

The Soccer-Playing Unicorn: Mitigating Gender Bias in AI-Created STEM Teaching Materials

Claudia Hess and Sibylle Kunz

IU International University of Applied Sciences, Erfurt, Germany

claudia.hess@iu.org

sibylle.kunz@iu.org

Abstract: Artificial Intelligence (AI) tools are increasingly being used in education for various purposes. In particular, AI chatbots such as ChatGPT, with their user-friendly interfaces are being explored in education to co-create teaching materials, provide advice and guidance to educators, simulate classroom scenarios, and offer personalized recommendations to students on how to study and approach subjects. With all the enthusiasm for these new opportunities, one should be aware of the risks due to potential biases in the generated content or the responses. These biases can be associated with factors such as gender, race, religion, or political orientation. As a consequence, educators who are using AI chatbots to (co-)create teaching materials need to have the knowledge and the strategies to mitigate such biases. This paper focuses on one particular type of bias, namely gender bias, and on specific disciplines, namely Science, Technology, Engineering and Mathematics (STEM). Gender bias in STEM education is particularly problematic because it may reinforce existing stereotypes about girls and women in STEM and contribute to their underrepresentation in STEM fields. To raise awareness of these risks of gender bias in AI-co-created STEM teaching materials, this paper identifies risks of gender bias by analysing potential usage patterns of AI chatbots by educators when creating teaching materials. An example of such a risk is if the AI chatbot generates educational materials that primarily portray men as STEM professionals and underrepresent women. This would exacerbate the lack of female role models in STEM. Therefore, strategies are developed that educators can apply to mitigate these risks. These strategies will be demonstrated using practical examples. This will allow them to break the vicious cycle of perpetuating stereotypes in STEM education. In addition, these examples demonstrate how AI chatbots can be used to make STEM education more inclusive, which may include co-creating educational materials tailored to individual interests and learning styles.

Keywords: STEM Education, AI Chatbots, Generative AI, Teaching Materials, Gender Bias

1. Introduction

Tools based on Artificial Intelligence (AI) are increasingly being used in education. Of particular interest is generative AI, which is designed to generate human-like text-based content. With their convenient user interfaces, AI chatbots such as ChatGPT dialogue with humans in natural language. A wide range of application scenarios in education are currently being explored (Kasneci *et al.*, 2023; İpek *et al.*, 2023). For example, educators can use AI chatbots to co-create teaching materials, get advice on how to make their lessons more engaging, and even simulate and practice difficult situations in the classroom (Hess and Kunz, 2023). Students can also benefit from interacting with AI chatbots. For instance, they can receive personalized recommendations on how to approach a subject or study for an exam.

While the use of AI chatbots in education offers new opportunities for educators and learners, there are potential risks due to bias that may be present in the responses generated. Bias refers to unfair and unbalanced viewpoints or prejudices toward certain groups, ideas, or stereotypes, and may be related to gender, race, religion, or political orientation, for example. Educators using AI chatbots in teaching must be aware of the risk of bias and actively seek to eliminate it. This is especially important in the context of science, technology, engineering and mathematics (STEM) which are still predominantly associated with males. And this gender bias is reproduced by AI chatbots. To help educators use AI chatbots responsibly, this paper outlines concrete strategies to mitigate gender bias risks in AI-generated STEM materials.

The paper is organized as follows. Section 2 outlines the methodology of our work. Section 3 describes the theoretical foundations of gender bias in STEM education. Section 4 examines gender bias in generative conversational AI systems. This sets the stage for Section 5, which discusses the potential risks of gender bias in AI-co-created learning materials. Section 6 presents strategies for addressing these risks. Section 7 concludes the paper.

2. Methodology

The paper aims to raise educators' awareness of gender bias in AI-created content, and in consequence of the risks of perpetuating these biases in educational materials and lessons. Moreover, it seeks to provide teachers with pragmatic strategies to mitigate these risks. Two research questions structure the presented research:

1. What gender bias risks must educators be aware of when co-creating STEM educational materials with the help of AI chatbots?
2. What strategies can educators use to mitigate these gender bias risks?

To answer the first question, research from three different areas is combined. Risks are identified at the intersection of these three areas as depicted in Figure 1. First, existing gender bias in STEM education is identified based on a literature review of studies from around the world. Second, the results of recent studies on gender bias in the content generated by AI chatbots are analysed. Third, activities in the creation of learning materials that can be supported by AI chatbots are identified.

