From Actuators to Action: Deriving Key Principles and Strategies for Social Justice Grounded Robotics Education

Yifei Zhu, Colorado School of Mines

I'm a PhD candidate in robotics at the Colorado School of Mines. I have an interdisciplinary focus on human–robot interaction, where I use mixed-methods approaches to design robot interaction and communication. I also study how emerging technologies such as robots are designed, deployed, and understood within broader social contexts.

Dr. Juan C. Lucena, Colorado School of Mines

Juan Lucena is Professor and Director of Humanitarian Engineering Undergraduate Programs at the Colorado School of Mines (CSM). Juan obtained a Ph.D. in Science and Technology Studies (STS) from Virginia Tech and a MS in STS and BS in Mechanical and Aeronautical Engineering fro

Dr. Tom Williams, Colorado School of Mines

Tom Williams is an Associate Professor of Computer Science at the Colorado School of Mines, Director of the MIRRORLab, and author of Degrees of Freedom: On Robotics and Social Justice (2025).

Tom earned joint PhDs in Computer Science and Cognitive Science from Tufts University, and is an internationally recognized scholar of Human-Robot Interaction whose work examines cognitive, social, and moral dimensions of Human-Robot Interaction.

Tom is the winner of Early Career awards from the National Science Foundation, the Air Force Office of Scientific Research, and NASA.

Dr. Alison J Kerr, Colorado School of Mines

Alison J. Kerr has a doctoral degree in Industrial-Organizational Psychology from The University of Tulsa. Dr. Kerr's research interests include training development and evaluation as explored across a variety of academic disciplines and organizational se

Dr. Katie Winkle, Uppsala University

Katie Winkle is an assistant professor at the Department for Information Technology at Uppsala University. She previously held a Digital Futures postdoctoral fellowship at KTH Royal Institute of Technology and completed her PhD in Robotics and Autonomous Systems at the Bristol Robotics Laboratory. Her research sits at the intersection of robotics, design, and the social sciences to tackle technical and societal challenges relating to human-machine interaction. Her award-winning work on feminist human-robot interaction has gained significant attention within the field and is helping to define new, critical directions of research on the design of robots and their interaction behaviours.

Tiera Tanksley, University of California, Los Angeles

Dr. Tiera Tanksley is a Senior Researcher in education and information studies at UCLA's Center for Resilience and Digital Justice. Her scholarship, which theorizes a critical race technology theory (CRTT) in education, extends conventional education research to include socio-technical and techno-structural analyses of AI and EdTech. Specifically, her research examines anti-Blackness as "the default setting" and "organizing logic" of both schools and school-based technologies, and analyzes the socio-emotional, mental health and academic ramifications of algorithmic racism in the lives and schooling experiences of Black youth. Her work simultaneously recognizes Black youth as digital activists and civic agitators, and examines the complex ways they subvert, resist and rewrite algorithmically biased technologies to produce more just and joyous digital experiences for Communities of Color. In 2020, Dr. Tanksley founded the Race, Abolition and Artificial Intelligence summer program - a critical science and technology program that prepares young people to have more critical, agentic and algorithmically-conscious relationships with digital technologies that they encounter both within and beyond the classroom context.

From Actuators to Action: Deriving Key Principles and Strategies for Social Justice Grounded Robotics Education

Yifei Zhu[1], Ali Kerr[1], Juan Lucena[1], Tiera Tanksley[2], Katie Winkle[3], and Tom Williams[1]

[1]Colorado School of Mines, Golden, CO, {zhu1,akerr,jlucena,twillams}@mines.edu
[2]UCLA, Los Angeles, California, tctanksl@g.ucla.edu
[3]Uppsala University, Uppsala, Sweden, katie.winkle@it.uu.se

1. ABSTRACT

The purpose of this critical theory paper is to analyze and synthesize important and relevant prior frameworks, including Engineering for Social justice, feminist Human-Robot Interaction, Matrix-Guided Technology Power Analysis and Critical Race Pedagogy, to inform social-justice centered Robotics curricula. As Robotics garners increasing research and industry interest, and as robotic technology is increasingly deployed in ethically fraught domains such as healthcare and education, it is necessary to train roboticists to develop justice-oriented agents that help remediate rather than exacerbate disparities. Unfortunately, opportunities to develop justice-oriented robotics praxis are severely limited. Indeed, current training resources and tools are insufficient in aiding such development, as socially just robotics curricula need to prepare students to collaborate with communities, challenge dominant perspectives in robotics, and examine how racism, sexism, classism, and other systems of power that shape robotic technology. More broadly, such training must also prepare students to critically engage with societal topics such as labor, identity, and power. While existing computing ethics frameworks can guide analysis of algorithms and software, robotics students additionally need to develop the necessary skills and competency to approach and evaluate robots as embodied technology. Thus, in this work, we synthesize and analyze the insights from leading frameworks in critical tech teaching and design. We argue that such curricula must incorporate key principles in robot design, facilitate considerations of various stakeholders, encourage students to engage with larger social, socio-technical systems and power dynamics. We further provide relevant strategies for educators to consult, such as community-oriented ethics analysis and speculative exercises, which we also translate into learning objectives. Finally, we introduce the implications of our specific research process for other researchers and educators, share perspectives on the current state of robotics education, and describe planned future work.

