Gender biases in robots for education*

Laura Cesaro[1,†], Andrea Franceschini[1,†], Silvana Badaloni[1,2,†], Emanuele Menegatti[1,†] and Antonio Rodà[1,†]

[1]Department of Information Engineering, via Gradenigo, 6, 35131 Padova, Italy
[2]Elena Cornaro Center on Gender Studies, University of Padova, Italy

Abstract
Educational robotics is increasingly spreading in schools, also with the aim of fostering young women's interest in STEM disciplines, particularly in programming and Artificial Intelligence. However, it is crucial to design and select robots that resonate emotionally with female students to overcome gender stereotypes that traditionally deter them from computer science disciplines. This study explores the hypothesis that educational robots should be specifically tailored to meet the expectations and interests of female students. An experiment was conducted with 211 participants, equally divided by gender, who were asked to evaluate images of 16 different educational robots using a semantic differential scale. The results reveal differences between males and females in the attitudes and opinions towards educational robots. While both genders generally rated the robots as more masculine than feminine, female participants tended to provide higher overall scores, except for specific robots. Additionally, robots that were perceived as more feminine were often rated as simpler whereas masculine robots are associated to the words intelligent and creative, reflecting established societal stereotypes. These insights suggest that educational robots should be designed to appeal to both girls and boys, avoiding reinforcing gender stereotypes and ensuring inclusivity in STEM education. Further research is necessary to explore these attitudes and their implications for fostering a more balanced interest in STEM among both genders.

Keywords
gender biases, gendered innovation, gendered robots, artificial intelligence, educational robotics,

1. Introduction

Women continue to be severely underrepresented in professional roles related to Artificial Intelligence-based applications and this gap is particularly pronounced in leadership and advanced technical roles [1]. The lack of diversity in AI can contribute to the design of biased algorithms and products that don't adequately consider the needs and the point of view of women and other underrepresented groups. Given the importance that these applications are assuming in our society (from healthcare to criminal justice to news and information), this imbalance risks exacerbating existing gender inequalities. The situation appears even more

3rd Workshop on Bias, Ethical AI, Explainability and the Role of Logic and Logic Programming (BEWARE24), co-located with AIxIA 2024, November 25-28, 2024, Bolzano, Italy
*Corresponding author.
†These authors contributed equally.
laura.cesaro@phd.unipd.it (L. Cesaro); andrea.franceschini.2@unipd.it (A. Franceschini); silvana.badaloni@unipd.it (S. Badaloni); emanuele.menegatti@unipd.it (E. Menegatti); antonio.roda@unipd.it (A. Rodà)

concerning in perspective, given that the percentage of women choosing to study STEM disciplines in general, and computer science in particular, continues to be low [2]. Some studies have shown that playful and hands-on interventions, such as educational robotics [3] and gamified approches [4], can be a promising means to introduce females to STEM disciplines, starting from childhood. However, it is crucial to design and select robots that resonate emotionally with female students to overcome gender stereotypes that traditionally deter them from computer science disciplines. It is well known that users tend to attribute, more or less explicitly, a gender to robots and gender stereotypes can influence the approach to educational robots in various ways [5]. However, the effect that certain external characteristics of educational robots (shape, color, material, etc.) can have on users remains little studied.

To the best of our knowledge, this is the first study that analyzes how the external appearance of robots influences the attitudes and opinions of female and male users, comparing a selection of 16 robots used in educational contexts. The main research questions are

Q1 Are there differences in attitudes and opinions towards educational robots across genders?

Q2 Are there educational robots that are perceived as more masculine or feminine?

After an introduction on the concept of gendered innovations and the aims of educational robotics, we will present and discuss the preliminary results of an experiment that involved 211 participants.

2. Background

2.1. Gendered innovations

Diversity equity and inclusion dimensions of science and technology have become a highly visible and debated theme worldwide, impacting society at every level, in many fields of knowledge. In particular we will focus on AI and Robotics and we will analyze the relationship between AI and Robotics and our societal goals, in particular, concerning the aspect of gender equality [6].

While the concept of bias is very broad, gender-related biases are considered an essential aspect of fairness [7]. We will focus on the gender dimension, for the following reasons [8, 9]:

- numerous studies have shown that gender biases are deeply rooted in our society, so the risk that datasets are biased concerning gender is high
- gender biases concern more or less half of the population
- in comparison with other types of bias (racial, social, etc.), it's easier to define the categories subject to possible discrimination
- promoting studies on gender biases in AI and robotics can facilitate the introduction of gender knowledge into computer science/engineering courses with a twofold advantage:
    1. increasing the degree of involvement of female students,

2. making male students aware of stereotypes that risk discriminating against their female counterparts.