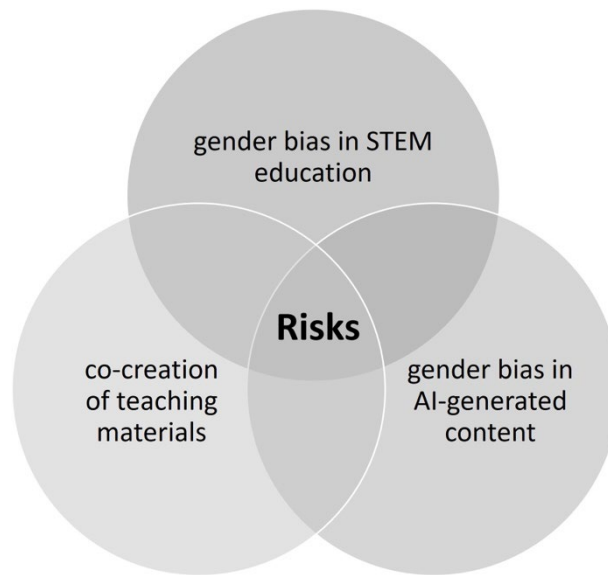


Figure 1: Risks at the intersection of three research areas

This theoretical foundation provides educators with the knowledge to understand the reasons for these risks. As postulated in the second research question, strategies to address the identified risks are defined. It should be possible to apply these strategies in a pragmatic way during the co-creation of STEM teaching materials. Therefore, concrete examples are given.

3. Gender Bias in STEM Education

Despite of many initiatives, girls and women still face stereotypes about their abilities and roles in STEM at school, at work, and in their personal environments, including family and friends (Makarova, Aeschlimann and Herzog, 2019; Raabe, Boda and Stadtfeld, 2019). Gender bias can manifest itself in STEM education in different forms, some of them being more explicit and others more implicit. They can be categorized into three categories, the underrepresentation of women in STEM teaching materials, predominantly male-centred tasks and an underestimation of girls' competencies in STEM, which will be discussed in the following.

3.1 Underrepresentation of Women in STEM Teaching Materials

Studies conducted globally have examined teaching materials across different countries, revealing a consistent trend of gender bias. The findings show that women are underrepresented in STEM teaching materials, and this lack of representation manifests in the following ways. First of all, males are more frequently portrayed in STEM teaching materials, in both text and visuals, than women (Papadakis, 2018; Dele-Ajayi *et al.*, 2020). In the data analysed by Kerkhoven *et al.* (2016) for instance, only 25% of the persons shown with a science-related profession were women. At the same time, more women than men (63.9% versus 36.1%) were depicted as teachers. In addition, the few pictures that portray girls and women in STEM show them rather as passive

observers (Blickenstaff, 2005) or as digital consumers instead of digital producers (Papadakis, 2018). This male-centred perspective and the lack of female role models convince learners of a “science is for men” picture (Kerkhoven *et al.*, 2016; Ellemers, 2018) which discourages girls in pursuing a career path in STEM.

3.2 Male-Oriented Tasks

There is the risk that teaching materials, such as the tasks, are more related to boys’ interests than to girls’ interests. Female learners are rarely attracted by such exercises (Steffen *et al.*, 2023). It is important to consider a female-responsive design of learning activities that takes the interests of girls and women into account (Hess *et al.*, 2023). If this is not considered, there is the risk of reflecting and reinforcing male-centred perspectives in STEM teaching materials. Teachers should reflect about this every time they are designing a lesson and teaching it.

3.3 Underestimation of Girls’ Competencies in STEM

A well-researched problem that contributes to the low number of girls and women in STEM is that their competencies regarding STEM topics are often underestimated. This applies to teachers, parents and to the girls themselves (e.g. Eccles, 2009; Weber, 2012; Ellemers, 2018). Over the last decades, many studies have demonstrated that girls tend to underestimate their own competencies in STEM subjects. At the same time, boys’ abilities in STEM are often overestimated (Eccles, 2009). Consequently, girls often lack the self-confidence that would be required for taking up a STEM-course of study. Teacher attitudes towards girls’ competencies therefore play an important role, too.

4. Gender Bias in AI Chatbots

In the following, generative conversational AI systems are introduced and analysed for gender bias.