2. MOTIVATION AND INTRODUCTION

Robots are increasingly entering people's everyday lives, including diverse and ethically fraught settings. For instance, robots are now expected to help students learn [1, 2], provide care for older adults [3, 4], assist communication for people with complex communication needs [5, 6], and provide support during therapy sessions [7]. Though robots can improve human wellbeing, they also come with well-known risks, including exacerbating socioeconomic inequality [8, 9], amplifying systemic biases [10, 11], and further reinforcing existing power imbalances [12]. Moreover, assistive robots in particular present risks such as encouraging over-trust in robotic technology [13], deteriorating social relationships [14], and harming human dignity [15]. To ensure that robots provide meaningful benefits to communities, robot design and deployment must integrate assessment of both potential benefits and harms. This includes not only technical evaluations of safety and efficiency, but also ethical analyses of how robotic systems may reinforce, disrupt, or reconfigure existing social structures. As such, it is evident that roboticists must be able to thoughtfully engage with *social justice* principles, to understand how issues of race, gender, class, power dynamics and socio-technical systems shape both the design and the societal impact of robots; and roboticists should take an active role in guiding robot development and deployment toward justice-oriented outcomes [16, 17]. For educators, this raises key questions: **How can socially just engineering practices be instilled in robotics students; and what are the key principles that should be instilled in such students?** This work thus explores this research question and presents principles and tools to help robotics educators implement justice-oriented curricula.

However, answering these questions is difficult for a number of reasons. First, computing ethics education has traditionally focused on topics from moral philosophy and professional ethics [18, 19, 20]. While these foundations are important, they are insufficient for addressing the unique challenges posed by embodied interactive technologies like robots. Robots can operate in shared physical and social environments. As such, robots need to be designed in a way that incorporates the needs and desires of the community members that robots may directly interact with [21], and with sensitivity to their potential psychological and societal impacts [22, 23]. This means moving beyond generalized ethical principles to engage with the lived experiences, cultural values, and social dynamics of those communities. Second, robots are not just tools. They are often created to depict humans; yet conventionally, predominantly White and Asian men of higher socio-economic class are included in the robot design process [24]. A justice-oriented approach to robotics must actively challenge these dominant perspectives and seek to expand the benefits of robot deployment. To achieve this, robotics research and design must be sensitive to factors such as race, gender, class, colonialism, and robots' broader historical context [25, 26]. Finally, because robots are developed for specific use cases and contexts, such as education, caregiving, and social companionship, their design and deployment can impact not only the existing labor structure in their domains of deployment, but also how people fundamentally work and interact with each other in those domains. Therefore, roboticists need a genuine understanding of issues of labor that are fundamental to people's identity, livelihoods, and power in society [27, 28]. Without this understanding, roboticists risk perpetuating existing labor injustices or exacerbating inequalities in the name of efficiency or innovation. This can include robots that perpetuate malignant racial stereotypes [29], perpetuating sexist behaviors [30, 31], exacerbate economic and social inequality [32, 33, 34], and carry out oppressive agendas of imperialism and colonialism [35, 36].

Training roboticists to grapple with the complexity of these intersectional socio-technical issues is especially challenging. One prominent reason is that Humanities and Social Sciences classes that highlight and discuss these interactions are among the least available in science and engineering curricula. As such, robotics educators must find ways within the context of robotics curricula to provide students with knowledge, skills, and attitude development to enhance their growth as ethical, responsible, socially just, and inclusive roboticists. However, as discussed above, the intersectional awareness and knowledge required for roboticists is complex and broad. Therefore, in this work we aim to provide educators with relevant tools to conceptualize and design justice-oriented robotics curricula. In particular, we aim to understand:

How can Critical Theories of Feminism, Race, and Social Justice jointly inform a framework for advancing how students understand, value, and engage in socially just and equitable design, and provide them with knowledge, skills and attitudes to guide robot designs in equitable ways?

Specifically, in this work, we first identify four key frameworks relevant to social justice-oriented robotics curriculum. We then conduct a sereis of workshops with the leading experts in the aforementioned areas, during which we identify eleven relevant principles and eight strategies for designing social justice-oriented robotics curricula that help pursue those principles. While the framework presented in this manuscript is more theory-based, it is a necessary first step and will guide and inform subsequent projects that involve course design and piloting in specific classrooms.

3. KEY THEORETICAL FRAMEWORKS

In this section, we will first describe the four key frameworks in which our approach was grounded: Engineering for Social Justice, Feminist Human-Robot Interaction, Matrix-Guided Technology Power Analysis, and Critical Race Pedagogy.

3.1 Engineering for social justice

Engineering for Social Justice (E4SJ) is a set of engineering practices that aims to enhance human capabilities through equitable distribution of opportunities and resources while reducing imposed risks and harms [37]. In particular, E4SJ provides five guiding practices:

1. *Listening contextually* — Engineers need to listen to and empathize with different communities' perspectives, struggles, concerns, desires, and preferences.

2. *Identifying structural conditions* — These communities' perspectives must be understood through the lens of the structural conditions (e.g., racial, gendered, socioeconomic) that constrain the communities' opportunities, desires, and aspirations.

3. *Acknowledging political agency / mobilizing power* — Engineers must understand how communities' political power and agency (as well as their own) can be mobilized and leveraged when developing engineering solutions.

4. *Increasing opportunities and resources* — Engineers should work with communities to

identify the opportunities (e.g., health, education, housing, and employment) that could be improved by leveraging engineering solutions.

5. *Reducing imposed risks and harms* — Engineers should work with communities to identify how the solution's potential risks affect community members.

