

# A Case for Diverse Social Robot Identity Performance in Education

Lux Miranda  
Department of Information Technology,  
Uppsala University,  
Uppsala, Sweden  
lux.miranda@it.uu.se

Ginevra Castellano  
Department of Information Technology,  
Uppsala University,  
Uppsala, Sweden  
ginevra.castellano@it.uu.se

Katie Winkle  
Department of Information Technology,  
Uppsala University,  
Uppsala, Sweden  
katie.winkle@it.uu.se



Figure 1: Human-like robot platforms such as Furhat require designers to select from a variety of voices, faces, and languages—which leads users to ascribe identity characteristics such as gender and ethnicity.

## ABSTRACT

Educational outcomes for students belonging to disadvantaged social identities are unavoidably influenced by overlapping systems of inequity which arise along lines such as gender, ethnicity, and age. Robot platforms like Furhat require designers to select features which are interpreted by users as these same kinds of social identity. Prior work has posited that social robots might be intentionally designed to leverage these social identities in a “norm-breaking” fashion with the aim of disrupting social stereotypes in STEM education. However, research in HRI has been largely limited to the examination of gender only. We present a 2x2, between-subjects study in which 161 participants aged 9-12 are shown a robot-delivered lecture presented by a group of three separate robot personas with varying gender and ethnicity performances. We find that participants place greater trust in the persona groups with high gender diversity. Incorporating ethnic diversity seems to have little impact on our quantitative interaction metrics, however we do find evidence to suggest diversity in robots’ language capabilities may be important for trustworthiness. In all, the study contributes nuance to the discussions on the implications of (norm-breaking) social identity performance when using robots to pursue more equitable STEM education.

## CCS CONCEPTS

• **Social and professional topics** → Gender; Geographic characteristics; Race and ethnicity; K-12 education; • **Human-centered computing** → Empirical studies in HCI; HCI theory, concepts and models; • **Computer systems organization** → Robotics.

## KEYWORDS

robot identity; diversity; intersectionality; gender bias; human-robot interaction; trustworthy HRI; STEM education; feminist HRI

## ACM Reference Format:

Lux Miranda, Ginevra Castellano, and Katie Winkle. 2024. A Case for Diverse Social Robot Identity Performance in Education. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24 Companion)*, March 11–14, 2024, Boulder, CO, USA. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3610978.3640768>

## 1 INTRODUCTION













Students’ educational experiences are constantly influenced by the inequities created by the intersection of overlapping structures of disadvantage which arise along lines of social identity such as gender, ethnicity, and class [4, 5, 14]. Despite this, most work concerning diversity in STEM (and indeed most work in HRI) considers gender only [9]. Strategies for tackling such inequities in STEM include educating about their existence and restructuring the educational and professional STEM environments to be more inclusive [5]. Exposure to diverse role models is another common strategy [7].

Previous work has posited that social robots might be intentionally designed to demonstrate “norm-breaking” social identities with the aim of disrupting social stereotypes in STEM education [19], not



This work is licensed under a Creative Commons Attribution-ShareAlike International 4.0 License.

**Table 1: Summary of robot personas and associated identity cues across the four treatments. Note languages spoken are in addition to English (the primary language used during robot presentations).**

	Low gender diversity			High gender diversity		
<b>Low ethnic diversity</b>						
	<b>Jonas</b> he/him speaks Swedish	<b>Martin</b> he/him speaks Swedish	<b>Gustav</b> he/him speaks Swedish	<b>Jonas</b> he/him speaks Swedish	<b>Lo</b> they/them speaks Swedish	<b>Hillevi</b> she/her speaks Swedish
<b>High ethnic diversity</b>						
	<b>Jonas</b> he/him speaks Swedish	<b>Omar</b> he/him speaks Arabic (Syrian)	<b>Mustafa</b> he/him speaks Somali	<b>Amany</b> she/her speaks Arabic (Syrian)	<b>Lo</b> they/them speaks Swedish	<b>Mustafa</b> he/him speaks Somali

unlike similar ideas seen in toys and other media, e.g. “Computer Engineer Barbie” [8]. Furthermore, robot platforms like Furhat require roboticists to choose characteristics which unavoidably cue for social identities such as gender, ethnicity, and age. This both bestows a responsibility to the roboticist to accomplish this in a socially harmonious way, whilst also enabling great possibilities for exploring a wide diversity of social robot identity performances.

