

Diversity Equity and Inclusion in Embodied AI : Reflecting on and Re-imagining our Future with Embodied AI

Original

Diversity Equity and Inclusion in Embodied AI : Reflecting on and Re-imagining our Future with Embodied AI / Zaga, Cristina; Lupetti, MARIA LUCE; Cila, Nazli; Lee, Minha; Huisman, Gijs; Fosch Villaronga, Eduard. - (2022).
[10.3990/1.9789036553599]

Availability:

This version is available at: 11583/2987031 since: 2024-03-18T12:03:23Z

Publisher:

University of Twente

Published

DOI:10.3990/1.9789036553599

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

Diversity Equity and Inclusion in Embodied AI

→ CRISTINA ZAGA, MARIA LUCE LUPETTI, NAZLI CILA,
MINHA LEE, GIUS HUISMAN, EDUARD FOSCH VILLARONGA

Reflecting on and Re-imagining
our Future with Embodied AI

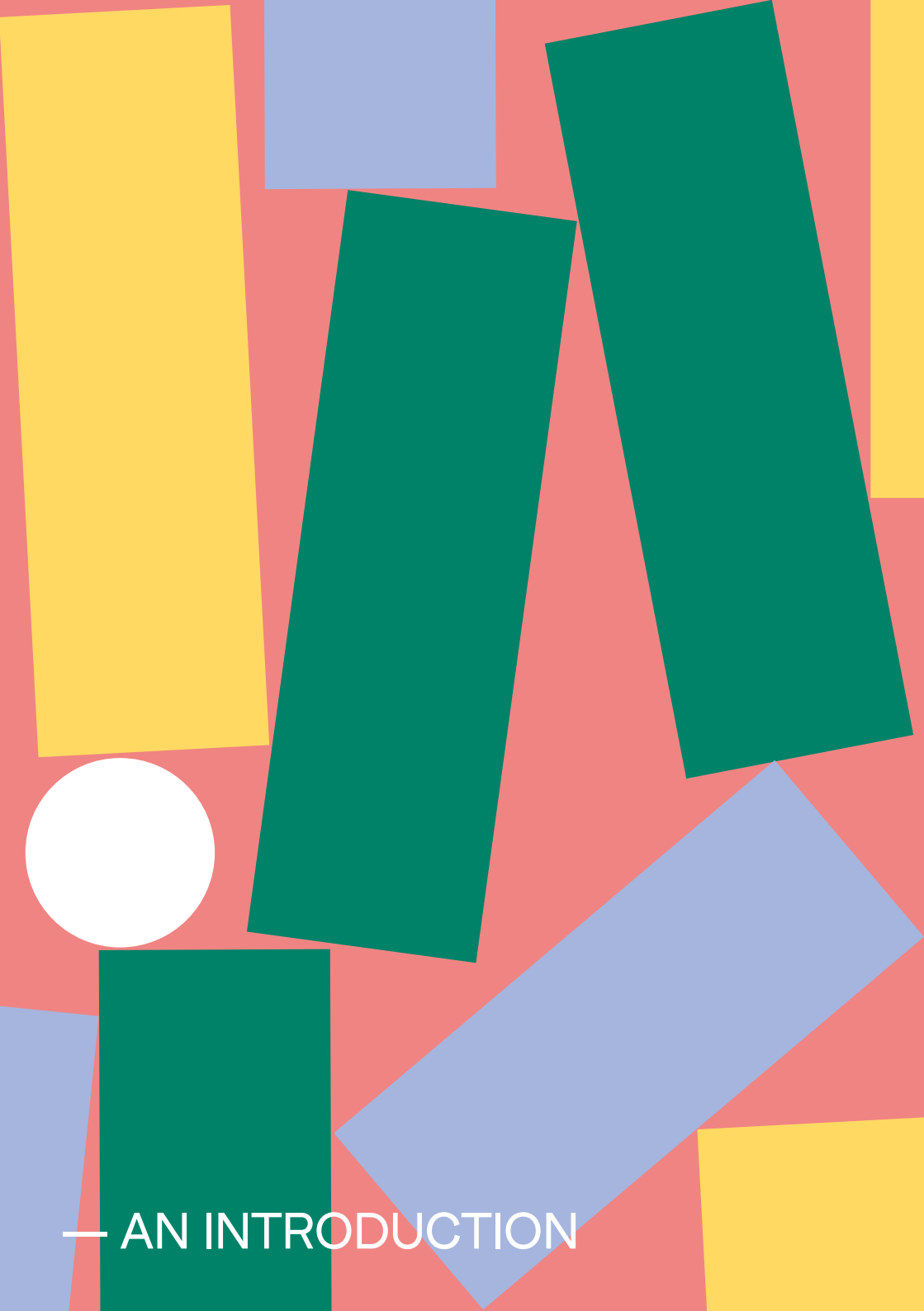
DIVERSITY EQUITY AND INCLUSION
IN EMBODIED AI: REFLECTING ON AND RE-IMAGINING
OUR FUTURE WITH EMBODIED AI

Technology often works silently as a double-edged sword. On the one hand, it has the power to amplify inequality, reinforce existing stereotypes, and further push people into categories that do not represent them. On the other, its design may help devise strategies and counternarratives that redirect ongoing discourses towards a fairer and more inclusive society.

However, giving voice to and operationalizing these reflections is difficult, especially in new complex fields such as robotics and artificial intelligence (AI). To this aim, this book collects the reflections, insights, and tools resulting from the Diversity, Equity, and Inclusion for Embodied AI (DEI4EAI) project.

This book is intended for students, researchers, designers, developers, and societal stakeholders working with embodied AI and interested in contributing to more equitable and just futures. All those things dubbed as ordinary are in fact, so cultural: they represent values, beliefs, and narratives that influence how we collect and use data, how we craft algorithms, how we define agency, how we mold AI embodiment, how we design interaction, and how we define embodied AI intervention. Although in different roles and capacities, designers, researchers, and broader stakeholders like policymakers and communities are responsible for reflecting on how their values, perspectives, biases, and stereotypes may affect embodied AI technology.

This is important because siloed practices influence our capacity to assess the risks and harms of our actions. To prevent designing harmful and inadequate technology, there is the need to inspect narratives, practices, and methods with reflexivity and openness to shift mindsets.



— AN INTRODUCTION

Reflecting on and Re— imagining our Future with Embodied AI

EMBODIED AI: A BEACON OF INNOVATION, OR AN OPPRESSIVE MIRROR OF OUR POWER AND SOCIETAL STRUCTURES?

Smart objects, robots, and conversational user interfaces constitute what we, academics and technologists, call embodied AI. Embodied AI has many conflicting definitions (Ziemke 2004). For this book, we describe embodied AI systems as a set of physically realized machines that have a body with various shapes, biologically or not biologically informed by theories of cognition (i.e., conscious and unconscious processes by which knowledge is accumulated, such as perceiving, recognizing, conceiving, and reasoning) (Franklin 1997). Embodied AI systems are

designed to collect and sense data (e.g., vision, audio, speech, etc..) from the environment, learn from the sensors in their body, and communicate accordingly through various modalities (e.g., voice, sounds, lights, movements) and through actions (Deng, Mutlu, and Mataric 2019; Ziemke and Lowe 2009). Smart assistants (Kěpuska and Bohouta 2018), socially assistive robots (Feil-Seifer and Mataric 2005), autonomous cars (Hussain and Zeadally Secondquarter 2019), and

many other embodied systems (Kong and Jeon 2006; Piyare and Lee 2013; Belpaeme et al. 2018; Allen et al. 2001; Cassell 2001; Dellon and Matsuoka 2007) are profusely studied, and there are many advocates for them to become an integral part



The figure is “oppressive mirror of our power and societal structures” by Dall-e Mini, by Crayon

of our everyday life (Belpaeme et al. 2018; Broekens, Heerink, and Rosendal 2009; Bharti et al. 2020; Pradhan, Findlater, and Lazar 2019). Some of these systems are starting to become consumer products for our households and cities (e.g., smart assistants like Google Home or Alexa, some autonomous transportation or logistic systems); others — besides the hype and excitement that they generate (Benson 2020) — are not yet ready for deployment in society (e.g., socially assistive robots).

Embodied AI and AI systems in general are often referred to as the beacon of innovation and progress. Indeed, AI is well-positioned to advance tools and “solutions” to address current social, economic, and environmental challenges humanity faces (Tomašev et al. 2020). As the Oxford Initiative on AI and Social Development goals (“AI4SDG” n.d.) shows, there are many examples of the usefulness of AI and embodied AI. From supporting responsible and sustainable agriculture, to systems for reducing educational inequality. From low-cost diagnostic devices in healthcare to rapid evaluation of small toxic particles in the air. From robot rescuers to robot nurses supporting overburdened healthcare systems.

However, AI and embodied AI technology are presently often harmful (Crawford et al. 2019).

‘The current state of our relationship with artificial intelligence (AI): the technology has advanced to achieve astounding feats, but its value system is lingering behind. In turn, this is symptomatic of stagnation in our socio-moral framework. Even though social movements in the last century have introduced many nuances to complex issues such as race, gender, and power, their mirror images in AI development remain overwhelmingly simplistic, reductionist, and sometimes laughably clueless’

AI and embodied systems can be discriminatory on gender, race, and abilities. Systems might misgender people or not recognize gender identities beyond the binary categories of male and female (Keyes 2018). Computer vision algorithms do not recognize and fail to classify people of color and label them as a potential danger to society (Buolamwini and Gebru 23–24 Feb 2018). The decisions and predictions made by AI can be “weapons” to increment power asymmetry

and gatekeeping access to services and opportunities. For example, excluding women from job opportunities or not providing access to credit or education. The roles, interaction models, and features we imbue embodied AI can reinforce stereotypes, augment biases, and become renewed forms of coloniality. Smart assistants are, for example, often modeled onto female submissive personas and reinforce homogeneity in language and culture (Lee et al. 2021). The social, health, and educative interventions we set embodied AI often reinforce discriminatory and outdated models, nudging to “normative” human behaviors and policing what it means to be human. For example, embodied AI systems for care focus on the interaction paradigm deemed to be ableist. Robots, IOTs, and other systems are often imbued with technoableist principles, making them more subject to scrutiny.

This is the privilege hazard: the phenomenon that makes those who occupy the most privileged positions among us - those with good educations, respected credentials, and professional accolades - so poorly equipped to recognize instances of oppression in the world. They lack of what Anita Gurumrthy, executive director of IT for Change, has called “the empiricism of lived experience”. And this lack of lived experience - this evidence of how things truly are - profoundly limits their ability to foresee and prevent harm, to identify existing problems in the world, and to imagine possible solutions. [...] So it matters deeply that data science and artificial intelligence are dominated by elite white men because it means there is a collective privilege hazard so great that it would be a profound surprise if they could actually identify instances of bias prior to unleashing them into the world.

Why is this happening? Why does a technology that serves humanity bring more issues than it solves? Why do “*algorithmic systems tend to become inaccurate, absurd, harmful, and oppressive*” (Alkhatib 2021)? Why do robots perpetuate race and gender stereotypes (Perugia et al. 2022; Hundt et al. 2022)? Is AI an evil twin of humanity (Rosenberg and Unanimous 2022)? Why are technologies meant to provide services and assistance in discriminating and policing the lives of people with little power (MacInnes 2021)? Do engineers, computer scientists, and designers have bad intentions?

↑ Catherine D’Ignazio and Lauren F. Klein, The power chapter in Data Feminism, pp. 29 The MIT Press 2020 (D’Ignazio and Klein 2020)

Research shows that the reasons are multifaceted and that we must pay close attention to the complex and tangled power issues and societal and ethical implications (Verma 2019; Birhane and Cummins 2019; Crawford 2021). We are witnessing and perpetuating the beliefs that AI technology is a superior solution to human problems. Technological solutionism is ingrained in our practices: we think that most of our challenges can be tackled with AI solutions.

At the same time, most of the discourse around AI fails to acknowledge its non-neutrality. Current mainstream narratives about embodied AI technology consider it neutral (i.e., technology is just a tool), overly positive (utopian, e.g., robots will make better decisions than humans, hence there is no need to act), or overly negative (dystopian, e.g., robots will take over the world and destroy us). In reality, technology is neither positive, negative, nor neutral. AI technology is a vehicle of values and worldviews.

We often fail to recognize how AI and embodied AI has socio-political and ethical implications. AI technology and embodied AI are social and political

artifacts (Winner 1980) (i.e., human-made objects that embody forms of authority and subordination, socio-cultural structures) and socio-technical artifacts that mediate our experience in the world. Foremost, we overlook how AI is an expression of power, oppressing — to say it with the words of the philosopher Rosy Braidotti (Braidotti 2022) — a multitude of otherness: sexualized others, racialized others, able others, etc... Embodied AI is made by humans immersed in a

particular culture, political and economic system, and with particular agendas. AI and embodied AI reflects our human world (Arista et al. 2021), our values (Birhane et al. 2022), and our biases (D'Ignazio and Klein 2020), and therefore, it is an expression of our world and of who rules our world (Crawford 2021). At this particular moment, AI and embodied AI is the expression of a minority of academic institutions



The figure is "algorithms and systematic oppression" by Dall-e Mini, by Crayon

and companies in the so-called WEIRD societies (white, educated, industrialized, religious, and democratic). We are subject to what D'Ignazio and Klein call *privilege hazard*: those with privilege cannot easily take the perspective of those who are oppressed as they lack their lived experience (D'Ignazio and Klein 2020).

