The Eurasia Proceedings of
Educational & Social Sciences (EPESS)
ISSN: 2587-1730

**The Eurasia Proceedings of Educational & Social Sciences (EPESS), 2023**

**Volume 33, Pages 26-33**

**IConSE 2023: International Conference on Science and Education**

# Increasing the Effectiveness of Fake News Detection: An Educational Program for High School Students Using Interactive Neural Network Training and Collective Intelligence

**Rafał Olszowski**
AGH University of Krakow

**Abstract**: This paper presents the project results designed to provide high school students with essential ICT tools to identify and counteract fake news and disinformation commonly found on the Internet, especially on platforms like X/Twitter. Additionally, it introduces an educational program that utilizes interactive neural network training and collective intelligence to combat fake news. For this project, there was developed an IT framework enabling the collective training of a specialized neural network. In order to conduct a quasi-experiment, we engaged three research groups of high school students, each containing app. 10-15 members. Through a set of comprehensive workshops, the students were trained to recognize harmful online content. After this training, students actively participated in data classification on various topics, laying the foundation for the neural network's training model. The presented results underscore the efficacy of this immersive method in imparting digital literacy and enhancing the group intelligence. Moreover, the results highlight the promising potential of machine learning in assisting youth to navigate the complex digital terrain safely and responsibly. The final phase of the conducted research involved testing the trained neural network in detecting disinformation, particularly in the topics of 5G technologies and immigration problems in Poland.

**Keywords:** Fake news, Disinformation, Education, Neural networks, Collective intelligence, High school

## Introduction

The digital era, while offering numerous benefits, has given rise to a complex problem: the proliferation of disinformation and fake news (FN). This phenomenon has garnered international concern due to its potential for widespread manipulation and its detrimental effects, particularly across social media platforms. One of the most daunting tasks in today's digital public sphere is the filtration of 'information noise' to extract meaningful content from the cacophony of online discourse. The pervasive nature of such noise poses a dire threat to the integrity of our information ecosystems, undermining the ability of users to differentiate between reality and falsehood. High school students represent a demographic particularly susceptible to these pitfalls, often lacking the necessary critical acumen to navigate the complexities of online information. This research paper introduces an innovative initiative designed to equip students with the requisite competencies and tools to navigate and counteract the tide of disinformation prevalent in the online environment.

The educational part of the project was carried out by a Polish non-governmental organization – Instytut Aurea Libertas from April to December 2022. Concurrently, the research aspect was conducted by social scientists affiliated with AGH University of Science and Technology in Krakow, taking place in June and October of the same year. The project explored a novel methodology that harnessed artificial intelligence (AI)—notably neural networks—in synergy with Collective Intelligence (CI). The efficacy of this approach in mitigating the spread of fake news was evaluated via a quasi-experimental design. High school students were engaged in the collective training of a neural network to identify and filter deleterious content on social media. The initiative concentrated on two prevalent topics of misinformation: spurious claims associated with 5G technology and negative stereotypes concerning Ukrainian migrants in Poland. The objective was to bolster the digital literacy

of students and to establish a replicable model for broader educational application. The approach demonstrated its value, achieving marked improvements in the participants' knowledge and competencies in identifying fake news and enhancing the overall efficacy in countering disinformation.

## Literature Review

The escalating spread of fake news (FN) has emerged as a critical issue in recent years, gaining particular urgency amidst the Covid-19 pandemic, social upheaval, and armed conflicts globally. Fake news is characterized by Allcott and Gentzkow (2017) as *a news item that is deliberately and verifiably false*. Similarly, Kshetri and Voas (2017) define it as any media message disseminated through media channels that contains false information, regardless of the methods and motivations behind it. The proliferation of fake news poses a significant threat to free speech and the integrity of ethical journalism, eroding public trust in institutions. Moreover, the rise and ubiquity of social media platforms have dramatically amplified the spread of false information.