While recent work has suggested that robots should not represent socially salient traits such as gender and race [17], we argue that diverse, human-like social robot identity performances of these characteristics — when mindfully executed — might yet have the potential to affect positive social change when it comes to inequities in STEM education, and we demonstrate that this may also allow for more trustworthy robots. In this work, we explore how diversity in robot social identity performance influences participants’ interest in robotics and computer science, their social outlooks concerning gendered biases, their perceptions of how welcoming these fields are to people of diverse identities, and the intersectional identities held by the participants themselves.

### 1.1 Research questions

1. (How) do children differently perceive diverse robot identities in the context of their ability to help with learning?
2. (How) does the level of diversity in gender and ethnicity of human-like robot personas impact children’s perceptions and biases towards:
  - a. Robotics?
  - b. Social identity?
3. (How) does the level of diversity in gender and ethnicity of human-like robot personas impact established human-robot interaction metrics relevant for the educational setting, such as social trust and competency trust?

## 2 METHODS

We employ a 2x2, between-subjects study in which 161 participants (72 female, 86 male, 3 nonbinary) aged 9-12 are shown a robot-delivered presentation on the topic of machine learning and algorithmic bias (adapted from the MIT AI Ethics Curriculum for Middle School Students [10]). The presentation consists of the Furhat robot giving a 15 minute lecture accompanied by slides during which it cycles through and embodies three different personas, each of which each gives a different part of the lecture (with approximately 5 minutes presentation time per persona). The presentation given by each “persona team” is the same. To avoid priming, the presentation does not specifically discuss topics related to diversity (or lack thereof) in social robot identities. The full robot presentation script (along with a repository link for our code and the robot presentation slides) is available in the supplementary materials. Participants are provided with a questionnaire which includes a demographic survey and questions designed to measure social biases, interest in and perception of computer science and robotics, and how much social and competency trust the participants place in the personas. Social trust is a measure of participant’s trust that a particular robot would engage in pro-social behavior, while competency trust concerns trust the the robot is competent in its designated purpose [13].

The two variables we manipulate are diversity in gender and ethnicity performance of the robot personas. Each variable has two levels: “high,” or “low.” In the low gender diversity treatments, all personas are cued as men. In the high gender diversity treatments, one persona is cued as a man, one as a woman, and one as nonbinary. In the low ethnic diversity treatments, all personas are cued as ethnically Swedish. In the high ethnic diversity treatments, one persona is cued as ethnically Swedish, one as ethnically Syrian, and one as ethnically Somali. Table 1 identifies the name, pronouns,

Metric	Pre- or post-hoc	Statement (English)	Statement (Swedish)
Interest in robotics or computer science			
I1	Both	I am interested in learning more about robotics and/or computer science.	<i>Jag är intresserad av att lära mig mer om robotteknik och/eller datavetenskap.</i>
I2	Both	I would enjoy working with robots and/or computer science in the future.	<i>Jag skulle gärna arbeta med robotteknik och/eller datavetenskap i framtiden.</i>
Gender bias			
G1	Both	Girls find it <b>harder</b> to understand robots and computer science than boys do.	<i>Tjejer har <u>svårare</u> att förstå robotar och datavetenskap än killar.</i>
G2	Both	Girls find it <b>easier</b> to understand robots and computer science than boys do.	<i>Tjejer har <u>lättare</u> att förstå robotar och datavetenskap än killar.</i>
Perception of welcomeness			
W1	Both	Girls are just as welcome as boys to work in robotics and computer science.	<i>Tjejer är lika välkomna som killar att arbeta med robotteknik och datavetenskap.</i>
W2	Both	Everyone is welcome to work in robotics and computer science, no matter where they are from.	<i>Alla är välkomna att arbeta med robotteknik och datavetenskap, det spelar ingen roll var de kommer ifrån.</i>
Social trust			
ST1	Post	I feel that I can trust the robot presenters	<i>Jag känner att jag kan lita på robotpresentatörerna.</i>
ST2	Post	I feel that the robot presenters are honest	<i>Jag känner att robotpresentatörerna är ärliga.</i>
Competency trust			
CT1	Post	I feel that the robot presenters know a lot of things	<i>Jag känner att robotpresentatörerna vet många saker.</i>
CT2	Post	I feel that the robot presenters are smart	<i>Jag känner att robotpresentatörerna är smarta.</i>

**Table 2: Statements provided to participants on the questionnaire. Participants are asked to rate their agreement all statements using a 5-point Likert scale.**

language, and Furhat face used to create each of the three personas used within each condition. To introduce and reinforce exposure to the three personas, the presentation begins with introductions to each persona, ends with goodbyes from each persona, and a slide containing the pictures, names, and pronouns of each persona are displayed during the introduction and goodbye. Personas introduce themselves with minor variations on the sentence “Hi, my name is [name], my pronouns are [pronouns], and I speak English and [other language],” followed by a brief quip in the other language such as “Good morning!” or “Nice to meet you all!”