Taking the community-focused approach advocated by the E4SJ framework is critical to the development of technology that can bring genuine benefits to society. Technologies need to be designed with specific communities in mind because different communities have inherently different values and thus different goals. Navigating this tension requires one to clearly specify the community that's being designed for, to have an awareness of that specific community, and to develop a local understanding of that community (or, at least, explicitly building on relevant prior findings). Failure to integrate community values and perspectives risks inequitable technology that only prioritizes the values and needs of those who are already empowered by technology, such as engineers. In this work, we aim to understand how the aforementioned set of E4SJ principles can be applied in cultivating the next generation of roboticists. Achieving this goal, however, requires adopting specific practices for analyzing structural conditions, analyzing political agency, and analyzing connections between engineering solutions, structural conditions, opportunities, and resources. To achieve this, We argue that Feminist Analysis can provide important complementary tools that can help adapt these practices in the robotics context.

3.2 Feminist Human-Robot Interaction

Winkle et al. proposed the Feminist Human-Robot Interaction (FHRI) framework [38], which adapts D'Ignazio and Klein's Data Feminism to the field of robotics [39], using its definition of power: "The current configuration of structural privilege and structural oppression, in which some groups experience unearned advantages — because various systems have been designed by people like them and work for people like them – and other groups experience systematic disadvantages – because those same systems were not designed by them or with people like them in mind." It is worth noting that while feminism can carry the conventional definition that is antagonistic of patriarchy, in this context, feminism primarily evokes sensitivity towards power structure and injustices, which includes but is certainly not limited sexism. Data Feminism encourages the examination of how data is collected, and how data science tools are used, in relation to who data science serves and ultimately empowers. Through similar lenses of power, Winkle questions who gets to design and who gets to benefit and be empowered from the robot design process. The FHRI framework highlights that designers are not simply makers; They also hold power within our society that is wielded through and embodied in their robotic designs. FHRI provides guidelines for how robot designers can examine and challenge existing power structures in HRI both in their robot designs and in their research practices. More specifically, to examine power, FHRI suggests that we need to consider the existing power structure (i.e., how power currently operates in the world), and the context and ecology of the deployment of technology. To challenge power, Feminist HRI suggests that we prioritize a participatory approach, highlight invisible labor, value and prioritize diverging perspectives, elevate users' emotional state, and rethink social and power hierarchies.

While FHRI encourages a particular framework for analyzing power relations, we argue that roboticists also need to more fundamentally consider the different types of power that are wielded at different levels of analysis (for example, what are the different ways that such power dynamic can interact with user demographics? and how would such power dynamic shape the development of robotics technology?). We argue that insights from Black Feminist Thought can facilitate this multi-level analysis of power in the robotics context.

3.3 Matrix-Guided Technology Power Analysis

A key framework from Black Feminist Thought that can help us reason about power in robotics is the *Matrix of Domination* [40]: a framework for analyzing how intersecting systems of race-, class-, and gender-based oppression (*inter alia*) manifest and can be resisted. Critically, the Matrix of Domination analyzes four key domains of power: interpersonal, cultural, disciplinary, and structural: a broader and more nuanced set of domains than is typically considered in robotics.

Leveraging this framework, Williams presents Matrix-Guided Technology Power Analysis (MGTPA) as a framework that aims to understand, categorize, and disseminate the power that technologists wield, especially in robotics [41]. This framework is leveraged by Williams in *Degrees of Freedom: On Robotics and Social Justice* [42] to demonstrate how Roboticists, through the ways that they choose to design and use robotic technologies, hold significant power in domains beyond those typically discussed in the field of HRI. Specifically, MGTPA leverages the Matrix of Domination to highlight how roboticists wield not only Interpersonal power, but also Structural, Disciplinary, and Cultural power. MGTPA thus highlights the necessity of examining power comprehensively and the relevance of the Matrix of Domination for the field of Robotics.

While FHRI and MGTPA provide key analysis tools that allow E4SJ to guide robotics curricula, there remains a key gap in how to effectively integrate and teach these tools in the curricular modules used in the training of new roboticists. To achieve this goal, we leverage Critical Race Pedagogy.

3.4 Critical Race Pedagogy

Tanksley presented an Abolitionist pedagogy (ACRP), which draws on Critical Race Theory to critically engage with and challenge anti-Black racism within STEM education [43]. Specifically, Tanksley argues that by incorporating and prioritizing abolitionist ideologies in classrooms, STEM students, who are future engineers and designers, can approach technology more thoughtfully, and more importantly, they will be equipped with knowledge and tools to challenge racism and inequality in existing socio-technical systems. ACRP aims to facilitate a justice-oriented infrastructure, and encourages educators to design curricula that can guide students to critically evaluate and probe the technology they might encounter. Specifically, CRP presents five tenets of CRT-informed CS pedagogy:

1. Acknowledge anti-black racism is pervasive and deeply embedded in our society, and thus has been codified into technology.

2. Encourage scholars to critically examine the existing dominant narratives, and to push back against the idea that technology always serves everyone equally.

3. Commit to actively pursuing social justice by pushing back against all forms of oppression that are produced on the basis of gender, race or others
4. Recognize that students of color are holders of knowledge and have valuable insights that stem from their life experiences.
5. Prioritize an interdisciplinary approach that consults scholarship in all relevant fields.

Motivated by these four guiding theoretical frameworks, we set out to identify convergences between these frameworks, and in doing so, identify overarching principles informed by these frameworks, and strategies for achieving those principles.