To deal with the integration of gender principles into research an important approach consists in developing "gendered innovation". This term refers to the integration of gender dimension into innovation across disciplines [1]. How can scientific theories be re-designed taking into account the gender dimension? Only a complete redefinition of the method and of the research model with new ways of observation can re-design science in a gender perspective [10]. Science does not proceed by progressive and continuous accumulation of truth but thanks to attempts to disprove proposed theories.

So the next step should be to address the problem of how the gender dimension can be taken into account in the content of AI and Robotics. While for AI there is a large awareness on the use of ethical issues in developing algorithms mitigated from gender biases [9, 10], for Robotics the questions are still open on gendering social robots [11]: does gendering robots enhance acceptance by humans? Does it reinforce gender stereotypes that amplify social inequalities?

The challenge is to study and to understand how gender can be embodied in robots and how this aspect can be perceived by students of different gender.

2.2. Educational robotics

Educational Robotics (ER) are gaining popularity in education, allowing teachers to provide students with engaging and stimulating learning experiences aimed at acquiring the skills needed for the future. Several teachers are incorporating innovative technologies and teaching methods into their instructional activities [12, 13]. The utility of ER in enhancing problem-solving skills [14], embodied learning [15], promoting participation and motivation, and improving subject-specific learning outcomes [16] is demonstrated in different studies conducted in primary schools [17].

ER activities are less represented in lower secondary schools, a segment that includes students aged between 11 and 14 years. This school level has been under-explored about ER but it is essential for promoting STEM disciplines, particularly among girls. It is widely recognized that social norms, a sense of belonging, and personal efficacy, influence the educational and career choices of girls and boys from an early age [18]. Girls experience a confidence gap, or dream gap [19], that prevents them from imagining themselves, as adults, in male-dominated occupations. In certain countries, girls pursue STEM careers as much as boys. However, this is not the case in many other countries. In Italy, this gap is evident in the choice of high school: only one-third of the total number of students enrolled in a technical-scientific high school are girls [2]. This shows how initiatives are needed to support schools in fighting gender stereotypes. This can be achieved by promoting positive attitudes, self-efficacy beliefs, and active behaviours towards STEM disciplines among female students in (lower and upper) secondary schools [20] by introducing factors that could make ER more appealing to girls [21], and offering workshop activities instead of structured lessons [22].

[1] https://genderedinnovations.stanford.edu
[2] https://dati.istruzione.it/espscu/index.html?area=anagStu

3. Experiment

To investigate the attitudes and opinions towards a selection of educational robots across genders, we set up a survey study presenting a list of robot images and requesting a subjective evaluation by means of a semantic differential scale.

3.1. Materials and procedure

The selection criteria for the educational robots have been the subject of many studies [23], covering physical characteristics, hardware functionality, operating software systems, expected outcomes, and other specific aspects. Additionally, educational robots can be classified based on their operational environment (land, air, water), type of locomotion, difficulty of programming, construction requirements (kit-based, pre-built, or ready-to-use), and more [24]. For this research, we selected robots from two databases: the IEEE Robotics Guide [3] and the database created within the European project EARLY [4]. This preliminary selection phase led to the identification of 34 educational robots. Then, we considered specific requirements when choosing which robots to include in the survey. In particular, we required that the robots be:

- suitable for ages 11-14
- programmable via visual and text-based languages
- priced under 500 euros
- not requiring soldering (for kit-based robots)
- currently in production

Out of 34, only 9 met the estabilished criteria. To address the underrepresentation of social and animaloid robots in the initial selection, we expanded our search to include well-known examples from these categories. This led to the addition of several popular models, bringing the total number of selected robots to 16.

To evaluate attitudes and opinions towards robots we selected a list of 10 word pairs starting from a longer list of 48 words identified by Bidin et al. [25]. Moreover, we added the two dimensions "feminine" and "masculine" that are particularly relevant for this study, obtaining the following 12 word pairs: *complex - simple*, *negative - positive*, *boring - fun*, *uncreative - creative*, *uncurious - curious*, *unfeminine - feminine*, *unintelligent - intelligent*, *uninteresting - interesting*, *unmasculine - masculine*, *unpleasant - pleasant*, *dangerous - safe*, *sad - cheerful*.