These phenomena are arguably amplified in embodied AI because it is physically present (Howley et al. 2014), with a 'body' that occupies space while interacting with us. Embodied AI is meant to converse with us and share tasks (Cila 2022). To interact with embodied AI, people easily imbue embodied AI with the illusion of life, which is, of course, a deception. And while a certain degree of deception is necessary to ease our interaction with AI machines, the potential ways in which embodied AI shapes our reality and interaction with the world is far greater than disembodied AI.

It does not matter if embodied AI is purposefully designed to resemble a human's body or behavior: humans are hard-wired to anthropomorphize anything that moves or behaves and ascribe animacy, intelligence, intention, and agency. In turn, humans are compelled to establish relationships of trust and companionship. For these reasons, embodied AI is designed to have human-like roles (helper, teacher, companion, coach, coworker) and to enter into partnerships. A discriminatory, oppressive, and harmful partnership could linger if we do not tackle underlying issues.

When designing embodied AI, data and algorithms for embodied AI to interact with us and our environment need to be considered. Still, it is not only about "bad" data and datasets or harmful algorithms. How we imagine and design the embodiment and anthropomorphization-related issues need to be carefully assessed. Further, uncharted territory is the phenomena of the ascription of intention and social traits and how it impacts human-AI relationships. Ultimately, how we design agency, collaboration, and conversation narratives matters. And we need to assess the values embedded in embodied AI behaviors without forgetting the social and assistive paradigms we design for embodied AI.

THE DEI4EAI PROJECT

Designing embodied AI is not a matter of easy fixes. Are there ways to turn the tide and put AI at the service of social justice? The complexity of the socio-technical issues we outlined above requires interdisciplinary and transdisciplinary processes and awareness of power structures. At the same time, addressing embodied AI's issues about justice, equity and inclusion require reflexivity and self-awareness.

Many are taking action to turn the tide in industry, academia, and non-profit organizations. Institutes like AI Now (Crawford et al. 2019) or Mozilla Foundation (Foundation and Mozilla Foundation 2019) and initiatives like ("Better Images of AI" n.d.) are examples of how industry and non-profit organizations embrace reflexivity and activism to tackle justice issues in AI. In academia, the emergent fields of Responsible AI (Dignum 2019) and Fairness, Accountability, and Transparency in computing are contributing to informing academic knowledge and policy-making concerning AI in general. These initiatives, academic and non-academic, focus on many facets of dis-embodied AI, like datasets, algorithms, sensing (e.g., computer vision), and language models. Issues related to embodied AI are less explored.

We have noticed that the field of embodied AI has only taken the very first steps and lacks a coherent body of literature and ways to assess social justice issues. While issues analyzed in general AI and HCI fields apply to embodied AI, general works so far haven't focused on the central tenements of embodied AI: agency and human-likeness, human-agent relations, the social intervention of embodied AI (i.e., the roles and scope of the interactions). We, as a collective of researchers and activists working in four technical universities in the Netherlands, decided to focus on the tensions and issues related to the sub-field of Embodied AI. We took action to broaden participation by assessing power and inclusiveness practices in embodied AI and by challenging current development practices in an open conversation with academics and communities. We hoped to engage in a meaningful conversation with the embodied AI community: listening and co-creating in a spirit of reflexivity. We started with the following questions: who gets to develop and make decisions about embodied AI? Whose perspective



The figure is “non-neutrality” by Dalt-e Mini, by Crayon

is included, whose excluded? Whose culture/worldview/identity is taken into account? Who is harmed? Who gets access to resources? What are the effects on human self-determination?; questions on inclusion and justice in embodied AI: What interests and goals are imbued in embodied AI projects? Who is embodied AI for? What values, biases, and stereotypes are we encoding? Whose worldview and life experiences are we potentially

obliterating or invaliding? How can we develop with justice in mind? (Costanza-Chock 2020). Concretely, we have founded the Diversity, Equity, and Inclusion for Embodied AI initiative (DEI4EAI). Between April 2021 and April 2022, we organized four international workshops and extra events focusing on gender, race, ableism, design methods, and futuring. Workshops were free and open access both for academic and non-academic participants. We specifically focused on raising awareness, stimulating reflexivity, and imagining future practices relating to gender, race, and ableism, integrating perspectives of equity, inclusiveness, diversity, and justice. As embodied AI connects various disciplines and perspectives with a solid drive to tackle societal challenges, we took an interdisciplinary and transdisciplinary perspective and borrowed from various epistemologies and methodologies. Theory-wise, we take from post-human feminism perspectives. Methodologically, we were inspired by critical/speculative design and responsible futuring practices. These methodologies and approaches offer tools and techniques that enable co-sense making and co-creation and help define common ground.

We departed from our work on imagining conversational futures with AI, on how to challenge narratives of human-autonomous car, and our reflexive practices when designing embodied AI for health and disabilities. And finally, the practices that help us inspect responsibilities, the human-agent relationship, and imagine embodied AI futures worth wanting. Our goal was to learn with students, researchers, experts, and non-academic communities.

We believe that assessing, critiquing, and changing our practices is what academics and society need to design the technologies of the future we want to live in. Reflecting and re-imagining just futures for embodied AI, however, is only one small step to challenging the current status quo in embodied AI. This book, then, is a call for more initiatives to inspire systemic change toward a more just, equitable, and inclusive future in Embodied AI.

BOX 1.1

DEI is a term used to describe organizational policies and programs that advocate the representation and inclusion of traditionally excluded individuals, including people of races, abilities and disabilities, genders, sexual orientations, and cultures. DEI is a term that companies and institutions have adopted to signal initiatives to extend the

participation of so-called (by them) “minorities” at various levels (e.g., hiring, funding, access to spaces, etc.) Recently, there has been a call for a paradigm shift due to the realization that DEI initiatives may bring more issues than they solve. Many organizations, academics, and initiatives are shifting towards equity, diversity, and inclusion (EDI).

BOX 1.II

Others are adding justice to EDI (JEDI). What does the acronym shift signify? It signals a call to put equity and justice at the forefront to tackle systemic structures of oppression for BIPOC (i.e., black, indigenous, people of color), LGBTQAI+ (lesbian, gay, bisexual, transexual, queer, intersex, and asexual), women, disabled people, the politically-socially-economically disadvantaged and intersections thereof. Why is the shift happening? The term DEI is becoming associated with practices that engage in diversity, equity, and inclusion superficially and with actions unable to reach social change (i.e., the alteration of mechanisms within the social structure, characterized by changes in cultural symbols, rules of behavior, social organizations, or value systems). Many initiatives

focus only on the diversity facet of DEI and appear to reinforce the socio-political and economic structures in the current status quo. Diversity-washing (e.g., hiring people from minority groups but not giving them real functions or prospects only to gain fake trust with the public), inclusiveness-washing (e.g., taking minor action to signal that the organization embraces inclusiveness values without making any real, meaningful changes within their organizations to further the goals of those groups), and the inability to act beyond a public display of solidarity have been common issues. Changing hiring pipelines with diversity recruitment and promoting inclusiveness with short-term interventions put the strain of DEI actions on the marginalized people the initiatives wish to support.

BOX 1.III

Focusing only on establishing policy-driven initiatives (e.g., diversity officers) and expanding the diversity pool of entry-level hires is a myopic solution: without reflexivity about practices, values, and power dynamics within an organization, change is not in reach. The shortcomings of DEI apply to companies, institutions, and academic groups working on embodied AI. We know that AI has a diversity crisis, and diversity needs to be tackled. However, hiring a diverse workforce or adding diverse officers to embodied AI team are initiatives that can backfire if organizations fail to nurture an environment of reflexivity both within the workforce and within management. We (and many others with us) argue that we need to acknowledge the limitations of

current DEI approaches, such as the limited focus on the systemic disadvantages, socio-economic inequality, and power imbalances involved with access to resources, decision-making in data, and design matters, and actively seek ways to tackle dismantle barriers and act on current power dynamics. The DEI4EAI initiative is closer to JEDI perspectives than DEI perspectives. However, the acronym has been criticized for its reminiscence of pop culture personas (see Star Wars Jedi) and its lack of power awareness. Decolonized researchers and practitioners have proposed the shift toward universal values of Belonging, Dignity, and Justice (BDJ).

BOX 1.IV

Our work strive to work on social justice in embodied AI as activists and as a matter of good scientific practice. Our goal was and is to establish a community that allows maximum participation to individuals with various identities and socioeconomic statuses and supports the dismantling of the unjust power structures that cause injustice. When we started to take

action, we chose the acronym DEI to be in an open dialogue with current social justice academic efforts in the Netherlands and Europe, which predominantly focus on sole DEI. Should we be able to continue our project, we will opt for an appropriate acronym that is aware of the power and colonial structure and focused on relational ethics.

BOX 2

We are a research collective based in the Netherlands from various disciplines, cultures, gender identities, socio-economical statuses, and politics. Our work focuses on the broadening of participation in the development, evaluation, and policy of embodied AI. We are committed to promoting a diverse, inclusive, and just embodied AI culture. We take a post-humanist, transdisciplinary approach: Going beyond disciplines and working with all relevant stakeholders on equal footing to define a desirable future. We take a perspective that allows a holistic and ecosystemic net of “otherness” to be in a relational collaboration, instead of considering the human (white, male, cisgender, socially and economically powerful) to be the center and measures of things. We define otherness broadly: sexualized others, gender others, racialized others, socio-economical others, age others,

able others, political others, and non-human others (things and AI). Within a post-humanist perspective and transdisciplinary methods, we take a critical race and intersectional feminism lens to interpret and understand the complex social and political dynamics of embodied AI in society and the development of embodied AI. We are academics, aware of the privileged and powerful role that we have gained at the moment: We come from different positions of privilege and marginalization. We have had a range of experiences navigating issues of social justice. Our experiences and outlooks cannot and do not represent everyone who shares a particular identity. We hope to engage in a meaningful conversation with the embodied AI community: broadening participation, listening, co-creating in a spirit of reflexivity to challenge the status quo, and contributing to change in science and society.

An often forgotten aspect is that embodied AI designers must deal with the law, including the EU Charter of Fundamental Rights (EU CFR). The EU CFR includes Article 1 on dignity, Arts. 7–8 on private life and protection of personal data, Art. 21 on non-discrimination, and Art. 23 on equality between women and men. These articles translate into direct obligations for roboticists and AI designers to develop safe systems, respect user privacy, not discriminate, and not generate or reinforce stereotypes.

THE MEANING OF DISCRIMINATION AGAINST CERTAIN GROUPS

International human rights treaties also include explicit prohibitions on harmful and wrongful stereotyping at the

international level. Member States are the recipients of international treaties, but there is a growing understanding that these rules bind businesses (including start-ups working on robotics). Without an exhaustive aim, here below are some provisions laid down by the United Nations that explain what discrimination is against particular user groups:

→ **DISCRIMINATION AGAINST WOMEN.** The United Nations Convention on the Elimination of All Forms of Discrimination against Women described the meaning of discrimination against women as ‘any distinction, exclusion or restriction made based on sex which has the effect or purpose of impairing or nullifying the recognition, enjoyment or exercise by women, irrespective of their marital status, on the basis of

equality of men and women, of human rights and fundamental freedoms in the political, economic, social, cultural, civil or any other field.’

→ RACIAL

DISCRIMINATION. The United Nations International Convention on the Elimination of All Forms of Racial Discrimination describes in its Art. 1.1. the meaning of racial discrimination: ‘any distinction, exclusion, restriction or preference based on race, color, descent, or national or ethnic origin which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise, on an equal footing, of human rights and fundamental freedoms in the political, economic, social, cultural or any other field of public life.’

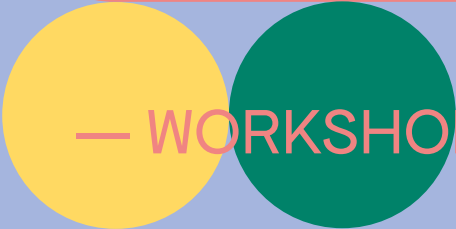
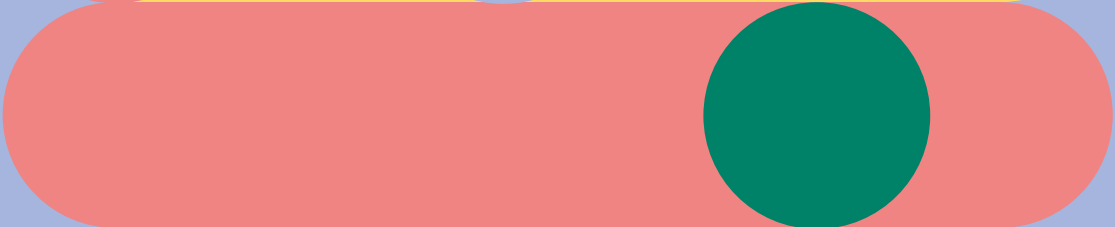
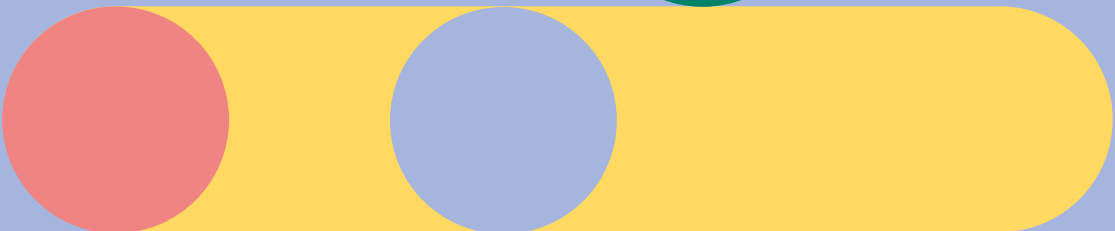
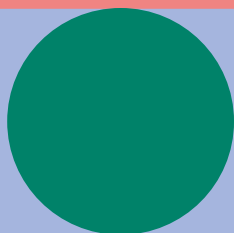
→ DISCRIMINATION

BASED ON DISABILITY. The United Nations Convention on the Rights of Persons with Disabilities explains that discrimination on the basis of disability means ‘any distinction, exclusion or restriction on the basis of disability which has the purpose or effect of impairing or nullifying the recognition, enjoyment or exercise, on an equal basis with others, of all human rights and fundamental freedoms in the political, economic, social, cultural, civil or any other field. It includes all forms of discrimination, including denial of reasonable accommodation’.

