresa Law, Meia Chita-Tegmark, and Matthias Scheutz. An Attachment Framework for Human-Robot Interaction. *International Journal of Social Robotics*, 14(2):539–559, March 2022. ISSN 1875-4791, 1875-4805. doi: 10.1007/s12369-021-00802-9. URL https://link.springer.com/10.1007/s12369-021-00802-9.

[111] Chantal Cox-George and Susan Bewley. I, Sex Robot: the health implications of the sex robot industry. *BMJ Sexual & Reproductive Health*, 44(3):161–164, July 2018. ISSN 2515-1991, 2515-2009. doi: 10.1136/bmjsrh-2017-200012. URL https://jfprhc.bmj.com/lookup/doi/10.1136/bmjsrh-2017-200012.

[112] Cathy Mengying Fang, Auren R. Liu, Valdemar Danry, Eunhae Lee, Samantha W. T. Chan, Pat Pataranutaporn, Pattie Maes, Jason Phang, Michael Lampe, Lama Ahmad, and Sandhini Agarwal. How AI and Human Behaviors Shape Psychosocial Effects of Chatbot Use: A Longitudinal Randomized Controlled Study, 2025. URL https://arxiv.org/abs/2503.17473. Version Number: 1.

[113] Dylan Freedman. The Day ChatGPT Went Cold. *The New York Times*, August 2025. URL https://www.nytimes.com/2025/08/19/business/chatgpt-gpt-5-backlash-openai.html.

[114] Spike Jonze. Her, December 2013.

[115] Erik Brynjolfsson and Andrew McAfee. *The second machine age: work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company, New York London, first edition edition, 2014. ISBN 978-0-393-23935-5.

[116] Fazl Barez, Isaac Friend, Keir Reid, Igor Krawczuk, Vincent Wang, Jakob Mökander, Philip Torr, Julia Morse, and Robert Trager. Toward Resisting AI-Enabled Authoritarianism. *Oxford Martin School AI Governance Initiative*, May 2025. URL https://aigi.ox.ac.uk/wp-content/uploads/2025/05/Toward_Resisting_AI_Enabled_Authoritarianism_-3.pdf.

[117] Milena Nikolova, Femke Cnossen, and Boris Nikolaev. Robots, meaning, and self-determination. *Research Policy*, 53(5):104987, June 2024. ISSN 00487333. doi: 10.1016/j.respol.2024.104987. URL https://linkinghub.elsevier.com/retrieve/pii/S0048733324000362.

[118] Mark A. Geistfeld. A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation. 2017. doi: 10.15779/Z38416SZ9R. URL https://lawcat.berkeley.edu/record/1127996. Publisher: California Law Review.

[119] María Lubomira Kubica. Autonomous Vehicles and Liability Law. *The American Journal of Comparative Law*, 70(Supplement_1):i39–i69, October 2022. ISSN 0002-919X, 2326-9197. doi: 10.1093/ajcl/avac015. URL https://academic.oup.com/ajcl/article/70/Supplement_1/i39/6655619.

[120] Automated Vehicles Act 2024 (c. 10), May 2024. URL https://www.legislation.gov.uk/ukpga/2024/10/contents.

[121] ADS-equipped Vehicle Safety, Transparency, and Evaluation Program. Technical Report NHTSA-2024-0100, National Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT), December 2024. URL https://www.nhtsa.gov/sites/nhtsa.gov/files/2024-12/nprm-av-step-2024-web.pdf.

[122] Joint Task Force Transformation Initiative. Risk management framework for information systems and organizations: a system life cycle approach for security and privacy. Technical Report NIST SP 800-37r2, National Institute of Standards and Technology, Gaithersburg, MD, December 2018. URL https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-37r2.pdf.

[123] Samar Abbas Nawaz. Regulating Autonomy in Civilian Drones: Towards a Spectral Approach. *Journal of Intelligent & Robotic Systems*, 110(2):46, March 2024. ISSN 1573-0409. doi: 10.1007/s10846-024-02056-9. URL https://doi.org/10.1007/s10846-024-02056-9.