Fake news frequently targets topics crucial to public debate, thereby influencing the formation of social and political opinions among voters, opinion leaders, and politicians. Disinformers employ half-truths, edited videos, and selective reporting to propel their narratives, rendering fact-checking occasionally ineffective. Here are some subjects that have been particularly vulnerable to fake news in recent years:

1. **Health and disease** (e.g., coronavirus pandemic, vaccines): Misinformation proliferated during the pandemic, with claims of unconventional cures and prevention methods. Additionally, numerous false assertions regarding COVID-19 vaccines, such as altering DNA or containing tracking microchips, have been disseminated (Carmichael & Goodman, 2020).
2. **Street riots and social unrest**: Fabricated images and videos, sometimes repurposed or AI-generated, have been used to exaggerate chaos and violence in protests and riots, often discrediting the actual events and participants (Bahl, 2023; Lee, 2020).
3. **Public figures in controversial situations**: Technologies like deepfakes create misleading images or videos of public figures in fabricated scenarios, leading to public deception. An instance includes manipulated images of former US President Donald Trump in fictitious legal troubles (Aleguas, 2023).
4. **Conspiracy theories**, both old and new, have flourished in the age of social media, encompassing various topics like 5G technology being accused of mind-control intentions (Ahmed et al., 2020).
5. **Military conflicts**, such as the war in Ukraine, have been marred by the spread of disinformation and fake news, with conflicting parties making unverified accusations and spreading falsehoods about events like the Bucha massacre (Marchant de Abreu, 2022).

Empirical evidence indicates that exposure to high levels of information noise can diminish an individual's ability to effectively identify relevant and accurate information (Bessi et al., 2016; Vosoughi et al.,2018). Studies in social psychology and communication reveal that humans tend to act irrationally and struggle to distinguish between truth and falsehood when overwhelmed by a surplus of misleading information. Such studies suggest that, under conditions of cognitive overload and information noise, human accuracy in detecting deception is only slightly better than random guessing, with accuracy rates typically ranging from 55% to 58% (Zhou & Zafarani , 2020).

A review of the literature on current trends in combating FN shows that various detection techniques are evolving independently. For instance, technology that analyzes the language used in FN, including vocabulary and syntax analysis, has garnered considerable attention. Meanwhile, the study of FN propagation patterns on social media forms an entirely separate domain. Only recently have there been efforts in academic literature to systematize various FN combat strategies, acknowledging their potential for mutual reinforcement. The most noteworthy attempts, in my view, were made by Zhou and Zafrani in *A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities* (Zhou & Zafrani, 2020), and by Collins and his team in *Trends in combating fake news on social media – a survey* (Collins et al., 2020).

In Zhou and Zafrani's work, authors divided the methods for combatting fake news into four main groups: knowledge-based (manual and automated fact-checking), style-based (analysis of text from lexical, syntactical, sentiment, and rhetorical perspectives), propagation-based (examination of how users distribute fake news), and source-based analyses (evaluation of the reliability of the news source). Fact-checking, often employed in the knowledge-based approach, involves the comparison of the information derived from the news content that is

being verified (like its claims or statements) with known facts. Style-based techniques, conversely, evaluate the intention of the news, looking for distinguishable patterns in the content (text and images) of fake news versus real news. The propagation-based method follows the trail of how fake news is disseminated by internet users. Lastly, source credibility analysis determines the trustworthiness of those involved in creating and distributing the content, necessitating the establishment of standards for reliable and unreliable authors or publishers (Zhou & Zafrani, 2020).

In the second paper by Collins and his team, they present a comprehensive review of various fake news detection models. These models include techniques such as expert fact-checker approach, natural language processing, machine learning, recommendation system, deep learning, graph-based methods, and crowdsourcing (Collins et al., 2020). Collins et al. suggest that future anti-fake news systems could benefit from a hybrid model, which integrates two or more techniques, asserting that such a system can achieve an impressive detection accuracy rate of 87% (Collins et al., 2020).

Thus far, no model for fake news detection has been developed that principally utilizes the concept of Collective Intelligence (CI) as its operative mechanism. Collective intelligence is defined as the widespread capacity of a group to address problems, primarily through the aggregation of data, ideation, and decision-making processes. This capacity arises from the synergistic interaction and competition among numerous individuals (Woolley, 2010). In recent years, the CI phenomenon has garnered considerable interest from researchers, with collectives engaging in generating potential solutions, evaluating and refining them, synthesizing and organizing knowledge and insights, and ultimately arriving at joint decisions. While Shabani and Sokhn (2018) made some efforts to leverage CI in the detection of fake news, these initiatives did not achieve optimal results due to the collective's limited engagement in content classification.