RESPONSIBLE RESEARCH AND INNOVATION

Implementing these human rights in practices, teams, and designs may be challenging, mainly because many laws lack concrete guidance for developers and strive to account for diversity and inclusion. For instance, sex and gender considerations have not traditionally been considered sensitive or essential in related EU legal frameworks, such as the General Data Protection Directive (GDPR), the Medical Device Regulation, or the Safety Machinery Directive. Amid this regulatory turmoil, the responsible research and innovation (RRI) framework can guide researchers in ensuring that science, research, and innovation have positive, socially acceptable, and desirable outcomes. The European Commission defines RRI as “an approach that anticipates and assesses potential implications and societal expectations concerning research and innovation, intending to foster the design of inclusive and sustainable research and innovation.” Through the lens of RRI, the principles of inclusion, anticipation, reflection, and responsiveness typically guide the research and innovation (R&I) processes and could prove to be instrumental in achieving more inclusive and diverse embodied AI. Provided that RRI is power aware.

We organized a series of four workshops, each including guest talks and interactive sessions. The workshops revolved around gender, ableism, race, and methods. These topics are essential to the community due to the ulterior adverse impacts missing these considerations in general for society but especially for vulnerable communities.



— WORKSHOPS OVERVIEW

— DEI4EAI & GENDER

15/09/2021—ONLINE

Recent years have seen a growing number of calls for considering gender during designing or evaluating software, websites, or other digital technology. Bias, stereotypes, and gender norms are often embedded in technology implicitly and explicitly with massive societal impact. For example, we learned that voice assistants might not recognize certain accents, image recognition algorithms embedded in IoTs may mislabel people based on assumed gender, and embodied AI, like robots, can be non-inclusive in design, e.g., robots assistants with “female” voices. This workshop explored questions as: How do we make sure that we can make room for designing and developing while being



Different screenshot from dr. Eduard Fosch Villaronga's talk about Gendering Algorithms. The pictures show the consequences of missing diversity considerations in EA. Among them, a passage related to identity and the impossibility for algorithms to objectively identify something inherently subjective as gender.

mindful of the biases, stereotypes, and values we have about gender? How do we integrate these reflections into our processes rather than confining them as an afterthought? What practical actions can we take in our daily practices to incorporate

diversity, equity, and inclusion into the design process? In this half-day workshop, participants learned about gender in embodied AI together with experts, artists, colleagues, and societal stakeholders. We used methods from critical design to 1) create a hands-on understanding of our current practices and narrative and 2) compile a concrete, desirable future scenario, providing practical pointers to implement design processes with diversity, equity, and inclusion in mind.

INVITED SPEAKERS

Catherine D'Ignazio
— Lisa Mandemaker

024

— DEI4EAI & RACE

Bias in AI comes in many forms. Voice assistants may not recognize certain accents, image recognition algorithms may mislabel people based on assumed race and gender, and embodied AI, like robots, can be non-inclusive



Abdo Hassan, invited speaker and activity moderator of the workshop on AI and Race, introducing winter school students to the theme of decolonization in AI and data practices.

in design, e.g., robots with white bodies. Thus, there is a need to think of ways to make room for greater diversity in the design and deployment of AI systems on this day, for ethics in AI should be an integral process rather than an afterthought. In this workshop, we aimed to unpack how AI and data-driven systems

can cause harm to marginalized groups using a framework rooted in decolonial theory. Once we decoupled the AI technologies from their colonial, exclusive/oppressive systems, we borrowed from critical thought and world-building to imagine new avenues for community-in-the-loop design. The workshop was held in person in Delft as part of the 4TU winter school on Humans & Technology.

INVITED SPEAKERS

Cameron Lee Taylor
— Abdelrahman (Abdo)
Hassan

— DEI4EAI & ABLEISM

Ableism can be defined as discrimination against disabled people in favor of non-disabled people. Notions of ableism in embodied AI intertwine with how we see ‘bodies.’ Whether it be bodies represented in the design of social robots or bodies to which AI systems refer to or trained in general, they nearly always refer to ‘fully-abled bodies.’ However, what should be defined as ‘abled bodies’ is up for discussion as there are as many bodies as there are people, all with different historicity and lived experiences. In ableist thinking, there is a tendency to think in terms of curing or fixing disabilities, a view that makes disability an individual issue, with the disabled person being solely responsible. Studying, questioning, and designing embodied AI systems from a Diversity, Equity, and Inclusion perspective requires moving away from ableist notions of curing or fixing disability to ways of thinking that do justice to the diversity of bodies. Close collaboration with communities and individuals that

identify as disabled persons is essential here. In this workshop, we learned more about ableism in embodied AI, together with professionals living with disabilities.



Wheelchair by Credits Vitor Camilo on Unsplash.

INVITED SPEAKERS

Kristen Parisi
— Simon Dogger

— DEI4EAI & DESIGN METHODS

During our workshops, we learned and discussed gender, race, ableism, and their implications in the design of embodied AI. Thanks to the inspiring contribution of speakers and participants from different countries and disciplines, we shed light on the impact of our design actions on marginalized groups. The fourth and last workshop discussed the designers' responsibilities in Embodied AI design. Stemming from an intersectional feminist

perspective, the activities of this final workshop raised awareness of the complexity of designing with diversity, equity, and inclusion in mind. As mainstream approaches in the design of technologies tend to marginalize 'diverse' populations, participatory design is often argued for as a necessary practice. Traditional participatory methods, however, are themselves not

free from inclusivity issues. These, in fact, often fail to address the complexity and challenges that certain marginalized groups may face, hindering their ability to participate as equal partners in the design process. Conscious about the impossibility of defining a "DE&I" toolkit, we discussed desirable practices for designers engaged in developing embodied AI systems.

C. Leira by Abran Maldonado.

INVITED SPEAKERS

Lonneke van Kampen —
Valentina Migliarini —
Laura Forlano — Future
Wake — Abran Maldonado



— DEI4EAI PROJECT

WHAT CAN WE DO AS DESIGNERS

The design field is increasingly engaging with the challenge of better using and revising its methods and tools to address diversity, equity, and inclusion issues (Erete et al., 2018). Mainstream approaches to design technologies tend to marginalize populations characterized by diversity (Erete et al., 2018), whether they are women, an ethnic minority, or persons with disabilities. Furthermore, marginalization takes different forms: we both lack *diversity in who makes* the AI systems and in *who benefits or carries their costs*. In this regard, a recent report by AI Now Institute (West et al., 2019) revealed how women represent only 15% of AI research staff at Facebook and just 10% at Google. In academia, the situation is not much better: only 18% of authors at leading AI conferences are women, and more than 80% of AI professors are male, which is reflected in journal editorial boards. For instance, the editorial board of Artificial Intelligence only counts 13% of women (Forsch-Villaronga et al., 2022). Even worse is the situation regarding race: only 2.5% of Google's workforce is black, while Facebook and Microsoft are each at about 4%.

To tackle the lack of diversity in the making of AI systems, the industry, as well as public institutions, are engaged in *inclusivity initiatives* aimed at achieving gender equality, as well as ethnic diversity in the workforce. Discriminations based on disabilities, sexual orientation, and other forms of diversity are also discouraged through the explicit commitment of companies to their visions as well as hiring policies:

"We are actively working to build a culture that values diversity, equity, and inclusivity. We are intentionally building a workplace where people

feel respected and supported—regardless of who you are or where you come from. We believe this is foundational to building a great company and community. Hugging Face is an equal opportunity employer and we do not discriminate on the basis of race, religion, color, national origin, gender, sexual orientation, age, marital status, veteran status, or disability status.” Hugging Face

While one can assume that more diversity and inclusivity in the workforce developing AI would lead to ‘better’ AI systems, i.e., systems that are more ‘sensitive’ towards diverse human conditions and identities, a problem remains. The uneven distribution of benefits and costs of AI systems is also bound to the different socio-economic and power conditions that people experience, designers and engineers included. A common answer to this issue is to practice [participatory design](#) (Aizenberg and Van Den Hoven, 2020; Wolf, 2020; Rocasolano, 2022).

Because of their scope to open the design process and make it more inclusive, participatory activities are increasingly acknowledged as a necessary practice. Quoting Charlton (1998), we should not design any AI system that could impact minorities without actively involving them in the process. The inclusion of people with diverse abilities, socio-cultural backgrounds, and diverse ethnicity (to name a few aspects) allows designers to become conscious and considerate of people’s heterogeneity, as well as to abandon the counterproductive idea of ‘normalness’ (Patston, 2007). However, to create inclusive moments of participation, we, as designers and researchers, need to reflect on our role deeply: do we have the capacity and tools for delivering a participatory design process that is mindful of diversity, equity, and inclusion?

Examples of these practices are variegated and can enter the design process of AI systems at different levels, from conceptualization to dataset curatorship. At the conceptualization level, opening the process to non-experts might help envision possible consequences of a specific AI application, as well as to [define the moral boundaries](#) of where and when a certain technology should be used or not (considerations that are lately more and more asked for in the AI ethics discourse (Cavalcante Siebert, 2022). Often,

society is confronted with the unintended and undesirable consequences of AI used for controversial applications. Whether it is predicting the possibility for a person to commit a crime in the future or assessing the quality of a teacher's performance, AI is gradually more involved in services that promise to optimize aspects of life that are hardly quantifiable. The participation of diverse people in the process could help anticipate some of these possible drawbacks and question the underlying assumptions.

Even more practically, non-experts can be involved in [collaboratively creating better datasets](#) representative of real-world diversity. In this line of thought, Google now involves users in various ways to improve its datasets. *Crowdsourcing*, for instance, is an app designed to involve users in improving Google services' accuracy by performing quick tasks, like checking the accuracy of image recognition and translation algorithms. Volunteers test the crowdsourced datasets worldwide through initiatives like the [Inclusive Design Competition](#). People are invited to use Open Images, a publicly available image classification dataset that is majority-sampled from North America and Europe, to train a model that will be evaluated on images collected from different geographic regions across the globe (Doshi, 2018). This way, diverse people can actively contribute to enriching AI's understanding of the world, so the thinking goes.