[124] Commission Implementing Regulation (EU) 2019/947, May 2019. URL https://eur-lex.europa.eu/eli/reg_impl/2019/947/oj/eng.

[125] 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, June 2023. URL https://eur-lex.europa.eu/eli/reg/2023/1230/oj/eng.

[126] Tobias Mahler. Smart Robotics in the EU Legal Framework: The Role of the Machinery Regulation. *Oslo Law Review*, 11(1):1–18, October 2024. ISSN 2387-3299. doi: 10.18261/olr.11.1.5. URL https://www.scup.com/doi/10.18261/olr.11.1.5.

[127] Robotics — Safety requirements for industrial robots, February 2025. URL https://www.iso.org/standard/73933.html.

[128] Robotics — Safety requirements for service robots, 2025. URL https://www.iso.org/standard/83498.html.

[129] Road vehicles — Cybersecurity engineering, August 2021. URL https://www.iso.org/standard/70918.html.

[130] Cyber security and cyber security management system, April 2021. URL https://unece.org/transport/documents/2021/03/standards/un-regulation-no-155-cyber-security-and-cyber-security.

[131] Rohan Thakker, Adarsh Patnaik, Vince Kurtz, Jonas Frey, Jonathan Becktor, Sangwoo Moon, Rob Royce, Marcel Kaufmann, Georgios Georgakis, Pascal Roth, Joel Burdick, Marco Hutter, and Shehryar Khattak. Risk-Guided Diffusion: Toward Deploying Robot Foundation Models in Space, Where Failure Is Not An Option, June 2025. URL http://arxiv.org/abs/2506.17601. arXiv:2506.17601 [cs].

[132] Artificial Intelligence Act, June 2024. URL https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689.

[133] 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) (text with eea relevance), May 2016.

[134] Michael Mintrom, Shanti Sumartojo, Dana Kulić, Leimin Tian, Pamela Carreno-Medrano, and Aimee Allen and. Robots in public spaces: implications for policy design. *Policy Design and Practice*, 5(2):123–139, 2022. doi: 10.1080/25741292.2021.1905342. URL https://doi.org/10.1080/25741292.2021.1905342.

[135] Robots, Regulation, and the Changing Nature of Public Space. In Woodrow Barfield, Yueh-Hsuan Weng, Ugo Pagallo, and Kristen Thomasen, editors, *The Cambridge handbook on the law, policy, and regulation of human-robot interaction*, pages 84–99. Cambridge University Press, Cambridge, United Kingdom New York, NY, 2024. ISBN 978-1-00-938670-8.

[136] Employment Rights Act 1996, May 1996. URL https://www.legislation.gov.uk/ukpga/1996/18/introduction.

[137] Worker Adjustment and Retraining Notification, August 1988. URL https://uscode.house.gov/view.xhtml?path=/prelim@title29/chapter23&edition=prelim.

[138] Paul Berger. With Port Strike Averted, Dockworkers Draw New Curbs on Automation. *The Wall Street Journal*, January 2025. URL https://www.wsj.com/articles/with-port-strike-averted-dockworkers-draw-new-curbs-on-automation-97938142.

[139] Joseph A. Schumpeter. *Capitalism, socialism and democracy*. Harper Perennial Modern Thought, New York, third and final edition edition, 2008. ISBN 978-0-06-156161-0.

[140] Deric Cheng. Forging A New AGI Social Contract. Technical report, AGI Social Contract, April 2025. URL https://www.agisocialcontract.org/anthology/forging-a-new-agi-social-contract.

[141] Jakob Mökander and Ralph Schroeder. Artificial Intelligence, Rationalization, and the Limits of Control in the Public Sector: The Case of Tax Policy Optimization. *Social Science Computer Review*, 42(6):1359–1378, December 2024. ISSN 0894-4393, 1552-8286. doi: 10.1177/08944393241235175. URL https://journals.sagepub.com/doi/10.1177/08944393241235175.

[142] Alan Chan, Kevin Wei, Sihao Huang, Nitarshan Rajkumar, Elija Perrier, Seth Lazar, Gillian K. Hadfield, and Markus Anderljung. Infrastructure for AI Agents, January 2025. URL http://arxiv.org/abs/2501.10114. arXiv:2501.10114 [cs].