## Method

The methodology implemented in this project was multifaceted, encompassing: (1) the technical and substantive development of the educational program and IT tools, and (2) the execution of a quasi-experiment with high school students.

The project's educational materials included three workshop scenarios designed to enhance the detection of fake news in topics particularly vulnerable to misinformation during the study period, video tutorials, IT software for the collective training of a neural network, and IT software for identifying fake news on social media.

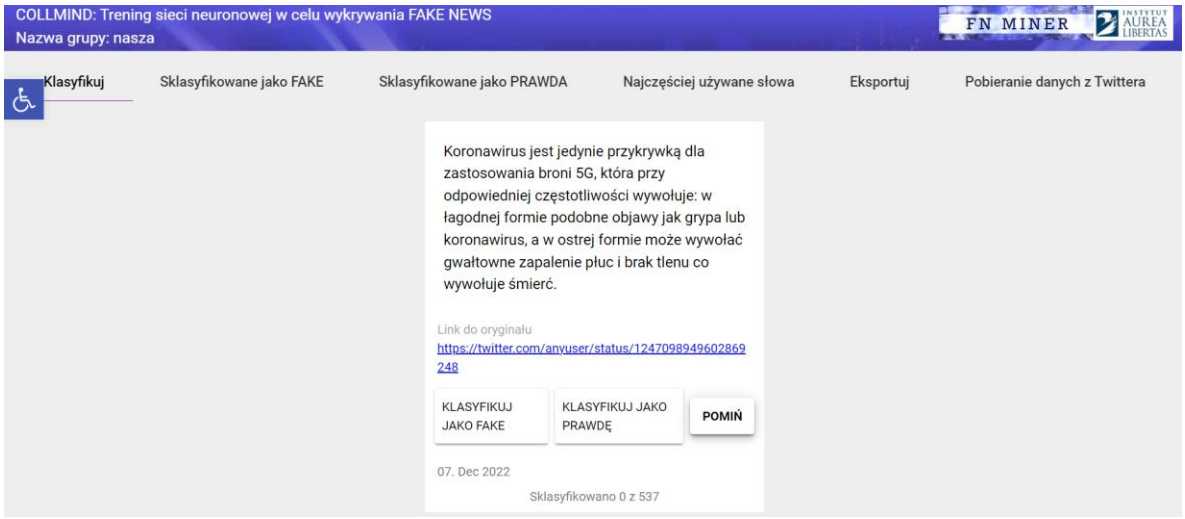



Figure 1. The software used to train the neural network in the project: A classification screen.

The project commenced with a series of interactive workshops aimed at educating participants on the nuances of harmful online content, with a particular emphasis on three critical topics of disinformation. These sessions were led by experts in digital literacy, media studies, and artificial intelligence, ensuring a multifaceted understanding of the issues. The objective was to furnish students with the ability to identify and comprehend the ramifications of online disinformation.

At the core of the initiative was the development of an IT framework facilitating the collective training of a neural network. The neural network training software was developed using Node.js version 6.2.10, React, Polymer.js, JavaScript, HTML, and CSS. Compilation and management of sources were handled through NPM, PM2 was utilized for backend server operations, and PostgreSQL served as the database platform. For fake news detection, the software was created within the Orange Data Mining suite, employing a script that utilized text processing tools (such as lowercase transformation, HTML parsing, URL removal, tweet tokenization, regular expression filtering) and machine learning models (including neural networks and Naive Bayes). The platform was intentionally designed for ease of use, enabling students to efficiently classify data on the chosen topics. This classified data was instrumental in training the neural network to identify and mark disinformation pertinent to the specified subjects.