In the context of STEM education, the robot personas utilized in the high-diversity conditions are motivated by continued racialized, gender-based inequities in STEM [6, 16]. Varying gender as well as ethnicity in a 2x2 design allows us to take a more intersectional approach in line with feminist HRI principles [18], allowing us to begin to more expansively understand social robot identity performance and the way it may be perceived by different children.

The choice of the Syrian and Somali ethnicities is motivated by several factors. Firstly, to allow for a reasonable “cultural distance” between each ethnicity, we choose each ethnicity to originate from a separate continent (in this case: Europe, Africa, and Asia). Further, we rely on Sweden’s immigration statistics to select ethnicities from the most common nationalities of people living in Sweden with an international origin. This lends an additional level of relevancy when interfacing with the local community, as the selection represents ethnicities which people are more likely to belong to themselves or encounter day-to-day. According to 2022 statistics, the top two United Nations geoscheme subregions producing foreign-born

persons living in Sweden are western Asia and eastern Africa [11]. Within these two subregions, the top source countries are Syria and Somalia, respectively. Thus, given that these two countries’ populations consist of a majority respectively-eponymous ethnic groups (Syrian and Somali), we select them as the two ethnicities to represent in addition to the ethnically Swedish personas.

We recognize that, in attempting to design robots with such social identities, we cannot manipulate the identities directly. We may, however, control cues which lead to a higher probability of a given identity being ascribed by an interactant [9]. Thus, to influence participant’s ascription of gender and ethnicity onto the personas, we employ several cues.

To manipulate perception of gender, the most direct cues are the personas’ statement of their name and pronouns in their introduction. This is further reinforced by the selection of Furhat character faces, which employ a variety of gendered visual cues such as the presence or absence of makeup or facial hair. Third is the selection of the voice. We utilize Microsoft Azure text-to-speech voices. These designers of these voices designate an explicitly binary “male” or “female” gender for each voice, meaning that they feature variations in pitch, timbre, and tone which are typically (but not necessarily) perceived as either more male-like or female-like.

To manipulate the perception of ethnicity, the primary cue used is the additional language spoken by the persona in addition to English. That is, Swedish is used to cue for Swedish ethnicity, Somali for Somali ethnicity, and a Syrian variety of Arabic (the Microsoft Azure implementation of text-to-speech language code *ar – SY*) is used to cue for Syrian ethnicity. This is further reinforced by the

selection of the persona's name (chosen to be relatively common names among members of each ethnic group) as well as the Furhat character face chosen to roughly exhibit morphological features commonly (but not necessarily) possessed by members of each ethnic group.

Details on the creation of the nonbinary persona deserve special attention in this text due to the relative newness of the concept of truly nonbinary (and not simply "gender-neutral") robots' entry into the HRI discourse [12]. In humans, there is no particular way in which one must look or sound in order to be perceived as nonbinary. Androgyny is not necessarily indicative of nonbinary gender – nonbinary people may look, sound, and behave masculine, feminine, both, neither, androgynous, or none of the above. Self-identification as being nonbinary (as with any gender) is the only necessary and sufficient condition to cue for nonbinary gender [3]. Given that robots do not hold internal identities such as this, a cue with perhaps the next-highest probability of nonbinary gender is usage of they/them pronouns. As it is also common for nonbinary people to undertake gender-neutral names [3], we further reinforce the design through selecting a relatively common gender-neutral name within the persona's designated ethnic group.

We argue that these two cues (the gender-neutral name and they/them pronouns) should be sufficient for a persona to be perceived as nonbinary among participants who are already accustomed to interacting with nonbinary people, regardless of the other cues exhibited by the persona. Recent work exploring the creation of a nonbinary robot in a similar fashion supports this [12] and proposes that this may be a beneficial design strategy in helping normalize non-cisgender identity through increased public exposure.

We cannot, however, assume that every member of our study population is accustomed to interacting with nonbinary people and successfully perceiving them as nonbinary. We may expect that some participants may have little-to-no experience conceptualizing nonbinary gender or may otherwise come from a cultural upbringing with only a binary gender conception. In order to account for these possibilities, we intentionally select the Furhat character face and voices used for the nonbinary persona to lean in a direction such that, if a participant is not able to ascribe a nonbinary gender to the persona, they are likely to instead ascribe it a woman gender. This way, the norm-breaking aspect of the "high gender diversity" treatment group is maintained, as two out of three of the "AI-expert" personas would be perceived as women.

