AI and Narrative Scripts to educate adolescents about Social Media Algorithms: Insights about AI overdependence, trust and awareness

Emily Theophilou[1][0000−0001−8290−9944], Francesco Lomonaco[2][0000−0002−2295−1443], Gregor Donabauer[2,3], Dimitri Ognibene[2][0000−0002−9454−680X], Roberto J. Sánchez-Reina[1][0000−0002−6068−1229], and Davinia Hernàndez-Leo[1][0000−0003−0548−7455]

[1] Universitat Pompeu Fabra, Barcelona, Spain
emily.theophilou@upf.edu
[2] Università degli Studi di Milano-Bicocca, Milano, Italy
[3] University of Regensburg, Regensburg, Germany

Abstract. Social Media Artificial Intelligence algorithms provide users with engaging and personalized content. Yet, the personalization of algorithms may have a negative impact on users who lack AI literacy. The limited understanding of SM algorithms among the population suggest that adolescents are more likely to place blind trust in the information they consume, exposing them to negative consequences (misinformation, filter bubbles and echo chambers). We therefore propose an intervention with a narrative scripts approach to raise awareness of AI algorithms in SM. To foster an authentic learning experience and question adolescents' trust in AI, we deploy a low-accuracy AI image classifier. A quasi-experimental study was conducted among 144 high-school students in Barcelona, Spain. The results show that the narrative scripts intervention improved students' awareness of SM algorithms and shaped more critical attitudes towards them. A comparison of students' choices between human predictions and those produced by a low-accuracy AI classifier shows a lack of AI overdependence. Information about predictions' source did not affect students' trust or learning about AI. These findings contribute towards SM algorithms education and share insight into the effect of deploying low-accuracy detectors in learning technology interventions.

Keywords: Social media algorithms education · Low-accuracy image classification · Adolescents AI trust · AI overdependence.

1 Introduction

The technological advances in machine learning algorithms have seen significant development in the generation, curation and delivery of content. Techniques such as user profiling, data mining, and tracking users' actions identify the patterns in

user behavior with the aim of delivering content preferences and increasing user engagement [36]. With algorithms influencing the information and content users consume, new concerns emerge about their adverse effects on users' experiences. The filtering of content based on user's preferences and online behavior can eventually lead to negative consequences such as polarization, the creation of echo chambers, reality distortion, and social media addiction [19, 14].

The invisibility of AI-based algorithms deepens the problem, as users are often unaware of the mechanisms behind the content they consume [27] and are unaware of its homogeneity. According to [3], young adults are aware of the consequences of SM algorithms, such as filter bubbles, but fail to take action against them. Conversely, [27] suggest that younger generations, who have grown up surrounded by personalized AI systems, tend to trust AI but are unaware of its potential negative consequences. This highlights the need for interventions that educate adolescents on SM algorithms and provide strategies for dealing with them [36, 27].

Additionally, it emphasizes the potential risk of relying too heavily on algorithms and prompts questioning of the uncritical trust placed in AI. Therefore, this study proposes an intervention to raise awareness of the algorithms behind SM platforms for high school students. In particular, the study analyzes the implementation of the narrative scripts [16], an authentic learning approach, situated in teaching about filter bubbles and body image distortion. To foster an authentic learning experience towards AI algorithms, the study utilizes and evaluates a low-accuracy image classifier. The implementation of the AI classifier allowed us to challenge adolescents' trust towards AI and explore AI overdependence. The study addresses a threefold objective (a) assess the narrative scripts approach as an intervention tool for SM algorithms education, (b) challenge adolescents trust in AI and enhance their awareness with the deployment of a low accuracy AI classifier, and (c) investigate if the use of low-accuracy AI classifier can result in adverse educational outcomes.

2 Social media algorithms and adolescents

SM use among adolescents has become increasingly prevalent in recent years, with a notable increase after the COVID-19 lockdown [9]. Visual platforms such as YouTube, TikTok, and Instagram serve as key sources of social interaction, entertainment, and self-expression [37]. These types of platforms typically utilize proprietary personalization algorithms to curate and present content that matches a user's preferences. SM algorithms aim to improve navigation/online experience by displaying content that is both relevant and engaging to each individual user. As a result, no two users have the same SM experience [8]. In spite of the effectiveness/efficacy of AI algorithms, there is a potential for adverse outcomes, such as the creation of echo chambers that reinforce a homogeneous content while creating a (statistically) distorted or biased perception of reality [14, 20].

The negative effects of AI algorithms can generate harmful effects that may impact the users' attitudes, behaviors, and well-being [6, 32]. Concerning adolescents' body image, filter bubbles can shape young users' exposure to idealistic body images affecting their perceptions of a healthy body [10, 14] and potentially lead to lower body satisfaction [5]. It is therefore important to recognize the functions and impacts of algorithms on social media platforms and interact with them consciously and critically [11]. A study investigating how Instagram's body type-specific filter bubbles are understood by female adolescents found that the effect was attributed to individual preferences rather than an adverse effect of an AI algorithm [35]. Even though content recommendation is based on individual preference, preference magnification and the exclusion of contrasting content can create a disclosed environment that can isolate the user. This demonstrates how an increase in algorithms and AI technologies can significantly impact young people's understanding of the world around them [34]. As [33] note, children are growing up with AI applications, including chatbots and recommendation tools, without necessarily understanding the basic principles behind them. Despite their lack of awareness of AI technologies, adolescents tend to trust AI [27] and have uncritical attitudes towards AI-based assistants [31].