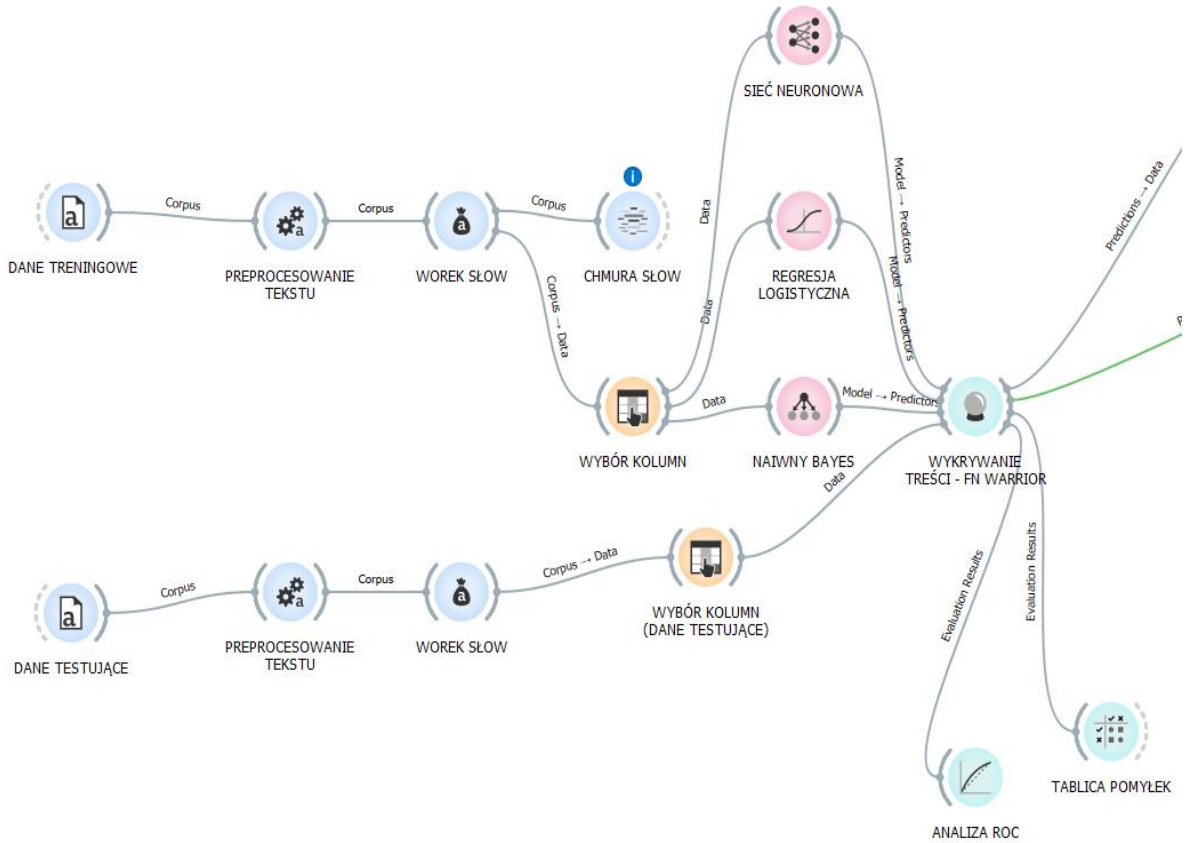



Figure 2. The script used to detect fake news prepared in the Orange Data Mining environment: the main workflow.

The quasi-experiment involved the recruitment of three research groups from high school students, each comprising approximately 10-15 members. The study included conducting educational workshops, facilitating data classification by the students, and evaluating the effectiveness of the trained neural network. Recruitment took place at two secondary schools in Krakow, Poland, with the study being conducted in June and October 2022. The diverse backgrounds of the recruited students enriched the data classification process. Selection criteria were based on their interest in the project, frequent use of social media, and eagerness to learn new techniques.

The main task of the recruited groups was to conduct workshop scenarios concerning detecting fake news in the field of 5G technology and stereotypes regarding immigration. The dataset for classification comprised approximately 1,000 posts sourced from the Twitter platform, which were collected via software interfaced with Twitter's API and selected based on specified keywords.

It is important to note the variation in classification categories adopted for the two distinct topics due to their inherent specificities. The topic of 5G networks falls within the broad realm of business, albeit with significant social implications. For this topic, the classification categories were defined as: (a) fake news and (b) true information. In contrast, the topic of immigrants pertains to public affairs discourse, and the classification categories were designed to identify: (a) negative, harmful stereotypes indicative of disinformation, and (b) the

absence of such stereotypes. During the study, the following research tasks were undertaken: (1) verifying the effectiveness of the proposed educational method; (2) conducting qualitative research through individual in-depth interviews with participating students; and (3) assessing the effectiveness of collective effort in detecting fake news using the provided software, according to the tested scenario.