4.1 Sexism and Racism by AI Chatbots

Since the public release of ChatGPT by OpenAI in November 2022, generative conversational AI systems such as ChatGPT gained huge attention in public, in companies as well as in research. They are capable of generating human-like content in response to questions asked in natural language, the so-called prompts. Based on these capabilities, people can interact and have conversations with AI chatbots in a natural way. The chatbot remembers the entire conversation and builds its responses based on previous input and the context given throughout the interaction.

Despite of various efforts by the providers of such generative conversational AI systems, recent studies have shown that these systems exhibit some degree of bias, ranging from obvious to more subtle forms (Gross, 2023; Urchs *et al.*, 2023; Zhou and Sanfilippo, 2023). Bias in generative conversational AI systems is characterized by the existence of systematic misrepresentations and attribution errors, or by omitting or distorting facts which, in consequence, privileges particular groups of people and their beliefs (Ferrara, 2023). A systematic misrepresentation could be that the AI chatbot consistently highlights male figures in STEM fields while downplaying or omitting female scientists’ contributions. An example of an attribution error could be that the AI chatbot attributes the success of a female STEM scientist to luck instead of acknowledging her competence. Such bias in AI-generated content poses a significant problem because the language we encounter and use shapes our understanding of the world and influences our actions.

Several factors contribute to this gender bias in AI chatbots (Nadeem, Abedin and Marjanovic, 2020). The underlying models are trained using large amounts of data from various sources such as websites or social media platforms (Ferrara, 2023; Ray, 2023). Biases present in this training data can be absorbed by the AI model and subsequently amplified in its responses. Bias can also be introduced and reinforced during the learning process, as the AI system learns to generalize and respond in new contexts.

4.2 Studies on Gender Bias in AI Chatbot Conversations

Recent studies have examined AI chatbots for gender bias. The different types of bias identified in these studies can be grouped into the following six categories. Figure 2 provides an overview of these categories.

Difficulties with gender-neutral language. This category refers to AI chatbots’ difficulties in using gender-neutral language. An example from the study performed by Gross (2023) is that the AI ignored gender-diverse identities

and overlooked nongendered pronouns like "they". Urchs et al. (2023) found differences between English and German answers. While English results were grammatically correct, ChatGPT struggled with gender-neutral language accuracy in German.

Reinforcement of stereotypes. This category emphasizes the reinforcement of gender stereotypes, particularly in roles and professions. An example from Gross's (2023) study where the AI chatbot failed to present diverse representations is when ChatGPT portrayed an economics professor as a male.

Neglect of gender-diverse identities. This category highlights the problem of AI chatbots struggling with the complexities of gender diversity. Examples include mislabelling of individuals with gender-diverse identities (Gross, 2023).

Assignment of characteristics based on gender. This category addresses the challenge of AI chatbots making assumptions about an individual's behavior or preferences based on gender identity. An example is assigning technology-related skills to a male person (Urchs et al., 2023).

Inappropriate distinction of female and male perspectives. This category includes statements in which the AI chatbot struggles to distinguish between female and male perspectives, often providing uniform responses. An example is an AI chatbot's statement about the lack of male professors at universities (Urchs et al., 2023).

Inappropriate content and recommendations. This category addresses cases in which AI chatbot responses include inappropriate recommendations related to gender roles, thereby promoting outdated and discriminatory views of individuals' roles and priorities based on gender. For instance, Zhou and Sanfilippo (2023) report an incident where the AI chatbot suggested that women should prioritize marriage over a career.

Difficulties with gender-neutral language	<ul style="list-style-type: none"> Using 'he' or 'she' when 'they' would be more appropriate Referring to a group as 'guys' instead of a gender-neutral term like 'everyone'
Reinforcement of stereotypes	<ul style="list-style-type: none"> Depicting a nurse as female and a doctor as male Describing a primary caregiver in a family setting as a woman by default
Neglect of gender-diverse identities	<ul style="list-style-type: none"> Ignoring a non-binary individual in a prompt and assigning a binary gender Representing a character as either male or female, overlooking gender diversity
Assignment of characteristics based on gender	<ul style="list-style-type: none"> Automatically assuming competence in technology-related prompts for a male character Assigning emotional sensitivity traits to female characters without context
Inappropriate distinction of female and male perspectives	<ul style="list-style-type: none"> Providing the same response for a question about challenges in male-dominated fields as in female-dominated fields Example discussion of the "shortage of male STEM professionals"
Inappropriate content and recommendations	<ul style="list-style-type: none"> Advising women to prioritize appearance over skills in a professional context Advise women to marry instead of pursuing a career

Figure 2: Gender bias in AI chatbot conversations

Another finding from these studies is that some AI chatbots are specifically trained to respond to gender-related queries in a politically correct manner. Comparing responses from female, male and neutral perspectives, Urchs et al. (2023) found that certain words in a prompt seem to trigger a "gender template" in GPT3.5, resulting in responses that emphasize the importance of fairness, equality, and diversity. An example is the claim in the generated content that the aspects described are independent of gender.