A study conducted among the Finnish population to assess awareness of AI algorithms found that adolescents aged 15 to 19 have a positive attitude towards algorithm-driven recommendations despite low algorithmic awareness levels [11]. These findings highlight the potential issue that younger generations may not fully understand how algorithms work or where they are present. Yet, adolescents tend to trust and have a favorable view of their outputs and recommendations. Trust in information technology tools involves the acceptance of potential vulnerabilities that might affect an output [24]. However, AI algorithms outputs are often found biased or unfair [28, 1]. Hence an insufficient knowledge among adolescents regarding algorithms' functionality and potential vulnerabilities leaves them open to accepting falsified information without being critical about it. This could raise concerns about the overreliance on algorithm-generated recommendations without critical evaluation which could potentially contribute to the propagation of systemic biases [2].

Overdependence on algorithms and AI can be a significant concern, mainly if individuals rely too heavily on these technologies without understanding their limitations. Additionally, overreliance on algorithms and AI can lead to a lack of critical thinking skills, as individuals may not question the results they receive or understand how these results were generated. Overall, the literature suggests that adolescents' knowledge towards SM algorithms is often lacking, yet they tend to trust AI algorithms and overdependence on these technologies can lead to adverse effects. Consequently, it is crucial to develop interventions to challenge adolescents' blind trust towards AI algorithms and raise awareness of potential consequences that can arise and provide them with strategies to counteract them.

2.1 Social media algorithms education

Interventions aimed at enhancing information literacy and resistance to manipulation can help exert greater control over users' online environment [18]. The implementation of AI literacy courses in school curriculums has been advocated by researchers and educators in the last few years and saw the design, development and evaluation of different curriculum proposals [17, 21, 33]. According to [17], an AI curriculum should have four stages depending on the student's age. For younger students (kindergarten and primary school), the initial stage should involve a playful exploration of AI topics to build awareness. As students grow older and move into middle school, they should engage in more critical thinking approaches through experimentation and familiarisation with AI topics. By reaching high school, students should cover more advanced AI topics to promote independent thinking and apply their knowledge. Overall, [17] model suggests that AI education should be tailored to a student's developmental stage, focusing on building awareness, critical thinking, and applied knowledge as they progress through their education.

While [17] model suggests a gradual exposure to AI concepts, [21] explores an approach that targets middle school students without prior experience on computer science-related topics. Both approaches have common aspects that see students familiarization with the dynamics behind AI algorithms and explore machine learning topics at different paces. Besides the content covered in the AI literacy curriculum, it is also important to consider the methods and types of learning being encouraged. For example, [17] applied methods such as discovery and inquiry learning alongside techniques of storytelling and educational robotic tools whereas [21] saw the incorporation of hands-on games, discussions, and building projects that utilise AI functionalities. Both types of interventions have proven successful in raising awareness of AI algorithms and demonstrate that AI literacy can take different forms, regardless of the educational tools used.

The use of algorithms to support and illustrate their functionality, can help students understand the impact of algorithms on users' actions, behaviors, and well-being [26]. Demonstrations of how SM invincible algorithms work can help understand algorithms at work [12] and promote awareness of their capabilities and limitations [4]. Interventions that deconstruct SM algorithms and demonstrate them to students at work saw a rise in the last few years. For instance, [23] replicated machine learning mechanisms through interactive classroom activities to raise awareness of negative algorithmic consequences, such as echo chambers.

In addition to demonstrating how algorithms function, previous research has investigated the potential of using algorithmic predictions to increase awareness of SM risks. For instance, a study replicated AI-based labeling of image classifications in a controlled photo-sharing platform. It found that it helped adolescents develop a critical stance towards body image representations on SM [29]. However, this raises the question of whether demonstrating how machine learning algorithms work requires genuine AI predictions or if human-made predictions can suffice in conveying the message. We believe that the type of activity to be performed determines the appropriate approach. When conducting experimen-

tal work, AI algorithms may be used to deceive participants in order to explore their decision-making or preferences [39, 38]. Conversely, in educational settings, AI predictions can be used to demonstrate the workings of algorithms in an attempt to spark students' curiosity, and provide insights into the capabilities and limitations of AI. Advanced interventions can even offer a hands-on environment where students can manipulate algorithm parameters and experience authentic learning opportunities.

2.2 The present study

This study proposes the design of an intervention aimed at adolescents to raise awareness of how SM algorithms work, their potential consequences and strategies to overcome them. Based on the literature presented, the intervention will cover AI media topics as suggested by [36], interactive classroom activities that replicate machine learning algorithms as outlined by [23], and the use of real machine learning algorithms predictions to prompt critical questioning of AI algorithms. To provide an engaging educational experience towards SM education, we implement the proposed activities within the Narrative Scripts environment [16, 22]. In particular, the NS approach sees the deployment of a simulated SM platform that delivers learning material to students through educational scripts guided by narratives. The use of the narrative script approach will provide an opportunity to frame the learning material within a storyline to explore AI algorithms' negative consequences in social media through the eyes of a fictional user.