We again note that this uncertainty of whether participants ascribe the intended gender to each persona is a general one which applies also to the intended ethnicities we cue for, and we consider this a study limitation. To promote greater rigor in future work, we recommend an additional validation step where a sample of the target participant population is polled to see which identities are generally ascribed.

## 2.1 Participants

We carried out our study at a local (Swedish) international school which delivers a bilingual curriculum in English and Swedish, making it a popular choice with international families. Our study activities, the robot-led presentations, and experimental measure

completion were integrated into the timeslot of a typical 55-minute classroom activity for two different age groups (year 4/age 9-10 and year 6/age 11-12) with four class groups each, such that one year 4 and one year 6 class group each saw one of our four experimental conditions (determined at random by a dice roll). For statistical purposes in our intersectional analysis, we group participants into two gender groupings (male or female/nonbinary) and two nationality groupings (Swedish only or inter/multinational) based on their answers to the demographic survey. Female and nonbinary participants are classified together, as the number of nonbinary participants is too small to achieve a reasonable level of statistical power. We are moreover principally concerned with gender as a lens of analysis in the context of robotics and computer science where male-ness is the stereotypical norm and non-male-ness is not, thus grouping gender in this way remains conducive to answering our research questions. Table 3 identifies the total number of participants, grouped according to these classifications, who saw each of the experimental conditions.

## 2.2 Experimental measures

The questionnaires provided to participants are bilingual English-Swedish and consist of three parts: a demographic survey, pre-hoc questions answered before viewing the personas' presentation, and post-hoc questions answered after viewing.

The demographic survey collects the participants' self-identified age, gender, and nationality. Typically, when utilizing intersectionality as an analytical framework, it is ethnicity (not nationality) which is used as an important axis of analysis for understanding the unique ways in which groups and individuals may suffer from systemic inequities. While it is standard practice in much of the world to collect ethnicity information from study participants for analytical purposes, in Sweden this is considered sensitive data protected by law and is strictly regulated. The official position is that it cannot and should not be compiled. [1] The collection of nationality and country-of-origin data, however, is a standard practice which is often used instead. Hence, we follow local best practices and choose to do the same. We note that this has the effect at generally being a loose proxy of ethnicity for countries which primarily consist of one ethnic group, but is less so for more multicultural nations.

Questions measuring the participants' interest in robotics and computer science, bias towards or against girls' ability to engage with robotics and computer science, and perceptions of robotics and computer science as a field welcome (or not welcome) to all genders and nationalities are measured pre-hoc and post-hoc. Finally, questions measuring the participants social trust and competency trust in the robot personas, drawn from a subset of a larger Social/Competency Trust and Likability survey used in previous child-robot interaction studies, [13, 15] is measured post-hoc. As in these prior studies, all questions (apart from the demographic survey) are measured as a 5-point Likert scale asking respondents to rate their agreement to statements using the scale: "Absolutely not" (0 points), "No" (1 point), "Not sure" (2 points), "Yes" (3 points), and "Absolutely yes" (4 points). The survey questions and associated metrics are exhibited in Table 2.

Treatment	Gender	Age 9-10		Age 11-12		Total
		Inter/multinational	Swedish only	Inter/multinational	Swedish only	
High gender/ High ethnic	F+NB M	5 3	5 3	2 7	4 6	16 19
High gender/ Low ethnic	F+NB M	3 4	8 6	3 7	4 2	18 19
Low gender/ High ethnic	F+NB M	7 5	4 7	6 2	3 5	20 19
Low gender/ Low ethnic	F+NB M	3 6	9 8	6 9	3 6	21 29
Total		36	50	42	33	161

**Table 3: Cross-tabulated count of participants divided by demographic group and treatment received.**

Questions are written to strike a balance between redundancy (for resiliency to question interpretation) and avoiding survey fatigue from presenting the young participants with too many questions. For the gender bias metric, we ask participants to rate agreement with both “Girls find it **harder** to understand robots and computer science than boys do” as well as “Girls find it **easier** to understand robots and computer science than boys do.” Although inferring from historically stereotypical precedent, we might expect any bias to typically lean *against* girls’ abilities, asking the question in this way allows us to also capture participants in *favor* of girls’ abilities, which we may then still register as some amount of gendered bias.

