Are Algorithms Enough?
Analyzing Fake News Solutions Designed by Students

Milica Milenkovic
University Union – Nikola Tesla
Belgrade, Serbia
e-mail: micagmekic@gmail.com

Essi Häyhänen
University of Vaasa, School of Marketing and Communication
Vaasa, Finland
e-mail: essi.hayhanen@gmail.com

Joni Salminen
University of Vaasa, School of Marketing and Communication
Vaasa, Finland
e-mail: jonisalm@uwasa.fi

Bernard J. Jansen
Qatar Computing Research Institute
Hamad Bin Khalifa University
Doha, Qatar
e-mail: bjansen@hbku.edu.qa

***Abstract*— We examine the fake news problem, emphasizing user-led solutions. Students were assigned a task to design a technology-based solution to combat online misinformation defined as fake news. Eighteen teams participated, a total of 89 students. Analysis reveals that 38.9% of teams devised algorithm-focused solutions, 27.8% proposed human-focused solutions, while 33.3% designed solutions that incorporated both algorithmic and human-centered approaches to addressing the misinformation problem. We identified a fundamental assumption regarding the effectiveness of Artificial Intelligence and algorithms, highlighting technological sophistication. These findings contribute to the ongoing discourse on combating fake news and provide directions for future research and development of effective technological interventions by considering human factors.**

Keywords- fake news; solutions; human factors; design.

I. INTRODUCTION

Fake news is fabricated and untruthful information spread deliberately to deceive a readership or viewers. It often resembles real news stories but contains false information and may mix real and fake sources, quotes, and information. While the Internet has enabled people to stay informed about global events, it has also become a breeding ground for spreading false and malicious news [1]. Consequently, fake news has become a global issue that affects people in various aspects of life, such as healthcare, transportation, education, and business.

Researchers have explored many techniques for detecting fake news [2]. These detection approaches often focus on social media, as fake news is frequently distributed on social media, also known as social media platforms or social network sites. While the consumption of news through social media is increasing due to its speed, accessibility, and affordability, social media also provides a platform for the widespread dissemination of fake news, which intentionally contains false information for political purposes, trolling, or other nefarious objectives [3]. The popularity of social media has led to a shift in how people access news, with traditional sources of journalism being replaced by online social media sources. In particular, the rapid rotation of news on social media can make it difficult to determine its reliability [4].

The creation and dissemination of fake news can potentially deceive users and influence their opinions, leading to undesirable consequences for society [5]. Online social networks like Twitter have increased the spread of false information and fake news. This misinformation can lead to harmful consequences for individuals who believe in inaccurate claims and articles. Therefore, the prevalence of false information and deceitful content in the form of articles, posts, videos, and URLs on popular social media platforms has raised concerns among journalists and editors, among other stakeholder groups, emphasizing the need for tools and processes to aid in content verification [6].

Detecting fake news is essential to prevent panic and confusion [7]. To tackle the issue of fake news, researchers tend to develop algorithms, models, and systems that aim to distinguish between real and fake news and help scientists and the public access accurate information [8]. However, the computing research community has not thus far been successful at delivering definitive solutions to the fake news problem. Despite impressive results in laboratory settings, these results are rarely implemented in real systems, and when they are, they are unable to address the full scope of issues relating to users' behaviors, psychology, and sociology, i.e., factors that cannot easily be affected using algorithms alone. So, there is a need to better understand the full scope of the fake news problem.

The current research approaches the problem through *co-creation with students* [9]. In our study, university undergraduate students who are experts in neither Machine Learning (ML) or fake news research (apart from general knowledge as users of online media) propose solutions to the fake news problem – the naïveté of these solutions can reveal novel angles about the problem, as business students are not "pre-programmed" by the current paradigms and models of approaching the problem, thereby potentially observing aspects that *might* be hidden or left with little attention in the computational research on fake news.

We analyze students' proposed solutions through six Research Questions (RQs), justified as follows:

RQ1: How many student teams designed algorithm-focused and how many human-focused solutions? This question is relevant as computational literature tends to focus on algorithms rather than human factors, potentially limiting the scope of solutions. We examine the division of the proposed solutions into these two broad categories.

RQ2: What themes can be observed in the students' fake news solutions? We conduct an open coding [10] of the material submitted by the student teams in order to inductively identify central themes in the solutions. This inquiry is of a qualitative nature.