As a component of the intervention, our proposal involves the demonstration of actual predictions made by machine learning algorithms to stimulate critical thinking in relation to AI algorithms. The deployment of AI algorithms for educational purposes is a challenging task however it can provide an authentic learning scenario where students are exposed to their realistic capabilities. While AI has shown impressive performance in several benchmarks [15], the accuracy of AI predictions depends upon the quality of data available for model training and relies on task-specific datasets that are difficult to locate. However, when replicating accurately human behaviors, such AI machinery would go undetected in learning settings. While the employment of a low-accuracy algorithm can hinder the behavioral trust of participants [39], it could potentially work as a tool to question adolescents' uncritical trust towards these systems and possibly foster critical thinking towards their use. Given the limited research exploring the potential impact of low-performing AI algorithms in educational contexts, this study aims to address this gap and provide valuable insights. Hence, the following research questions have been formulated:

RQ1: What effect does an intervention on SM algorithms within the NSs have on adolescents' awareness and attitudes towards SM algorithms?

RQ2: Does the exposure to a low accuracy detector affect students' awareness, attitudes and trust towards AI in general (as vs human predictions)?

RQ3: Can we use low accuracy AI recommendations to evaluate the impact of the educational activity on students AI overdependence?

3 Methodology

3.1 Participants and study design

The study was conducted across high schools in Barcelona. In total, 144 students participated (n = 144; 54.86% male, 45.14% female; Ages 12 to 19, mean age = 14.8, SD = 1.71). The study was done as part of a SM literacy workshop and was conducted in two sessions. It took place in the spring semester of the academic year 2022, with schools participating in one session per month. As part of the research protocol, students and parents were briefed on the research objectives and the purpose of the workshop prior to its commencement. Both students and their families were requested to sign an electronic consent form to consent to their participation in the study.

3.2 Procedure

Our learning approach saw a combination of the different stages proposed by [17]. Students initially became involved in playful storytelling techniques to engage with the context of AI in social media, and saw the deployment of interactive classroom activities to engage into critical thinking approaches to become familiar with algorithms (see figure 1). Students were introduced to topics related to filter bubbles, recommender systems and image classification under the theme of body image.

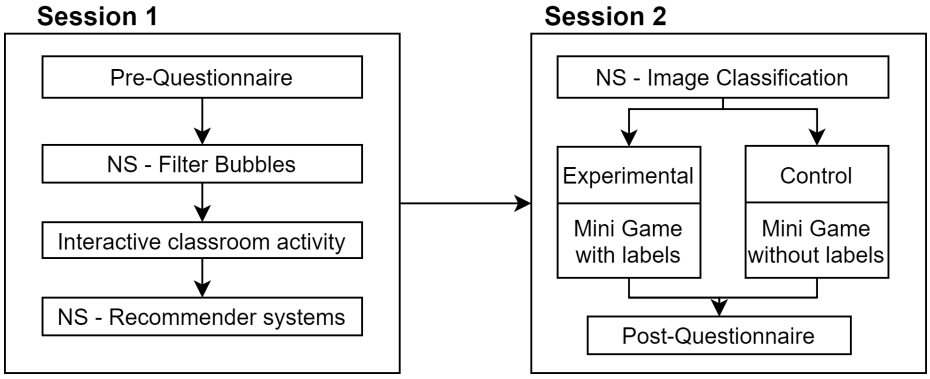

Fig. 1. Learning design and procedure of the study.

During the first session of the workshop, students completed the pre questionnaire and then accessed material through the Narrative Script platform. Then the teacher initiated an interactive classroom activity where students were exposed to a filter bubble scenario and the negative effect it can have on one's body image. After the activity, students returned to the Narrative Scripts platform and were introduced to the topic of recommender systems.

The second session revolved around a mission to help the fictional CEO of the platform to employ a new "AI agent" with the goal to help monitor the types of images uploaded to the platform. To begin with, students were explained how an image classification system works. Then they were given access to a mini-game with the goal of evaluating an image classification system.

During the mini-game, a randomized controlled experimental design was applied to evaluate the potential effects of an AI classifier with low accuracy. To achieve this, we presented two predictions to students: one generated by a low-accuracy algorithm and the other by humans. The game interface displayed an image of a person, two sets of predictions (human and AI), the predictions (gender and BMI), and a button to select the decorations they found more suitable (see figure 2). In the control group, students were not informed whether the prediction was made by AI or human, while it was reported to the experimental group (see figure 1). The conditions were randomly assigned. Students were shown ten images and the set of predictions was randomly placed each time. After the game was over, students completed the post questionnaire.

Fig. 2. The interface of the mini-game. The interface on the left corresponds to the experimental group, where participants could see which prediction was made by AI or humans. The right interface corresponds to the control group. The images have been blurred in this paper for privacy reasons.