We prepared an online survey asking participants to rate each robot according to the above defined semantic differential scale. We asked the participants to provide their subjective evaluation of each word pair on a 5-point Likert scale only showing the two terms at the respective ends of the scale [5].

[3] https://robotsguide.com/
[4] https://edurobots.eu/
[5] See https://www.dei.unipd.it/~roda/gender-robot/question.png for a question example

Figure 1: List of the robots selected for this study.

3.2. Results

We received 211 valid responses, with 109 respondents identifying as female and 102 as male. Table 1 reports the breakdown of the respondents' professions, showing a majority of students

Table 1
Jobs breakdown

	Female	Male	Total
Employed	9	8	17
Freelance	1	1	2
Student	46	68	114
Teacher	47	14	61
Other	6	11	17
All	109	102	211

Table 2
Summary of the ANOVAs calculated with the formula *value* $\sim$ *robot* $\times$ *word_pair* $\times$ *gender*.

	DoF	Sum Sq	Mean Sq	F	p	sig.
robot	15	775.2	51.681	48.7352	<2.2e-16	< 0.001
word_pair	11	2186.4	198.768	187.4371	<2.2e-16	< 0.001
gender	1	49.2	49.233	46.4265	9.784e-12	< 0.001
robot:word_pair	165	1844.5	11.179	10.5416	<2.2e-16	< 0.001
robot:gender	15	46.0	3.063	2.8887	0.0001415	< 0.001
word_pair:gender	11	59.7	5.430	5.1208	4.550e-08	< 0.001
robot:word_pair:gender	165	152.2	0.922	0.8696	0.8846439	
Residuals	19872	21073.3	1.060			

followed by a considerable cohort of teachers. In total, we collected 40512 scores, equals to 1 response for 211 participant $\times$ 16 robots $\times$ 12 word pairs.

To verify the statistical significance of the differences between mean values and the existence of interactions between independent variables, particularly whether the gender variable significantly influenced the evaluation of robots, we calculated a three-way ANOVA, with gender, robots and word pairs as independent variables. The results in table 2 show a $p < .001$ for all the interactions except for the last one, indicating that female and male participants have different attitudes and opinions towards the selected educational robots. The main differences are detailed below.

First of all, female respondents on average attributed significantly higher scores ($m = 3.42, sd = 1.17$) than males ($m = 3.32, sd = 1.10$), demonstrating a rather positive attitude towards these educational robots.

Figure 2 shows the mean scores calculated for each robot across all the word pairs, including both male and female respondents. We consider this as a "global score" of appreciation given that the valence of the word pairs was always presented in the same direction – i.e., left, and lower values, for negative, right, and higher values, for positive. We see that the robots that scored the lowest were Edison ($m = 2.91, sd = 1.08$), Thymio ($m = 3.09, sd = 1.05$), and the DJI drone ($m = 3.18, sd = 1.11$), while those that scored the highest were Aibo ($m = 3.58, sd = 1.15$), Paro ($m = 3.58, sd = 1.21$), and Spike ($m = 3.53, sd = 1.06$). Overall, female participants gave higher scores than the males to all the robots except for Aibo, Petoi, and RVR, which received a higher scores from males (with statistical significance according to a Tukey test). Interestingly, among the three robots preferred by males compared to females, there are two with dog-like appearances.