RQ3: What are the central assumptions that the student solutions rely on? By analyzing the assumptions underlying the solution, i.e., what factors are required for the solutions to work, we aim to shed light into the conditions, especially those extending beyond the technological realm, of practically workable fake news solutions. The literature does not often discuss such conditions (see Section 2).

RQ4: What are the central risks that the student solutions involve? We asked the students to identify key risks in their solutions; here, we analyze the types of risks they identified.

RQ5: How realistic are the student solutions? What are the more realistic aspects? What are the less realistic aspects? With this RQ, we aim to dissect the parts in the student solutions that have a degree of possibility of succeeding and, symmetrically, the parts that are likely to fail (according to our assessment). For this, we conduct a critical inquiry into the content of the solutions.

RQ6: What metrics do students propose to measure if fake news solutions work? Finally, we address the question of how the students would suggest measuring the success of fake news solutions. Prior research has focused on technical metrics (e.g., F1 score, accuracy, precision, recall) to evaluate research contributions–however, these metrics focus on the internal performance of the models while ignoring their ecological validity, i.e., how well the models would contribute to the solution of the fake news problem when implemented. Naturally, this question is beyond the computational paradigm based on ML, but we precisely argue here that fresh ideas on evaluating fake news solutions can be fruitful and interesting for the research community. So, we analyze the metrics proposed by the students.

Our analysis is based on a qualitative interpretation of the student-based solutions. We do not attempt to present definitive facts about solving the fake news problem, as we believe doing so is extremely complex. What we aim to do instead is to shed light on the more rarely discussed aspects of the fake news problem – those not directly associated to the creation of better systems, models, and tools but instead indirectly affecting the implementation in actual user environments. While most research in this field is of technical nature, focused on factors like algorithm selection, hyperparameter optimization, training and test splits, and so on [11], we believe that the current study complements these technical views by offering a perspective closer to the *everyday user* of social media platforms. As such, we believe this inquiry has value for the research community.

Section 2 reviews prior research on this topic. Section 3 outlines the methodology for data collection and analysis. Section 4 presents the results, followed by a discussion of findings and future work in Section 5.

II. Literature Review

As discussed above, the fast spread of fake news is a significant concern. This has motivated researchers to introduce solutions for automatically classifying news items [12]. Much research has focused on Artificial Intelligence (AI). Previous research has concentrated on classifying online reviews and publicly accessible social media-based posts [13]. Automated fake news identification technologies, such as ML models, are essential in the current body of research [14]. Current techniques primarily rely on Natural Language Processing (NLP) and ML models [15]. While traditional ML methods have been used to detect fake news, genetic algorithms are potential due to converging to near optima with low computational complexity [15].

Many news agencies publish news on their websites, but not all are trustworthy. Therefore, before quoting any news from a website, it is necessary to evaluate the reputation of the news resource using a trusted website classifier. Mughaid et al. [2] proposed using the world rank of news websites as the main factor for news accuracy, along with a secondary factor that compares the current news with fake news to determine its accuracy [2]. Thus, the source of the news is considered a crucial factor in determining fake news.

According to Shu et al. [3], existing detection algorithms focus on clues within news content (e.g., text, semantics, images), which may not always be effective as fake news is often intentionally written to mimic true news; e.g., by making it sound professional and convincing. Therefore, it is necessary to explore auxiliary information to improve detection [3]. For example, sophisticated techniques are used to deliberately modify text or images to create fake news. Giachanou et al. [5] proposed a multimodal system that combined textual, visual, and semantic information to detect fake news. They utilized Bidirectional Encoder Representations from Transformers (BERT) to capture the underlying meaning of text. For a visual representation, they extracted image tags using the Visual Geometry Group-16 (VGG-16) model. The semantic representation was calculated using cosine similarity between the title and image tags embeddings [5].

Nikam and Dalvi [16] proposed a method for classifying fake news on Twitter using a web-based Graphical User Interface (GUI). They developed an ML model that compared tweets to genuine sources to identify fake news, using the Naïve Bayes (NB) and Passive Aggressive (PA) algorithms with Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction [16].

Sheikhi [12] presented a system for detecting fake news articles based on content-based features and the Whale Optimization Algorithm-Extreme Gradient Boosting Tree

(WOA-xgbTree) algorithm. The proposed system can be applied in different scenarios to classify news articles. The approach consisted of two main stages: first, the useful features were extracted and analyzed, and then an xgbTree algorithm optimized by the WOA was used to classify news articles using the extracted features [12].