3.3 AI Model: Architecture and Training

To generate the AI predictions we developed a machine learning model based on multiple pre-trained convolutional neural networks (CNN) and fine tuned it on a dataset that consists of face images annotated with their associated BMI values. The pretrained models we use as part of our fully-connected BMI prediction model are adopted from the DeepFace library [30] and we adapt four different versions (VGGFace, age, gender, race) as hidden components. We also used the hidden gender model out-of-the-box for gender prediction in our experiments. The training of the BMI model was performed using the Reddit-HWBMI

dataset [13]. During the training stage the model was underfitted on the data, leading to BMI outputs that are in the majority of the cases higher than the ground truth (so predictions are skewed towards higher BMI values). To get AI-predictions for the experiments we use MTCNN to crop images around a person's face and resize this cutout to a dimension of 224x224 (model input size). Afterwards we feed them inside our BMI model which predicts the person's BMI as a scalar. We use these values during the experiments as the AI condition.

3.4 Measurements

Social media algorithms (SMA) awareness: To measure participants' awareness of algorithmic functions in SM, the question "When I navigate social media, algorithms influence the content I see" was included as part of the pre and post questionnaire. It was measured with a 5-point Likert scale ranging from 1 - Very untrue to 5 - Very true. Moreover, the question "I have heard of the concept Artificial intelligence algorithms" was included to measure students' experience with AI algorithms.

SMA attitude: Attitude items were formulated based on the AI divide study by [11], and questions were developed to measure students' attitudes towards algorithmic functions and effects. In particular, two items were formulated to measure students' attitudes towards algorithm recommendations in SM and the addictive consequences of SM use. Both items were measured with a 5 Likert scale ranging from 1 - I love it and 5 - It frustrates me and were included in the pre and post questionnaires.

Trust towards AI: To measure adolescents' trust towards AI, four items were extracted and adjusted from the trust in technology questionnaire [25]. The items were based on the Institution-Based Trust questions and covered the dimensions of competence, trusting intentions, integrity and benevolence. The items were measured with a Likert scale ranging from 1- Strongly disagree and 5 - Totally agree. Trust towards AI was measured as part of the post questionnaire.

AI overdependence: AI overdependence occurs when individuals excessively rely on AI systems, disregarding their limitations and potential errors. In our study, we measured this variable by observing students' choices of predictions during the mini-game phase, which served as an indicator of the extent to which students exhibited excessive dependence on AI-generated predictions.

3.5 Data analysis

To minimize the effect of considering data from students not paying attention to the images and classifications shown to them, a misleading image labeling was given to them halfway through the experiment (a female was classified as a male (photo 6)). Consequently, we excluded participants who responded incorrectly

to the image during the analysis. This resulted in 23 students being removed and the final sample size being 121 students.

To calculate the human predictions of the BMI values, an internal study was conducted before the main study with 39 participants (n = 39; 54% male, 46%female; Ages 19 to 61, mean age = 26.3, SD = 8.2). The participants were asked to classify a set of images from the VIP dataset [7] regarding the gender and BMI they believed the person in the image had.

4 Findings

The final sample included the data of 121 students (n = 121; 52.9%male, 47.1% female; Ages 12 to 19, mean age = 14.6, SD = 1.69). In regards to RQ1, the intervention showed to be effective in raising adolescents' awareness of SM algorithms. A paired t-test showed significant differences between the pre and post-questions ($p \leq .01$) as the participants' awareness was higher in the post-questionnaire (mPost = 3.72) than in the pre (mPre = 3.27). The intervention also influenced students' attitudes towards AI in SM with a significant effect on algorithms recommending them content in SM (mPre = 2.21, mPost = 2.61,$p \leq .01$) and becoming hooked in SM (mPre = 3.21, mPost = 3.63, $p \leq .01$). Figure 3 displays the findings concerned with the effect of the intervention on students' awareness and attitudes of AI in SM. An independent t-test showed significant differences between the two genders previously hearing about AI algorithms, with male students (mMales = 2.77) reporting hearing more about it than female students (mFemales = 2.21). However, no significant differences were found between the two genders regarding how SM algorithms work. Concerning AI attitudes, female adolescents shared a more significant initial concern of becoming hooked in SM than male adolescents (mFemales = 3.5, mMales = 3.08, p = $p \leq .05$). However, this difference decreased after the intervention and showed no significant differences (mFemales = 3.8, mMales = 3.5, $p \geq .05$).

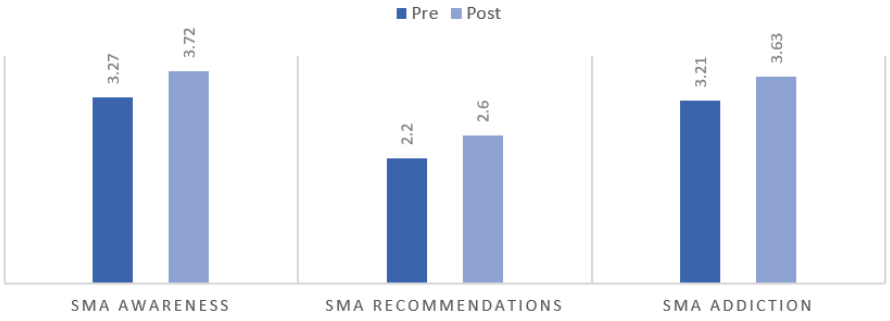

Fig. 3. Students' SMA awareness and attitudes before and after the intervention.