Participation, however, is not a fix (Sloane et al., 2020; Ayling and Chapman, 2021). Participatory design practices themselves have inclusivity issues. First participatory design work is [not free](#), whether we acknowledge it or not, it is a form of labor, and we should explicitly account for this when involving people in the design process, especially if they are marginalized groups. Even initiatives that are in principle designed for a good cause that is improving inclusivity, may result to be exploitative, as, in the case of the app Crowdsourcing mentioned above, that is also criticized by its users because of its exploitative nature:

“They should pay us, we basically work for free for Google, I would accept also to be paid in Google Rewards. But working for free NO” (Crowdsourcing app reviewer)

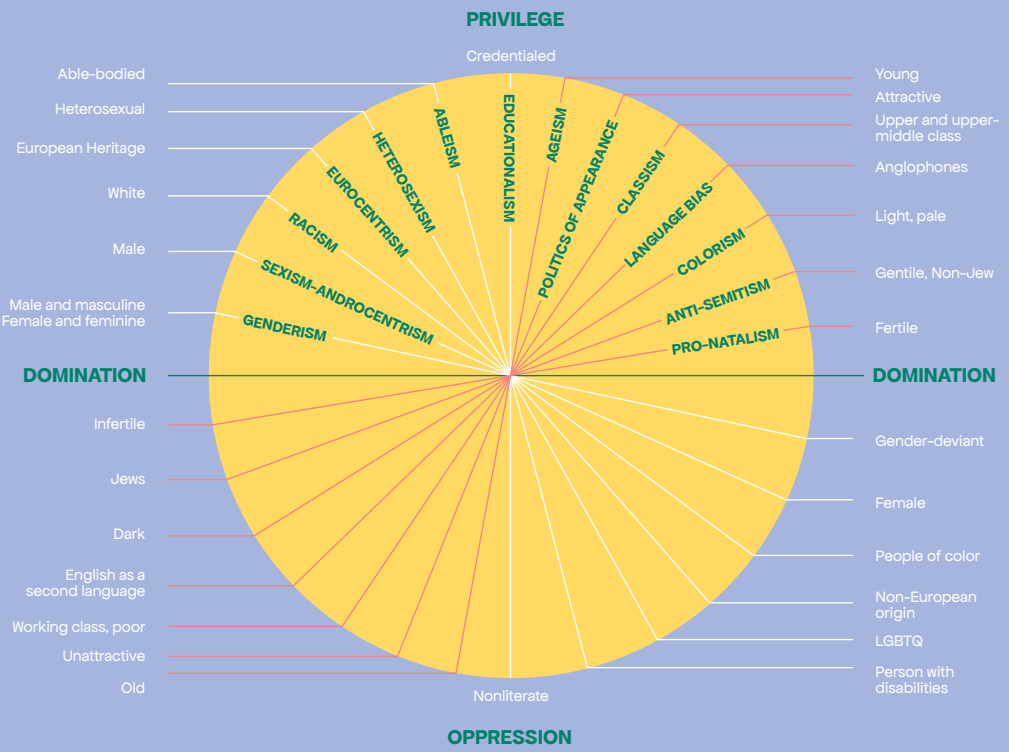
store/apps/details?id=com.google.android.apps.village.boond
 ↓ Crowdsourcing app review: <https://play.google.com/store/apps/details?id=com.google.android.apps.village.boond>

Second, participatory design is a **lengthy process** that, if practiced properly, hardly fits with the speed of dominant digital product development processes, guided by the ‘move fast and break things’ mantra. Third, conducting participatory design activities is a **skillful job**, and AI designers and developers usually lack necessary expertise. Last but not least, participatory design is—in most of the cases—**exclusionary**. This last point might sound counterintuitive as the very value of these practices is to open the design process to participants other than the designers and developers. Yet, the ones who get to be involved, and are envisioned as potential users, usually are people belonging to a similar socio-economic status as the designers. This leaves out of sight a variety of groups that might be affected by a designed system in unpredictable ways. For instance, when developing a recommender system, a designer would hardly think of the people eventually involved in the invisible sphere of human labor employed to label datasets.

Participatory practices, then, are necessary for inclusivity but also not sufficient. Even when in place, traditional participatory methods might fail to achieve the aspired inclusiveness because they fail to address the complexity and radical challenges that certain marginalized groups may face, which hinder their ability to participate as equal partners in decision-making and design processes (Ayling and Chapman, 2021).

Inclusivity is a complex issue that asks us to challenge power structures at the institutional level where inequalities consolidate, perpetuate and accentuate (Amis et al., 2018). To give a concrete example of how these issues are radicalized at the institutional level, we may look at the case of the SIGCHI R.A.C.E. initiative (Grady et al., 2020). In 2019, a group of volunteer researchers all identifying as racial minorities started an official SIGCHI Diversity and Inclusion team committed to making the community more inclusive of diverse perspectives. Only one year later the whole group resigned, declaring that they were the object of public defamation from members (especially Caucasian) of the community (Siobahn et al., 2020).

This anecdote surfaces the need for radical



↑
 Intersecting axes of privilege, domination, and oppression. By Erete et al. 2018, adapted from (Patston, 2007).

change that can only happen by admitting the existence of discriminatory practices within our institutions and design processes. There is a fundamental need for learning about and promoting the work of marginalized people, and ‘making them leaders’ (Siobahn et al., 2020). Achieving inclusivity, then, asks foremost for **acknowledging positions of power**, especially to the ones who have it. As a matter of fact, power structures are ‘invisible’ to most people, especially to the ones who benefit from privileged positions in society (Sanders and Mahalingam, 2012; Atewologun and Sealy, 2014). In this regard, several frameworks and tools have been developed for allowing people to become more aware of their (lack of) privileges. Erete and colleagues (2018), for instance, provide a framework (figure 1) that allows one to examine individuals’ experiences and identities in relation to power and privilege.

Together with acknowledging the fact that factors like gender, race, abilities, and other socio-demographic

factors determine a different access to opportunities in life, we also need to create venues for understanding how these inequalities unfold and are exacerbated by AI systems. As designers, we need to [listen to the voice of the marginalized](#). To get a better idea of why, as designers, we should step back and listen, we can look at the striking example of AI applications for disabilities. The lived experience of Laura Forlano, Associate Professor of Design at the Illinois Institute of Technology, as Type 1 diabetic, for instance, confronts us with the consequences of neglecting the knowledge and felt experience of the people for which an AI system is designed for. In her words:

[“The AI system is keeping me alive, but it is also ruining my life”](#)

With this phrase, she summarizes the struggle of living with an AI-based insulin dose adjustment system. The automated pump, in fact, represents for her a significant step forward compared to her previous situation in which she was regularly going to sleep hoping that she would wake up in the morning and not fall into a diabetic coma, because of a severe glucose low during the night. Yet, the frequent occurrence of malfunctioning and alert signals from the pump makes her now live in a continuous state of alert and anxiety. Listening to her story, then, one could argue that the design of the AI system was left halfway: a basic life-saving function was provided and considered sufficient. The neglected user experience, however, results in tremendous consequences on the user’s wellbeing.

Another interesting example is the case of cochlear implants for deaf people, especially children. Most, including doctors, see cochlear implants as unique opportunities for giving children more options to participate in social life by normalizing their communication (Shew, 2020). Yet, the Deaf community often express a different perspective on the topic:

[“What is there to fix? We’re happy with the way we are. We don’t view it as problem.”](#)

For the ones who belong to the Deaf community and take pride in such identity, the cochlear implant is even

↓ Caroline Praderio Interview with Brandon Edquist,
2017. <https://www.insider.com/why-deaf-people-turn-down-cochlear-implants-2016-12#:~:text=Deaf%20people%20assert%20that%20deaf.and%20actively%20protect%20the%20technology.>

seen as an obstacle to their culture, as children may feel discouraged to learn sign language and develop themselves as part of the Deaf community (Shew, 2020).

Finally, a related–yet different–example is the one of autonomous wheelchairs. These are being developed as potential solutions to the dangers associated with the use of power wheelchairs, especially in care facilities. Often, in fact, power wheelchairs are being banned from care facilities because of the difficulty of controlling them appropriately which often causes accidents (Braze Mobility, 2018). Autonomous wheelchairs could potentially prevent the occurrence of such accidents. Yet, listening to the voices of everyday wheelchair users surfaces a different perspective that might be easily neglected:

“The power wheelchair is one of the few things that he has total physical control over, and giving up control is almost unthinkable”

These examples clearly surface the complexity of designing inclusive AI systems, especially when it comes to optimizing the life of marginalized people that is hardly understood by the mainstream culture and related design practices. As a response to such complexity, academia is increasingly opening towards ‘alternative’ research practices revolving around the felt experience of the researcher, usually addressed as *first-person methodologies* (Varela and Shear, 1999). These methods are characterized by authors writing or performing in the first person, becoming themselves one of the objects of research. The results of these methods are usually narrative texts where generalization of insights is built from single cases extended over time (Bochner, 2012). The scope of these methods is usually to subvert deep-rooted assumptions through personal stories (Ogbonnaya-Ogburu et al., 2020). As such, they are particularly suited to address diversity issues. These, as most social science inquiry, aspire to surface truth, not literal, rather emotional truths that are not intended to be received, but encountered and collaboratively constructed (Bochner, 2012). First-person methodologies, then, allow us to build a collective understanding about the plurality of truths we can encounter in life.

↑ Interview to a person on a wheelchair. Braze mobility, 2018. The Ethical Implications of Autonomous Wheelchairs. <https://brazemobility.com/the-ethical-implications-of-autonomous-wheelchairs/>

First-person methods invite us to self-reflection and, as such, become central to the scope of understanding positions of power and listening to the voices of the marginalized. Participatory practices, then, should emphasize even more their interest into personal stories of both designers and the public.

It must be noted, however, that even when these desirable methods are practiced, due to the complexity of diversity issues, designers can still encounter resistance and adverse reactions. On the one hand, activities aimed at raising awareness about power and privileges (related to socio-demographic factors) can be very confronting and generate resistance (Atewologun and Sealy, 2014). In fact, even if motivated by genuine intention to be inclusive and sensitive towards diversity, a white, highly educated, fit and straight man might feel offended, or poorly represented if named 'privileged'. Such terms, in fact, may clash with the image of the self that a person has, that hardly can be captured through simple socio-demographic data. Thereafter, together with finding new ways of including the marginalized in the process, designers should also consider building a more [inclusive language](#) that would allow both the ones that experience privileges as well as the ones who don't, to have constructive conversations. On the other hand, even when appropriately working on one diversity issue, i.e., gender equality, we might end up further marginalizing other communities, such as ethnic minorities. As a matter of fact, factors like race, class and gender are interrelated and most of the time experienced together (Costanza-Chock, 2018). Yet the tendency for designers and developers is to consider inequality on a single-axis, focusing on one issue at the time, which leaves unattended the problems of certain groups of people who are intersectionally disadvantaged under white supremacist heteropatriarchy, capitalism, and settler colonialism (Costanza-Chock, 2018). This, again, is due to the interdependent nature of diversity issues. As designers, we should acknowledge such complexity and strive for interventions that address inequality as a network of contributing factors, along with being ready for criticism.

To conclude, designing for diversity, equity and inclusion is a complex challenge that requires existing approaches and methods to be revised. Yet, a lot can and should be done. We, as designers, must respond to this challenge, being humble; being ready to provide others with platforms to tell their story, and collectively challenge the idea that there is one normal way of being.



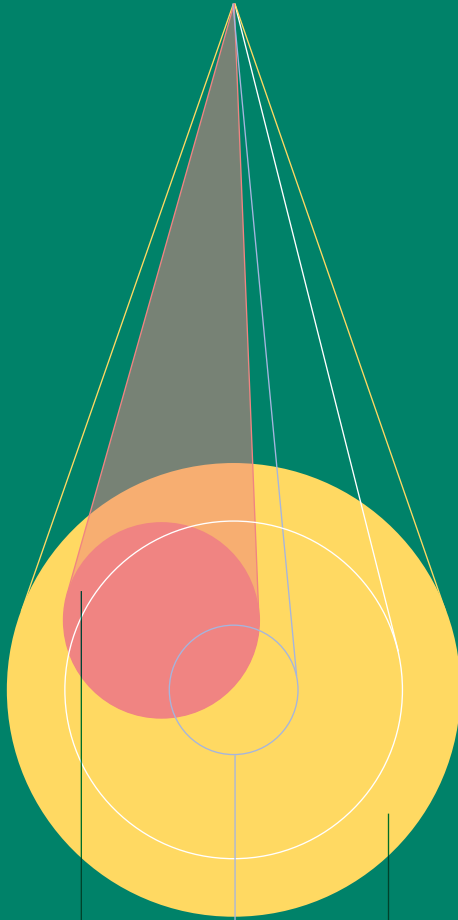
— FUTURES TOOLS

Tools and techniques to promote reflexivity with the embodied AI community

— BY CRISTINA ZAGA,
MARIA LUCE LUPETTI,
NAZLI CILA, MINHA LEE,
GIJS HUISMAN, EDUARD
FOSCH VILLARONGA,
ANNE ARZBERGER

TOOL 1

NOW



PREFERRED
FUTURES

PROBABLE
FUTURES

POSSIBLE
FUTURES

TEAM 1

TEAM 1

TEAM 2

TEAM 2

TEAM 3

TEAM 3

Reflecting on the implicit assumptions

WHAT IS THE TOOL ABOUT

The tool offers a two-step activity for tangible reflections related to how we design embodied AI and imagine possible, probable, and desirable futures. The first activity of the tool presents current scenarios and embodied AI artifacts that embed specific narratives, biases, and stereotypes. The second activity uses the future cone metaphors to support participants in reflecting on how the future embodied AI will interact with the current infrastructure, who will be involved, and how it might be co-opted. The tool can be used to reflect on implicit assumptions we have about different topics. In our workshops, we used it to reflect on gender and ableist assumptions.

HOW TO USE IT

The tool can be made available online for collaboration, e.g., through Miro, or also printed. Participants are invited to split in groups, read the instructions on the left and work in the assigned space on the right. Teams are asked to note down on post-it their reflection and share them with the other teams. The activity takes 40 to 75 minutes, depending on the number of participants.

WHEN TO USE IT

We suggest using the tool at the very beginning of the design process to map current issues and challenges and kick off a reflection about future visions. It is suitable to establish disciplinary and interdisciplinary understanding.

FOR WHO

Students, researchers, practitioners, and policymakers working in embodied AI.

Mapping privileges

WHAT IS THE TOOL ABOUT

This tool is designed to let people reflect on personal position of privilege, or lack of. The tool is based on reknown models of privileges and power, such as the one by Erete et al. (2018) that shows intersecting axes of privilege, domination, and oppression. Our tool invites participants to position themselves into binary axes generally associated with privilege, i.e., white color of the skin vs. black. The tool is purposefully forcing people into binary categories that hardly represent the complexity of reality. By doing so, it aims at provoking discussions and reflections on how we personally experience privilege and how also we attribute it to others.