### 2.3 Data and analysis

Following the conclusion of the experiment and collection of paper forms, each form is manually transcribed into a digital dataset. During this transcription, several labeling decisions are made. Participants’ nationalities are represented as a list of all places which were written in the nationality free-response field. As we are more concerned with participants’ self-perception and inner construction of identity rather than their legal nationality, all places identified by the participants are included regardless of the place’s geopolitical status as a sovereign and/or widely-recognized nation (resulting in the inclusion of places such as Kurdistan, Palestine, Taiwan, etc.).

Gender is also a free-response field. Many participants, out of either misunderstanding or own interpretation, chose to write their pronouns in this field. Thus, all instances of entries such as “girl,” “woman,” “female,” “she/her,” etc. receive the gender label of “F” for female/feminine, instances of “boy,” “man,” “male,” “he/him,” receive the gender label of “M” for male/masculine, and any participant either explicitly identifying with a nonbinary gender or providing pronouns different from she/her or he/him receive the “NB” label for nonbinary.

We further make labeling decisions with regards to how the Likert scale questions were answered. It was sometimes the case that participants circled two adjacent scores for a given question; These instances are coded as the mean of the two scores. For example, if a participant circled both “Yes” (3 points) and “Absolutely yes” (4 points), the value is recorded as 3.5. We completely discard the

answers of any participant who responded with “Yes” or “Absolutely yes” to *both* G1 and G2, as this produces a logical paradox and casts uncertainty on whether the participant was able to fully understand each question of the survey.

The collected dataset is almost entirely complete with very few missing responses to questions — three participants did not disclose nationality, one of whom also did not answer one of the welcome questions. Thus, only 0.26% of the data is missing. For the responses which are missing, we perform statistical imputation using *datawig*, a Python package which utilizes machine learning to obtain greater imputation accuracy than traditional statistical methods [2]. Imputation of missing data, particularly with Likert-scale survey data, is known to reduce biases which arise when simply deleting entries with missing data [20]. The *datawig* imputer is used with all default parameters.

Following imputation, we compute metrics to be used for analysis. These are a social trust score obtained from summing SCT1 and SCT2, a competency trust score obtained from summing CT1 and CT2, pre- and post-interaction interest scores obtained from summing I1 and I2, pre- and post-interaction welcome scores obtained from summing W1 and W2, pre- and post-interaction scores measuring the amount of gendered bias (regardless of which gender is favored) by summing G1 and G2, and differences for all pre- and post-interaction metrics obtained from subtracting the respective pre- measure from the post-. Additionally, a gender bias indicator is computed from subtracting G2 from G1; a positive number indicates a level of bias against girls, a negative number indicates a level of bias against boys, and 0 indicates no measured bias.

For all analysis statistically comparing two distributions, we first conduct normality tests. If both distributions are normal, we proceed with a student’s t-test. If either distribution is not normal, then we instead utilize a Mann-Whitney U test.

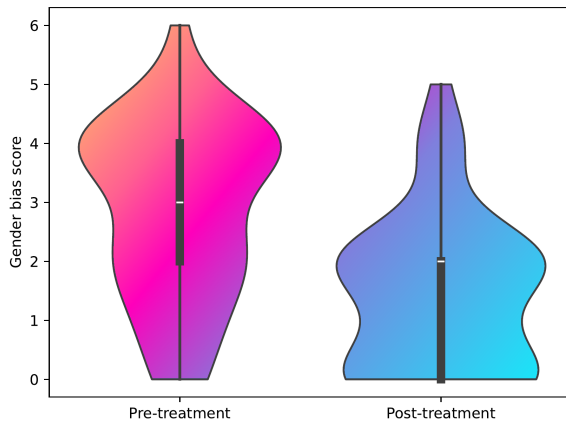
## 3 RESULTS

We were first concerned with measuring any significant differences in the answers between the two age groups, as children’s rapid development may possibly change the way in which the two cohorts respond. Our first test measures each of the experimental measures

between the two age groups. We find no significant difference between the two groups' responses except for a slight difference in the pre- and post- interest scores ( $0.04 < p < 0.05$ ) with the 11-12 year-olds registering an average interest 0.5 points lower than the 9-10 year-olds. Thus, we separate the two age groups when analyzing the interest scores, and allow them to be joined in the analysis of all other measures.