## Results and Discussion

The project produced highly encouraging outcomes. The participating high school students showed an increased sensitivity towards the issue of disinformation and its various manifestations across social media platforms. This new-found awareness was evident in their ability to discern between factual and misleading content, which was a significant aim of the project. They demonstrated a deep comprehension of the selected disinformation topics i.e. fake news concerning 5G technology and stereotypes about Ukrainian migrants in Poland. This understanding was manifested in the precision of their data classification efforts. Each student, having been exposed to these topics during the training workshops, was able to accurately classify data which was then utilized to train the neural network. This showed their ability to translate theoretical knowledge into practical application.

### The Effectiveness of the Proposed Educational Method

The effectiveness of the proposed method was examined in two ways. The first was an evaluation of the acquired knowledge and competences in detecting fake news, carried out among students participating in the quasi-experiment. This was done using a Pre-Test conducted before the workshop, and a Post-Test carried out at the end of the work. The aim of the study was to check how much students improved their competences in recognizing fake news. The tests were anonymous, the answers below are given in random order. The test result was highly satisfactory: an average increase of 39.6% in the competencies of the participating students was observed**.** Below, Table No. 1 presents the individual increase in competences in Group No. 2, as an example.

Table 1. Increase in the individual competences in recognizing fake news in Experimental Group No. 2.

| No. | PRE-TEST result (points, max = 10) | POST-TEST result (points, max = 10) | Increase of Competences |
|---|---|---|---|
| 1 | 6 | 7 | 14% |
| 2 | 6 | 7 | 14% |
| 3 | 5,5 | 7,5 | 26,6% |
| 4 | 5 | 8 | 37,5% |
| 5 | 5 | 8 | 37,5% |
| 6 | 4,5 | 8 | 43,7% |
| 7 | 4,5 | 8 | 43,7% |
| 8 | 4,5 | 8,5 | 47% |
| 9 | 3,5 | 8,5 | 58,8% |
| 10 | 3,5 | 9 | 61,1% |

### Qualitative Research Based on Individual In-depth Interviews

The qualitative study was carried out based on Individual In-depth Interviews conducted with students participating in the project. Each interview lasted about 30 minutes and concerned young people's opinions and conclusions regarding participation in the study and awareness of the threats related to the spread of fake news on the Internet.

During the interviews, a notable shift in the participants' awareness was observed due to interactions within the group. Collaborating on content classification to train the neural network necessitated the confrontation of various opinions and viewpoints, requiring mutual listening to reach a consensus. The change in participants' awareness revolved around a better understanding of the complexities associated with fake news (FN) and disinformation, recognizing that seemingly straightforward statements often have underlying meanings and are open to multiple interpretations. According to the majority of participants, a better understanding of the issues with FN and disinformation resulted in enhanced group work effectiveness and, consequently, an increase in the group's collective intelligence in ICT-supported fake news identification.

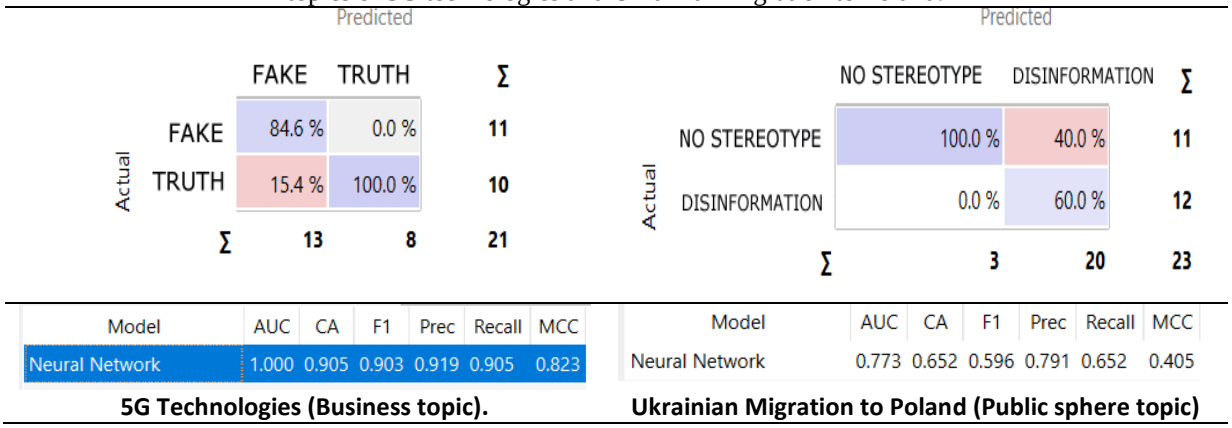
Additionally, a sense of collective effort towards a common good and achieving a common goal was cultivated through engagement in opposing harmful online content. Observing the outcomes of collective efforts, such as the trained neural network reflecting the group's collective opinion, was an intriguing experience for participants. It fostered a sense of agency and awareness that collectively they could achieve more than individually. About 80% of the project participants felt that working collectively to address the problem brought them satisfaction from the achieved results, which they couldn't have accomplished individually. The synergy effect attained through group work allowed for the realization of a result beneficial for all, epitomizing the product of collective intelligence.

