Is This Fake or Credible? A Virtual Learning Companion Supporting the Judgment of Young Learners Facing Social Media Content

Farbod Aprin[1](✉), Sven Manske[1], Irene-Angelica Chounta[2], and H. Ulrich Hoppe[1]

[1] RIAS – Rhein-Ruhr-Institut für angewandte Systeminnovation e.V, Bürgerstr. 15, 47057 Duisburg, Germany
{fa,sm,uh}@rias-institute.eu

[2] Department of Computer Science and Applied Cognitive Science, University of Duisburg-Essen, 47057 Duisburg, Germany
irene-angelica.chounta@uni-due.de

Abstract. Today's young generation uses a variety of social media as part of their everyday life. This can be seen as an enrichment, but it is also the source of various threats: Filter bubbles and echo chambers are general phenomena that counter-balance free communication and exchange and can lead to toxic radicalization confirmation bias and polarization. Conspiracy theories, fake news are two phenomena that can change the social media user perspective about the facts. This paper reports on the development of a web-based learning environment that includes a dedicated "learning companion" to help students in raising their understanding, resilience, and critical thinking related to such social media threats. The point is not to protect young people from such threats through a type of censorship but to help them develop their own strategies to identify and counteract such influences. The web-based learning environment mimics Instagram as a well-known social medium. The actual companion is realized as a browser plugin with an underlying architecture that supports xAPI logging as well as the connection to intelligent backend components.

Keywords: Learning companion systems (LCS) · Intelligent tutoring system · Recommendation · Resilience · Increase awareness · Fake news · Social media · AI-based learning support systems · Chatbot · xAPI

1 Introduction

Besides positive aspects of social media like facilitating the exchange of information and opinions, young people have some general issues, mainly because of toxic content like cyberbullying, hate speech, fake news, and conspiracy theories. Cyberbullying has become increasingly common using social media channels like WhatsApp, especially among teenagers [1, 2]. Near half of the teachers stated that students face harassment

W. Zhou and Y. Mu (Eds.): ICWL 2021, LNCS 13103, pp. 52–60, 2021.
https://doi.org/10.1007/978-3-030-90785-3_5

through their mobile phones and the Internet, and some teachers were themselves victims of cyberbullying [3].

Fake news is presented as real news, and because dis/misinformation, and even blatant propaganda continue to spread, it can undoubtedly distort one's worldview. The most common fake news narratives were more broadly shared on Facebook than the most popular mainstream news stories. Most US Americans who see fake news stories report that they believe them [4]. Coupled with "deep fake" videos and other manipulated images and audio files, the truth is not so easy to identify independent of user's age groups.

Other issues originate from the social media companies' goal to retain users on their platforms. To increase user engagement social platforms present and filter the posts that are in line with prevailing opinions to draw them more into their platform [5]. Eli Parisier calls this phenomenon "filter bubble" [6]. Parisier explains that recommendation algorithms may create a situation where users increasingly receive information that confirms their previous beliefs and leaves no space for doubt. The "echo chamber" phenomenon is a similar issue in social media where users only face opinions that indicate and reinforce their argument, leading to distancing themselves from contradicting beliefs and allowing extreme views to be amplified [7]. As a result, users become isolated from information that opposes their perspectives.

The spiral of silence theory indicates that if people realize that their private opinion differs from the perceived majority, they are less likely to comment publicly [8]. Although the spiral of silence theory has been discussed controversially for traditional media, there is some evidence for the effects in the context of social media [9].

A serious mini-game platform and information booklet have already been prepared and distributed to help teenagers critically reflect on digital advertising by Media Awareness Network (MNet) [10]. MNet made available various free games for teenage users, some of which are specifically aimed at petitioning the question of cyberhate through exploring bias and prejudice and supporting critical thinking skills. The aim is to encourage teenagers to check information and look for alternative viewpoints to increase their awareness [11]. In the Reality Check game[1], students will discover how to find evidence, like finding where a story initially came from and matching it to other sources, plus, how to use tools like fact-checking websites and reverse image searches.