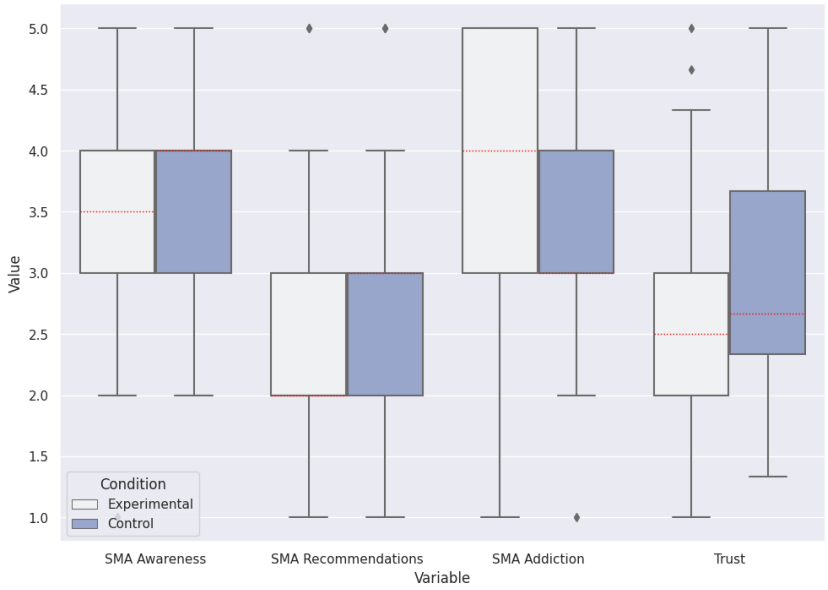

Fig. 4. Students performance under each condition in the post questionnaire.

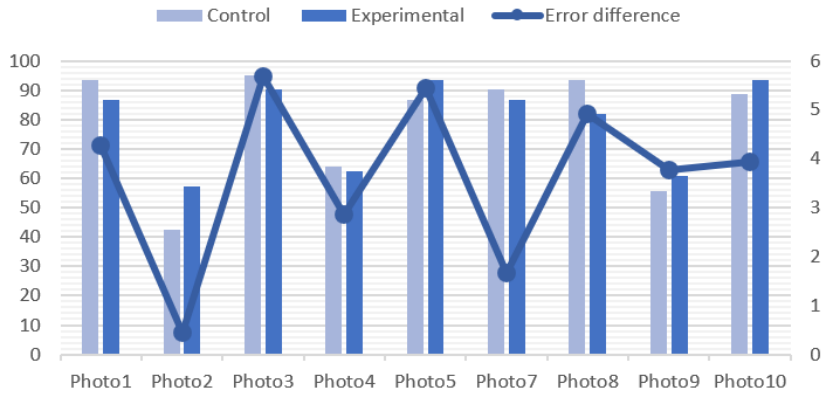

Fig. 5. Percentage of participants' selections of human-generated predictions during the mini-game. In addition, the graph visually compares the participants' choices with the error rate of the AI detector in relation to the human inputs. An error difference less than 1 indicates a high agreement between AI and human-prediction.

Regarding RQ2, the exposure to a low-accuracy image classification did not affect students' learning. The final sample saw 61 students under the control condition and 60 under the experimental. An independent t-test showed significant differences between the two groups' initial awareness of how SM algorithms work ($\text{mControl} = 3.65$, $\text{mExperimental} = 3.27$, $p \leq .05$). Therefore, we conducted an

independent t-test on the students' learning gains instead, where no significant differences were found ($p \geq .05$). Concerning AI attitudes and AI trust, no significant differences were found between the two conditions. Figure 4 displays the impact of the intervention on students' awareness and attitudes per condition.

Regarding RQ3, an exploratory analysis was conducted to examine students' selections during the mini-game. Findings showed that students under both conditions selected similar predictions. Students' human prediction selection under the control condition was 78% in comparison to experimental condition students, where it was 80%. An independent t-test showed no significant differences ($p \geq .05$). Figure 5 depicts students' human-generated prediction selection for each photo.

5 Discussion and conclusion

This study utilized a narrative scripts approach to educate adolescents about social media AI algorithms and challenge their trust using a low-accuracy image detector. Three research questions were formulated to examine algorithmic awareness, trust, and over-dependence. RQ1 assessed the effectiveness of the narrative scripts approach as an intervention tool. RQ2 investigated the impact of a low-accuracy detector on adolescents' learning and trust. Lastly, RQ3 explored adolescents' AI overdependence. The findings provide valuable insights into the educational outcomes influenced by a low-accuracy image classifier.