**The Effectiveness of Collective Content Classification: Measuring the Level of Fake News Detection**

The final phase of the conducted research involved testing the trained neural network in detecting disinformation. For testing, we chose two topics that were particularly susceptible to the spread of fake news during the period when the project was conducted. The first one was related to 5G technologies, which was at that time a business-related topic particularly susceptible to the spread of conspiracy theories. The second one concerned disinformation and negative stereotypes related to the increased immigration of Ukrainian citizens to Poland. As a matter belonging to public sphere, it was a hot topic because of the war and social migration from Ukraine to Poland in 2022. In order to conduct the effectiveness evaluation, two sets of 21-23 social media posts were extracted from classified databases and set as testing data. Then a performance test was carried out.

Despite these difficulties, the neural network was capable of identifying and flagging disinformation related to the chosen topics effectively. It performed with a considerable degree of accuracy, suggesting that the collective training approach adopted in this project was successful. The students were not only able to train the network but also helped in refining it to function optimally.

Table 2. Confusion matrices and performance scores for the tested neural networks detecting fake news on the topics of 5G technologies and Ukrainian migration to Poland.




| Model | AUC | CA | F1 | Prec | Recall | MCC |
|---|---|---|---|---|---|---|
| Neural Network | 1.000 | 0.905 | 0.903 | 0.919 | 0.905 | 0.823 |

**5G Technologies (Business topic).**

| Model | AUC | CA | F1 | Prec | Recall | MCC |
|---|---|---|---|---|---|---|
| Neural Network | 0.773 | 0.652 | 0.596 | 0.791 | 0.652 | 0.405 |

**Ukrainian Migration to Poland (Public sphere topic)**

There is an eye-catching difference in the effectiveness of both models. The model associated with 5G networks has extremely high efficiency, and in the case of True Negative it is 100% effective. The situation was slightly different in the case of the model regarding migrants. Here, the absence of a negative stereotype towards migrants was detected in every case, but the detection of a negative stereotype/disinformation was only possible in 60%. It seems that the model was not able to effectively detect many disinformation statements connected with negative stereotypes, due to the greater controversiality and complexity of the issue than in the case of the 5G networks.

## Conclusion

The presented study has made a compelling case for an interactive, experiential approach to teaching high school students about digital disinformation. It was proved, that by actively involving students in the process of training a neural network, they not only gained a more profound understanding of the nature of online disinformation but also collectively contributed to the creation of a tool that could be instrumental in mitigating this widespread issue. This is visible both at the level of individual people's knowledge test results, and collective intelligence released by a shared commitment to accomplish useful work leading to better

understanding of the complex problems associated with disinformation. This immersive learning experience enhanced students' understanding and perspective of the examined topics, and highlighted the potential of artificial intelligence and machine learning as powerful tools in combating digital disinformation. Collective involvement in joint preparation of a neural network model helped to increase participants' involvement in combating disinformation, making the learning experience more meaningful and impactful for them.

The revealed differences between the effectiveness of model trained to detect fake news regarding 5G, and the model that detects disinformation regarding migrants from Ukraine also lead to important conclusions. It can be assumed that it is much easier to detect disinformation in non-political areas, where "ground truth" can be more easily defined. However, in the case of issues belonging to the public sphere, statements are the subject of greater controversy and create a wider field for interpretation. This shows in which domains fake news detection using trained artificial intelligence can work better, which is a valuable observation for the future.