There are two main strategies to control toxic content in social media. The first one is to detect and eliminate postings of poisonous content ("censorship"). This strategy is currently applied in social media. On the other hand, some threats like hate speech are laborious to track and control their distribution since nuances in cultures and languages make it hard to present a well-defined distinction between hate speech and dangerous speeches. The detection is not an easy task for social media providers [6]. The second strategy is to help users develop their own mechanisms to identify and counteract such influences.

[1] Reality Check: The Game https://mediasmarts.ca/digital-media-literacy/educational-games/reality-check-game, 2021.

In this paper, we present an approach that aims to increase the resilience and awareness of social media users within the European project Courage[2] funded by Volkswagenstiftung.

The Courage is characterized by providing games and an educational platform to simulate social media environments with inherent challenges. In this paper we report on the development of a virtual companion that supports and educates young learners in the prepared environment. The main focus is on strengthening self-protection through analytical/critical thinking, empathy, and ultimately resilience rather than providing external protection through censorship.

2 Learning Companion Systems and ITS

As a specific version of intelligent tutoring systems (ITS), learning companion systems (LCS) personalize the support and adaptive feedback through an explicit, and possibly human-like agent interacting with the learner [12]. The agent or companion guides the learner step by step and typically adopting a non-authoritative role. The interface may be composed of multimedia, interactive buttons, menus, text, voice, animation, diagrams, virtual reality, or other interactive techniques.

LCS interfaces usually include natural language processing (NLP) to facilitate communication between LCS and the student. Tracing the student's interactions with an LCS is used as a part of student modeling. For example, the LCS may ask the student to explain the reasons behind their answers as a reflective question for each step during their task, which might consequently lead to more robust learning, in line with work on self-explanation at the time. Learners might generate many explanations and articulate the reasons behind their answers that refine their understanding (self-regulated learning strategies) [13].

According to the definition of an adaptive system, one of the key points for a tutoring system that includes LCS to be adaptive is to respond to learner actions flexibly depending on the context and history. This response can be implemented in the conversation as feedback to the user textual input or give proper response to the user interaction in various ways, like suggesting a multimedia instruction as a recommendation [14]. Student modeling as a basis for adaptive feedback in LCS tutorial dialogues can significantly increase learning gains in low and high prior knowledge students [15]. A LCS can play many roles in an instructional context. For instance, the role of an executive, suggesting new ideas for the learners to consider, or a critic, challenging the learner's proposals [16].

For answering the question "how knowledgeable should the companion agent be to reach the learner's expectation and motivate the student to continue collaborating with the agent?", Hietala and Niemirepo [17] found that the learners lose their motivation if they use a strong and knowledgeable companion all the time. Especially in the beginning, a companion that makes mistakes like humans is more effective. Still, for a challenging task or for dealing with a new issue, both introvert and extrovert learners prefer knowledgeable and robust learning companions.

[2] A social Media Companion Safeguarding and Education Students couragecompanion.eu.

Our virtual learning companion (VLC) supports features such as role-playing for the users, gives adaptive feedback based on previous user interaction, judges the answers, and asks knowledge activation questions. In addition, providing analysis and recommendations, comprising receiving input and displaying the information, are core features of the VLC system.

3 System Environment and Implementation

3.1 Conceptual Architecture

To meet the objectives to help learners develop strategies to counteract the negative influences social media, we provide a virtual companion that will work with two different environments. First, in a controlled social network environment, where the scenario controls the content and reaction of the participants. Second, the companion should also provide basic support in real, open environments like Instagram as an example that is frequently used by the target age group.

Currently, the interaction is situated in a controlled environment that is implemented using the PixelFed[3] framework. PixelFed is an open-source social media network, like Instagram, that is suitable for multimedia. Our PixelFed version has been enhanced with a logging component to archive user actions for analytical purposes.

In conjunction with PixelFed, the VLC is developed as a Chrome browser plugin. It intervenes with questions and suggestions while the student interacts with the environment's artifacts. Learners will interact with the social media environment guided by the tasks they receive from the companion as a chatbot (cf. Fig. 1).

To prepare and enable such scenarios, educational designers identify example cases and materials by creating new cases from harvested real-world examples. We designed a companion scenario for the controlled environment that includes fake news and conspiracy theory samples mostly related to the pandemic for the first trials.