The use of the narrative scripts as an intervention tool effectively increased awareness of SM algorithms and fostered critical attitudes towards them. Following the intervention, students demonstrated improved awareness of how SM algorithms work. Additionally, the intervention led to students becoming more critical regarding algorithm-recommended content and the addictive nature of SM. This aligns with previous work that saw interactive interventions as an effective approach towards raising awareness of AI algorithms [21, 23]. This finding carries significant importance in today's rapidly evolving AI-driven world. The increased awareness and critical thinking skills towards AI algorithms can empower students to make informed decisions while navigating in online platforms and ultimately help prevent harmful situations.Female adolescents' self-reported awareness of AI algorithms was lower than that of male students, consistent with previous findings [11]. However, their understanding of how the algorithms work was not significantly different from male participants. This suggests that gender overconfidence may contribute to the lower self-reported awareness among females, as suggested by [11]. In terms of AI attitudes, female adolescents exhibited a significantly more critical view of social media addiction compared to males. This could be attributed to their higher tendency towards social media addiction [40], leading to greater scrutiny.

Regarding RQ2, the study found that students' learning and attitudes improved regardless of their exposure to low-quality image classification. This suggests that the intervention itself was effective in enhancing student outcomes, independent of the additional factor of low-quality image classification. Thus,

the implementation of low-accuracy image classification for demonstrative purposes does not have a negative impact on learning about SMA. However, the use of low-accuracy image classifications to challenge students' trust remains questionable. Our results indicated that students in the experimental condition had slightly lower levels of trust compared to the control group, but no significant differences were observed. This could be attributed to the specific AI algorithm used and the sample size of our study.

Finally, RQ3 examined students' AI overdependence by comparing low accuracy predictions to human-made predictions. Our findings indicate that students' selections were similar in both conditions, suggesting that students in the experimental condition did not demonstrate AI overdependence. Notably, in photo 2, where the AI classifier's accuracy was close to human predictions, students in the control condition blindly preferred the AI prediction, while those in the experimental condition chose the human prediction. This contrasts with the findings of [39], where participants showed a stronger preference for AI decisions over human decisions. Although this may indicate potential AI aversion, no significant differences were observed, warranting further investigation of secondary variables that might influence the results. To delve deeper, we analyzed students' selections and trust levels under the experimental condition, but no significant differences were found. This could be attributed to the accuracy level of our deployed AI algorithm. While low-accuracy algorithms can be used to raise awareness of AI limitations, more research is needed to determine the ideal accuracy level that can challenge students' beliefs without promoting AI aversion.

To conclude, we remark on a few limitations that future researchers should take into consideration. The use of low-accuracy classifiers to challenge adolescents' trust in AI does not affect learning outcomes; however, we find that such classifiers need to be closer to adolescents' day-to-day activities to make them more critical of their outcomes. Also, accuracy levels should not be at the lowest to not promote AI aversion. Finally, the design of our study evaluated the low-accuracy detector through a mini-game interface. With no previous work having a similar design, we find that this could have potentially influenced the results and further work needs to be done to establish its accuracy. Lastly, our study had a low sample size.

This study's intervention effectively raised participants' awareness and critical thinking about social media AI algorithms, promoting responsible use of social media. Low-accuracy classifiers can be used as educational tools to provoke discussions about AI overdependence and trust among adolescents without impacting learning outcomes.

Acknowledgements This work has been partially funded by the Volkswagen Foundation (COURAGE project, no. 95567). TIDE-UPF also acknowledges the support by AEI/10.13039/ 501100011033 (PID2020-112584RB-C33, MDM-2015-0502), ICREA under the ICREA Academia programme (D. Hernández-Leo, Serra Hunter) and the Department of Research and Universities of the Government of Catalonia (SGR 00930).