4.3 Approaches to Mitigating Gender Bias in AI Systems

Several strategies have been developed to reduce gender bias in AI systems (see, for example, Nadeem, Abedin and Marjanovic, 2020 for an overview). These include adherence to principles of fairness and ethics in AI development, application of bias reduction approaches in algorithms, and diversity in AI development. Ferrera (2023) emphasizes the importance of humans providing input, giving feedback and overseeing the training process. Beattie et al. (2022) discuss an additional training of AI chatbots inspired by counter-stereotyping. The idea is that an AI chatbot can be trained with “nice” conversations, just as it can learn toxic behaviour from “bad” conversations.

It is important to note that while these strategies can help to reduce and mitigate bias, they may not be able to eradicate it altogether. Moreover, LLMs continue to learn as they interact with their users. Therefore, it is necessary to conduct regular audits of the AI models, including the application of fairness metrics that evaluate the generated content for potential bias, and to retrain LLMs with curated data if necessary.

5. Risks of Gender Bias in AI-co-created STEM Teaching Materials

5.1 Co-Creation Activities

Educators can use AI chatbots as co-creators when developing teaching materials and preparing classes. The cycle of exercise development by Kunz and Hess (2023) shows typical activities for designing STEM teaching materials. It encompasses the activities required to plan a task, let the learners work on it, provide feedback to the learners, grade the results, and finally, to optimize the task.

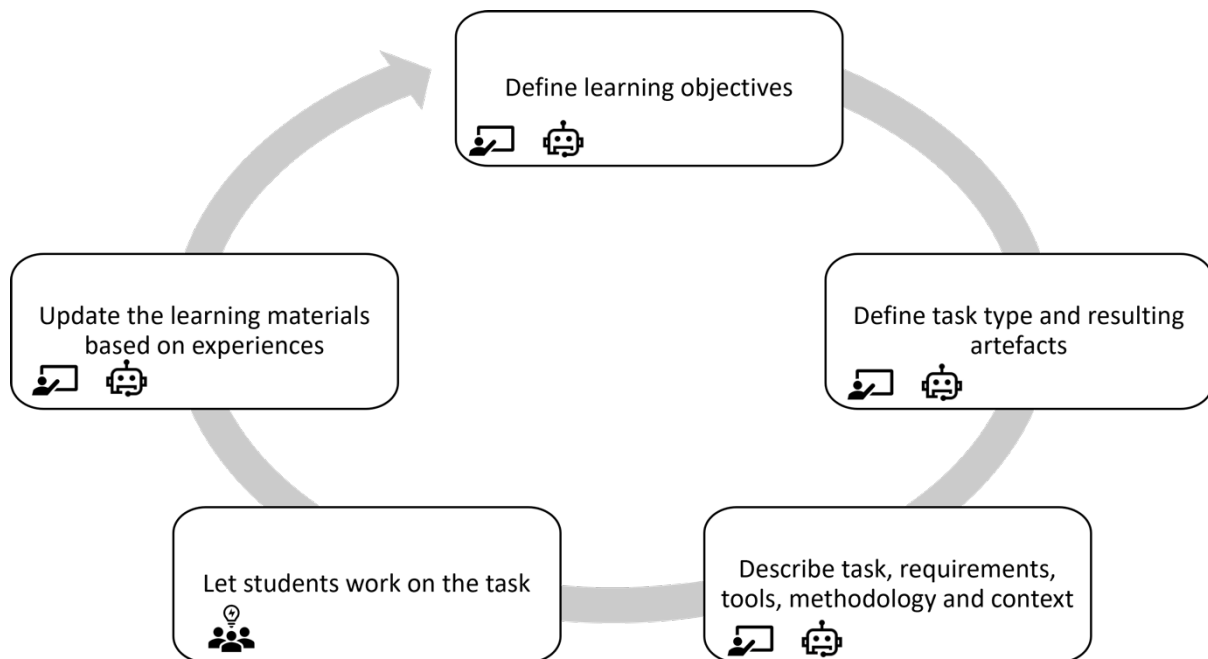


Figure 3: Cycle of exercise development

In this paper, a simplified version of this cycle is used as shown in Figure 3. The focus is on activities that are particularly well supported by AI chatbots.