HOW TO USE IT

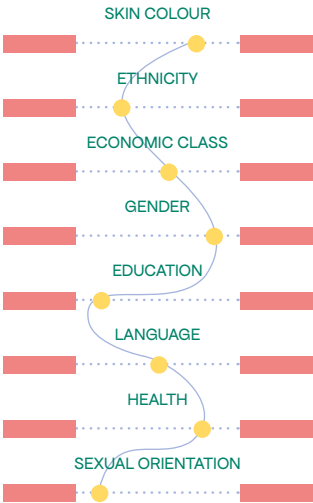
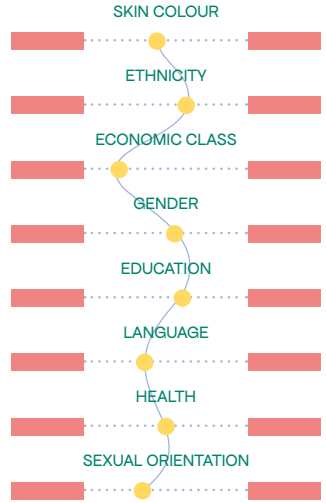
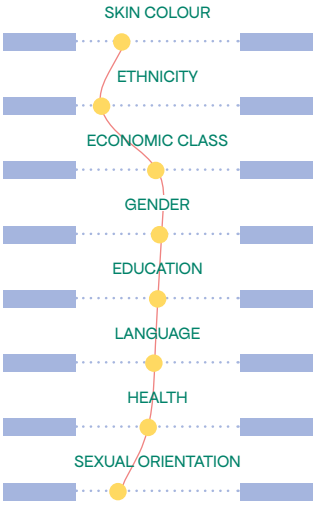
The tool can be prepared as a card, either printed or set up in a digital collaboration environment, i.e., Miro. The activity is organised into three steps. First participants need to split into groups of maximum four people and each person has to fill in one card individually. After that, people share how their 'curve' looks like (see in the example image) and discuss how this matches or not with their felt experience of privilege or exclusion. Last, all participants share highlights of the discussion emerged in their group. The whole activity is guided by a moderator who can, optionally, decide to turn the activity also into a sort of satiric competition about who holds more privilege.

WHEN TO USE IT

The tool is particularly suited for workshop settings where people need to familiarize with the concepts of diversity, privilege and power. The activity is best suited at the beginning as it facilitates jumping into the topic while also functioning as an icebreaker.

FOR WHO

Students, researchers, practitioners, policymakers with an interest in diversity, equity and inclusion. It is not necessary to be an expert in the field nor to have any technical knowledge.





Punkbots collages against status quos

WHAT IS THE TOOL ABOUT

The tool is meant to support the overturn of the status quo of robot design using déornament activities inspired by punk techniques by the Letterist international. These activities help participants to grow awareness and implement critical reflection in their design process. Participants are going to be presented with one HRI scenario (photograph and description) and two bots: a StatusQuo Bot and Punkchabot. The bots are designed with DALL-E and presented to the participants through a power-point presentation. The bots are prompting the participants to first describe the status quo of the images and then overturn it. To do so, they use a collage technique developed by Letterist International to reverse storylines and images, collating and pasting images that change their meaning.

HOW TO USE IT

The activity is better suited for in-person workshops, but can easily be organized online. Materials are prints of status quo pictures, post-its and magazines and scissors for collages. The activity is designed for 45-75 minutes but can be shortened if necessary. To facilitate the activity, moderators should be ready with questions to prompt participants and have ready examples of détournement from Letterist international. Please beware of accessibility issues for disabled participants.

WHEN TO USE IT

This tool is meant to support transdisciplinary discussion and to create common ground and consciousness-raising among participants.

FOR WHO

The tool is meant for professionals, academics, and non-professional communities. It is not necessary to be an expert in the field nor to have any technical knowledge.

Exploring spaces between categories — A biased classifier

WHAT IS THE TOOL ABOUT

Personal values, interests and experiences can impede designers' intentions in creating fair and inclusive solutions for people of all gender. Counteracting bias perpetuation and augmenting limitations of human cognition, this tool aims to break with stereotypical expectations and thinking into binary categories. By surfacing our unconscious associations and the narrow of ways of categorizing things, a classification algorithm (i.e., Teachable Machine) can be used to help us challenge gender norms and stereotypes.

HOW TO USE IT

Step 1 (for groups) — Visualize yourself — Each participant creates 3-4 cards that represent personal interests, personality and identity. This can be done by cutting out images from magazines and gluing them on paper cards.

Step 2 — Curate the Dataset — Each person or team has to create its own dataset. For this activity, either the facilitator or the individual participant can setup a board on Miro (see example online). First two categories are set, for instance “male” and “female”. Next the participant can intuitively classify images from magazines or images found online in either one of the binary categories. This can be done as a team of max 4, or individually.

Step 3 Train the classifier — Download the collected images, keeping them separated into categories, and pass them on to the classification algorithm. You can use any classifier for this experiment, however we recommend Teachable Machine. Once the images are uploaded you can start the training.

Step 4 Explore, challenge and learn — Try out any kind of object, design, prototype, or photo, including the identity cards, to explore how the “biased” classifier we created sees the world. For this, simply upload the images of interest or hold your object in front of the camera and observe to what extent the classification result reflects your view of the object. Question why the algorithms reacts in a certain way and how can you influence that.

**MATERIALS
(FOR
GROUPS)**

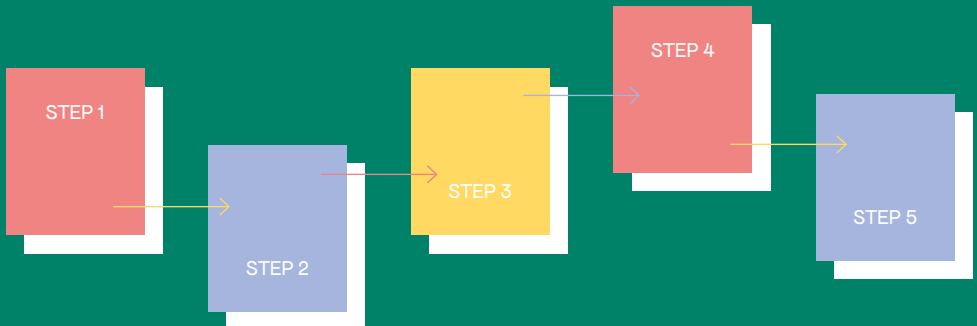
Magazines (e.g. science,cars,nature,baking,etc.), Scissors, Glue, Paper; mandatory: Laptop (with camera), internet to access Miro, for material preparation, and Teachable Machine, for running a classification algorithm.

**WHEN TO
USE IT**

This tool can be used when designing artifacts to be mindful and reflect on the unconscious biases that can be projected into objects. From early design process doodles, to more elaborate design sketches or even prototypes, looking through the eyes of the algorithm you trained can be a powerful moment of confrontation. The same reflexive potential can be used as a way of introducing people with diverse background and expertise to issues of diversity and biases into automated systems, in collaborative work settings, i.e. multistakeholders workshops.

FOR WHO

The tool can either be used in a workshop with multiple people, best grouped in teams of 4, or applied in individual sessions. For using the tool in a workshop include the “group” marked activities. Depending on the amount of groups, more than one facilitator might be necessary. If used by individuals, no facilitator is necessary.



To conclude this book, we want to offer a set of reflections distilled from our final internal event in which we looked back at the experiences of the workshops. What have we learned with students, researchers, policymakers, and communities? In this section, we argue for collective actions focused on re-imagining our futures with embodied AI and conclude with a vision of desirable future(s) horizons.

FINAL REFLECTION

Calling for Embodied AI Justice and Futuring

— BY CRISTINA ZAGA

The series of workshops on embodied AI and gender, ableism, and race helped us get a broader picture of the complex and interwoven social, economic, political, and moral issues related to how we imagine and realize human-embodied AI interactions (and relationships).

Considering the conspicuous volume of work related to social justice and AI, and current activism in Academia, we (naively) thought that the embodied AI community and the community of people using the technology were ready for an in-depth analysis of facets of justice, diversity, inclusion, and equity.

In practice, we noticed that we needed to take a step back and lay the groundwork to make the desired reflexivity possible for our team and the people who joined us during our events. We observed a challenge hindering

our tangible sense-making: our activities could be adversarial as they relate directly to personal lived experiences. At the same time, it was hard to fully take the perspective of another, put ourselves in their shoes, and in



“As automation grows, we need to take a frequent intermissions to look into the mirror and examine images it reflects.”
Showeria Zhang - The wicked Queen’s Smart Mirror, pp.22 in Arista et al., Designing Human Futures with Machines. Against Reduction. The MIT Press 2021 (Arista et al. 2021)

so doing, critically assess the values embedded in our practices. **Common ground and common vocabulary were (sometimes) missing.** While participants all shared an interest in expanding their knowledge and reflecting on their ways of working, the discussion often lacked a compass to navigate issues of oppression, domination, and privilege. It became apparent that **the metaphor of AI (and embodied AI) as a magnified mirror that reflects our image of societies was more salient than ever.** Yet, we lack ways to see and interpret these magnified reflections.

Our reflections started at an individual level, with our role as researchers, designers, developers, and scientists. First, we wanted to act on the individual awareness as a call to action to inspect and eventually challenge ways of working. Raising awareness and consciousness among those who study, design, and develop embodied AI is necessary to examine current practices.

However, we realized that inviting to re-assess values and assumptions underlying the design of embodied AI, surfaces discussion on power asymmetries and oppressive dynamics. These topics are often tacit and difficult to talk about, particularly in a wealthy, mostly-white, cis-gender, European context, like the Netherlands, where our initiative was situated. The hard realization that plurality of political/economic interests, cultural perspectives, and narratives inform our perception of embodied AI is particularly confrontational when we realize that it is implicitly shaping our attitudes and decision-making when —for instance — imagining scenarios in which embodied AI can be applied when molding personas onto embodied AI agents, when designing bodies and behaviors of robots, chatbots, smart assistants, etc...

Another apparent yet hard reality to reconcile is that there is not one coherent perceived reality of embodied AI we can all agree on. Our perspective is the product of our role in the socio-economical and cultural structure we are embedded in. And, our decisions, although impactful, are subordinated to political and economic power structures that are often hard (if not impossible) to challenge.

To say it with the works of Brewer and Collins (1992), our lived experience provides the most reliable form of knowledge about oppression and power. Although uncommon in disciplines such as computer science or engineering, we learned that surfacing and openly discussing our lived experiences ease the sense-making process and increase the salencies of the social justice issues to examine. Still, the plural intersection of our experiences of power and oppression is difficult to untangle. And, we must avoid the tendency to prioritize one social justice issue over the other. Equity, inclusion, and diversity have different dimensions and shapes, and one issue should not be prioritized over others.

We learned the hard way that intersectionality, while being the best way to account for the complex net of the otherness of our society (Braidotti 2022), is hard to realize, particularly with heterogeneous working groups. There is never enough intersectionality for some, while it is hard to be concrete when broadening horizons for others. Researchers might make decisions with good intentions and in the spirit of intersectionality, but they might not be able to tap into the lived experience of others, eventually contributing to adverse societal implications. We have noticed that intersectionality can be perceived as an impossible balancing act that might lead to unavoidable paternalism, particularly when we involve societal stakeholders.

At the same time, engaging in dialogues with those affected by design decisions is an imperative that comes with many barriers. We identified barriers and enablers in each of our workshops.

During our workshop about gender, we learned that many scholars are working on understanding how gender plays a role in embodied AI. There is a line of work in HCI (Keyes 2018; Spiel, Keyes, and Barlas 2019) and data science (D'Ignazio and Klein 2020) looking at feminism and trans-feminism that can serve as a source for theoretical substantiation. Still, in embodied AI, there is less attention to the matter, particularly how we design embodied AI with gender equity and inclusion in mind. The only example of research on gender and embodied AI is the work of the scholars involved in the GenR workshop ("RO-MAN22 GenR Workshop" n.d.) series and special issue.

We noticed that the focus is primarily on women's issues and a widespread difficulty in framing gender as a spectrum rather than a binary dichotomy (male vs. female). In this landscape, various communities have fringes (and fringes of the fringes) active in the field of embodied AI and gender. Still, each community comes with a specific language and frame of reference. An enabler to find common ground and tangibly work on future perspectives for gender in embodied AI is to make live experiences, milieu, and frame of reference visible in co-creation activities and collectively shape a compass to navigate the complexity.

During our workshop examining ableism and embodied AI, it became apparent that — citing Lupetti’s words from our section about what designers can do — participation is not enough. Involving disabled people to work with us in the development of embodied AI is fundamental, but the actual implementation is often tainted by patronizing and paternalistic tendencies. Without the active involvement of disabled people in setting up research questions and the mission/vision of our work, we are only asking them to validate our perspectives rather than embracing and helping make their vision flourish. As a result, the disabled’s perspectives are critical of embodied AI technology that offers easy fixes and accommodation for the caregivers rather than the actual disabled.

One of the persistent barriers to an integral partnership between the disabled and the researcher is bureaucracy, funding, and ethical clearance. We need to take action to facilitate researchers to integrate disabled communities in project proposals by offering funding for them to participate and easing the processes to get ethical clearance from IRBs and ethical committees.