Huang and Chen [17] presented a deep learning-based fake news detection system. The proposed system preprocessed news articles and analyzed them using various training models [17]. To detect fake news, Huang and Chen [17] introduced an ensemble learning model called the Embedding Long Short-Term Memory (LSTM), Depth LSTM, Linguistic Inquiry and Word Count (LIWC) Convolutional Neural Network (CNN), and N-gram CNN. Moreover, they optimized the weights of the ensemble model using the Self-Adaptive Harmony Search (SAHS) algorithm [17].

Singhal et al. [18] argued that detecting fake news requires a multimodal approach. As most multimodal fake news detection systems rely heavily on subtasks such as event discrimination and correlation analysis, they proposed SpotFake, a multimodal framework for fake news detection that does not require any additional subtasks. Their approach leveraged both textual and visual features of an article, using BERT to extract text features, and VGG-19 to extract image features [18]. Gundapu and Mamidi [7] used an ensemble of three transformer models (BERT, ALBERT, and XLNET) for evaluating the reliability of information related to the COVID-19 pandemic shared on social media [7].

As in most NLP tasks, transformers represent the state-of-the-art in fake news detection (note that our review does not include Generative AI or Generative Pre-trained Transformer (GPT) models, as these were not broadly available at the time of the review). The accuracy achieved by these models is impressive, which is one reason why the research community should expand its scope of examination – it is unlikely that the algorithm will get much better from this point. Instead, we expect decreasing marginal returns, which is why it is logical to pursue other aspects of the problem, including the *implementation* and *application* aspects.

Overall, the predictive results are impressive. It is not evident how the researchers can continue improving them over time, as it appears we are already at the >90% performance. Thus, the domain requires new, fresh ideas to explore. Some of these ideas can originate from externality, outside the computing research community, for example, from students. To this end, we move forward to the empirical part of this study, addressing our research questions. Before that, we briefly summarize our methodology.

III. Methodology

A. Data Collection

Students were given an assignment to design a technology-based solution to fake news, which was defined as online misinformation. The students carried out the task in teams, and there were 18 teams (in total, the course had 89 students, so each team had 4.9 students on average). Each team submitted their solution in a slideshow presentation. The content of these slideshow presentations was coded into a spreadsheet and then analyzed to address the research questions. Students were informed that their contributions could be used as a part of ongoing research on fake news.

B. Data Analysis

For the analysis, first, we evaluated students' presentations. Each solution was assigned a unique ID (S01-18), and we analyzed the solutions individually. This procedure was as follows:

Step 1: Determine whether a solution is:

Algorithm-focused (i.e., the solution relies mainly on the technical aspect of the algorithm and technology to deal with the problem of misinformation).

Human-focused (i.e., the solution relies mainly on humans doing the activities of finding, filtering, and decision-making about misinformation)

Mixed-focused (i.e., the solution combines aspects of both algorithms and humans)

Step 2: Evaluate the level of realism (1-7) of the solution, 1 being not realistic at all, and 7 being very realistic. If a solution involved many different stakeholders aligning in their thinking, too much technical sophistication, too many aspects of misinformation covered in the solution, that would make the solution unrealistic, with a mark of 1 or 2. A specific solution focused on a specific form of misinformation, a specific platform, or a development of a new platform for a specific niche need, and has a clear view about its functioning. Such a solution was marked with 6 and 7; the difference between the two is the number of stakeholders involved in the solution being built and implemented or the likely amount of work needed to develop the solution. For example, S01 has a realism of 6, meanwhile, S09 has 7. This is because S09 is a feature integration with already existing social media platform(s), meanwhile, S01 proposes a standalone platform that gets information to and from different individuals and organizations and is more difficult to implement, as opposed to working directly with an already existing platform to improve this segment of misinformation handling.

Step 3: Evaluate the level of clarity (unknowns) of the solution, one, meaning there are a lot of unknowns, and 7, there are little unknowns about the solution presented. Too many unanswered questions on how the solution will be built, who the stakeholders are, and how it will be implemented, with what technology leads to marks 1 and 2. Opposing, having the most clarity and the least number of unanswered questions leads to marks 6 and 7. For example, S19 is a highly specific, niche solution for removing AI-generated misinformation, giving information on methods and technology that will be used, the implementation route, and the precise limitations of the solution.

Step 4: Extract assumptions, risks, and metrics from the students' presentations.