4. METHODOLOGY AND PARTICIPANT BACKGROUND

To identify our key set of principles and strategies, we convened the leading experts on the four key theoretical frameworks described above. Together, the group engaged in iterative discussions and workshops to synthesize the four key frameworks in the context of robotics education [44]. Workshop sessions were recorded or included note-taking that authors can revisit. Specifically, we began by conducting a series of workshops introducing each of the frameworks relevant to this project. During these workshops, the creators of the E4SJ, FHRI, MGTPA and ACRP frameworks led presentations on their frameworks, describing the frameworks' development, and their key insights and recommendations. These creators are also authors of this work, and they all actively pursue social justice through research and educational praxis at the intersection of robotics, engineering education, and critical technology studies. As workshop participants, they bring fluency in their expert domain and their lived experiences as educators. Three participants are based in the US and one participant is based in Sweden. Then, workshop participants discussed the relevance of these insights to robotics curricula. Through this process, participants drew connections between frameworks and identified both common threads (for example, both E4SJ and MGTPA highlight the importance of acknowledging and analyzing power structures and power dynamics) and distinct insights (for example, FHRI advocates for explicit consideration of how the physical embodiment of robots can affect deployment).

Following this step, the authors collaboratively conducted iterative thematic analysis to find additional connections and points of divergence. Specifically, we identified key **principles** encouraged by each proposed framework, identified key **strategies** used to answer those questions under each framework, and discussed existing tools and activities that are used or can be used in pedagogical contexts to pursue each of those strategies. When discussing principles derived from the four key frameworks, the authors discussed each principle in depth, until reaching agreement about each principle's relevancy to the context and goals of justice-centered robotics curricula. Similarly, when discussing strategies suggested by each of the four key frameworks, each strategy was discussed and connected back to relevant principles. Conceptually, we synthesize and present the principles as the important areas that articulate what justice-centered robotics education should attend to, and the strategies as actionable pedagogical approaches that educators can leverage to highlight these areas. Finally, the team synthesized these insights into a single overarching framework. In this step, we revisited the connections between our four key frameworks, and their associated principles and strategies. The authors discussed principle or strategies with similar definitions in order to discover

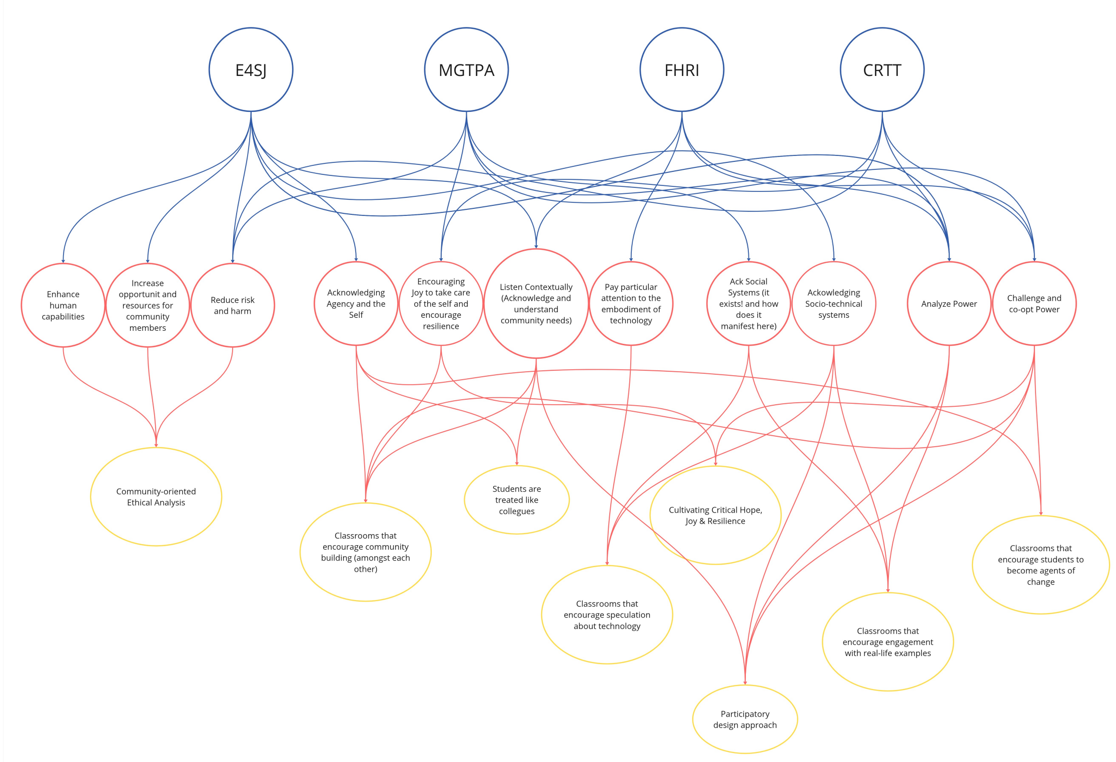

Figure 1: Figure showing the thematic coding mindmap that illustrates connections between prior work, emerging principles, and relevant tools. The upper-level blue codes represent four prior works, the middle-level red codes represent eleven principles, and the lower-level yellow codes represent the strategies. The red codes also show proximity that illustrates the four clusters used in this paper

possible redundancies that could be eliminated to yield a less complex framework. Overall, this process yielded eleven key principles grounded in our four frameworks, and eight key strategies for achieving those principles.

5. RESULTS

In this section, we will now discuss the eleven key principles and eight key strategies that resulted from this research effort, Fig 1 visually shows the connection between prior work discussed above and the key principles and strategies. Additionally, Table 2 shows how each prior work informs and ties into each principle.