### 3.1 Baseline effects

Next we consider any baseline changes in the pre- and post-measures across all persona treatments and all participants. Analysis reveals an overall reduction in the gender bias measure ( $p = 0.02$ ). On further investigation, we find that 70% of participants did not change their gender bias score after the presentation. Among the 30 percent who did, there is an average of a 1.5-point reduction in gender bias (corresponding to 25% of the maximum value of 6 points). The percentage of students who changed their answer is relatively constant across treatments ( $30 \pm 8\%$ ). The distribution of gender bias scores pre- and post- is exhibited in Figure 2. There were no significant baseline differences among any other measure ( $p \gg 0.05$ ).



**Figure 2: Baseline change across all treatments in gender bias (Table 2 W1+W2) pre-treatment and post-treatment ( $p = 0.02$ ). Figure shows only the 30% of participants whose post-treatment response differed from their pre-treatment response.**

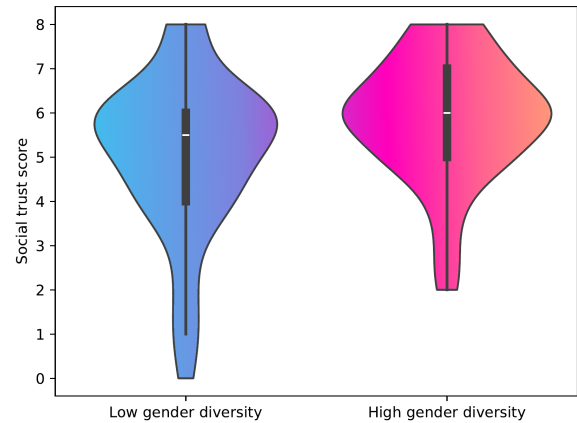
### 3.2 RQ1 & RQ2: Effects on perceptions of robotics and social identity

The different persona treatments did not have an effect on interest in robotics, gender bias, or perceptions of welcomeness ( $p \gg 0.05$ ). Further investigations examining responses with participants split by gender, nationality, and response language likewise found no significant difference among the treatments ( $p \gg 0.05$ ).

### 3.3 RQ1 & RQ3: Effects on social and competency trust

The high gender diversity treatments scored higher social trust ( $p = 0.003$ ) and slightly higher competency trust ( $p = 0.03$ ) scores from participants versus the low gender diversity treatments. The score distributions are exhibited by Figure 3 and Figure 4, respectively.

Participants who responded in Swedish gave a mean of 0.71 points lower social trust ( $p = 0.01$ ) and a mean of 1.02 points lower competency trust ( $p = 0.02$ ) to the high ethnic diversity treatments versus the low ethnic diversity treatments.



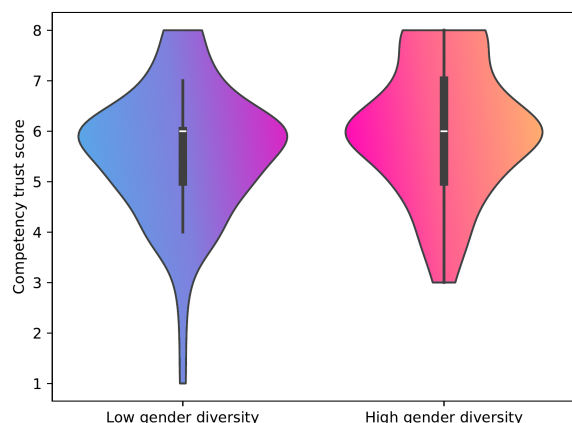
**Figure 3: Higher social trust scores (Table 2 ST1+ST2) are given to the persona treatments with high gender diversity versus the treatments with low gender diversity ( $p = 0.003$ ).**

## 4 DISCUSSION AND CONCLUSION

In this study we provided a classroom lecture to children aged 9-12 using a team of three human-like robot personas while varying the gender and ethnic diversity of the teams in a between-subjects study design. Before and after, we measured the participants' interest in robotics, gendered biases towards who is good at robotics and computer science, and perceptions of the fields' welcomeness to different social identities. Afterwards, we measured social trust and competency trust for each of the persona groups, and analyzed all metrics in the intersectional context of participants' demographics.

Regardless of the treatment received, participants were not unlikely to show a substantial decrease in measured gendered bias. We hypothesize that the reason for this is that the author who ran the data collection is a feminine-presenting woman exhibiting expertise in robotics and computer science to the participants through the act of introducing the study, managing the robot, and collecting the questionnaires. Exposure to this may have been enough for some students to reconsider and lower their bias in the post-treatment questionnaire.