Step 5: Assign themes/taxonomy to each solution. Answering the questions: "What is common to the solutions – algorithm-focused or human-focused?", "How will the solution be implemented?", "What kind of app/solution will

 ISBN: 978-1-68558-163-3

be developed?", "What misinformation aspects is the solution tackling?" Through qualitatively analyzing the solutions, we developed the themes presented in this paper.

Steps 1-5 were carried out by the lead author and verified by a senior research team member. After finishing the evaluation, we could unify the assumptions, risks, and metrics. Looking at similar occurrences with different wording, we were able to craft common assumptions. For example, we first separated all the Risks, and Metrics per presentation. Then, we started comparing them to each other, which then helped us quantify the results. For example, the most common metrics for success of the solution is Engagement rate, mentioned in 12 out of 18 solutions.

After finishing the evaluation and analysis of the results, we addressed the research questions. For RQ1, we have the exact number for each type of solution. For descriptions of the solutions, we specified the most realistic and clear solution in each type and the ones on the opposite side of that scale. For RQ2, we enlisted and described all the themes (whose development we explained above) and elaborated the main taxonomies that they belong to. We also mentioned their share of occurrence in the solutions. For RQ3, we grouped the central assumptions that the solutions relied on (as indicated by the students). For RQ4, we grouped the central risks that the solutions involved (as indicated by the students). For RQ5, referencing the previous step, evaluation, we were able to assess realism. Based on that and the decision-making behind the evaluations, we outlined the most and least realistic solutions. For RQ6, we identified and grouped the metrics that the students proposed.

IV. RESULTS

A. *RQ1: How many teams designed algorithm-focused and how many human-focused solutions?*

Inspecting the results, we observe that seven teams (38.9%) designed Algorithm-Focused Solutions (**AFS**), five (27.8%) Human-Focused Solutions (**HFS**), and six (33.3%) of them designed a Combined-Focus Solution (**CFS**) that contained both algorithm and human-focused approaches to solving the misinformation problem. Solutions in algorithm-focused approaches were the following (quotes indicate direct quotation from the student team's presentation):

AFS01: "The platform evaluates the news post's veracity before it can be published. The […] algorithms will check from all over the Internet the accuracy of the news that they want to publish. If the AI cannot find a solution fast enough, people moderators will then look it up and decide."

AFS02: An application that integrates with social media, that has "better algorithms & filters that are ranking fake news visibility down and/or leading the reader via link to a confirmed site that has the right information."

AFS03: An application that integrates with social media consists of an "algorithm that scores the publications and identifies topics, phenomena, words, punctuation, vocabulary, abbreviations, and the presence of references. If the score is below the credibility limit, the publication will be investigated further."

AFS04: An application that integrates with Instagram, has an "algorithm using ML to detect the misinformation posts, stop the spread of them, and eliminate fake news content from Instagram. Methods used are an AI-based algorithm that detects the fastest spreading/biggest fake news on the platform based on e.g., the shares, (negative) comments, and reports by users on a certain post."

AFS05: A standalone social media platform that has an "algorithm that uses website crawlers. It crawls through different website sources. It recognizes keywords and this way connects related articles. It is designed to be connected to each user's posts and shows related articles according to what the user has posted. This way, other users can compare different sources to find the most reliable information and make decisions based on them."

AFS06: A standalone news platform "to help ensure that you are always using credible sources for research. The platform consists of peer-reviewed, reliable, and trustworthy articles that have been fact-checked."

AFS07: An application that integrates with social media to remove AI-generated misinformation.

In turn, solutions in human-focused approaches were:

HFS01: An integration that consists of an extensive user verification process (social security number and verifying the account with a video holding the ID) to eliminate fake accounts. The users would be encouraged to report suspicious activity, so the moderators could check and act.

HFS02: A standalone platform based on combatting misinformation with a user rating system, including informing users about the news considered misinformation, so that they would react by providing the real information.

HFS03: An educational platform that teaches users how to recognize misinformation. The platform would give different examples from different platforms on the Internet. This is not a news-sharing platform or integration with social media platforms, this is a training platform.

HFS04: A platform called Truth Seekers. Gamification of fact-checking. Users are presented with stories and articles that circulate on social media platforms. Users earn points by reporting stories with suspected misinformation with the help of an AI-powered fact-checking tool.

HFS05: An educational platform that allows users to practice and develop media literacy skills by answering problems that volunteering "creators" upload. The idea is to offer users the possibility to learn and educate themselves and simultaneously attract creators to the platform by offering a small profit every time they create hard enough questions that other users rate as helpful.