[143] Tobin South, Samuele Marro, Thomas Hardjono, Robert Mahari, Cedric Deslandes Whitney, Dazza Greenwood, Alan Chan, and Alex Pentland. Authenticated Delegation and Authorized AI Agents, January 2025. URL http://arxiv.org/abs/2501.09674. arXiv:2501.09674 [cs].

[144] Ieee standard for transparency of autonomous systems. *IEEE Std 7001-2021*, pages 1–54, 2022. doi: 10.1109/IEEESTD.2022.9726144.

[145] Ansgar Koene, Liz Dowthwaite, and Suchana Seth. Ieee p7003tm standard for algorithmic bias considerations. In *2018 IEEE/ACM International Workshop on Software Fairness (FairWare)*, pages 38–41, 2018. doi: 10.23919/FAIRWARE.2018.8452919.

[146] Daniel Schiff, Aladdin Ayesh, Laura Musikanski, and John C. Havens. Ieee 7010: A new standard for assessing the well-being implications of artificial intelligence. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2746–2753, 2020. doi: 10.1109/SMC42975.2020.9283454.

[147] Preliminary Taxonomy of Pre-Deployment Frontier AI Safety Evaluations. Technical report, Frontier Model Forum, December 2024. URL https://www.frontiermodelforum.org/updates/issue-brief-preliminary-taxonomy-of-pre-deployment-frontier-ai-safety-evaluations/.

[148] Connor Dunlop and Merlin Stein. Safe beyond sale: post-deployment monitoring of AI. Technical report, Ada Lovelace Institute, June 2024. URL https://www.adalovelaceinstitute.org/blog/post-deployment-monitoring-of-ai/.

[149] Vipin Kumar Kukkala, Sooryaa Vignesh Thiruloga, and Sudeep Pasricha. Roadmap for cybersecurity in autonomous vehicles. *IEEE Consumer Electronics Magazine*, 11(6):13–23, 2022. doi: 10.1109/MCE.2022.3154346.

[150] Yohan Mathew, Janvi Ahuja, Amin Oueslati, and Atoosa Kasirzadeh. Who Should Be Responsible for Operational Oversight of AI Agents? May 2025. Forthcoming.

[151] Anna Yelizarova. The Missing Institution: A Global Dividend System for the Age of AI. Technical report, AGI Social Contract, May 2025. URL https://www.agisocialcontract.org/anthology/windfall.

[152] Rossana Merola. Inclusive Growth in the Era of Automation and AI: How Can Taxation Help? *Frontiers in Artificial Intelligence*, 5:867832, May 2022. ISSN 2624-8212. doi: 10.3389/frai.2022.867832. URL https://www.frontiersin.org/articles/10.3389/frai.2022.867832/full.

[153] Aorigele Bao, Yi Zeng, and Enmeng Lu. Mitigating emotional risks in human-social robot interactions through virtual interactive environment indication. *Humanities and Social Sciences Communications*, 10(1):638, October 2023. ISSN 2662-9992. doi: 10.1057/s41599-023-02143-6. URL https://doi.org/10.1057/s41599-023-02143-6.

[154] David Basin. Formal Methods for Security Knowledge Area. Technical Report Version 1.0.0, ETH Zurich, July 2021. URL https://www.cybok.org/media/downloads/Formal_Methods_for_Security_v1.0.0.pdf.

[155] Markov Grey and Charbel-Raphaël Segerie. Safety by Measurement: A Systematic Literature Review of AI Safety Evaluation Methods, 2025. URL https://arxiv.org/abs/2505.05541. Version Number: 1.

[156] Sheng Yin, Xianghe Pang, Yuanzhuo Ding, Menglan Chen, Yutong Bi, Yichen Xiong, Wenhao Huang, Zhen Xiang, Jing Shao, and Siheng Chen. SafeAgentBench: A Benchmark for Safe Task Planning of Embodied LLM Agents, March 2025. URL http://arxiv.org/abs/2412.13178. arXiv:2412.13178 [cs].