5.2 General Risks

In all these activities, there is a risk of introducing gender bias. The repeated performance of gender in STEM education perpetuates the image of science as a male domain. There are several general risks that are not specific to any particular activity. First, teachers using AI chatbots must be aware that responses might be gender biased, even if the prompts are phrased in a gender-neutral way, such as using nongendered pronouns. This bias can manifest in the language itself, i.e., the chatbot using “she” or “he” instead of “they”. Alternatively, it can be reflected in the (fictitious) persons and their characteristics. Second, there is a risk that teachers “believe” AI

chatbots’ responses without fact-checking and without questioning gender aspects, due to their eloquent and convincing answers. When these answers become part of educational materials, they reinforce gendered views.

Moreover, educators must be aware that bias can also be introduced when they use the AI chatbot to translate a task description in another language. Using machine translation could be an easy way to provide learners who are non-native speakers with the opportunity to participate in a course because they could use the educational material in their native language. However, studies have shown that such machine translators can easily produce biased texts (e.g. Prates, Avelar and Lamb, 2020).

5.3 Risks Specific to Certain Co-Creation Activities

Besides of these general aspects, there are some particularities in how gender bias can be introduced in certain phases of the cycle of exercise development. To discuss them in a systematic approach, we consider all activities that can be supported by an AI chatbot and map to them the risk of gender bias in STEM education (introduced in Section 3) and AI chatbots’ potential gender bias (discussed in Section 4.2).

In the phase of defining learning objectives, there is the risk to underestimate girls’ competencies. Teachers should be aware that AI chatbots might generate learning objectives that vary in their difficulty when girls or boys are explicitly mentioned in the target group. Depending on the AI chatbot being used, a “gender template” as described by Urchs et al. (2023) could be applied that emphasizes fairness and diversity. Nonetheless, the text might exhibit some bias.

When defining the task type and the resulting artifacts, the bias in AI chatbots answers can be more subtle. Based on the teaching materials that the AI chatbots has consumed during its training and its knowledge about “typical” interests of students interested in STEM, i.e., male learners, the selection of the tasks could be biased and focus on rather technical aspects instead of showing the diversity of topics in STEM subjects.

AI chatbots can aid educators by describing tasks, requirements, tools, and methodologies, enhancing engagement with tailored background stories as shown by Kunz and Hess (2023). They emphasize the need to carefully craft the prompts when asking the AI chatbot for background stories that are particularly relevant to girls’ interests. One example showed that the background story for 12–19-year-old female students resembled a child-book with very flowery descriptions. There is also a risk of gendered views. For example, different personality traits are emphasized for men and women, and may even be omitted for gender-diverse individuals as Gross (2023) has shown. Moreover, there is the risk of presenting women only as teachers in STEM.

6. Strategies to Deal with Gender Bias in AI-Generated Teaching Materials

Based on discussion in the previous section on the risks that arise due to potential gender biases in the teaching materials co-created with an AI chatbot, this section provides educators with an idea of how to mitigate these risks. Therefore, four strategies will be described, and example prompts be given. They fall into two categories as presented in Figure 4. First, there are strategies for identifying bias and second, there are strategies for helping educators to amplify existing perspectives in their teaching materials.