Finally, two solutions were mixed-focused approaches:

CFS01: A standalone AI-based fact-checking platform would enable users to upload content for verification against trusted sources. It would connect users with educational and research institutions to counter misinformation.

CFS02: A Social Media Integration tool with a Trust Factor, assigned to each account and posted on a scale of 1 to 100. The Trust Factor is based on user ratings and other factors, like its use in online games. This solution is highly realistic and effective.

B. RQ2: What themes can be observed?

While building themes, we focused on commonalities between solutions and their differentiating factors. The themes that we outlined are shown in Table 1.

TABLE I. FAKE NEWS SOLUTION THEMES

Description	Subthemes
Themes about the basic functionality of the solution (applicable to 61% - 11 solutions)	➢ User Verification (27%) ➢ User Rating System and/or Reporting System (44%) ➢ User Education (27%)
Themes about the type of the solution (applicable to all 18 solutions)	➢ SM Integration (39%) - Non-Specified SM Integration 28% - Specific SM Integration (Meta, Instagram) 11% ➢ Standalone Platforms (61%) - Non-Specified Standalone Platform (22%) - Standalone SM Platform (22%) - Standalone Educational Platform (11%) - Standalone News Platform (6%)
Themes about how the misinformation is viewed	➢ A multidimensional view of misinformation (72% - 13 solutions) ➢ Single-dimensional view of misinformation (28% - 5 solutions)
Themes about additional features/functionalities of the solution	➢ Online Game (S16) ➢ Collaboration with Law Enforcement (S15)

Having assumptions, risks, and metrics as the RQs prevented a possibility of a large pool of themes.

We will start elaborating on a taxonomy group that applies to all 18 solutions - **The type of the Solution.** This thematic categorization has two major themes in it: *Social Media Integration* and *Standalone Platforms*. Of these solutions, 61% were categorized as Standalone Platforms, while 39% were Social Media Integration applications. These are implemented leveraging existing platforms' infrastructure. This approach is less challenging in terms of business development. Standalone platforms offer more freedom but require a multi-dimensional approach, making them more difficult to build. Despite the complexity, most student solutions (approximately 2/3) fell into the standalone platform category. The other taxonomy group - **The basic functionality of a Solution** applies to 61% of the solutions (11 out of 18). This one focuses on how the solution works, and what the basic prerequisites are whether the solutions tackle multiple dimensions of misinformation, or whether they specialize and focus only on conspiracy theories, fake accounts, and bots, or catching clickbait content. The last and optional (additional) taxonomy group, closer to a simple tag or category, contains non-essential features, that are unique to the solution. **Additional features/functionalities of the solution**. These themes can be used as descriptors for the solutions. *Online Games* and *Collaboration with Law Enforcement*. Both solutions were rated 5 for realism and clarity.

C. RQ3: What are the central assumptions?

The assumptions, limitations/risks were presented together in students' solutions (see Table 2) because the assumption that something will work directly implies the risk to the solution. For example, assuming that the solution will have enough users (and then a growing base of users) is what is needed for the solution to work implies that not having enough users is a risk to the solution working.

TABLE II. THE MOST COMMON ASSUMPTIONS

Assumption	Occurrences	%
AI and algorithms working well - technological sophistication	7	39%
User participation & engagement	4	22%
Having enough trustworthy sources for the algorithm to use	4	22%
Good UI & UX	3	17%
Users willing to learn	3	17%
Having access to external platforms	2	11%
Transparency and credibility	2	11%

To elaborate further on the major assumption, we start with *AI and algorithms working well*. All algorithm-focused solutions have this assumption. This assumes, firstly, that the AI and the algorithm will do what they are built to do: detect various forms of misinformation. Secondly, a proper infrastructure is in place for the solution to be operational and efficient. As the idea of algorithms is to be able to connect to different sources of information, it is of utmost importance that the sources have a *high score of trustworthiness and reliability*. Three out of four solutions that highlighted these assumptions are standalone platforms. In turn, having *enough users, participation, and engagement* will ensure that the standalone application stays 'alive'.

Good User Interface (UI) & User Experience (UX) focus on standalone platforms, where users being able to find their way around the information-sharing or fact-checking platform is of great importance. Even though one would think that *user willingness to learn* would be tied to the Standalone Education Platform, this is not the case. Both occurrences are assumptions where the solution is not an education platform, but depends on users to rate, and/or follow true/real information. For that to happen, users need to be willing to find out the truth, even when it is not convenient. Also, for some solutions to work (as integrations with other Social Media (SM) platforms), they need to be *compatible for integration* (both technologically and business-wise). To build trust with users, *transparency and credibility* are necessary for the solution/integration to have enough users in the first place. We can identify relations between the assumptions. Here are some unique assumptions tied to the most realistic and clear solutions:

Currently, it is difficult to detect misinformation or verify; therefore, solution S09, which is an algorithm-focused solution, also assumes *that algorithms and AI are more objective than humans* regarding this issue.