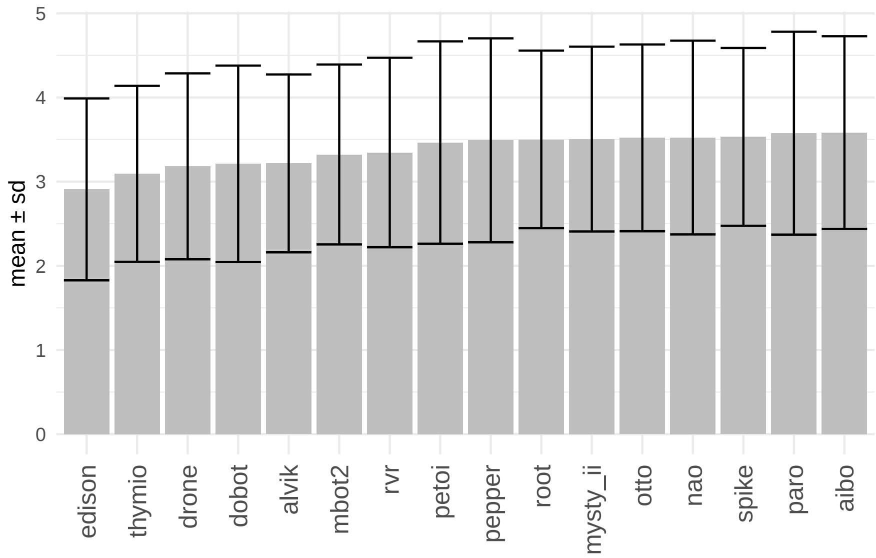

Figure 2: Mean values across genders and terms, split by robots.

Figure 3 shows the average score of each word pair, where a lower value corresponds to the word with the negative valence, and vice-versa for the higher values. We see that, overall, the safety evaluation achieved a relatively high score ($m = 3.83, sd = 1.06$), and likewise did positivity ($m = 3.66, sd = 0.93$), interest ($m = 3.65, sd = 1.09$), and curiosity ($m = 3.59, sd = 1.07$). These result imply a general good attitudes towards these robots. Interestingly, they are considered significantly more masculine ($m = 3.15, sd = 1.07$) than feminine. ($m = 2.60, sd = 1.02$).

Figure 4 gives an intuitive representation of the average scores of each robot for each word pair; a darker shade corresponds to a higher score, e.g. the light shade of Spike in the *unfeminine/feminine* column corresponds to an average score equals to 2.48 (less feminine), whereas the darker shade of Root in the *dangerous/safe* column corresponds to 4.46 (very safe). One prominent feature is the *dangerous/safe* column where we find Alvik, Edison, Paro, Thymio, and Root scoring considerably more safe than Nao, Petoi, Pepper, Dobot, and the DJI drone. Although we cannot find a consistent explanation for all of these, we think that there might be some influence from factors outside the mere appearance of these robots. For example, we sometimes associate the word “drone” with a weapon, and indeed the DJI drone is the lowest scoring in the safety scale. Petoi resembles Boston Dynamics’ Spot which is also associated with the Defence sector. Nao, with its athletic and articulated build, could be perceived as a potentially harmful sci-fi robot. In figure 4, we also note the *unfeminine/feminine* and *complex/simple* which highlight that robots considered more feminine are also perceived as more simple. There seems to be however a coincidence of higher femininity and simplicity scores for

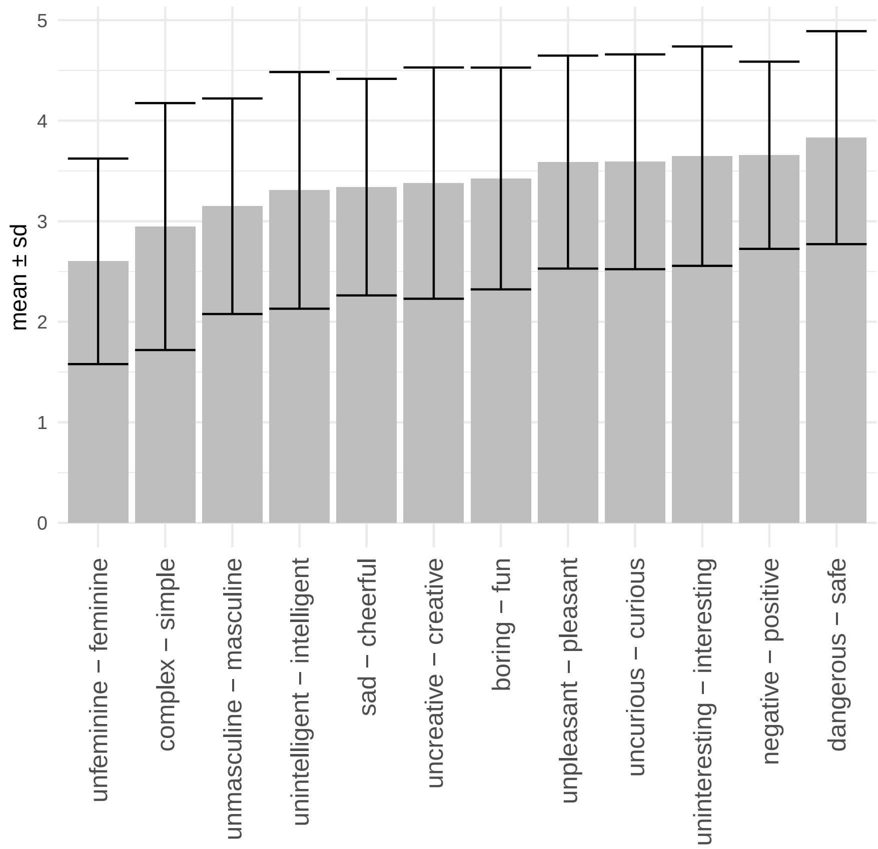

Figure 3: Mean scores across genders and robots, split by terms.