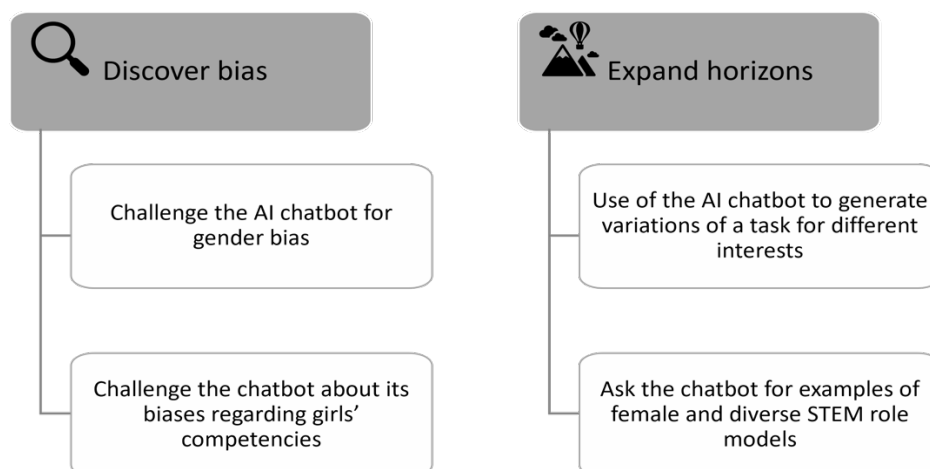


Figure 4: Strategies to deal with gender bias in AI-created teaching materials.

6.1 Challenge the AI Chatbot for Gender Bias

Teachers can use the AI chatbot to become aware of potential biases. This could give teachers hints on which aspects they should consider when deciding on aspects such as class organization or when explaining the task. Moreover, if teachers are aware that (some) girls might find some aspect especially appealing, they could make use of it, e.g., mentioning it when talking to their students. The following example shows two prompts that can be used to ask the AI chatbot explicitly for girls' and boys' interests.

1. "Are there any parts within the lesson that we designed above, that are especially interesting to girls?"
2. "Are there any parts within the lesson that we designed above, that are especially interesting to boys?"

Building up on the case study presented in Hess and Kunz (2023), we used this prompt in the context of a computer science lesson that introduces basics about data analysis and data visualization. First, ChatGPT states that the lesson designed is not inherently biased towards any particular gender and fosters the need to create inclusive lessons that consider diverse interest and learning styles (the "gender template" discussed in Section 4.2). When comparing the answers, there are several aspects that are included in both, although worded slightly different. First, the aspect of problem solving is named both for girls and boys. Second, both answers stress the importance of real-world applications and suggest that students be allowed to select topics they are passionate about. Moreover, the use of visuals, diagrams, and interactive elements is recommended.

However, there are some differences in the answers. The first is about class organization. ChatGPT argues that many girls enjoy social interaction and collaboration, and therefore a group setting may be appealing to them. In contrast, ChatGPT suggests "friendly competition" for boys, as some boys thrive in competitive environments. These aspects have been analysed in different studies and can be found in various pedagogical strategies (e.g., Kulturel-Konak, D'Allegro and Dickinson, 2011; Dancstep and Sindorf, 2018).

The second difference that can be noted in the answers is how ChatGPT describes the special interest in the practical tasks of collecting data and designing the chart. For girls, the creative aspect of designing charts was highlighted. Teachers should, for example, encourage them to think about what colours to use. For boys, ChatGPT emphasizes that it is a hands-on activity. Teachers could use these two notions to better "sell" the task to their students and gain their interest, however without explicitly targeting one gender.

6.2 Challenge the Chatbot About its Biases Regarding Girls' Competencies

The second strategy also challenges the AI chatbot to discover potential bias, this time with regard to the underestimation of girls' competencies in STEM subjects. When co-creating teaching materials with an AI chatbot, educators should explore whether this problem shows up in ChatGPT's answers. With the help of different prompts, one can find out whether and to which extent ChatGPT will reproduce prejudices about girls' competencies regarding STEM topics. To this end, one can use the following prompts:

1. What difficulties could students have during this lesson?
2. What difficulties could girls have during this lesson?

Applied in the sample scenario described above, the first prompt resulted in a list of potential difficulties that students might encounter during a lesson on data analysis and data visualization. Examples are difficulties in properly collecting and organizing the data, in using software for data entry and visualization, in understanding the information conveyed by the charts or in presenting findings. There was no gender bias in the responses, nor were there examples of girls or boys struggling with particular aspects. The answer to the second prompt emphasizes that girls, like all students might face difficulties, but that these are not gender-specific. A gender bias could not be seen in the answer.

6.3 Use of the AI Chatbot to Generate Variations of a Task for Different Interests

The next strategy uses the AI chatbot to enrich existing teaching materials with alternative tasks that are more likely to appeal to the student's age group. This could also include the idea that educators do not try to avoid all potential bias in their teaching materials and generate gender-neutral materials, but to make use of the bias and the stereotypes inherent in generative AI. Teachers can use the AI chatbot to create variations of a certain task that are more likely to appeal to girls or boys in an efficient way. The reasoning can be, as discussed by Kunz and Hess (2023), "fighting fire with fire" and attract girls to STEM subjects with the help of "girl-specific topics". The

chatbot could be asked to create alternatives to an existing task, for example, with topics related to sports, nature, animals, comics or fantasy worlds. You could also provide TV or book series as input in the sense of “create a task as Ladybug & Cat Noir would have to do it”. A further possibility would be to provide the AI chatbot with a female STEM professional as an example and ask it to create a background story for a task based on this role model.