By defining a niche focus, S10 provides more certainty in combatting misinformation. The goal is to use rapidly scaling news in terms of engagement and check whether it is fake news or not. *The assumption here is that negative news gets more comments and engagement, and that there is more misinformation among negative news.*

S19 has a basic assumption that one team (developers of the solution) will do better than the developers of solutions that spread fake content and misinformation.

D. RQ4: What are the central risks?

The common risks are presented in Table 3.

TABLE III. SELF-IDENTIFIED RISKS IN FAKE NEWS SOLUTIONS

Risk	Occurrences	%
Having enough users to participate	6	33%
Technology is a risk factor on its own	4	22%
The ability of the algorithm to be able to predict topics that are constantly changing	3	17%
The ability of the algorithm to determine the tone of the text	3	17%
Bias	3	17%
UI & UX	2	11%
Technology compatibility	2	11%
People are not following the links to the sources to find reliable information	2	11%
People not believing the information no matter how trustworthy	2	11%
Many people have multiple accounts	2	11%
Profitability	2	11%

Not having enough users is the most common risk to both standalone and integration solutions (in terms of people using the features of flagging the information).

Another major concern is *technology on its own* and AI development must be constant and consistent. The worry that technology can generate significantly more misinformation than users in any time frame is overwhelming. That is why it is a positive sign that one of the solutions is strictly directed at combating the misinformation generated by AI (S19).

The risks to implementing are AI-generated content and users already being on the platforms, therefore flagging the new ones would not delete the old bot accounts and old content that has been circulating, and the resources for the race against AI fake news generation may be scarce. For the algorithm-focused solutions, which are integrations with other platforms, this is not a concern, but rather the technical abilities of the algorithm to properly detect misinformation, adjust to the changing landscape, and develop ML capacity.

Another concern is *bias*. In human-based solutions, it is the bias of either moderators or users themselves. For algorithm-based solutions, it is the bias of the users whose content is used to train ML models, and the individuals building the ML capability. However, bias is not perceived as a large risk factor, as only three solutions outlined it as a risk factor.

In addition to the five most common risks, it is vital not to overlook whether the solutions are compatible and can be integrated with other platforms, considering that people tend to have multiple accounts. Winning over trust is difficult, as well as sparking curiosity which is important for fact-checking. Also, the solution, whether integration or standalone, must be commercially viable.

E. RQ5: How realistic are the student solutions?

Overall, the students' solutions were deemed more realistic than not. The decisions on whether a solution is realistic or not and to what extent was made taking into consideration the assumptions, risks, metrics, manpower needed, technological sophistication needed, number and scale of stakeholders involved, and the level of input from the users of the solution. In the more realistic aspects of the solutions, we identified the following:

4 realistic solutions (S09, S10, S17, and S19) proposed integrations with existing social media platforms. These solutions involved the development of algorithms to detect potential misinformation, the introduction of a "trust factor" for information sources, and the mitigation of AI-generated misinformation. These solutions were considered highly realistic due to their technological focus, as they specifically addressed the misinformation problem within platforms that were already engaged in combating misinformation.

3 realistic solutions (S01, S11, and S13) involved creating their own platforms: news, educational, and fact-checking. These solutions utilized AI and ML models, along with human input for reliability. Manual fact-checking was necessary due to language nuances and human social nature. Building an educational platform for spotting misinformation was considered realistic.

Most of these solutions have a theme of a *Multidimensional view of misinformation* (6 out of 7), meaning they focus on multiple aspects of misinformation (fake news, sensationalism, propaganda, click-bait, conspiracy theories, etc.). One solution has a user rating system (as users are sources of information), introducing a 'trust factor', which will be based on an algorithm.

In the unrealistic solutions, two findings emerged:

Two solutions had a score of 2. One solution proposed an extensive user verification process using multiple personal documents, including a social security number, which was deemed unrealistic (S02).

The other had an unrealistic expectation of integrating social media apps and their features (S14). Too many features, too many requests from users, and too many stakeholders were seen as unrealistic.