3.2 PixelFed and Browser Plugin

To create a companion that can interact with simulated social media independently, we developed a Chrome extension as a plugin to detect the user's interaction with the social media environment. The closed environment is an Instagram-like environment enriched with images and their captions. The user can right-click on images or highlight the image caption to send them to the Chrome extension and activate the companion to respond to it.

As depicted in Fig. 2, the companion has two main sections: first, a frontend with a Chrome extension communicating with the PixelFed environment; second, backend microservices with a middleware mediating the frontend communication. WIT.AI[4] in the backend is responsible for managing the companion conversation via chatbot and detecting the user's intent for each conversation step. In the frontend, ReactJS[5] allows

[3] PixelFed is a decentralized open-source Instagram-like photo-sharing network based on the Activity Pub protocol, making it suitable for applying it in research experiments, https://pixelfed.social/. (2021).

[4] WIT-AI is a tool to Build Natural Language Experiences https://wit.ai, 2021.

[5] React JS is A JavaScript library for building user interfaces https://reactjs.org, 2021.

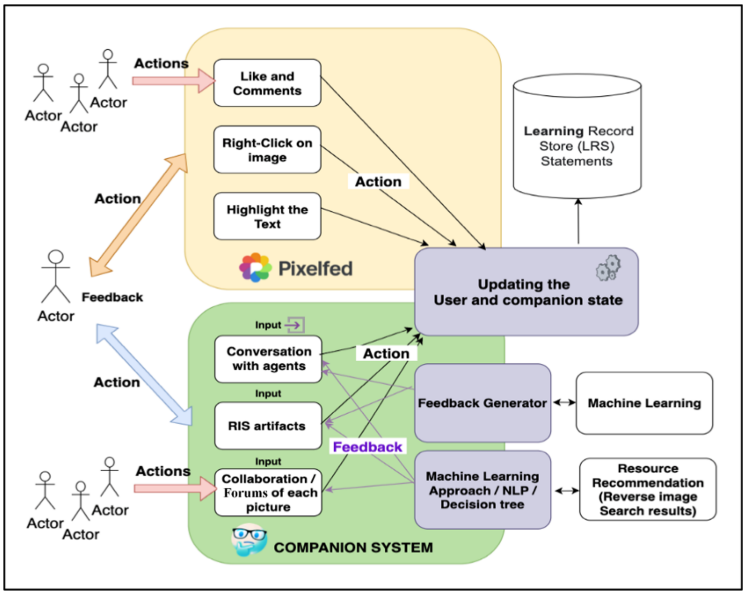

Fig. 1. User Interaction with companion and controlled environment architecture (focused on a single actor).

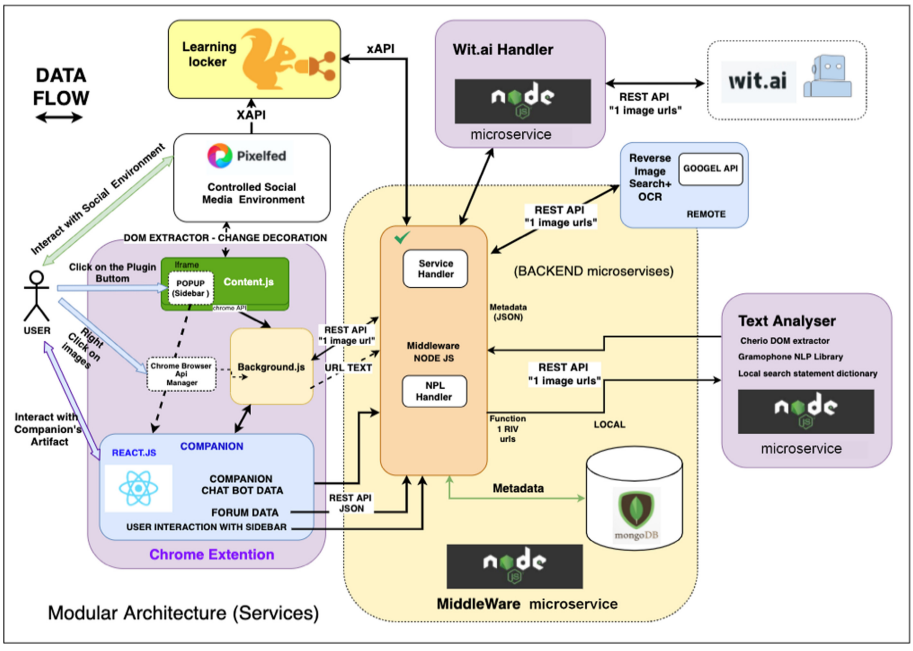

Fig. 2. Dataflow in technical architecture of the companion for controlled environment.

for a modern, responsive web design with flexible and reusable UI components in the standard structure of Chrome extension[6].