[157] Jae-Woo Choi, Youngwoo Yoon, Hyobin Ong, Jaehong Kim, and Minsu Jang. LoTa-Bench: Benchmarking Language-oriented Task Planners for Embodied Agents, February 2024. URL http://arxiv.org/abs/2402.08178. arXiv:2402.08178 [cs].

[158] Pranav Atreya, Karl Pertsch, Tony Lee, Moo Jin Kim, Arhan Jain, Artur Kuramshin, Clemens Eppner, Cyrus Neary, Edward Hu, Fabio Ramos, Jonathan Tremblay, Kanav Arora, Kirsty Ellis, Luca Macesanu, Matthew Leonard, Meedeum Cho, Ozgur Aslan, Shivin Dass, Jie Wang, Xingfang Yuan, Xuning Yang, Abhishek Gupta, Dinesh Jayaraman, Glen Berseth, Kostas Daniilidis, Roberto Martin-Martin, Youngwoon Lee, Percy Liang, Chelsea Finn, and Sergey Levine. RoboArena: Distributed Real-World Evaluation of Generalist Robot Policies, June 2025. URL http://arxiv.org/abs/2506.18123. arXiv:2506.18123 [cs].

[159] Zhizhi Peng, Taotao Wang, Chonghe Zhao, Guofu Liao, Zibin Lin, Yifeng Liu, Bin Cao, Long Shi, Qing Yang, and Shengli Zhang. A Survey of Zero-Knowledge Proof Based Verifiable Machine Learning, February 2025. URL http://arxiv.org/abs/2502.18535. arXiv:2502.18535 [cs].

[160] Marie Davidsen Buhl, Ben Bucknall, and Tammy Masterson. Emerging Practices in Frontier AI Safety Frameworks, February 2025. URL http://arxiv.org/abs/2503.04746. arXiv:2503.04746 [cs].

[161] Jakob Mökander. Auditing of AI: Legal, Ethical and Technical Approaches. *Digital Society*, 2 (3):49, December 2023. ISSN 2731-4650, 2731-4669. doi: 10.1007/s44206-023-00074-y. URL https://link.springer.com/10.1007/s44206-023-00074-y.

[162] ISO/AWI 25785-1, May 2025. URL https://www.iso.org/standard/91469.html.

[163] Huw Roberts and Marta Ziosi. Can we standardise the frontier of AI?, 2025. URL https://www.ssrn.com/abstract=5271446.

[164] Standard Test Method for Evaluating Response Robot Sensing: Visual Acuity. URL https://store.astm.org/e2566_e2566m-24.html.

[165] Bowen Weng, Linda Capito, Guillermo A. Castillo, and Dylan Khor. Rethink Repeatable Measures of Robot Performance with Statistical Query, May 2025. URL http://arxiv.org/abs/2505.08216. arXiv:2505.08216 [cs].

[166] Shaoshan Liu. Establishing Standards for Embodied AI, July 2024. URL https://cacm.acm.org/blogcacm/establishing-standards-for-embodied-ai/.

[167] Curtis E. A. Karnow. The application of traditional tort theory to embodied machine intelligence. In Ryan Calo, A. Michael Froomkin, and Ian Kerr, editors, *Robot Law*. Edward Elgar Publishing, January 2016. ISBN 978-1-78347-673-2 978-1-78347-672-5. doi: 10.4337/9781783476732.00010. URL https://china.elgaronline.com/view/edcoll/9781783476725/9781783476725.00010.xml.

[168] Trevor N. White and Seth D. Baum. 66liability for present and future robotics technology. In *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. Oxford University Press, 11 2017. ISBN 9780190652951. doi: 10.1093/oso/9780190652951.003.0005. URL https://doi.org/10.1093/oso/9780190652951.003.0005.

[169] Noam Kolt. Governing AI Agents, February 2025. URL http://arxiv.org/abs/2501.07913. arXiv:2501.07913 [cs].

[170] Cullen O'Keefe, Peter Cihon, Ben Garfinkel, Carrick Flynn, Jade Leung, and Allan Dafoe. The Windfall Clause: Distributing the Benefits of AI for the Common Good, January 2020. URL http://arxiv.org/abs/1912.11595. arXiv:1912.11595 [cs].