Of all unrealistic and less realistic solutions (scales 2, 3, 4), 30% were human-based solutions, 30% were technology-based, and 40% were mixed-based.

F. RQ6: What metrics do students propose?

Table 4 presents the most common measures, followed by a discussion on both quantitative and qualitative metrics.

The most common metric is *engagement rate*, regardless of whether the solution is a standalone platform or an integration with existing ones. Engagement rate has various meanings. For some solutions it is the number of likes and shares, or if a solution is commenting on the information and providing real information, in that case, comments on their own are a metric. This metric is quantitative, and the metrics such as the *number of reported misinformation instances* (fake news, clickbait, misleading, etc.), and the *number of users* or verified users are the most common quantitative metrics found in students' solutions.

TABLE IV. FAKE NEWS SOLUTION METRICS

Metric	N	%
Engagement rate	12	67%
Feedback from users	11	61%
Number of reported misinformation instances	6	33%
The ability of users to identify misinformation	3	17%
Number of users	3	17%
Reliability and accuracy of the information shared	3	17%
User satisfaction	3	17%
Independent analysis and verification	3	17%
Number of new verified accounts	2	11%
Reliability score of users	2	11%
3rd party reviews	2	11%
User behavior tracking	2	11%
Manual reviews of sample content	2	11%
Click through rate	2	11%
Bounce rate	2	11%
Conduct experiments to test the effectiveness	2	11%

Other **Quantitative Metrics** that were present in the realistic solutions, with the least number of unknowns, are:

Several texts downloaded from the platform: the solution involves a standalone platform where a user can upload a piece of text and get it checked through different reliable sources (scientific journals, papers, etc.) and obtain a reference that they can download.

Decrease in engagement: people react to fake news more than to actual news, which means that, if the algorithm does the job well, the platforms will have less misinformation, and, therefore, less engagement.

An appropriate user engagement metric, which was not observed in other solutions, is *Time spent on the platform*. Building on the above metric (even though these two metrics were not presented in the same students' solution) is deeply connected to this one. As engagement drops, an assumption would be that people spend less time, but a direct indicator that the algorithms work, and that people are not prone to get into any arguments over misinformation is the opposite curve on the graph for time spent on the platform.

Relevance, authority, and accuracy as metrics of the sources of information. This one is part of the solution which is a separate standalone news platform. Having these scores assigned to different sources would help the assessment of whether a news article from a particular source about a topic should be posted on this standalone news platform.

Moving onto **Qualitative Metrics**, from our list, 61% of solutions had *Feedback from users* as a necessary metric. The feedback from users would consist of any form of feedback provision through forms or surveys. There are also metrics for *third-party reviews & independent analysis and verification* including feedback from outsider bodies and organizations on the platform, algorithm, and the system implemented. Next, we have *user satisfaction* with the platform, feature (integration), and outcome. The outcome, in this sense, means the circulation of news and decreased proportion of misleading information.

The ability of users to identify fake news, as a metric, needs to track users' success in identifying different types of indicators that help determine if the news is fake.

Consumer education as a metric stems from solutions that have a training/education aspect to the platform or a feature. Users can learn on the spot through current news and examples of current news categorized as fake.

To summarize, the metrics are engagement rate, feedback from users, number of reported misinformation instances, the ability of users to identify misinformation, number of users, reliability and accuracy of the information, user satisfaction, and independent analysis and verification.

V. DISCUSSION AND CONCLUSION

This paper has examined user-led solutions to the problem of fake news. Our findings highlight a reliance on AI effectiveness and underscore the need for technological sophistication. These insights contribute to the ongoing discourse on combating fake news and suggest future research should integrate human factors in developing technological interventions.

A. Theoretical Implications

The findings involve several theoretical implications. First, identifying algorithm-focused, human-focused, and combined-focus solutions provides theoretical implications for combating fake news [19]. This underscores the importance of a balanced approach that combines technological advancements and human judgment.

Second, categorizing solutions as social media integration or standalone platforms emphasizes the need to adapt interventions to different technological contexts when addressing fake news [20]. This allows customization based on the specific misinformation problem and target audience.

Third, the assumption that AI and algorithms effectively combat fake news highlights technological sophistication. Because limited user engagement challenges the importance of user adoption and participation in successful interventions, this implies the need for user-friendly and accessible technological solutions. The prevalent use of engagement rate as an evaluation metric for both standalone platforms and Social Media Integration implies that interventions should prioritize metrics capturing user participation.