The project's success indicates that similar approaches, which integrate practical, technology-focused projects into the curriculum, could be fruitful in other educational contexts. Schools, educators, and policymakers should explore and adopt such methodologies to equip their students with the necessary skills to navigate the digital challenges that they will inevitably encounter in their future lives. The fight against disinformation is a collective effort, and young people, being digital natives, are an indispensable part of this battle. Empowering them with the right skills and tools is, therefore, a step in the right direction.

## Scientific Ethics Declaration

The author declares that the scientific ethical and legal responsibility of this article published in EPESS journal belongs to the author.

## Acknowledgements or Notes

* This article was presented as an oral presentation at the International Conference on Science and Education (www.iconse.net) held in Antalya/Turkey on November 16-19, 2023

*The educational project was conducted by Instytut Aurea Libertas and received funding from the European Social Fund under the Popojutrze 2.0 programme.

*The research was funded by Narodowe Centrum Nauki (National Science Centre, Republic of Poland), the research grant UMO-2018/28/C/HS5/00543 — "Collective intelligence on the Internet: Applications in the public sphere, research methods and civic participation models".

## References

Ahmed, W., Bath, P. A., & Demartini, G. (2017). Using Twitter as a data source: An overview of ethical, legal, and methodological challenges. In K. Woodfield (Ed.), *The ethics of online research (advances in research ethics and ıntegrity (*Vol. 2, pp. 79-107). Emerald Publishing Limited.

Aleguas, S. (2023) The fake arrest of Donald Trump: A deepfake odyssey. Retrieved from https://levelup.gitconnected.com/the-fake-arrest-of-donald-trump-a-deepfake-odyssey-db3a6c17eba6.

Allcott, H., Gentzkow, M. (2017) Social media and fake news in the 2016 election. *Journal of Economic Perspectives, 31*( 2), 211–236.

Bahl, V. (2023). Photo of elderly French protester being 'beaten' by police is AI-generated. Retrieved from https://www.france24.com/en/tv-shows/truth-or-fake/20230330-france-protests-old-man-beaten-by-police-likely-ai-generated.

Bessi, A., Zollo, F., Vicario, M. D., Puliga, M., Scala, A., Caldarelli, G., Uzzi, B., & Quattrociocchi, W. (2016). Users polarization on Facebook and YouTube. *PloS One*, *11*(8), e0159641.

Carmichael, F., & Goodman, J. (2020). Vaccine rumours debunked: Microchips, 'altered DNA' and more. Retirved from https://www.bbc.com/news/54893437.

Collins, B., Hoang, D.T., Nguyen, N.T., & Hwang, D. (2020). Trends in combating fake news on social media – a survey, *Journal of Information and Telecommunication*, *5*(2), 247-266

Kshetri, N., & Voas, J. (2017). The economics of "fake news". *IT Professional, 6*, 8–12.

Lee, J. (2020). *Were pallets of bricks strategically placed at US protest sites?* Retrieved from https://www.snopes.com/fact-check/pallets-of-bricks-protest-sites/.

Marchant de Abreu, C. (2022). *Debunking Russian claims that Bucha killings are staged. France 24.* Retrieved from https://www.france24.com/en/tv-shows/truth-or-fake/20220404-debunking-russian-claims-that-bucha-killings-are-staged.

Shabani, S., & Sokhn, M. (2018) Hybrid machine-crowd approach for fake news detection. *IEEE 4th International Conference on Collaboration and Internet Computing* (CIC), 299-306.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146-1151.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., Malone, T.W. (2010), Evidence for a collective ıntelligence factor in the performance of human groups. *Science, 330*(6004), 686–688

Zhou, X., Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys, 53*(5), 109.

## Author Information

**Rafał Olszowski**
1. University of Krakow,
al. Adama Mickiewicza 30 30-059 Kraków, Poland
2. Fundacja Instytut Aurea Libertas,
ul. Floriańska 3, 31-019 Kraków, Poland
3. MIT Center for Collective Intelligence
245 First Street, Cambridge, MA, United States
Contact e-mail: *rafal@omni.pl*

**To cite this article:**

Olszowski, R. (2023). Increasing the effectiveness of fake news detection: An educational program for high school students using interactive neural network training and collective intelligence. *The Eurasia Proceedings of Educational & Social Sciences (EPESS), 33,* 26-33.