Edison and Paro. In the case of Paro, we think that the soft and infantile appearance of the robot plays a role. In the case of Edison, especially if we compare it to other similar-looking robots who achieved different scores, we can only imagine that the orange colour is the differentiating factor, although this is not a particularly convincing explanation. Pepper and Dobot appear to be more complex, and are also given higher masculine scores, and we think this might be because they are perceived to have a more advanced functionality (social and industrial) than other robots, such as Lego Spike, Mysty II, and RVR that scored slightly more neutral.

4. Conclusion

In relation to the two main questions of this study (see 1), we have found indications of positive answers to both questions, and, although further research is needed, we have attempted some preliminary explanations.

With respect to question 1, we found that the average global scores given by female and male respondents to individual robots are different with a high level of statistical significance.

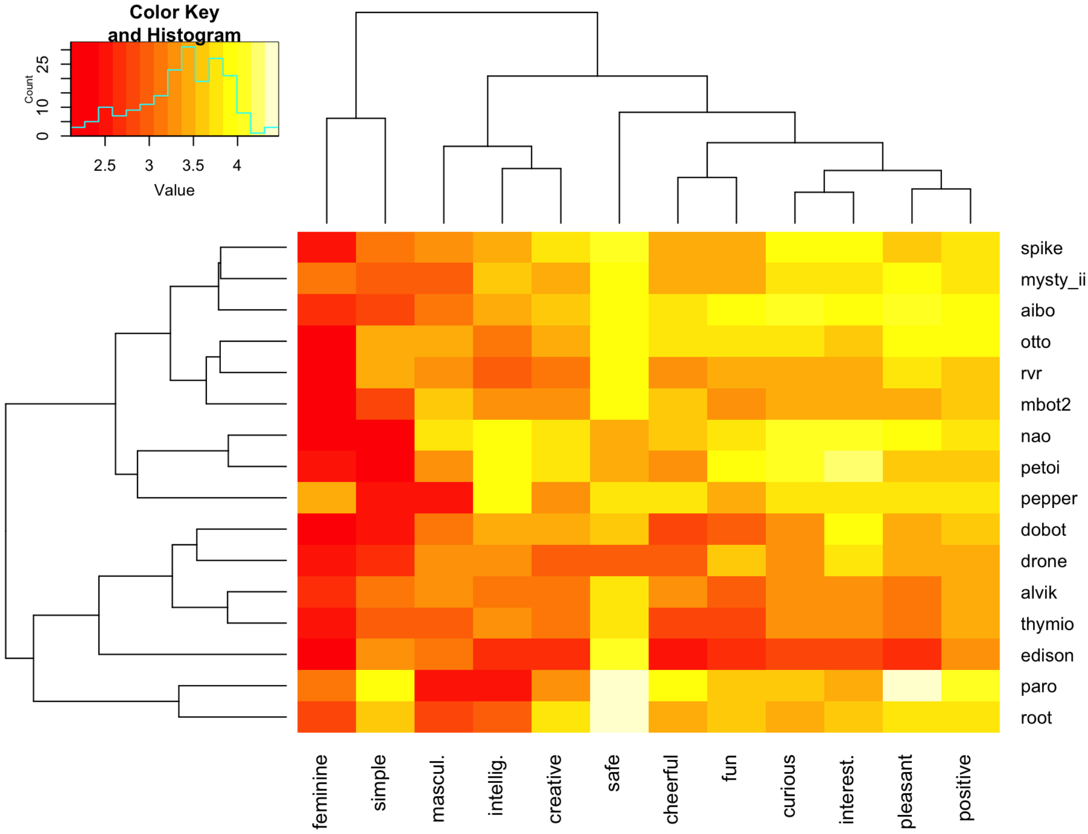

Figure 4: Heatmap and dendrogram comparing the scores of each robot for each word pair.