5.1 Principles

Our methodology produced eleven key principles, which fell into four clusters, each of which is one step farther removed from the robot design process as it is traditionally conceived (The mapping between each of these principles and the four frameworks is shown in Table 2):

Principle code	Engineering For Social Justice (E4SJ)	Matrix-Guided Technology Power Analysis (MGTPA)	Feminist Human-Robot Interaction (FHRI)	Abolitionist Critical Race Pedagogy (ACRP)
R1	X			
R2	X			
R3	X	X		X
R4	X			
R5				X
R6	X		X	
R7			X	
R8		X		
R9	X			
R10	X	X	X	X
R11		X	X	X

Table 1: Table showing how each principle relates to key prior work, each cross indicates that the specific prior work highlights the particular principle

- **Principle Cluster 1:** Ways that the goals of robot design should be framed,
- **Principle Cluster 2:** Ways that students should conceptualize and relate to different types of agents (stakeholders, robots, and themselves)
- **Principle Cluster 3:** Ways that students should engage with larger social and socio-technical systems
- **Principle Cluster 4:** Ways that students should analyze power within those systems.

5.1.1 Principle Cluster 1: Design

The first cluster of principles specify the ways that students should be taught to conceptualize the goals of robot design.

- *Enhance human capabilities (R1)*: Robots designs should aim to enhance one or more key human capabilities as delineated in E4SJ framework. Without intentionally considering human capabilities, robot designs risk serving only the designers, engineers or enterprise instead of improving human flourishing.

- *Increase opportunities and resources for community members (R2)*: Students should prioritize how technology can increase opportunities and resources for the communities that they aim to design for. While existing robot designs often consider the specific interaction with the intended user, it is also important to consider how robot designs can uplift the community that it's deployed in. For example, if a robot is designed to be stationed in a public library, instead of merely directing visitors to books or resources, it could encourage them to join reading groups or participate in library events, which can help nurture community building.

- *Reduce risk and harm (R3)*: Students should aim to reduce the potential risks and harms associated with robot designs. While robots can bring benefits such as productivity and companionship, they can also present harm such as unmediated disruption to the workforce and minimizing human connections. It is therefore critical to consider and minimize potential harm, to ensure robots can be deployed responsibly.

5.1.2 Principle Cluster 2: Agents

Goals for attending to different agents (stakeholders, students, robots): In this set of principles, we zoom out to attend to different agents in different ways, including the roboticist's self, the robot, and the stakeholders.

- *Acknowledging Agency and the Self (R4)*: Students should develop the ability to acknowledge their own agency as future roboticists, and to recognize how their own priorities may interact with their professional identities. Robots can convey and amplify the priorities of their makers. A roboticist with normal vision might not consider visually accessible designs for the robot; in contrast, a roboticist who has experience with disabilities may be more conscientious in ensuring accessibility in the robot design. It is therefore necessary to facilitate understanding of student agency and self to help students realize their potential biases and recognize their personal priorities.

- *Encouraging Joy to take care of the self and encourage resilience (R5)*: Joyful participation is the key to enabling students to speak up and subvert existing biases and unjust systems. Conventionally hierarchical class structures can be oppressive. A joyous environment can help subvert such hierarchy, and help students gain confidence and the agency to challenge bias or injustice that they may encounter. Moreover, such joy may protect students socially and emotionally by fostering a sense of community and belonging, which can help students remain resilient in the face of negative content and events.

- *Listen Contextually (Acknowledge and understand community needs (R6)*: Students should develop contextual listening skills to help receive, understand, and prioritize the needs and goals of stakeholders. For instance, robots are often designed to improve efficiency and alleviate shortages in certain resources. While these can be positive goals, contextual listening helps ensure that robots are not developed to address hallucinated needs. Importantly, contextual listening can help provide concrete answers to the question *"Is this a problem that a robot **should** solve?"*

- *Pay particular attention to the embodiment of technology (R7)*: Robotics students need to pay particular attention to the current or intended physical embodiment of robots, and to have the competency to investigate the implications of particular embodiments. Robots are distinct from tools such as chatbots or interfaces [45], as they often have physical appearances that resemble familiar social agents such as humans or animals, which facilitate robots becoming accepted into communities as social agents. As such, a robot's embodiment and appearance deserves careful consideration. For instance, the mere size of the robot can affect trustworthiness [46]. Therefore, it's important for roboticists to develop particular competence to investigate the effects of robot embodiments.

5.1.3 Principle Cluster 3: Systems

Goals for engaging with the larger social and socio-technical systems in which technology exists: In this set of principles, we zoom out further to acknowledge and think about the larger social and socio-technical systems in which our work is embedded.

- *Acknowledging Social Systems (R8)*: Students should have awareness of the social system that surrounds the design, research, production, and deployment of robots. Many application domains for robots, such as education, manufacturing, and healthcare, exist within complex social systems. For instance, researchers have developed socially assistive robots that can improve learning for students who need special education services [47]. However, existing social structures within special education create barriers in adopting assistive robots in this context. Special education teachers are often not financially compensated or logistically supported to complete optional professional development and improve programs, meaning that they face additional burden when adopting assistive robots in their classrooms. Roboticists need to therefore recognize relevant social systems and their impact on robotic technology.

- *Acknowledging Socio-technical systems (R9)*: Students should recognize how the social structures surrounding robotics interact with and shape technical design, deployment, and the ultimate impact of robotic systems. For example, as manufacturing robots are being deployed to complete repetitive tasks, the ultimate outcome of deployment depends on how they align with existing labor structures, including training and safety protocols. If roboticists do not take into consideration human workers' technical expertise and collaborative work patterns, then the introduction of such robots can cause accidents, disruptions, or face resistance [48]. Roboticists must therefore consider how technical choices will interact with broader social systems, and guide such interaction to ensure robots can deliver true benefits.

5.1.4 Principle Cluster 4: Power

Goals for analyzing and challenging power within those systems: Power is deeply embedded in robot design, from who defines a problem, to whose needs are prioritized, to who ultimately benefits from a robot's deployment. In this set of principles, we analyze and challenge power in order to loop back and effectively address the goals for robot design.