So let us say: artificial representation of diverse social identities in STEM are no replacement for the representation of real people in STEM with those identities. Neither are a replacement for



**Figure 4: Slightly higher competency trust scores (Table 2 CT1+CT2) are given to the persona treatments with higher gender diversity versus the treatments with low gender diversity ( $p = 0.03$ ).**

active work towards ending the systemic roots of social-identity-based disparities which continue to afflict our profession. That said, human-like robot personas may still be a useful tool in exactly this effort.

Participants placed higher social trust and higher competency trust in the high gender diversity treatment. This is in spite of the relatively norm-breaking nature of the personas. This suggests that gender diversity in human-like robot teams may be important in promoting trustworthiness, and that this diversity may be inclusive of nonbinary gender and the breaking of gendered norms without negative repercussions for other aspects of the interaction.

Qualitatively, we also note that the participants were sometimes acutely aware of the manipulation occurring – upon introduction of the personas in one of the low-gender diversity sessions, one student audibly groaned and exclaimed “They’re all dudes!”. It seems that, at least for this student, some level of gender diversity was an *expectation* in the situation of meeting multiple robot personas. Alongside the quantitative evidence above, this at least points to a need for interaction designers to carefully consider the range of genders presented by human-like robots, as a homogenous design may come at a detriment to the interaction.

Next, participants who responded in Swedish placed lower social and competency trust in the treatments consisting primarily of personas which did not also speak Swedish (the high ethnic diversity treatments). We lack the data to tease out confounding factors and determine whether this might be from implicit bias against the *ethnicity* that the personas are cueing for (that is, racism), or whether this can be more directly explained by a language barrier. Participants attending an English-focused school who choose to respond to a questionnaire in Swedish may be less confident with their English ability versus their peers who responded to the questionnaire in English – the school teachers we worked with indicated varying degrees of English speaking confidence within their Sweden-born

students and we observed different student groups typically utilizing Swedish or English within the classroom. Thus, those students more comfortable with Swedish may have faced language-related issues in being able to trust the personas, possibly struggling more to fully understand the personas’ English-language presentation. It is also possible that they feel less able to trust the personas for simple reason that they have a diminished ability to communicate with and learn from a robot which does not share the language which they are most comfortable with. Given that no difference in trust was found comparing participants’ self-identified nationality and trust scores across the ethnicity treatments, we reason that this latter, language-barrier explanation is more likely. This underscores the importance for interactants to be able to engage with robots in whichever language they are most comfortable with. To better understand this, further exploring the intersectional interplay between robot/user ethnicity and language is a potential avenue for follow-up research.

We find the lack of other effects resulting from manipulating the personas’ ethnic diversity a gaping, meaningful void. It suggests that the ethnicity constructed for robots may not play a significant role in many aspects of interaction. Given that having ethnicity-based diversity in addition to gender-based diversity has representational benefits [7], this may be a reason for interaction designers to more confidently incorporate ethnic diversity in their artificial agents where applicable, provided utmost care and cultural competency is taken in doing so.

Lastly, we would like to underscore the great importance and sensitivity of continuing to consider ethnicity and national origin in HRI research. As can be seen from the dataset, in the classrooms we visited were children from Israel and from Palestine; from Ukraine and from Russia; from every inhabited continent; children from across the street and children from across the world. The steady march of globalization continues, and our world gets a bit smaller every day. Every effort made to promote compassion and help the world be a little more peaceful – even the tiniest effort – is worth it.

## ACKNOWLEDGEMENTS

The work was partly funded by the Jacobs Foundation; the Marianne and Marcus Wallenberg Foundation and the Marcus and Amalia Wallenberg Foundation, partly via the Wallenberg AI, Autonomous Systems and Software Program-Humanities and Society (WASP-HS); and the Swedish Research Council (grant number 2020-03167).

## REFERENCES

- [1] Yamam Al-Zubaidi. 2022. Racial and Ethnic Statistics in Sweden: Has the Socialization Process Started Yet? *Scandinavian studies in law* 2022 68 (March 2022), 425–450. <https://doi.org/10.53292/92887d87.0ed13d6d>
- [2] Felix Bießmann, Tammo Rukat, Philipp Schmidt, Prathik Naidu, Sebastian Schelter, Andrey Taptunov, Dustin Lange, and David Salinas. 2019. DataWig: Missing Value Imputation for Tables. (2019).
- [3] Human Rights Campaign. 2016. *Supporting & Caring for Transgender Children*.
- [4] Patricia Hill Collins and Sirma Bilge. 2020. *Intersectionality*. John Wiley & Sons. Google-Books-ID: fyfDwAAQBAJ.
- [5] Sophie L. Kuchynka, Asia Eaton, and Luis M. Rivera. 2022. Understanding and Addressing Gender-Based Inequities in STEM: Research Synthesis and Recommendations for U.S. K-12 Education. *Social Issues and Policy Review* 16, 1 (2022), 252–288. <https://doi.org/10.1111/sipr.12087> \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/sipr.12087>
- [6] Sophie L. Kuchynka, Asia Eaton, and Luis M. Rivera. 2022. Understanding and Addressing Gender-Based Inequities in STEM: Research Synthesis and Recommendations for U.S. K-12 Education. *Social Issues and Policy Review*