The image-link is accessed from the frontend via a REST API to the backend middleware (NodeJS express[7]) modules. The reverse image search (RIS) API, applied to find similar images the user interacts with the environment from the external resources. Via, text analyzer module, we can extract the semantic keywords and the important section of text behind each RIS-link associated with retrieved extracted semantic keywords. We create and enrich the metadata for each image and store it in the document-oriented database (MongoDB[8]). We use Learning Locker[9] as Learning Record Store (LRS) to store action logs in the standard Experience API (xAPI) format.

4 Example Scenario

The experimental trials are carried out on desktop computers running Chrome browsers. The user has access to PixelFed environments enriched with limited numbers of images and captions containing fake/fact news. Also, the setup includes the Chrome extension that represents the virtual learning companion system in the browser.

4.1 Example Walkthrough

When users visit the PixelFed environment via a provided web link they will first see a short guide on how to open and use the browser plugin (Courage Chrome extension). After that, they log in to the companion system by clicking on the provided anonymous user-token link. Next, they will see a user-guide video on the sidebar equipped with the virtual tutor that explains how to interact with the environment artifacts to activate the companion.

The companion starts to communicate automatically with a chatbot-style conversation (e.g., "How are you?") and asks some general introductory questions (sex, age). After a few steps, the companion will ask to right-click on one of the images followed by knowledge activation questions and asks for the user's opinion about the selected image in the social media environment. After the response, the learner needs to answer some reflective questions that explain why they chose that specific answer. In the conversation, the companion asks if the user thinks it is fake news or fact, considering an image in social media and the caption text behind it. The user is limited to choose one of the provided choices. Then, the companion applies to stimulate reflection questions (e.g. "How sure are you?") according to a predefined decision tree.

Next, the companion will unlock a "Recommended" tab that contains reverse image search (RIS) links from the web. In these places, the same image is shown in a different context. The learners can visit these links and compare the keywords, metadata and

[6] Standard Chrome extension consists of three main modules: 1. Background and content script – https://developer.Chrome.com/docs/extensions/mv3/getstarted, Manifest file 2021.

[7] NodeJS is a JavaScript runtime built on Chrome's V8 JavaScript engine, 2021.

[8] MongoDB, document-based, distributed database built for modern application, 2021.

[9] Learning-Locker used to store learning activity statements generated by xAPI compliant learning activitieshttps://docs.learninglocker.net/welcome/, 2021.

abstract texts of each RIS link for the selected post. In this step, the learner can freely continue the instructions and come back to the conversation. Then, the user writes down their opinion as demanded in the chat and answers. Then chatbot will give new (adaptive) feedback. Next, the system will unlock the collaboration chart option Analysis Tab for the selected artifact. After interacting with metadata for around each image, the companion can give an alert or feedback as a pedagogical intervention. For example, if they judge or answer promptly in a short time without checking the recommended RIS or had poor participation in collaboration sections in comments.