- *Acknowledge Power (R10)*: Students should be able to acknowledge the power structures that affect their intended domain, community, and robot deployment. For instance, there has been much interest in developing policing robots that assist patrol, surveillance, and even arrests. The larger context of Policing is an institution that was created to (and continues to) enforce and reify boundaries of race [49], class [50, 51], and gender [52], and robots used by police are likely to continue to enforce these boundaries, and exacerbate existing power imbalance and injustice. Roboticists of the future therefore need to develop the skill to acknowledge and understand how power imbalance can shape robot development.

- *Challenge and Co-Opt Power (R11)*: Students should have the skills to challenge and subvert existing power imbalances, and to create and channel power to benefit social justice aims. In the case of policing robots described above, this could mean that roboticists resist the development of robots that reinforce existing surveillance patterns and instead direct their technical skills toward technology that supports community safety without continuing the existing punitive patterns of policing. By intentionally pursuing goals that redistribute power, roboticists can become leaders and advocates for systems that do not exacerbate oppression.

Principle code	Engineering For Social Justice (E4SJ)	Matrix-Guided Technology Power Analysis (MGTPA)	Feminist Human-Robot Interaction (FHRI)	Abolitionist Critical Race Pedagogy (ACRP)
R1	X			
R2	X			
R3	X	X		X
R4	X			
R5				X
R6	X		X	
R7			X	
R8		X		
R9	X			
R10	X	X	X	X
R11		X	X	X

Table 2: Table showing how each principle relates to key prior work, each cross indicates that the specific prior work highlights the particular principle

5.2 Emerging Strategies

Recognizing the four clusters of principles, we then started to identify and cluster key strategies employed by educators to pursue these principles. To perform clustering, we connect each identified strategy to the relevant principles. These connections are illustrated in Figure 1. This step of the proposed framework is still ongoing, as we intend to further iterate and polish the identified strategies in planned future work. Thus, in this paper, we provide a few examples of currently identified strategies, interpretation and how they connect to various principles to illustrate how principles can be translated into concrete strategies and actions.

5.2.1 Community-oriented Ethical Analysis

Students perform community-oriented ethical analysis of potential robot design and deployment. Such analysis should involve diverse stakeholders, and more specifically, students should evaluate if the intended robot design can enhance human capabilities **(R1)**, increase opportunities and resources for community members **(R2)**, and reduce risks and harm **(R3)**. Specifically in the context of robotics, such skill is critical as embodied technology carries much higher level of risks for causing harm. This process and set of priorities can help students move away from abstract ideals and principles, and move into community-oriented analysis that can help ensure a balanced consideration of both potential benefits and harms.

5.2.2 Classrooms that encourage community building amongst students

Students form a community with peers to help them recognize how their own agency and priorities affect their goals **(R4)**, to facilitate joyous participation **(R5)**, and to aid development of their skills in contextual listening **(R6)**. While communities often appear in the context of intended users or stakeholders, community building amongst students can also facilitate skill building. In particular, robotics courses often focus on learning specific components such as motion planning or perception, and students can become isolated. By improving community building amongst students, learning itself can become more joyous. and students can become more comfortable and supported to explore their own identities as roboticists together with peers.

5.2.3 Classrooms that encourage speculation about technology

Students are guided to speculate about a possible future their intended robot design could enable, which allows situated exercises around important aspects of the robot design, such as the particular embodiment of the robot **(R7)**, and the possible socio-technical implications of the intended robot design **(R8) (R9)**. Robotics is a type of emerging technology, as such, many robotics courses will envision a possible near future where certain types of robotics technology are not limited to classrooms or laboratories but are truly ubiquitous. In this way, the nascence of this field almost inevitably assigns a speculative nature to robotics courses. It is therefore important for students to learn the necessary skills to competently speculate about possible robot deployment and how robot designs and socio-technical systems can shape such deployment. This process should help students and become active narrators instead of passive listeners during the process.

5.2.4 Participatory design approach

Students are guided to follow participatory design principles and methodologies, where they will get to develop contextual listening skills **(R6)**, and through which process they will acknowledge the socio-technical context **(R9)** around robots and to recognize and challenge existing power structures around robot design **(R10) (R11)**. Participatory design emphasizes collaboration with community members as co-designers, not just as end-users or passive recipients of technology. In robotics classrooms, this can include design exercises, game-based design activities, or stakeholder interviews where students practice contextual listening of community priorities and needs, and making design decisions that center concrete community voices. For instance, students designing a healthcare robot might engage not only doctors, nurses, patients, but also family members to better understand not only functional requirements but also the socio-technical context of care. By situating robot design within a participatory process, students learn that design is not only a technical activity but also a blending of values and power. These classrooms can thus help train students to think critically about whose voices are amplified or silenced in design processes and to develop tools and strategies to redistribute power.

6. DISCUSSION

6.1 Theoretical and Practical implications

As noted earlier in this paper, our motivation for constructing this framework stems from the lack of social-justice centered pedagogical frameworks and curricula for robotics education. By drawing on distinct but complementary prior theories, we synthesize foundational principles and identify key areas of consideration. We recognize there are ongoing efforts in engineering education to depoliticize engineering, where the technical and social aspects are viewed as separate and independent of each other. In contrast to those trends, in this work, we detail how robotic systems are inherently socio-technical in the effort to make clear that a depoliticized approach is counter-productive, as other scholars have also cautioned against depoliticization [53]. We argue that this framework offers value not only to robotics educators, but also stands to provide a shared structure for critically examining how technical design, pedagogy, and societal impact are deeply intertwined.