The workshop on race shed light on the necessity to understand the idiosyncratic ways embodied AI might bring racism, colorism, and inequity forwards. The workshop on race was embedded in the 4TU Winter School on AI, and the primary audience was graduate students. Once again, the main barrier was a language and a knowledge gap. Students lacked terms and theoretical references on social justice issues, critical race theory, and intersectionality. And it makes sense because most bachelor’s and master’s programs focus on computer science and engineering subjects, with little space for the social sciences. Interdisciplinary and transdisciplinary efforts are needed to equip students with the necessary knowledge to engage in discussion. Training early-stage researchers and students will help future research and design endeavors and stimulate them to work in close partnerships with diverse community stakeholders. Lastly, we must provide the next generation of researchers and designers with insights related to embodied AI. Most of the research students belonged to the data science and algorithm fields.

Lastly, the workshop on design methods clarifies that we cannot think that design methods and toolkits can be an easy fix for our practices. We need to reflect on lived experiences and current practices and critically assess the participatory principles we employ. As we have seen in all the workshops, participatory approaches have fallen short. We must nurture practices to examine the power dynamics in research and design projects involving communities.

Originally, through the activities planned during the workshop series, and our transdisciplinary and future-oriented methods, we expected to go into a deep dive examining current narratives and analyzing the landscape of design practices to inform future practices critically. The plural complexity surrounding embodied AI design and development calls for deliberate explorations of how we make sense of our interactions with embodied AI. Through our workshops, we focused on making ideas and perspectives tangible. Instead of engaging in pure discussions, we visualized issues and tensions and sketched potential futures. In so doing, we aimed to productively take each other perspectives on values, power, agency, and responsibility. While we realized most of our objectives, we have dramatically reconsidered our ways of working, realizing that the road ahead to challenge the current status quo and designing embodied AI with a multitude of otherness in mind is longer than we thought.

In the following sections, we humbly offer our vision for the future in light of our preliminary learnings.

“We put so much investment in being saved by these objects we create, by these technologies. But our real resource is ourselves, our relationships, our stories, our narratives”.

The most important learning from the DEI4EAI initiative is that, to say paraphrasing Rua’s Benjamin words’, we need to invest in our relationship, listen to each other’s stories and perspectives and work for futures in which all of us can live in dignity, in a warm sense of belonging, with equal opportunities to thrive.

To this end, we believe the field of embodied AI and related research topics should embrace a series of paradigm shifts to re-imagine its futures. Or at least actively dialogue with perspectives that can help challenge the current status quo in embodied AI. Below, we provide a brief overview of the paradigm shifts we wish to nurture in the field of embodied AI. Then, we suggest ways to incorporate these perspectives into our practices

↑ Ruha Benjamin Race After
Technology: Abolitionist Tools for the New Jim
Code, 2020, retrieved from (CBC News 2019)



Nature and Technology by Delle Mini by Craiyon

First, we encourage a shift from a humanism-oriented perspective based on Cartesian dualism (i.e., humans, and in particular western white cis-gender males as sole autonomous, are the center of the universe, the intelligent agents of change and measure of whole things), to post-human feminism perspective, i.e., all humans, including the sexualized, racialized, able, age others are in a decentred relation with non-human forces as living symbiotic entities

relating to each other (Mackereth 2021; Braidotti 2022).

Many scholars have already embraced more than human approaches (Wakkary 2021), post-humanist practices (Forlano 2017), and indigenous epistemology of relationality in the design and development of technology, but we believe that we need a radical analysis of the the current **TRANS-HUMANIST*** and humanist perspective in embodied I to understand how to raise above them.

“ Our responsibility is in the relationship. Who is building them? Is it the kanaka or the human? The rock, the mineral, the rock and the human are engulfed. They birthed this program. Everything that comes with the kanaka—the human—his faults, his cellular structure, that gets folded in with the mineral. You need the volcanic activity, the structures that create the calcium. We have to interface with the spirit; if we disconnect and let the spirit just move us, we are not having a kinship. The human’s responsibility is to realize that the energy that makes up the god is in you somewhere. If it is not there, how is it possible to interface with sky, interface with the thing you are creating? The fact is that some of you is in it. And some of it is in you.”

Post-human feminism perspectives bring attention to a “we” mixed of diverse humans and non-humans, artificial agents included. And its ethos is in line with intersectional feminism, afro-futurism theories, and practices providing a fertile terrain to engage in social justice in design. Braoidotti shows that post-human-feminism can *“combine sharp critique with intense compassion and care for a damaged world”* and *“it has produced critical and creative forms of re-*

↑ Kekuhi Kēa'i'i'ikanaka oleohaililani,
Kēa'i'i'ikanaka oleohaililani, K. (March 1, 2019). JP AI Workshop,
University of Hawai'i at Mānoa, Honolulu, U.S.A., taken from
(Lewis et al. 2020)

naturalization of those who, being socialized as the devalored other of 'Man', have both had not the historical opportunity to express the forces, qualities, and values that constitutes them".

(Braidotti, 2022 pp.7-10). For embodied AI, taking a post-human perspective could mean seeing robots, chatbots, and other devices as non-human kins, not as mere tools or slaves of the human creators. In turn, seeing embodied AI as kin might open to interaction paradigms for ecosystemic flourishing rather than human subjugation of nature and dehumanizing and marginalized others. Or going beyond the human as the primary model to adhere to in body and intelligence and eventually surpass.

The world created through Western epistemology does not account for all members of the community and has not made it possible for all members of the community to survive, let alone flourish. The Western view of both the human and the non-human as exploitable resources is the result of what the cultural philosopher Jim Cheney calls an “epistemology of control” and is indelibly tied to colonization, capitalism, and slavery.”

Seeing embodied AI as kin does not mean seeing robots, chatbots, smart speakers as über-humans that should have rights and morals. AI as non-human kin does not mean to bear the technology with moral responsibilities: embodied AI that does not have responsibilities as they are designed and developed by humans. Instead, it nurtures a vision advocated by various authors (Birhane and van Dijk, 2020; Arista et al., 2021; Lewis et al., 2018) of embodied AI as mediating agents embedded in a relationship with humans, animals, and nature. Designing embodied AI within a post-human perspective could help us raise above narratives of domination, extreme rationality, and strict dichotomies and embrace epistemology of care, plurality, and relationality.

↑
 Lewis et al., Making Kin with the machines
 in Designing Human Futures with Machines. Against
 Reduction. The MIT Press 2021 (Arista et al. 2021)

▪ **TRANS-HUMANIST** From Wikipedia: Transhumanism is a philosophical and intellectual movement which advocates the enhancement of the human condition by developing and making widely available sophisticated technologies that can greatly enhance longevity and cognition. It also predicts the inevitability of such technologies in the future.

Ethics in this regard is not merely a methodology, a tool, or simply a matter of constructing a philosophically coherent theory but a down-to-earth practice that is best viewed as a habit—a practice that alters the way we do data science. Relational ethics is a process that emerges through the re-examination of the nature of existence, knowledge, oppression, and injustice. Algorithmic systems never emerge in a social, historical, and political vacuum, and to divorce them from the contingent background in which they are embedded is erroneous. Relational ethics provides the framework to rethink the nature of data science through a relational understanding of being and knowing.

Another paradigm shift is related to the ethics of embodied AI. The general drive to design AI and embodied AI with human values and morals in mind have stimulated philosophers, designers, engineers, computer scientists, and policymakers toward frameworks and best practices of responsible research and innovation (Stahl and Coeckelbergh 2016). In particular, scholars and practitioners are more and more advocating practicing responsible AI (Dignum 2017), people-centred AI (see, for instance, Google Pair’s toolkit), AI for social good (Floridi et al. 2018). In practice, these lines of research and development are focused on providing a framework and technical solution for AI explainability, fairness, and accountability.

These developments are important and welcome departures from ethical “what-if-scenarios” as they can support change in the way we design AI. Many are the reports (van Heteren Pieter Nel 2018; World economic forum 2022; Unesco 2020), positions (“AI for Social Good –” n.d., “AI for Good” n.d.), and toolkits (Lemonne 2021; “PAIR –” n.d.) from

governmental institutions and private companies are produced to support the responsible development of AI readily. However, a crucial criticism is that responsible and social good endeavors are primarily driven top-down by who has power and is less affected by the implication of technology. Doaa Abu-Elyounes and Karine Gentele (AI Now

↑
Abeba Birhane (2021), Algorithmic Injustice: a relational ethics approach, Patterns, 2(2), 100205. (pp. 8)



The threads and relationships of nature.
Figure from Unsplash.

Institute 2021) explain that *“The human rights of disadvantaged and marginalized groups are not properly taken into account” and “Most initiatives lack concrete design specifications that can address the needs and characteristics of minorities and marginalized populations”.* Another criticism is that terms like social good can be misleading: *“the soup term “AI for social good” creates even more confusion about the goals and methodologies of the endeavor as well as its values and stakeholders. The lack of precision from the perspective of technical practitioners tells us nothing about what good is, for whom and in what context and how it relates to other research domains.”* (Malliaraki 2019).

In short, social good projects and ethical approaches often appear to stem from the privileged few and offer little space for the ones oppressed by AI technology. Authors have noticed that social good projects, especially outside academia, are tainted by issues of ethics washing and shirking (AI Now Institute 2021; Floridi 2019); The pressing questions for those who want to engage in ethical and social good-oriented practices are: who decides what social good is? Who benefits from social good initiatives?

Inclusive and just ethics should prioritize understanding the historically complex and untangled relations between power and oppression. In line with post-humanist (Braidotti 2022) and algorithmic justice (Birhane and Cummins 2019) scholars, we believe that we should move towards relational ethics. In the DEI4EAI initiative, every workshop and discussion has touched upon the interconnected and relational nature of equity and inclusion matters. Working from a relational ethics perspective means decentering the powerful perspective when analyzing ethical and moral issues and embracing the perspective of the people (and non-humans) at the margin. Relational ethics could provide embodied AI a framework for examination and action based on relational ways of thinking and intending our society. It helps us focus on flows of solidarity and collective praxis of constructing social horizons (Braidotti 2022) situated in socio-political and cultural contexts. In concert with post-human perspectives, it can help us untangle and examine how embodied AI defines and redefines what bodies, identities, and relationships are possible and how the application and intervention of embodied AI alter social identities and relationships.

Our ways of working in Embodied AI

“So the big issue is that good robots need to confront the fact that humans are not all humans in the same way, to the same extent, and that there are really enormous differences of power and entitlement in terms of being humans, which makes a lot of humans de-humanized or not fully human. And consequently, they have a very ambivalent relationship to the robot”.

↑
The Good Robot Podcast - 28/9/2021 (Mackereith 2021)
Rosy Braidotti on the Posthuman -

Relational and post-human shifts are necessary to confront the reality that embodied AI perpetuates worlds and realities that might de-humanize and oppress others. What do these shifts mean for our work as researchers, designers, and policymakers? Below, we propose five ways to approach the paradigm shifts outlined above. The propositions are not meant to provide a recipe or a toolkit but to stimulate academics and practitioners to reconsider ingrained practices.

01 EPISTEMOLOGY CHANGE: RAISING ABOVE THE CONCEPT OF PARTICIPATION AND DEMOCRATIZING AI TOWARDS POST-HUMAN TRANSDISCIPLINARY WAYS OF KNOWING

We have witnessed an increasing drive towards extending participation in the design of embodied AI, along with a call to democratize AI. As we have seen multiple times in this book, participation is hard, costly, and often paternalistic. Democratization of AI is typically confined to efforts to make technology available and usable by a broader audience. The main criticism is that participation and democratization are used as marketing moves to practice ethics washing rather than focusing on extending access to decision-

making processes (Baur 2020). We believe that efforts to broaden participation are fundamental and that we need to move towards an equitable and just way to involve people impacted by and have lived experience of AI. A way we would like to explore in the future is engaging in transdisciplinary research and combining it with post-human perspectives. Transdisciplinarity embraces an engaged and society-oriented way of practicing research (Thompson Klein 2004), combining academic knowledge, lived experiences, and non-academic ways of knowing to extend knowledge production and decision-making to human and non-human communities. It entails focusing on societal challenges and problems rather than disciplinary research questions. Transdisciplinary practices focus on mutual and embodied social learning rather than mere participation. World-view awareness, power literacy and reflexive/dialogic skills (van der Bijl-Brouwer 2022) are typical transdisciplinary competencies that are well suited for collective sensemaking and action on the futures of embodied AI. Knowledge in transdisciplinary research is not produced only by academics but also by communities, the private sector, governmental institutions / political systems, and in some cases, the natural environment on an equitable footing, opening up to avenues where post-human and relational perspectives are possible.

02 REFLEXIVITY AND POWER LITERACY: RESPONSIBLE DESIGN MUST CONSIDER POLITICAL RESPONSIBILITY AND PRACTICE CONSCIOUSNESS-RAISING.

We believe that to practice responsible design in embodied AI and taking into account social justice, equity, and inclusion means critically examining and taking action on the field's social, political, and economic facets and drivers. A deep and committed work of consciousness-raising at the individual, organizational and societal levels is needed. In particular, responsibility means being aware of power structures and our role and relations in society. This work must be done in collaboration with those who have the lived experience of embodied AI, aware of the complex ecosystem of relationships we have with nature, animals, humans, and non-humans.