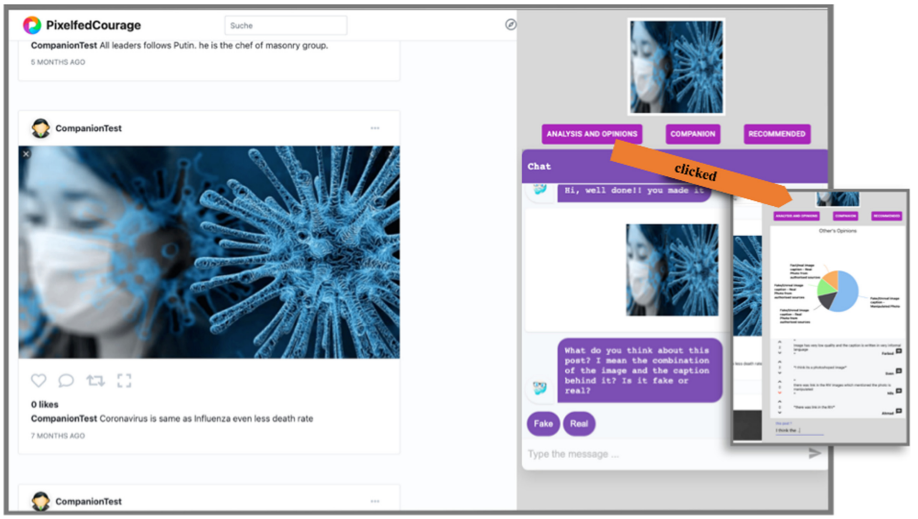

Fig. 3. Left: right-click on a Picture, then the companion will activate to react (ask their opinion about the selected picture). Right: companion unlocks the collaboration tab and ask if the user changes their mind after observing comments and others votes in the diagram.

Figure 3 shows how learners can see others' opinions in a collaborative forum and compare the answers. The best answers according to the user's votes goes to the top. For example, if the image is manipulated (fake) or real news. In the provided forum, the learner can find a better link from the comments. Via a pie chart, users can compare the answers of others to themselves. After observation of recommended metadata, the companion asks if learners would like to reconsider their judgments.

As researchers, it is interesting to know if learners change their minds after each step or if their direction of thinking and judgment changes after receiving feedback from the companion and interacting with recommended learning instructions? Therefore, we log all-important user interactions along with time stamps. Thus, we need the log for the answers, companion chats responses, interactions with an environment's artifacts.

One possible pedagogical intervention of the companion could be to warn or encourage learners. For example, the artifact has a clear orientation and status, and the user's initial opinion changes in contrast with their first judgment, when the learner observes that the majority has another mindset. In this case, the companion warns the users

about the "spiral of silence" phenomenon and encourages them to think again, and the companion present a small tutorial video (pedagogical intervention).

In the first scenario that we consider applying for the following school trials, we will select four pictures for the PixelFed environment related to pandemic situations. Pictures may include fake news and conspiracy theories, used to teach critical thinking as a skill or to increase their awareness. At the end of interacting with four pictures, the companion will ask the overall user's opinion and feedback.

4.2 Preliminary User Study

A first informal study was carried out with a small group of participants confronted with a controversial report on the coronavirus that included a picture shared on social media. The participants had to complete the classification "fake" or "real" for this item based on their assessment of credibility. After that, they wrote a few lines about their reflection on this task. The goal of the study was to check how users understood and handled the task and to identify possible misunderstandings. Fourteen participants aged 19–35 with varying cultural backgrounds contributed actively to the study. The education level ranged from bachelor students to postdoc.

As a result of the study, we found that more than half of the participants mentioned that they were unsure about their answers and would have preferred a broader range of answers reflecting levels of uncertainty, such as "probably fake," "not sure," "probably fact" in addition to just "fake" and "fact". Only three of the participants used external fact check tools or search engines to determine the truth. None of the participants was aware of Reverse Image Search tools to find a similar topic which is considered in the VLC tool and scenario. We envision that a real-time VLC can support learners in making more correct judgments and better usage of web resources.

5 Conclusion and Outlook

This paper presents our work to design and implement a virtual learning companion in a simulated social media environment. This companion assists learners in coping with the threats and toxic content of social media like Fake news. The companion is implemented as a browser plugin that can be added to Instagram-like social media environments for orchestrating the controlled tests. This companion interacts with learners via chat and triggers the user's action with the artifacts of the environment. The presented companion utilizes interactive multimedia to give learners feedback and intervene with a pedagogical reflection. The companion encourages the learners and provides knowledge according to a pre-defined decision tree.

In the next version, the companion will work with artifacts on open social media platforms such as Instagram, possibly repeated in the closed environment (with identical artifacts). Via interaction with the same metadata for the repeated artifact, the expert and other users can indicate their opinion around the unique conversion about controversial content and represent their different mindsets and backgrounds. This feature will give the users the power to view other parties' opinions without biased pre-filtering.