It is important that the task itself and the difficulty are comparable across all topics. Offering different topics to the students, they can choose a topic that matches their interests. With the help of an AI chatbot, the time-consuming task of designing variations including sample solutions can be facilitated.

6.4 Ask the Chatbot for Examples of Female and Diverse STEM Role Models

An AI chatbot can assist teachers in incorporating examples of successful female STEM role models into their educational materials, demonstrating to students that women have excelled in STEM fields. However, it is important to note that AI chatbots generate text instead of solely retrieving information from various sources. As a result, they might “hallucinate” and produce inaccurate results. Therefore, teachers must verify the accuracy of any factual information.

7. Conclusion

When teachers use AI chatbots to co-create educational materials, they need to be sensitive to gender bias. The materials may reinforce the stereotypes and inaccuracies that were present in the training data. This can impact how students perceive and understand a topic and perpetuate a culture that discourages girls and women from pursuing STEM disciplines because of their gender. The paper explored the risks of gender bias in AI-co-created teaching materials along the activities in the cycle of exercise development that could be supported by a generative conversational AI system. To provide educators with a concrete approach to address these risks, four mitigation strategies were developed which help educators to identify gender bias and to enhance teaching materials by a broader perspective. Applying these strategies, educators can create teaching materials that are likely to appeal to all students regardless of gender.

Future research should use real-world situations in which educators prepare teaching materials for specific lessons and specific target groups. One could perform all activities of the cycle of exercise development and systematically analyse the materials generated by the AI chatbot. It would be interesting to compare the results of different AI chatbots, especially also from different regions in the world and with a different cultural background.

References

- Beattie, H. *et al.* (2022) ‘Measuring and Mitigating Bias in AI-Chatbots’, in *2022 IEEE International Conference on Assured Autonomy (ICAA 2022)*. Piscataway, NJ: IEEE, pp. 117–123. Available at: <https://doi.org/10.1109/ICAA52185.2022.00023>.
- Blickenstaff, J.C. (2005) ‘Women and science careers: leaky pipeline or gender filter?’, *Gender and Education*, 17(4), pp. 369–386. Available at: <https://doi.org/10.1080/09540250500145072>.
- Dancstep, T. and Sindorf, L. (2018) ‘Creating a Female-Responsive Design Framework for STEM exhibits’, *Curator: The Museum Journal*, 61(3), pp. 469–484. Available at: <https://doi.org/10.1111/cura.12268>.
- Dele-Ajayi, O. *et al.* (2020) ‘Tackling Gender Stereotypes in STEM Educational Resources’, in *2020 IEEE Frontiers in Education Conference (FIE)*, pp. 1–7. Available at: <https://doi.org/10.1109/FIE44824.2020.9274158>.
- Eccles, J. (2009) ‘Who Am I and What Am I Going to Do With My Life? Personal and Collective Identities as Motivators of Action’, *Educational Psychologist*, 44(2), pp. 78–89. Available at: <https://doi.org/10.1080/00461520902832368>.
- Ellemers, N. (2018) ‘Gender Stereotypes’, *Annual Review of Psychology*, 69(1), pp. 275–298. Available at: <https://doi.org/10.1146/annurev-psych-122216-011719>.
- Ferrara, E. (2023) ‘Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models’. arXiv. Available at: <https://doi.org/10.48550/arXiv.2304.03738>.
- Gross, N. (2023) ‘What ChatGPT Tells Us about Gender: A Cautionary Tale about Performativity and Gender Biases in AI’, *Social Sciences*, 12(8), p. 435. Available at: <https://doi.org/10.3390/socsci12080435>.
- Hess, C. *et al.* (2023) ‘How does school lesson design contribute to girls’ perception of their future-readiness for STEM majors? – An empirical study using critical incidents technique’, *International Journal on Learning and Teaching*, 9(4).
- Hess, C. and Kunz, S. (2023) ‘“ChatGPT, Can you Motivate my Learners?” - Co-Creation of Inspiring STEM Lessons with AI Chatbots’, in *ICERI2023 Proceedings - 16th annual International Conference of Education, Research and Innovation*. IATED, pp. 6103–6110.