03 METHODOLOGICAL CHANGE: FAIRNESS IS MORE THAN A PROBABILISTIC SUPPORT DECISION; IT NEEDS A HOLISTIC PERSPECTIVE.

Fairness in AI, embodied AI, and machine learning refers to attempts to correct algorithmic bias against gender, race, disability, etc. Current approaches focus on statistical fairness to correct biased predictions or decision-making. Actions are usually taken on data at pre-processing, in-processing, and post-processing phases. As many scholars advocate (De-Arteaga et al. 2018; Birhane and Cummins 2019; Malliaraki 2019; Lewis et al. 2020), we should view fairness as a complex phenomenon embedded in our socio-economical and political structure. Fairness considerations should be considered at the beginning of an embodied AI project. We believe that to tackle bias in embodied AI, discrimination and marginalization need a holistic perspective, and interventions should be taken before algorithmic design. Understanding the complexity of a social problem should be the starting point of any attempt at fairness in embodied AI.

04 METHODOLOGICAL CHANGE: GOING BEYOND EXPLAINABILITY TOWARD SITUATED SENSE-MAKING

Explainability refers to approaches to ensure that humans can understand the decisions and actions of an AI. The premise of explainable AI is to offer a framework of tools that will make AI more transparent and trustworthy. In embodied AI, explainability is mediated by an agent's behaviors and the social context in which non-human and human agents interact (Wallkötter et al. 2021). The way an artificial agent moves, interacts and acts with us is parsed by humans trying to make sense of the interaction. We believe that to design more inclusive, diverse, and trustworthy embodied AI, our focus on explainability should shift from decision-making transparency towards how people make sense of AI intentions and social attitudes by parsing agents' behaviors, actions, and embodied cues (Miller 2019). We argue we should be shifting towards designing embodied AI interactions aware of the existence of a plurality of interpretations, the importance of context, and the implication of the illusion of life.

05 POST-HUMAN BODIES OF AI: WE SHOULD RETHINK THE BODIES AND BEHAVIORS OF AI TO AVOID DEHUMANIZING HUMANS.

We believe the field of embodied AI should critically examine the tendency to replicate the human form and behavior and take the “man” as the model and measure of things. As embodied AI seems to define and redefine what it means to be human discriminating based on gender, race, and ability, we should explore ways to depart from human-like forms and stereotypical behaviors. To take action, we should start with an in-depth analysis of what embodied AI represents, excludes, and reminisces.

06 FUTURING AND MAKING VISIONS TANGIBLE

When discussing the future(s) of (embodied) AI, we typically envision utopian/techno-optimist or dystopian/apocalyptic scenarios. These narratives permeated the social discourse of AI and the designer’s imagination. Current representation, bodies, behaviors, and roles of robots, conversational interfaces, smart speakers, and virtual agents reflect those narratives. We believe we should embrace futuring and speculative design methods to imagine just and equitable futures for embodied AI. A way to go in line with transdisciplinary and post-human perspectives is to engage in collaborative sense-making and ideation. We believe tangibilizing or ‘visibilizing’ (Matos Castaño et al., 2020; Schoffelen et al., 2015) is suitable. Tangibilizing consists in making thoughts, ideas, and reflections from concrete to tangible through visual scenarios and prototypes and, in so doing, supporting collective reflection and status-quo challenging. To this end, speculative design (Dunne and Raby, 2013), discursive design, design fiction or experiential futures (Candy et al., 2017), or other similar approaches provide us with techniques to translate imagination, abstract thoughts, and future visions into the present.

“And I think that tension between the two, and sort of the alliance of the non-humans and the fear of the dehumanized humans, towards technology that is structural, in certainly feminist cultures, anti-racist cultures, and anti-fascist cultures. So it’s a creative tension with the left historically taking a position against technology. And I see contemporary feminism as correcting that and trying to move a step further. So yes, good robots are possible, but some negotiations are necessary.”

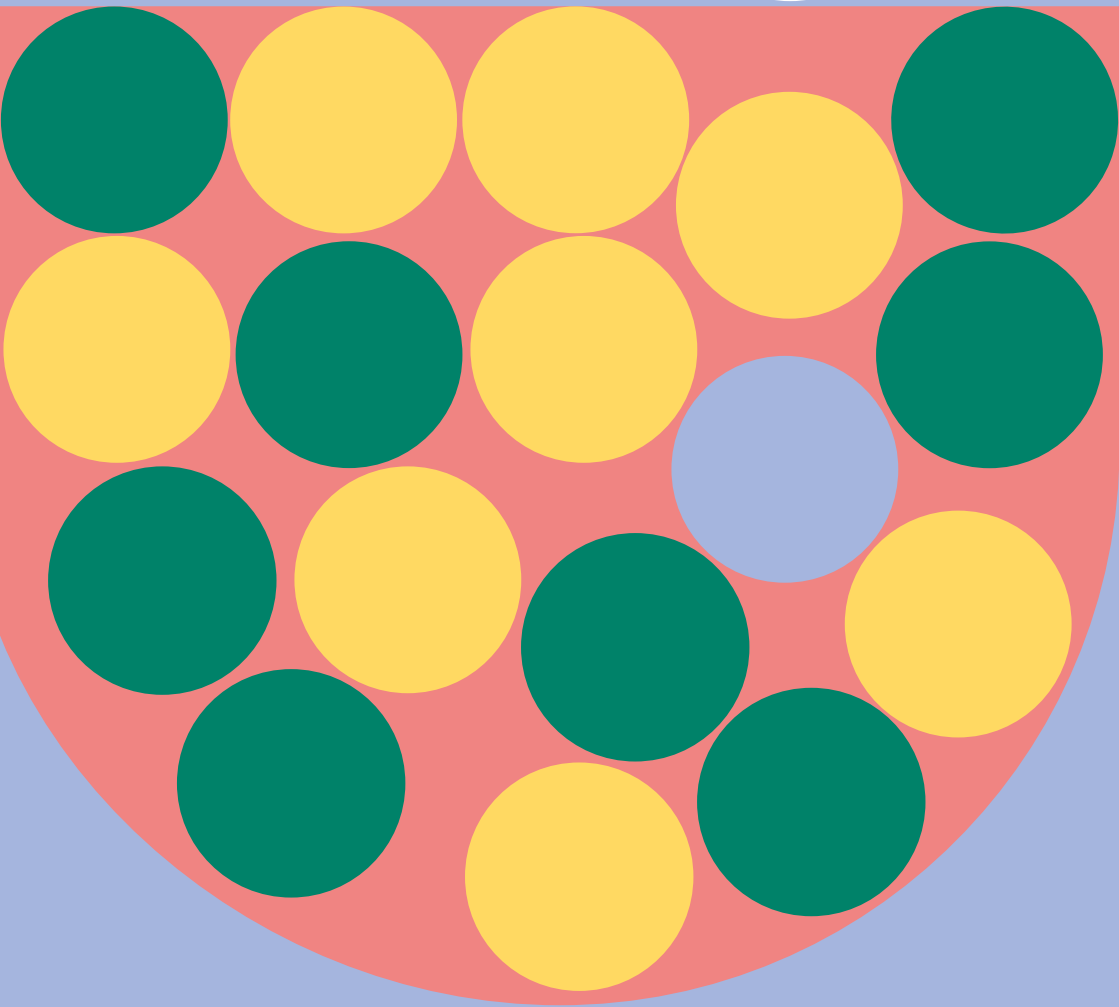
↑
The Good Robot Podcast - 28/9/2021
Rosy Braidotti on the Posthuman -

We got the chance to challenge our perspectives and engage in reflexivity about state of art in embodied AI and our collective design and development practices. While this book is a testament to our experiences during our year of activities, we believe we have only started to scratch the surface and hope to

continue our work.
We remain positive that
the futures of embodied
AI can be of social and
environmental justice.
We hope the book
is an inspiration to a
transdisciplinary community
of practice and serves
as a guide to take action,
question the status quo,
and engage in radical
change.

THANKS TO ALL THE
COLLABORATORS
OF THE DEI4EAI INITIATIVE.

— REFERENCES



A “AI4SDG.” n.d. Accessed July 4, 2022.
<https://www.aiforsdgs.org/>.

“AI for Good.” n.d. Microsoft. Accessed July 26, 2022. <https://www.microsoft.com/en-us/ai/ai-for-good>.

AI Now Institute. 2021. “A New AI Lexicon: Social Good – A New AI Lexicon – Medium.” A New AI Lexicon. November 4, 2021. <https://medium.com/a-new-ai-lexicon/a-new-ai-lexicon-social-good-c71edab270d3>

Aizenberg, E., & Van Den Hoven, J. (2020). Designing for human rights in AI. *Big Data & Society*, 7(2), 2053951720949566.

Alkhatib, A. (2021). To Live in Their Utopia: Why Algorithmic Systems Create Absurd Outcomes. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–9. CHI '21 95. New York, NY, USA: Association for Computing Machinery.

Allen, J. F., Byron, D. K., Dzikovska, M., Ferguson, G., Galescu, L., and Stent, A. (2001). Toward Conversational Human-Computer Interaction. *AI Magazine* 22 (4): 27–27.

Amis, J. M., Munir, K. A., Lawrence, T. B., Hirsch, P., & McGahan, A. (2018). Inequality, institutions and organizations. *Organization Studies*, 39(9), 1131–1152.

Arista, N., Costanza-Chock, S., Ghazavi, V., and Kite, S. (2021). *Against Reduction: Designing a Human Future with Machines*. MIT Press.

Atewologun, D., & Sealy, R. (2014). Experiencing privilege at ethnic, gender and senior intersections. *Journal of Managerial Psychology*.

Ayling, J., & Chapman, A. (2021). Putting AI ethics to work: are the tools fit for purpose?. *AI and Ethics*, 1–25.

- Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social Robots for Education: A Review. 3 (21): eaat5954.
- Benson, T. (2020). How Robots and A.I. Could Help Save Our Healthcare Workers and the Elderly. Inverse. Accessed March 25, 2020. <https://www.inverse.com/innovation/how-robots-ai-could-help-save-our-healthcare-workers-the-elderly>
- Better Images of AI. n.d. Accessed July 6, 2022. <https://betterimagesofai.org/>.
- Bharti, U., Bajaj, D., Batra, H., Lalit, S., Lalit, S., and Gangwani, A. (2020). Medbot: Conversational Artificial Intelligence Powered Chatbot for Delivering Tele-Health after COVID-19. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, 870–75.
- Birhane, A., and Cummins, F. (2019). Algorithmic Injustices: Towards a Relational Ethics. *arXiv [cs.CY]*. arXiv. <http://arxiv.org/abs/1912.07376>.
- Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R., and Bao, M. (2022). The Values Encoded in Machine Learning Research. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 173–84. FAccT '22. New York, NY, USA: Association for Computing Machinery.
- Bochner, A. P. (2012). On first-person narrative scholarship: Autoethnography as acts of meaning. *Narrative inquiry*, 22(1), 155–164.
- Braidotti, R. (2022). *Posthuman Feminism* | Wiley.
- Brewer, Rose M., and Patricia Hill Collins. 1992. “Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment.” *Contemporary Sociology* 21 (1): 132. <https://doi.org/10.2307/2074808>.
- Broekens, J., Heerink, M., & Rosendal, H. (2009). Assistive Social Robots in Elderly Care: A

Review.s Gerontechnology: International Journal on the Fundamental Aspects of Technology to Serve the Ageing Society. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.714.6939&rep=rep1&type=pdf>.

Buolamwini, J., and Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, edited by Sorelle A. Friedler and Christo Wilson, 81:77–91. Proceedings of Machine Learning Research. PMLR.

C Cassell, J. (2001). Embodied Conversational Agents: Representation and Intelligence in User Interfaces. *AI Magazine* 22 (4): 67–67.

Cavalcante Siebert, L., Lupetti, M. L., Aizenberg, E., Beckers, N., Zgonnikov, A., Veluwenkamp, H., ... & Legendijk, R. L. (2022). Meaningful human control: actionable properties for AI system development. *AI and Ethics*, 1–15.

CBC News. 2019. “How Algorithms Create a ‘Digital Underclass,’” September 4, 2019. <https://www.cbc.ca/radio/ideas/how-algorithms-create-a-digital-underclass-1.5269959>.

Charlton, J. I. (1998). Nothing about us without us. In *Nothing About Us Without Us*. University of California Press.

Cila, N. (2022). Designing Human-Agent Collaborations: Commitment, Responsiveness, and Support. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–18. CHI ’22 420. New York, NY, USA: Association for Computing Machinery.

Costanza-Chock, S. (2018). Design justice: Towards an intersectional feminist framework for design theory and practice. *Proceedings of the Design Research Society*.

Costanza-Chock, S. (2020). Design Justice. *Community-Led Practices to Build the Worlds We Need*. Edited by Sandra Braman. The Information Policy Series. The MIT Press.

Crawford, K. (2021). *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., Kak, A., et al. (2019). *AI Now 2019 Report*. New York: AI Now Institute.

D De-Arteaga, Maria, William Herlands, Daniel B. Neill, and Artur Dubrawski. 2018. "Machine Learning for the Developing World." *ACM Trans. Manage. Inf. Syst.*, 9, 9 (2): 1–14. <https://doi.org/10.1145/3210548>.

Dellon, B., and Matsuoka, Y. (2007). Prosthetics, Exoskeletons, and Rehabilitation [Grand Challenges of Robotics]. *IEEE Robotics & Automation Magazine / IEEE Robotics & Automation Society* 14 (1): 30–34.