References

1. Sprugnoli, R., Menini, S., Tonelli, S., Oncini, F., Piras, E.: Creating a Whatsapp dataset to study pre-teen cyberbullying. In: Proceedings of the 2nd Workshop on Abusive Language Online (ALW2), June 2019, pp. 51–59 (2019). https://doi.org/10.18653/v1/w18-5107
2. Aizenkot, D., Kashy-Rosenbaum, G.: Cyberbullying in WhatsApp classmates' groups: evaluation of an intervention program implemented in Israeli elementary and middle schools. New Media Soc. **20**(12), 4709–4727 (2018). https://doi.org/10.1177/1461444818782702
3. Eden, S., Heiman, T., Olenik-Shemesh, D.: Teachers' perceptions, beliefs and concerns about cyberbullying. Br. J. Edu. Technol. **44**(6), 1036–1052 (2013). https://doi.org/10.1111/j.1467-8535.2012.01363.x
4. Silverman, C., Singer-Vine, J.: Most Americans who see fake news believe it, new survey says. BuzzFeed (2016)
5. Bucher, T.: The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. Inf. Commun. Soc. **20**(1), 30–44 (2017). https://doi.org/10.1080/1369118X.2016.1154086
6. Rowland, F.: The filter bubble: what the internet is hiding from you. Portal: Lib. Acad. **11**(4), 1009–1011 (2011)
7. Colleoni, E., Rozza, A., Arvidsson, A.: Echo chamber or public sphere? predicting political orientation and measuring political homophily in twitter using big data. J. commun. **64**(2), 317–332 (2014). https://doi.org/10.1111/jcom.12084.
8. Schmidt, R.E.: Legacy of Elisabeth Noelle-Neumann: The spiral of silence and other controversies. Eur. J. Commun. **30**(3), 353–363, https://doi.org/10.1177/0267323115589265
9. Ross, B., Pilz, L., Cabrera, B., Brachten, F., Neubaum, G., Stieglitz, S.: Are social bots a real threat? an agent-based model of the spiral of silence to analyse the impact of manipulative actors in social networks. Eur. J. Inf. Syst. **28**(4), 394–412 (2019). https://doi.org/10.1080/0960085X.2018.1560920
10. De Jans, S., Hudders, L., Herrewijn, L., Van Geit, K., Cauberghe, V.: Serious games going beyond the call of duty: Impact of an advertising literacy mini-game platform on adolescents' motivational outcomes through user experiences and learning outcomes. Cyberpsychology **13**(2) (2019). https://doi.org/10.5817/CP2019-2-3.
11. Titley, G., Keen, E., Földi, L.: Three studies about online hate speech and ways to address it (2014)
12. Chou, C.Y., Chan, T.W., Lin, C.J.: Redefining the learning companion: the past, present, and future of educational agents. Comput. Educ. **40**(3), 255–269 (2003). https://doi.org/10.1016/S0360-1315(02)00130-6
13. Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P., Glaser, R.: Self-explanations: how students study and use examples in learning to solve problems. Cogn. Sci. **13**(2), 145–182 (1989). https://doi.org/10.1016/0364-0213(89)90002-5
14. Aleven, V., Beal, C.R., Graesser, A.C.: Introduction to the special issue on advanced learning technologies. J. Educ. Psychol. **105**(4), 929–931 (2013). https://doi.org/10.1037/a0034155
15. Katz, S., Albacete, P., Chounta, I.A., Jordan, P., Mclaren, B.M., Zapata-Rivera, D.: Linking dialogue with student modelling to create an adaptive tutoring system for conceptual physics. Int. J. Artif. Intell. Educ. **31**(3), 1–49 (2021)
16. Goodman, B., Linton, F., Gaimari, R.: Encouraging student reflection and articulation using a learning companion: a commentary. Int. J. Artif. Intell. Educ. **26**(1), 474–488 (2015). https://doi.org/10.1007/s40593-015-0041-4
17. Hietala, P., Niemirepo, T.: The competence of learning companion agents. Int. J. Artif. Intell. Educ. (IJAIED) **9**, 178–192 (1998)