References

1. Angwin, J., Larson, J., Mattu, S., Kirchner, L.: Machine Bias *. Auerbach Publications (3 2022). https://doi.org/10.1201/9781003278290-37
2. Banker, S., Khetani, S.: Algorithm overdependence: How the use of algorithmic recommendation systems can increase risks to consumer well-being. Journal of Public Policy and Marketing **38**, 500–515 (10 2019). https://doi.org/10.1177/0743915619858057
3. Burbach, L., Halbach, P., Ziefle, M., Calero-Valdez, A.: Bubble trouble: Strategies against filter bubbles in online social networks. In: Duffy, V.G. (ed.) Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Healthcare Applications. pp. 441–456. Springer International Publishing, Cham (2019)
4. Cai, C.J., Jongejan, J., Holbrook, J.: The effects of example-based explanations in a machine learning interface. International Conference on Intelligent User Interfaces, Proceedings IUI **Part F147615**, 258–262 (2019). https://doi.org/10.1145/3301275.3302289
5. Cataldo, I., Luca, I.D., Giorgetti, V., Cicconcelli, D., Bersani, F.S., Imperatori, C., Abdi, S., Negri, A., Esposito, G., Corazza, O.: Fitspiration on social media: Body-image and other psychopathological risks among young adults. a narrative review. Emerging Trends in Drugs, Addictions, and Health **1**, 100010 (2021). https://doi.org/10.1016/j.etdah.2021.100010
6. Coyne, S.M., Ward, L.M., Kroff, S.L., Davis, E.J., Holmgren, H.G., Jensen, A.C., Erickson, S.E., Essig, L.W.: Contributions of mainstream sexual media exposure to sexual attitudes, perceived peer norms, and sexual behavior: A meta-analysis. Journal of Adolescent Health **64**, 430–436 (4 2019). https://doi.org/10.1016/j.jadohealth.2018.11.016
7. Dantcheva, A., Bremond, F., Bilinski, P.: Show me your face and i will tell you your height, weight and body mass index. In: 2018 24th International Conference on Pattern Recognition (ICPR). pp. 3555–3560 (2018). https://doi.org/10.1109/ICPR.2018.8546159
8. Eg, R., Özlem Demirkol Tønnesen, Tennfjord, M.K.: A scoping review of personalized user experiences on social media: The interplay between algorithms and human factors. Computers in Human Behavior Reports **9**, 100253 (3 2023). https://doi.org/10.1016/j.chbr.2022.100253
9. Fernandes, B., Biswas, U.N., Tan-Mansukhani, R., Vallejo, A., Essau, C.A.: The impact of covid-19 lockdown on internet use and escapism in adolescents. Revista de Psicologia Clinica con Ninos y Adolescentes **7**, 59–65 (9 2020). https://doi.org/10.21134/RPCNA.2020.MON.2056
10. Fioravanti, G., Benucci, S.B., Ceragioli, G., Casale, S.: How the exposure to beauty ideals on social networking sites influences body image: A systematic review of experimental studies. Adolescent Research Review **7**, 419–458 (9 2022). https://doi.org/10.1007/s40894-022-00179-4
11. Gran, A.B., Booth, P., Bucher, T.: To be or not to be algorithm aware: a question of a new digital divide? Information, Communication and Society **24**, 1779–1796 (9 2021). https://doi.org/10.1080/1369118X.2020.1736124
12. Hamilton, K., Karahalios, K., Sandvig, C., Eslami, M.: A path to understanding the effects of algorithm awareness. In: CHI '14 Extended Abstracts on Human Factors in Computing Systems. p. 631–642. CHI EA '14, Association for Computing Machinery, New York, NY, USA (2014). https://doi.org/10.1145/2559206.2578883, https://doi.org/10.1145/2559206.2578883

13. Haritosh, A., Gupta, A., Chahal, E.S., Misra, A., Chandra, S.: A novel method to estimate height, weight and body mass index from face images. In: 2019 Twelfth International Conference on Contemporary Computing (IC3). pp. 1–6 (2019). https://doi.org/10.1109/IC3.2019.8844872
14. Harriger, J.A., Evans, J.A., Thompson, J.K., Tylka, T.L.: The dangers of the rabbit hole: Reflections on social media as a portal into a distorted world of edited bodies and eating disorder risk and the role of algorithms. Body Image **41**, 292–297 (2022). https://doi.org/https://doi.org/10.1016/j.bodyim.2022.03.007
15. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification (2015)
16. Hernández-Leo, D., Theophilou, E., Lobo, R., Sánchez-Reina, R., Ognibene, D.: Narrative scripts embedded in social media towards empowering digital and self-protection skills. In: De Laet, T., Klemke, R., Alario-Hoyos, C., Hilliger, I., Ortega-Arranz, A. (eds.) Technology-Enhanced Learning for a Free, Safe, and Sustainable World. pp. 394–398. Springer International Publishing, Cham (2021)
17. Kandlhofer, M., Steinbauer, G., Hirschmugl-Gaisch, S., Huber, P.: Artificial intelligence and computer science in education: From kindergarten to university. In: 2016 IEEE Frontiers in Education Conference (FIE). pp. 1–9 (2016). https://doi.org/10.1109/FIE.2016.7757570
18. Kozyreva, A., Lewandowsky, S., Hertwig, R.: Citizens versus the internet: Confronting digital challenges with cognitive tools. Psychological Science in the Public Interest **21**, 103–156 (12 2020). https://doi.org/10.1177/1529100620946707
19. Lazer, D.M.J., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S.A., Sunstein, C.R., Thorson, E.A., Watts, D.J., Zittrain, J.L.: The science of fake news. Science **359**(6380), 1094–1096 (2018). https://doi.org/10.1126/science.aao2998
20. Lee, E., Karimi, F., Wagner, C., Jo, H.H., Strohmaier, M., Galesic, M.: Homophily and minority-group size explain perception biases in social networks. Nature Human Behaviour **3**, 1078–1087 (8 2019). https://doi.org/10.1038/s41562-019-0677-4
21. Lee, I., Ali, S., Zhang, H., DiPaola, D., Breazeal, C.: Developing middle school students' ai literacy. In: Proceedings of the 52nd ACM Technical Symposium on Computer Science Education. p. 191–197. SIGCSE '21, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3408877.3432513, https://doi.org/10.1145/3408877.3432513
22. Lobo-Quintero, R., Sánchez-Reina, R., Theophilou, E., Hernández-Leo, D.: Intrinsic motivation for social media literacy, a look into the narrative scripts. In: Fulantelli, G., Burgos, D., Casalino, G., Cimitile, M., Lo Bosco, G., Taibi, D. (eds.) Higher Education Learning Methodologies and Technologies Online. pp. 419–432. Springer Nature Switzerland, Cham (2023)
23. Lomonaco, F., Ognibene, D., Trianni, V., Taibi, D.: A game-based educational experience to increase awareness about the threats of social media filter bubbles and echo chambers inspired by "wisdom of the crowd": preliminary results. In: 4th International Conference on Higher Education Learning Methodologies and Technologies Online (2022)
24. Mcknight, D.H., Carter, M., Thatcher, J.B., Clay, P.F.: Trust in a specific technology. ACM Transactions on Management Information Systems **2**, 1–25 (6 2011). https://doi.org/10.1145/1985347.1985353
25. McKnight, D.H., Choudhury, V., Kacmar, C.: Developing and validating trust measures for e-commerce: An integrative typology. Information Systems Research **13**, 334–359 (9 2002). https://doi.org/10.1287/isre.13.3.334.81