Deng, E., Mutlu, B., and Mataric, M. J. (2019). Embodiment in Socially Interactive Robots. *Foundations and Trends® in Robotics* 7 (4): 251–356.

D'Ignazio, C., and Klein, L. F. (2020). *Data Feminism*. MIT Press.

Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer Nature.

Dignum, Virginia. 2017. "Responsible Artificial Intelligence: Designing AI for Human Values," September. <http://dspace.daffodilvarsity.edu.bd:8080/handle/123456789/2181>.

E Erete, S., Israni, A., and Dillahunt, T. (2018). *An intersectional approach to designing in the margins*. *interactions* 25, 3 (May–June 2018), 66–69. <https://doi.org/10.1145/3194349>

F Feil-Seifer, D., and Mataric, M. J. (2005). Defining Socially Assistive Robotics. In *9th International Conference*

on Rehabilitation Robotics, 2005. ICORR 2005., 465–68.
ieeexplore.ieee.org.


Floridi, Luciano. 2019. “Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical.” *Philosophy & Technology* 32 (2): 185–93. <https://doi.org/10.1007/s13347-019-00354-x>.

Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, et al. 2018. “AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations.” *Minds and Machines* 28 (4): 689–707. <https://doi.org/10.1007/s11023-018-9482-5>.

Forlano, Laura. 2017. “Posthumanism and Design.” *She Ji: The Journal of Design, Economics, and Innovation* 3 (1): 16–29. <https://doi.org/10.1016/j.sheji.2017.08.001>.

Fosch-Villaronga, E. & Poulsen, A. (2022). Diversity and Inclusion in Artificial Intelligence. In: Custers, B., Fosch-Villaronga, E. (eds) Law and Artificial Intelligence. Regulating AI and Applying AI in Legal Practice. Information Technology and Law Series, vol 35. T.M.C. Asser Press, The Hague, 109–134. https://doi.org/10.1007/978-94-6265-523-2_6.

Franklin, S. (1997). Autonomous Agents as Embodied AI. *Cybernetics & Systems*. https://www.tandfonline.com/doi/abs/10.1080/019697297126029?casa_token=B4nnyRCrVBkAAAAA:hhFkye44dOeYcOR_5i6kG_E-VcLRIMhluSfyWXQamRVzUOifghx5Rg3bjhaADQ13TEURZrs2nOmWqQ.

 Grady, S. D., Wisniewski, P., Metoyer, R., Gibbs, P., Badillo-Urquiola, K., Elsayed-Ali, S., & Yafi, E. (2020). Letters accompanying the ACM Interactions blog post titled: Addressing Institutional Racism within Initiatives for SIGCHI’s Diversity and Inclusion. Retrieved on July 19th, 2022, from: <https://bit.ly/3IRVEWW>

Grady, S. D., Wisniewski, P., Metoyer, R., Gibbs, P., Badillo-Urquiola, K., Elsayed-Ali, S., & Yafi, E. (2020). Addressing

Institutional Racism within Initiatives for SIGCHI's Diversity and Inclusion.

H Heteren Pieter Nel, Michael Chui, Martin Harryson, James Manyika, Roger Roberts, Rita Chung, Ashley van. 2018. "NOTES FROM THE AI FRONTIER APPLYING AI FOR SOCIAL GOOD." *McKinsey Global Institute* (MGI). <https://www.mckinsey.com/~media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Applying%20artificial%20intelligence%20for%20social%20good/MGI-Applying-AI-for-social-good-Discussion-paper-Dec-2018.ashx>.

Howley, I., Kanda, T., Hayashi, K., and Rosé, C. (2014). Effects of Social Presence and Social Role on Help-Seeking and Learning. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, 415–22. ACM.

Hundt, A., Agnew, W., Zeng, V., Kacianka, S., and Gombolay, M. (2022). Robots Enact Malignant Stereotypes. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 743–56. FAccT '22. New York, NY, USA: Association for Computing Machinery.

Hussain, R., and Zeadally, S. (2019). Autonomous Cars: Research Results, Issues, and Future Challenges. *IEEE Communications Surveys & Tutorials* 21 (2): 1275–1313.

K Kēpuska, V., and Bohouta, G. (2018). Next-Generation of Virtual Personal Assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, 99–103. ieeexplore.ieee.org.

Keyes, O. (2018). The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. *Proc. ACM Hum.-Comput. Interact.*, 88, 2 (CSCW): 1–22.

Kong, K., and Jeon, D. (2006). Design and Control of an Exoskeleton for the Elderly and Patients. *IEEE/ASME Transactions on Mechatronics: A Joint Publication of the IEEE*

Industrial Electronics Society and the ASME Dynamic Systems and Control Division 11 (4): 428–32.

L Lee, M., Noortman, R., Zaga, C., Starke, A., Huisman, G., and Andersen, K. (2021). Conversational Futures: Emancipating Conversational Interactions for Futures Worth Wanting. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM. <https://doi.org/10.1145/3411764.3445244>.

Lemonne, Eric. 2021. “Ethics Guidelines for Trustworthy AI.” May 18, 2021. <https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>.

Lewis, Jason Edward, Angie Abdilla, Noelani Arista, Kaipulaumakaniolono Baker, Scott Benesiinaabandan, Michelle Brown, Melanie Cheung, et al. 2020. “Indigenous Protocol and Artificial Intelligence Position Paper.” Edited by Jason Edward Lewis. <https://doi.org/10.11573/spectrum.library.concordia.ca.00986506>.

Lewis, Jason Edward, Noelani Arista, Archer Pechawis, and Suzanne Kite. 2018. “Making Kin with the Machines.” *Journal of Design and Science*, July. <https://doi.org/10.21428/bfafd97b>.

MacInnes, P. (2021). Toyota Pauses Paralympics Self-Driving Buses after One Hits Visually Impaired Athlete. *The Guardian*, August 28, 2021. <https://amp.theguardian.com/technology/2021/aug/28/toyota-pauses-paralympics-self-driving-buses-after-one-hits-visually-impaired-athlete>.

M Mackereth, Kerry. 2021. “Rosi Braidotti on the Posthuman (Part 1).” My Site. September 28, 2021. <https://thegoodrobotpodcast.wixsite.com/the-good-robot/post/rosi-braidotti-on-the-posthuman-part-1>.

Malliaraki, Eirini. 2019. “What Is This ‘AI for Social Good’? – Eirini Malliaraki.” Medium. May 19, 2019. <https://eirinimalliaraki.medium.com/what-is-this-ai-for-social-good-f37ad7ad7e91>.

- Miller, Tim. 2019. "Explanation in Artificial Intelligence: Insights from the Social Sciences." *Artificial Intelligence* 267 (February): 1–38. <https://doi.org/10.1016/j.artint.2018.07.007>.
- Mozilla Foundation. (2019). Internet Health Report 2019. <https://doi.org/10.14361/9783839449462>
- O Ogbonnaya-Ogburu, I. F., Smith, A. D., To, A., & Toyama, K. (2020). Critical race theory for HCI. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-16)
- P "PAIR –." n.d. Google Research. Accessed July 26, 2022. <https://research.google/teams/brain/pair/>.
- Patston, P. (2007). Constructive Functional Diversity: A new paradigm beyond disability and impairment. *Disability and rehabilitation*, 29(20-21), 1625-1633.
- Perugia, G., Guidi, S., Bicchi, M., and Parlangeli, O. (2022). The Shape of Our Bias: Perceived Age and Gender in the Humanoid Robots of the ABOT Database. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, 110–19. HRI '22. IEEE Press.
- Piyare, R., and Lee, S. R. (2013). Towards Internet of Things (IOTS):Integration of Wireless Sensor Network to Cloud Services for Data Collection and Sharing. *arXiv [cs.NI]*. arXiv. <http://arxiv.org/abs/1310.2095>.
- Pradhan, A., Findlater, L., and Lazar, A. (2019). 'Phantom Friend' or 'Just a Box with Information.' *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW): 1–21.
- R Rocasolano, M. M. (2022). Human Rights, Big Data and Artificial Intelligence: Elements of a Complex Algorithm. In *Security and Defence: Ethical and Legal Challenges in the Face of Current Conflicts* (pp. 93-102). Springer, Cham.

“RO-MAN22 GenR Workshop.” n.d. Accessed July 22, 2022. <https://sites.google.com/view/ro-man-genr-workshop/home>.

Rosenberg, L, and A. I. Unanimous. (2022). Get Ready for Your Evil Twin. April 23, 2022. <https://venturebeat.com/2022/04/23/get-ready-for-your-evil-twin/>.

S Sanders, M. R., & Mahalingam, R. (2012). Under the radar: The role of invisible discourse in understanding class-based privilege.

Shew, A. (2020). Ableism, technoableism, and future AI. *IEEE Technology and Society Magazine*, 39(1), 40–85.

Sloane, M., Moss, E., Awomolo, O., & Forlano, L. (2020). Participation is not a design fix for machine learning. *arXiv preprint arXiv:2007.02423*.

Spiel, Katta, Os Keyes, and Pinar Barlas. 2019. “Patching Gender: Non-Binary Utopias in HCI.” In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–11. CHI EA ’19. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3290607.3310425>.

Stahl, Bernd Carsten, and Mark Coeckelbergh. 2016. “Ethics of Healthcare Robotics: Towards Responsible Research and Innovation.” *Robotics and Autonomous Systems* 86 (December): 152–61. <https://doi.org/10.1016/j.robot.2016.08.018>.

T Tomašev, N., Cornebise, J., Hutter, F., Mohamed, S., Picciariello, A., Connelly, B., Belgrave, D. C. M., et al. (2020). AI for Social Good: Unlocking the Opportunity for Positive Impact. *Nature Communications* 11 (1): 2468.

U Unesco. 2020. “ARTIFICIAL INTELLIGENCE and GENDER EQUALITY.” 2020. <https://unesdoc.unesco.org/ark:/48223/pf0000374174>.

- V Varela, F. J., & Shear, J. (1999). First-person methodologies: What, why, how. *Journal of Consciousness studies*, 6(2-3), 1-14.
- Verma, S. (2019). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. *Vikalpa* 44 (2): 97-98.
- W Wakkary, Ron. 2021. "Things We Could Design." <https://doi.org/10.7551/mitpress/13649.001.0001>.
- Wallkötter, Sebastian, Silvia Tulli, Ginevra Castellano, Ana Paiva, and Mohamed Chetouani. 2021. "Explainable Embodied Agents Through Social Cues: A Review." *J. Hum.-Robot Interact.*, 27, 10 (3): 1-24. <https://doi.org/10.1145/3457188>.
- West, S. M., Whittaker, M., & Crawford, K. (2019). Discriminating systems. *AI Now*.
- Winner, L. (1980). Do Artifacts Have Politics? *Daedalus* 109 (1): 121-36.
- Wolf, C. T. (2020). Democratizing AI? experience and accessibility in the age of artificial intelligence. *XRDS: Crossroads, The ACM Magazine for Students*, 26(4), 12-15.
- World economic forum. 2022. "A Blueprint for Equity and Inclusion in Artificial Intelligence." June 2022. https://www3.weforum.org/docs/WEF_A_Blueprint_for_Equity_and_Inclusion_in_Artificial_Intelligence_2022.pdf.
- Z Ziemke, T. (2004). Embodied AI as Science: Models of Embodied Cognition, Embodied Models of Cognition, or Both? In *Embodied Artificial Intelligence: International Seminar, Dagstuhl Castle, Germany, July 7-11, 2003. Revised Papers*, edited by Fumiya Iida, Rolf Pfeifer, Luc Steels, and Yasuo Kuniyoshi, 27-36. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Ziemke, T., and Lowe, R. (2009). On the Role of Emotion in Embodied Cognitive Architectures: From Organisms to Robots. *Cognitive Computation* 1 (1): 104-17.

A SPECIAL THANK YOU TO

Catherine D'Ignazio
Lisa Mandemaker
Cameron Lee Taylor
Abdelrahman (Abdo) Hassan
Kristen Parisi
Simon Dogger
Lonneke van Kampen
Valentina Migliarini
Laura Forlano
Future Wake
Abran Maldonado

Editors in Chief

Cristina Zaga and Maria Luce Lupetti

Authors

Cristina Zaga, Maria Luce Lupetti,
Nazli Cila, Minha Lee, Gijs Huisman,
Eduard Fosch-Villaronga Anne Artberger

The book is sponsored by

4TU NIRICT — Diversity Fund

A collaboration between

University of Twente, Delft University
of Technology, Eindhoven University
of Technology, Leiden University

DOI-number

10.3990/1.9789036553599

<https://doi.org/10.3990/1.9789036553599>

ISBN (PRINT) 978-90-365-5358-2

ISBN (DIGITAL) 978-90-365-5359-9

Art direction

Alice Lotti, Maria Luce Lupetti

Book design

Alice Lotti

Illustrations and Graphics

Alice Lotti

Paper

Fedrigoni Arena Smooth
120—300 gr

Typefaces

Apfel Grotezk (collettivo.it),
Savate (collectif-we.ch) by
Velvetyne Type Foundry

Edition

Printed in 100 copies

September 2022

Creative Commons

CC BY-NC-ND 4.0



ISBN 978-90-365-5359-9

UNIVERSITY OF TWENTE