- Review* 16, 1 (2022), 252–288. <https://doi.org/10.1111/sipr.12087> \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/sipr.12087>.
- [7] Anne E. Martin and Teresa R. Fisher-Ari. 2021. “If We Don’t Have Diversity, There’s No Future to See”: High-school students’ perceptions of race and gender representation in STEM. *Science Education* 105, 6 (Nov. 2021), 1076–1099. <https://doi.org/10.1002/sce.21677>
- [8] Cynthia Jo Martincic and Neelima Bhatnagar. 2012. Will Computer Engineer Barbie® impact young women’s career choices? *Information Systems Education Journal* 10, 6 (Dec. 2012), 4. <https://isedj.org/2012-10/N6/ISEDJv10n6p4.html>
- [9] Lux Miranda, Ginevra Castellano, and Katie Winkle. 2023. Examining the State of Robot Identity. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Stockholm Sweden, 658–662. <https://doi.org/10.1145/3568294.3580168>
- [10] Blakely H. Payne and Cynthia Breazeal. 2019. An Ethics of Artificial Intelligence Curriculum for Middle School Students. <https://www.media.mit.edu/projects/ai-ethics-for-middle-school/overview/>
- [11] Statistikdatabasen. 2023. Immigrations and emigrations by country of birth and sex. Year 2000 - 2022.
- [12] Michael Stolp-Smith and Tom Williams. 2024. More Than Binary: Transgender and Nonbinary Perspectives on Human Robot Interaction. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. Boulder CO USA.
- [13] Rebecca Stower, Natalia Calvo-Barajas, Ginevra Castellano, and Arvid Kappas. 2021. A Meta-analysis on Children’s Trust in Social Robots. *International Journal of Social Robotics* 13, 8 (Dec. 2021), 1979–2001. <https://doi.org/10.1007/s12369-020-00736-8>
- [14] Beverly Daniel Tatum. 2017. *Why are all the Black kids sitting together in the cafeteria?: And other conversations about race*. Hachette UK.
- [15] Caroline L. van Straten, Jochen Peter, Rinaldo Kühne, Chiara de Jong, and Alex Barco. 2018. Technological and Interpersonal Trust in Child-Robot Interaction: An Exploratory Study. In *Proceedings of the 6th International Conference on Human-Agent Interaction*. ACM, Southampton United Kingdom, 253–259. <https://doi.org/10.1145/3284432.3284440>
- [16] Mark West, Rebecca Kraut, and Han Ei Chew. 2019. *I’d blush if I could: closing gender divides in digital skills through education*. Technical Report. <https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1>
- [17] Tom Williams. 2023. The Eye of the Robot Beholder: Ethical Risks of Representation, Recognition, and Reasoning over Identity Characteristics in Human-Robot Interaction. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Stockholm Sweden, 1–10. <https://doi.org/10.1145/3568294.3580031>
- [18] Katie Winkle, Donald McMillan, Maria Arnelid, Katherine Harrison, Madeline Balaam, Ericka Johnson, and Iolanda Leite. 2023. Feminist Human-Robot Interaction: Disentangling Power, Principles and Practice for Better, More Ethical HRI. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Stockholm Sweden, 72–82. <https://doi.org/10.1145/3568162.3576973>
- [19] Katie Winkle, Gaspar Isaac Melsión, Donald McMillan, and Iolanda Leite. 2021. Boosting Robot Credibility and Challenging Gender Norms in Responding to Abusive Behaviour: A Case for Feminist Robots. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. ACM, New York, NY, USA.
- [20] Wei Wu, Fan Jia, and Craig Enders. 2015. A Comparison of Imputation Strategies for Ordinal Missing Data on Likert Scale Variables. *Multivariate Behavioral Research* 50, 5 (Sept. 2015), 484–503. <https://doi.org/10.1080/00273171.2015.1022644>