- İpek, Z.H. et al. (2023) 'Educational Applications of the ChatGPT AI System: A Systematic Review Research', *Educational Process: International Journal*, 12(3), pp. 26–55. Available at: <https://doi.org/10.22521/edupij.2023.123.2>.
- Kasneji, E. et al. (2023) 'ChatGPT for good? On opportunities and challenges of large language models for education', *Learning and Individual Differences*, 103(102274), pp. 1–9. Available at: <https://doi.org/10.1016/j.lindif.2023.102274>.
- Kerkhoven, A.H. et al. (2016) 'Gender Stereotypes in Science Education Resources: A Visual Content Analysis', *PLOS ONE*, 11(11), p. e0165037. Available at: <https://doi.org/10.1371/journal.pone.0165037>.
- Kulturel-Konak, S., D'Allegro, M.L. and Dickinson, S. (2011) 'Review Of Gender Differences In Learning Styles: Suggestions For STEM Education', *Contemporary Issues in Education Research (CIER)*, 4(3), pp. 9–18. Available at: <https://doi.org/10.19030/cier.v4i3.4116>.
- Kunz, S. and Hess, C. (2023) 'Teachers, Reload Your Toolboxes: Using a Task Design Framework and ChatGPT to Generate Motivational Exercises for Female Learners in Computer Science', in P. Dini (ed.) *Proceedings of Advances on Societal Digital Transformation (DIGITAL 2023)*, pp. 13–18.
- Makarova, E., Aeschlimann, B. and Herzog, W. (2019) 'The Gender Gap in STEM Fields: The Impact of the Gender Stereotype of Math and Science on Secondary Students' Career Aspirations', *Frontiers in Education*, 4. Available at: <https://doi.org/10.3389/feduc.2019.00060>.
- Nadeem, A., Abedin, B. and Marjanovic, O. (2020) 'Gender Bias in AI: A Review of Contributing Factors and Mitigating Strategies', in *ACIS 2020 Proceedings*. Available at: <https://aisel.aisnet.org/acis2020/27>.
- Papadakis, S. (2018) 'Gender stereotypes in Greek computer science school textbooks', *International Journal of Teaching and Case Studies*, 9(1), pp. 48–71. Available at: <https://doi.org/10.1504/IJTC.2018.090196>.
- Prates, M.O.R., Avelar, P.H. and Lamb, L.C. (2020) 'Assessing gender bias in machine translation: a case study with Google Translate', *Neural Computing and Applications*, 32(10), pp. 6363–6381. Available at: <https://doi.org/10.1007/s00521-019-04144-6>.
- Raabe, I.J., Boda, Z. and Stadtfeld, C. (2019) 'The Social Pipeline: How Friend Influence and Peer Exposure Widen the STEM Gender Gap', *Sociology of Education*, 92(2), pp. 105–123. Available at: <https://doi.org/10.1177/0038040718824095>.
- Ray, P.P. (2023) 'ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope', *Internet of Things and Cyber-Physical Systems*, 3, pp. 121–154. Available at: <https://doi.org/10.1016/j.iotcps.2023.04.003>.
- Steffen, A. et al. (2023) 'An Exploration of Critical Incidents Impacting Female Students' Attitude Towards STEM Subjects', in *6th International Conference on Gender Research (ICGR 2023)*, pp. 215–223. Available at: <https://doi.org/10.34190/icgr.6.1.994>.
- Urchs, S. et al. (2023) 'How Prevalent is Gender Bias in ChatGPT? -- Exploring German and English ChatGPT Responses', in *1st Workshop on Biased Data in Conversational Agents" (co-located with ECML PKDD 2023)*, arXiv. Available at: <https://doi.org/10.48550/ARXIV.2310.03031>.
- Weber, K. (2012) 'Gender Differences in Interest, Perceived Personal Capacity, and Participation in STEM-Related Activities', *Journal of Technology Education*, 24(1), pp. 18–33.
- Zhou, K.Z. and Sanfilippo, M.R. (2023) 'Public Perceptions of Gender Bias in Large Language Models: Cases of ChatGPT and Ernie'. arXiv. Available at: <https://doi.org/10.48550/arXiv.2309.09120>.