6.2 Limitation and future work

A unique aspect of this work lies in the the construction of the R4SJ framework. During this process, we brought together a research team composed of researchers from the fields of Human-robot interaction, abolitionist pedagogy, engineering education and organizational psychology. This team makeup enabled us to uniquely identify and synthesize perspectives from different disciplinary traditions. Nevertheless, we acknowledge that we may have benefited from involving an even more diverse set of stakeholders. Most notably, we note that student perspectives were largely missing from our research team, beyond the lead author. In future work, we plan to conduct semi-structured interviews and focus groups that will include both instructors *and students*, in order to adequately include and consider student perspectives and needs. We also acknowledge there may be a lack of practical implications of the frameworks. Although we intend for this work to lay a theoretical foundation, we plan to incorporate principles and strategies in actual curricula and classrooms to further extend this work from theory to practice.

A second limitation concerns the replicability of our specific methodology. In order to account for the complex socio-technical context that surrounds robotic technology, this work builds on a diverse set of frameworks and expertise. We acknowledge that it is challenging to assemble an interdisciplinary team like this, and that this difficulty may prevent other researchers and educators are interested in adopting the methodology used in this work in their own discipline and field. Therefore, while we believe the interdisciplinary approach was necessary for building this foundational framework, we plan to further refine our methodology and results in hopes to benefit more community members within engineering education.

7. CONCLUSION

This work presents a social-justice-centered pedagogical framework for robotics education that highlights the interplay between technical design and social context. By synthesizing insights from robotics, engineering education and abolitionist pedagogy, the framework offers educators a shared structure for critically examining how robotic is taught, designed, and evaluated. We argue that engaging with these tensions is essential for preparing roboticists to responsibly design technologies embedded in real-world social systems, and we see this framework as a foundation for ongoing interdisciplinary dialogue and empirical evaluation.

8. ACKNOWLEDGMENT

This work was supported by NSF grant SBE-2418199.

References

[1] Lai Poh Emily Toh, Albert Causo, Pei-Wen Tzuo, I-Ming Chen, and Song Huat Yeo. A review on the use of robots in education and young children. *Journal of Educational Technology & Society*, 19(2):148–163, 2016.

[2] Kai-Yi Chin, Zeng-Wei Hong, and Yen-Lin Chen. Impact of using an educational robot-based learning system on students' motivation in elementary education. *IEEE Transactions on learning technologies*, 7(4):333–345, 2014.

[3] Joost Broekens, Marcel Heerink, Henk Rosendal, et al. Assistive social robots in elderly care: a review. *Gerontechnology*, 8(2):94–103, 2009.

[4] Laura Stegner and Bilge Mutlu. Designing for caregiving: Integrating robotic assistance in senior living communities. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*, pages 1934–1947, 2022.

[5] Hideki Kozima, Cocoro Nakagawa, and Yuriko Yasuda. Interactive robots for communication-care: A case-study in autism therapy. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, pages 341–346. IEEE, 2005.

[6] Mohammed A Saleh, Fazah Akhtar Hanapiah, and Habibah Hashim. Robot applications for autism: a comprehensive review. *Disability and Rehabilitation: Assistive Technology*, 16(6):580–602, 2021.

[7] Carlos A Cifuentes, Maria J Pinto, Nathalia Céspedes, and Marcela Múnera. Social robots in therapy and care. *Current Robotics Reports*, 1:59–74, 2020.

[8] Virginia Eubanks. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press, 2018.

[9] Frank Pasquale. *The black box society: The secret algorithms that control money and information*. Harvard University Press, 2015.

[10] Nima Kordzadeh and Maryam Ghasemaghaei. Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*, 31(3):388–409, 2022.

[11] Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. Language (technology) is power: A critical survey of" bias" in nlp. *arXiv preprint arXiv:2005.14050*, 2020.

[12] Bryan Pfaffenberger. Social anthropology of technology. *Annual review of Anthropology*, pages 491–516, 1992.

[13] Bing Cai Kok and Harold Soh. Trust in robots: Challenges and opportunities. *Current Robotics Reports*, 1(4):297–309, 2020.

[14] Shanee Honig and Tal Oron-Gilad. Understanding and resolving failures in human-robot interaction: Literature review and model development. *Frontiers in psychology*, 9:861, 2018.

[15] Terran Mott and Tom Williams. Confrontation and cultivation: understanding perspectives on robot responses to norm violations. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 2336–2343. IEEE, 2023.

[16] Jon A Leydens, Kathryn Johnson, Stephanie Claussen, Jenifer Blacklock, Barbara M Moskal, and Olivia Cordova. Measuring change over time in sociotechnical thinking: A survey/validation model for sociotechnical habits of mind. In *2018 ASEE Annual Conference & Exposition*, 2018.

[17] Donna Riley and Lionel Claris. Power/knowledge: Using foucault to promote critical understandings of content and pedagogy in engineering thermodynamics. In *2006 Annual Conference & Exposition*, pages 11–1003, 2006.

[18] Deborah G Johnson. Computer ethics. *The Blackwell guide to the philosophy of computing and information*, pages 63–75, 2004.

[19] Ronald E Anderson. Social impacts of computing: Codes of professional ethics. *Social Science Computer Review*, 10(4):453–469, 1992.

[20] IEEE. Ieee code of ethics. `https://www.ieee.org/about/corporate/governance/p7-8`, 2025. Accessed: 2025-07-10.

[21] Belinda Louie, Elin A Björling, Annie Camey Kuo, and Patrícia Alves-Oliveira. Designing for culturally responsive social robots: An application of a participatory framework. *Frontiers in Robotics and AI*, 9:983408, 2022.