26. Ognibene, D., Donabauer, G., Theophilou, E., Buršić, S., Lomonaco, F., Wilkens, R., Hernández-Leo, D., Kruschwitz, U.: Moving beyond benchmarks and competitions: Towards addressing social media challenges in an educational context. Datenbank-Spektrum (2 2023). https://doi.org/10.1007/s13222-023-00436-3
27. Okkonen, J., Kotilainen, S.: Minors and Artificial Intelligence – Implications to Media Literacy, vol. 918. Springer Verlag (2019). https://doi.org/10.1007/978-3-030-11890-7_82
28. Pagano, T.P., Loureiro, R.B., Lisboa, F.V.N., Cruz, G.O.R., Peixoto, R.M., de Sousa Guimarães, G.A., dos Santos, L.L., Araujo, M.M., Cruz, M., de Oliveira, E.L.S., Winkler, I., Nascimento, E.G.S.: Bias and unfairness in machine learning models: a systematic literature review (2022)
29. Rodríguez-Rementería, A., Sanchez-Reina, R., Theophilou, E., Hernández-Leo, D.: Actitudes sobre la edición de imágenes en redes sociales y su etiquetado: un posible método preventivo (2022)
30. Serengil, S.I., Ozpinar, A.: Lightface: A hybrid deep face recognition framework. In: 2020 Innovations in Intelligent Systems and Applications Conference (ASYU). pp. 1–5 (2020). https://doi.org/10.1109/ASYU50717.2020.9259802
31. Serholt, S., Barendregt, W., Vasalou, A., Alves-Oliveira, P., Jones, A., Petisca, S., Paiva, A.: The case of classroom robots: teachers' deliberations on the ethical tensions. AI and Society **32**, 613–631 (11 2017). https://doi.org/10.1007/S00146-016-0667-2
32. Sherlock, M., Wagstaff, D.L.: Exploring the relationship between frequency of instagram use, exposure to idealized images, and psychological well-being in women. Psychology of Popular Media Culture **8**, 482–490 (10 2019). https://doi.org/10.1037/ppm0000182
33. Su, J., Ng, D.T.K., Chu, S.K.W.: Artificial intelligence (ai) literacy in early childhood education: The challenges and opportunities. Computers and Education: Artificial Intelligence **4**, 100124 (1 2023). https://doi.org/10.1016/j.caeai.2023.100124
34. Swart, J.: Experiencing algorithms: How young people understand, feel about, and engage with algorithmic news selection on social media. Social Media + Society **7**, 205630512110088 (4 2021). https://doi.org/10.1177/20563051211008828
35. Sánchez-Reina, J.R., Theophilou, E., Hernández-Leo, D., Medina-Bravo, P.: The power of beauty or the tyranny of algorithms. How do teens understand body image on instagram? Editorial Dykinson S.L (2021)
36. Valtonen, T., Tedre, M., Mäkitalo, K., Vartiainen, H.: Media literacy education in the age of machine learning. Journal of Media Literacy Education **11** (9 2019). https://doi.org/10.23860/JMLE-2019-11-2-2
37. Vogels, E.A., Gelles-Watnick, R., Massarat, N.: Teens, social media and technology 2022. Tech. rep., Pew Research Center (2022)
38. Warwick, K., Shah, H.: Can machines think? a report on turing test experiments at the royal society. Journal of Experimental and Theoretical Artificial Intelligence **28**, 989–1007 (11 2016). https://doi.org/10.1080/0952813X.2015.1055826
39. Zhang, G., Chong, L., Kotovsky, K., Cagan, J.: Trust in an ai versus a human teammate: The effects of teammate identity and performance on human-ai cooperation. Computers in Human Behavior **139**, 107536 (2 2023). https://doi.org/10.1016/j.chb.2022.107536
40. Žmavc, M., Šorgo, A., Gabrovec, B., Crnkovič, N., Cesar, K., Špela Selak: The protective role of resilience in the development of social media addiction in tertiary students and psychometric properties of the slovenian bergen social media addiction scale (bsmas). International journal of environmental research and public health **19**, 13178 (10 2022). https://doi.org/10.3390/ijerph192013178