Female respondents gave overall higher scores than males except for Aibo, Petoi, and RVR, but we cannot find a convincing explanation with the data we have collected and analyzed so far.

With respect to question 2, we found that these educational robots are, on average, perceived as having more masculine than feminine characteristics, and in fact we see that the overall score on the femininity scale is below the mid point of the 5-points Likert scale, if only by a small amount. The robots perceived as more feminine are Pepper, Paro, and Misty II, while those what scored higher for masculinity are the DJI drone and MBot. We also found an odd outlier in Edison: by its shape, it should align more with other similar robots by being perceived as more masculine, but in reality it seems to be perceived as more feminine. As we could not find satisfying explanations for this, nor could we clearly categorize other robots by perceived gender, we need further data and analysis.

Lastly, there seems to be an association between the feminine and simple attributes on the one hand, and between masculine, intelligent, and creative on the other hand. These associations mirror rather unfortunate gender stereotypes present in society, and we feel that it is important to highlight this issue so it can be accounted for, and hopefully addressed, when designing ER interventions but also in the design of AI algorithms used to control the robots behavior.

4.1. Limitations

As we see in table 1, there is a noticeable imbalance of genders between teachers and students who responded to our survey: teachers tend to be a majority female, while students comprise a higher proportion of males. The teaching profession at large is often more frequently taken up by women, so this imbalance may just reflect a societal imbalance.

Our survey asked respondents to evaluate the robots based on their appearance, and we only presented one picture per robot, often taken from promotional materials. Aspects that are not considered in this evaluation include, for example, size, movement, voice and sounds, and facial expressions, for robots that have a face. While it may be possible that some respondents were familiar with some of the robots, we think that, on the balance of things, we can assume this is a small proportion, and thus not affecting the results significantly.

Lastly, the survey was only circulated among Italian respondents. We do not know whether this might introduce cultural biases, and we hope to open up the survey to international respondents in the future to enable us to perform a cross-cultural analysis.

Acknowledgments

Partially funded from Next Generation EU, Mission 4 Pillar 1 CUP C53D23007970001.

References

[1] D. Zhang, S. Mishra, E. Brynjolfsson, J. Etchemendy, D. Ganguli, B. Grosz, T. Lyons, J. Manyika, J. C. Niebles, M. Sellitto, et al., The ai index 2021 annual report, arXiv preprint arXiv:2103.06312 (2021).

[2] J. Waite, A. Franceschini, S. Sentance, M. Patterson, J. Sharkey, An online platform for teaching upper secondary school computer science, in: United Kingdom and Ireland Computing Education Research conference., UKICER ’21, Association for Computing Machinery, New York, NY, USA, 2021, p. 1–7. URL: https://doi.org/10.1145/3481282.3481287. doi:10.1145/3481282.3481287.

[3] A. Sullivan, M. U. Bers, Investigating the use of robotics to increase girls’ interest in engineering during early elementary school, International Journal of Technology and Design Education 29 (2019) 1033–1051.

[4] A. Franceschini, A. Rodà, Play to learn: from serious games to just games, in: Proceedings of the 2023 ACM Conference on Information Technology for Social Good, GoodIT ’23, Association for Computing Machinery, New York, NY, USA, 2023, p. 117–127. URL: https://dl.acm.org/doi/10.1145/3582515.3609525. doi:10.1145/3582515.3609525.

[5] N. Reich-Stiebert, F. Eyssel, (ir) relevance of gender? on the influence of gender stereotypes on learning with a robot, in: Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction, 2017, pp. 166–176.

[6] R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Felländer, S. D. Langhans, M. Tegmark, F. Fuso Nerini, The role of artificial intelligence in achieving the sustainable development goals, Nature communications 11 (2020) 1–10.