[22] Ruth Stock-Homburg. Survey of emotions in human–robot interactions: Perspectives from robotic psychology on 20 years of research. *International Journal of Social Robotics*, 14(2):389–411, 2022.

[23] Anna Henschel, Guy Laban, and Emily S Cross. What makes a robot social? a review of social robots from science fiction to a home or hospital near you. *Current Robotics Reports*, 2(1):9–19, 2021.

[24] Katie Seaborn, Giulia Barbareschi, and Shruti Chandra. Not only weird but "uncanny"? a systematic review of diversity in human–robot interaction research. *International Journal of Social Robotics*, 15(11):1841–1870, 2023.

[25] Tom Williams. Race in the eye of the robot beholder: Against racial representation, recognition, and reasoning in robotics research. In *Proceedings of the 2022 Inclusive HRI Workshop on Equity and Diversity in Design, Application, Methods, and Community*, 2022.

[26] Yoyo Tsung-Yu Hou, EunJeong Cheon, and Malte F Jung. Power in human-robot interaction. In *Proceedings of the 2024 ACM/IEEE international conference on human-robot interaction*, pages 269–282, 2024.

[27] Jay Dixon, Bryan Hong, and Lynn Wu. The robot revolution: Managerial and employment consequences for firms. *Management Science*, 67(9):5586–5605, 2021.

[28] John Chung and Yong Suk Lee. The evolving impact of robots on jobs. *ILR Review*, 76(2):290–319, 2023.

[29] Andrew Hundt, William Agnew, Vicky Zeng, Severin Kacianka, and Matthew Gombolay. Robots enact malignant stereotypes. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 743–756, 2022.

[30] Katie Winkle, Ryan Blake Jackson, Gaspar Isaac Melsión, Dražen Bršćić, Iolanda Leite, and Tom Williams. Norm-breaking responses to sexist abuse: a cross-cultural human robot interaction study. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 120–129. IEEE, 2022.

[31] Jennifer Robertson. Gendering humanoid robots: Robo-sexism in japan. *Body & Society*, 16(2):1–36, 2010.

[32] Andrew Berg, Edward F Buffie, and Luis-Felipe Zanna. Robots, growth, and inequality. *Finance & Development*, 53(3):10–13, 2016.

[33] Sultan Almuaythir, Atul Kumar Singh, Mohammad Alhusban, and Ahmed Osama Daoud. Robotics technology: catalyst for sustainable development—impact on innovation, healthcare, inequality, and economic growth. *Discover Sustainability*, 5(1):486, 2024.

[34] 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, 2018.

[35] Ian GR Shaw. Robot wars: Us empire and geopolitics in the robotic age. *Security dialogue*, 48(5):451–470, 2017.

[36] Neda Atanasoski and Kalindi Vora. *Surrogate humanity: Race, robots, and the politics of technological futures*. Duke University Press, 2019.

[37] Jon A. Leydens and Juan C. Lucena. *Engineering Justice: Transforming Engineering Education and Practice*. John Wiley & Sons, Hoboken, NJ, 2017.

[38] Katie Winkle, Donald McMillan, Maria Arnelid, Katherine Harrison, Madeline Balaam, Ericka Johnson, and Iolanda Leite. Feminist human-robot interaction: Disentangling power, principles and practice for better, more ethical hri. In *Proceedings of the 2023 ACM/IEEE international conference on human-robot interaction*, pages 72–82, 2023.

[39] Catherine D'ignazio and Lauren F Klein. *Data feminism*. MIT press, 2023.

[40] Patricia Hill Collins. *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. routledge, 2022.

[41] Tom Williams. Understanding roboticists' power through matrix guided technology power analysis. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 46–56, 2024.

[42] Tom Williams. *Degrees of Freedom: On Robotics and Social Justice*. MIT Press, 2025.

[43] Tiera Chante Tanksley. "we're changing the system with this one": Black students using critical race algorithmic literacies to subvert and survive ai-mediated racism in school. *English Teaching: Practice & Critique*, 23(1):36–56, 2024.

[44] Rikke Ørngreen and Karin Levinsen. Workshops as a research methodology. *Electronic Journal of E-learning*, 15(1):70–81, 2017.

[45] Peter H Kahn Jr and Solace Shen. Noc noc, who's there? a new ontological category (noc) for social robots. *New perspectives on human development*, pages 106–122, 2017.

[46] Peter A Hancock, Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, Ewart J De Visser, and Raja Parasuraman. A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, 53(5):517–527, 2011.

[47] Shane Romero, Saad Elbeleidy, and Tom Williams. Understanding barriers to entry and invisible labor for educational care wizards. In *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*, pages 1287–1293. IEEE, 2024.

[48] Jo Ann Oravec. Robo-rage against the machine: Abuse, sabotage, and bullying of robots and autonomous vehicles. In *Good Robot, Bad Robot: Dark and Creepy Sides of Robotics, Autonomous Vehicles, and AI*, pages 205–244. Springer, 2022.

[49] Benjamin Bowling. *Violent racism: Victimization, policing, and social context*. Oxford University Press, USA, 1999.

[50] Alex S Vitale. *The End of Policing*. Verso Books, 2017.

[51] Kristian Williams. *Our enemies in blue: Police and power in America*. AK Press, 2015.

[52] Mike Rowe. *Policing, race and racism*. Willan, 2012.

[53] Erin A Cech. The (mis) framing of social justice: Why ideologies of depoliticization and meritocracy hinder engineers' ability to think about social injustices. In *Engineering education for social justice: Critical explorations and opportunities*, pages 67–84. Springer, 2013.