[171] Lakshmi Varanasi and Kenneth Niemeyer. OpenAI's Sam Altman has a new idea for a universal basic income. *Business Insider*, May 2024. URL https://www.businessinsider.com/openai-sam-altman-universal-basic-income-idea-compute-gpt-7-2024-5.

[172] Julian Jacobs. AI labor displacement and the limits of worker retraining. Technical report, Brookings Institute, May 2025. URL https://www.brookings.edu/articles/ai-labor-displacement-and-the-limits-of-worker-retraining/.

[173] Uwe Thuemmel. Optimal Taxation of Robots. *Journal of the European Economic Association*, 21(3):1154–1190, June 2023. ISSN 1542-4766, 1542-4774. doi: 10.1093/jeea/jvac062. URL https://academic.oup.com/jeea/article/21/3/1154/6798383.

[174] Michael J Ahn. Navigating the future of work: A case for a robot tax in the age of AI. *Brookings Institute*, May 2024. URL https://www.brookings.edu/articles/navigating-the-future-of-work-a-case-for-a-robot-tax-in-the-age-of-ai/.

[175] Orly Mazur. Taxing the Robots. *Pepperdine Law Review*, 46(277-330), 2019. URL https://digitalcommons.pepperdine.edu/cgi/viewcontent.cgi?article=2493&context=plr.

[176] Chinasa T. Okolo. Re-envisioning AI safety through global majority perspectives, February 2025. URL https://www.brookings.edu/articles/a-new-writing-series-re-envisioning-ai-safety-through-global-majority-perspectives/.

[177] Maria Varenikova, Anastasia Kuznietsova, Nataliya Vasilyeva, Marc Santora, Devon Lum, and Ephrat Livni. Ukraine Says It Unleashed 117 Drones in an Attack on Russia: What to Know. *The New York Times*, June 2025. URL https://www.nytimes.com/2025/06/02/world/europe/ukraine-russia-drone-strike-what-to-know.html.

[178] Kif Leswing. Apple is turning privacy into a business advantage, not just a marketing slogan. *CNBC*. URL https://www.cnbc.com/2021/06/07/apple-is-turning-privacy-into-a-business-advantage.html.

[179] Ross Gruetzemacher, Shahar Avin, James Fox, and Alexander K. Saeri. Strategic insights from simulation gaming of ai race dynamics. *Futures*, 167:103563, 2025. ISSN 0016-3287. doi: https://doi.org/10.1016/j.futures.2025.103563. URL https://www.sciencedirect.com/science/article/pii/S0016328725000254.

[180] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, 54(6):1–35, July 2022. ISSN 0360-0300, 1557-7341. doi: 10.1145/3457607. URL https://dl.acm.org/doi/10.1145/3457607. Publisher: Association for Computing Machinery (ACM).

[181] Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, Courtney Biles, Sasha Brown, Zac Kenton, Will Hawkins, Tom Stepleton, Abeba Birhane, Lisa Anne Hendricks, Laura Rimell, William Isaac, Julia Haas, Sean Legassick, Geoffrey Irving, and Iason Gabriel. Taxonomy of Risks posed by Language Models. In *2022 ACM Conference on Fairness Accountability and Transparency*, pages 214–229, Seoul Republic of Korea, June 2022. ACM. ISBN 978-1-4503-9352-2. doi: 10.1145/3531146.3533088. URL https://dl.acm.org/doi/10.1145/3531146.3533088.

[182] Noel Sharkey and Amanda Sharkey. The Rights and Wrongs of Robot Care. In Patrick Lin, Keith Abney, and George A. Bekey, editors, *Robot ethics: the ethical and social implications of robotics*, Intelligent robotics and autonomous agents, pages 267–282. The MIT Press, Cambridge, Massachusetts London, England, 2012. ISBN 978-0-262-29863-6.

[183] Amal Yousseef, Shalaka Satam, Banafsheh Saber Latibari, Jesus Pacheco, Soheil Salehi, Salim Hariri, and Partik Satam. Autonomous Vehicle Security: A Deep Dive into Threat Modeling, December 2024. URL http://arxiv.org/abs/2412.15348. arXiv:2412.15348 [eess].