[7] A. Fabris, A. Purpura, G. Silvello, G. A. Susto, Gender stereotype reinforcement: Measuring the gender bias conveyed by ranking algorithms, Information Processing & Management 57 (2020) 102377. doi:10.1016/j.ipm.2020.102377.
[8] S. Badaloni, A. Rodà, et al., Gender knowledge and artificial intelligence, in: Proceedings of the 1st Workshop on Bias, Ethical AI, Explainability and the role of Logic and Logic Programming, BEWARE-22, co-located with AIxIA, 2022.
[9] S. Badaloni, A. Rodà, M. Scagnet, An italian dataset for the analysis of gender stereotypes in textual documents., in: BEWARE@ AI* IA, 2023, pp. 73–78.
[10] S. Badaloni, F. A. Lisi, Towards a gendered innovation in ai, in: Proceedings of the AIxIA 2020 Discussion Papers Workshop co-located with the the 19th International Conference of the Italian Association for Artificial Intelligence (AIxIA2020), volume 1613, 2020, p. 0073.
[11] G. Beraldo, S. Di Battista, S. Badaloni, E. Menegatti, M. Pivetti, Sex differences in expectations and perception of a social robot, in: 2018 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO), IEEE, 2018, pp. 38–43.
[12] M. Malvezzi, D. Alimisis, M. Moro (Eds.), Education in & with Robotics to Foster 21st-Century Skills: Proceedings of EDUROBOTICS 2020, volume 982 of *Studies in Computational Intelligence*, Springer International Publishing, Cham, 2021. doi:10.1007/978-3-030-77022-8.
[13] L. Cesaro, G. Dodero, E. Menegatti, Integrating AI into school curriculum: A maker-oriented activity, in: R. Balogh, D. Obdržálek, M. Fislake (Eds.), Robotics in Education, Springer Nature Switzerland, Cham, 2024, pp. 367–378.
[14] F. Gratani, L. Giannandrea, A. Renieri, M. Annessi, Fostering Students' Problem-Solving Skills Through Educational Robotics in Primary School, in: M. Malvezzi, D. Alimisis, M. Moro (Eds.), Education in & with Robotics to Foster 21st-Century Skills, Springer International Publishing, Cham, 2021, pp. 3–14.
[15] S. Almpani, D. Almisis, Dance and robots: Designing a robotics-enhanced project for dance-based STEAM education using ENGINO, in: M. Malvezzi, D. Alimisis, M. Moro (Eds.), Education in & with Robotics to Foster 21st-Century Skills, Springer International Publishing, Cham, 2021, pp. 139–151.
[16] E. Fokides, D. Papadakis, V. Kourtis-Kazoullis, To drone or not to drone? Results of a pilot study in primary school settings, Journal of Computers in Education 4 (2017) 339–353. doi:10.1007/s40692-017-0087-4.
[17] E. Mangina, G. Psyrra, L. Screpanti, D. Scaradozzi, Robotics in the Context of Primary and Preschool Education: A Scoping Review, IEEE Transactions on Learning Technologies 17 (2024) 342–363. doi:10.1109/TLT.2023.3266631.
[18] N. Đapo, I. Čelebičić, L. Spahić, R. Binder-Hathaway, Gender Gap in the STEM Fields and Proposed Intervention Programmes, https://eca.unwomen.org/en/digital-library/publications/2021/8/gender-gap-in-the-stem-fields-and-proposed-intervention-programmes, 2020.
[19] L. Bian, S.-J. Leslie, A. Cimpian, Gender stereotypes about intellectual ability emerge early and influence children's interests, Science 355 (2017) 389–391. doi:10.1126/science.aah6524.
[20] C. Giberti, Differenze di genere in matematica: dagli studi internazionali alla situazione italiana, Didattica della matematica. Dalla ricerca alle pratiche d'aula (2019) 44–69. doi:10.

33683/ddm.18.5.2.
[21] B. K. Maigaard Kjær Pedersen, B. Charlotte Weigelin, J. C. Larsen, J. Nielsen, Using educational robotics to foster girls' interest in STEM: A systematic review, in: 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), 2021, pp. 865–872. doi:10.1109/RO-MAN50785.2021.9515538.
[22] B. M. D. Bagatini, Lavorare sul genere a scuola con coding e robotica educativa, Carocci editore, 2022.
[23] S. Krause, A.-L. Henk, F. Stolzenburg, Selecting an educational robot: A comprehensive guideline, 2024.
[24] S. Evripidou, L. Doitsidis, G. Tsinarakis, Z. Zinonos, S. A. Chatzichristofis, Selecting a Robotic Platform for Education, in: 2022 IEEE International Conference on Consumer Electronics (ICCE), 2022, pp. 1–6. doi:10.1109/ICCE53296.2022.9730568.
[25] S. A. Bidin, A. M. Lokman, W. A. R. W. M. Isa, A development of kansei words for robotic therapy and the elderly, Environment-Behaviour Proceedings Journal 6 (2021) 63–67.