Fake news researchers should start approaching the problem more holistically, engaging in cross-disciplinary research collaboration. The goal is for users to be more information literate and critical, which requires not only a technical environment, but users also need to be incentivized about the need to be information literate.

B. Limitations and Future Research

The limitations include the lack of actual verbal presentations from the students, which may contain additional information. This could impact the ratings of realism and clarity. Another limitation is the absence of user opinions and concerns from different platforms regarding misinformation. It would be interesting to know how a fact-checking feature or platform would work and whether the users would use it in the first place. How will it be connected to the organizations and institutions that do research or other trusted sources and which ML models and algorithms will be used? Also, we would love to see the results of a test on spotting the misinformation with users and their reasoning about what and why something is some type of misinformation. That way, we can make a bridge between the current level of ability of users to spot misinformation, and what is needed to improve that. Human factors in fake news detection merit much more research.

REFERENCES

[1] R. R. Mandical, N. Mamatha, N. Shivakumar, R. Monica, and A. N. Krishna, ‘Identification of Fake News Using Machine Learning’, in *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, Jul. 2020, pp. 1–6. doi: 10.1109/CONECCT50063.2020.9198610.

[2] A. Mughaid *et al.*, ‘An intelligent cybersecurity system for detecting fake news in social media websites’, *Soft. Comput.*, vol. 26, no. 12, pp. 5577–5591, Jun. 2022, doi: 10.1007/s00500-022-07080-1.

[3] K. Shu, S. Wang, and H. Liu, ‘Beyond News Contents: The Role of Social Context for Fake News Detection’, in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, in WSDM ’19. New York, NY, USA: Association for Computing Machinery, Jan. 2019, pp. 312–320. doi: 10.1145/3289600.3290994.

[4] R. Natarajan *et al.*, ‘Intelligent gravitational search random forest algorithm for fake news detection’, *Int. J. Mod. Phys. C*, vol. 33, no. 06, p. 2250084, Jun. 2022, doi: 10.1142/S012918312250084X.

[5] A. Giachanou, G. Zhang, and P. Rosso, ‘Multimodal Multi-image Fake News Detection’, in *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, Oct. 2020, pp. 647–654. doi: 10.1109/DSAA49011.2020.00091.

[6] S. Kaur, P. Kumar, and P. Kumaraguru, ‘Automating fake news detection system using multi-level voting model’, *Soft. Comput.*, vol. 24, no. 12, pp. 9049–9069, Jun. 2020, doi: 10.1007/s00500-019-04436-y.

[7] S. Gundapu and R. Mamidi, ‘Transformer based Automatic COVID-19 Fake News Detection System’. arXiv, Jan. 21, 2021. doi: 10.48550/arXiv.2101.00180.

[8] M. Zivkovic, C. Stoean, A. Petrovic, N. Bacanin, I. Strumberger, and T. Zivkovic, ‘A Novel Method for COVID-19 Pandemic Information Fake News Detection Based on the Arithmetic Optimization Algorithm’, in *2021 23rd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, Dec. 2021, pp. 259–266. doi: 10.1109/SYNASC54541.2021.00051.

[9] N. Zarandi, A. Soares, and H. Alves, ‘Strategies, benefits and barriers–a systematic literature review of student co-creation in higher education’, *Journal of Marketing for Higher Education*, pp. 1–25, 2022.

[10] J. M. Corbin and A. Strauss, ‘Grounded theory research: Procedures, canons, and evaluative criteria’, *Qualitative sociology*, vol. 13, no. 1, pp. 3–21, 1990.

[11] A. A. A. Ahmed, A. Aljabouh, P. K. Donepudi, and M. S. Choi, ‘Detecting fake news using machine learning: A systematic literature review’, *arXiv preprint arXiv:2102.04458*, 2021.

[12] S. Sheikhi, ‘An effective fake news detection method using WOA-xgbTree algorithm and content-based features’, *Applied Soft Computing*, vol. 109, p. 107559, Sep. 2021, doi: 10.1016/j.asoc.2021.107559.

[13] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, ‘FNDNet – A deep convolutional neural network for fake news detection’, *Cognitive Systems Research*, vol. 61, pp. 32–44, Jun. 2020, doi: 10.1016/j.cogsys.2019.12.005.

[14] S. Vinothkumar, S. Varadhaganapathy, M. Ramalingam, D. Ramkishore, S. Rithik, and K. P. Tharanies, ‘Fake News Detection Using SVM Algorithm in Machine Learning’, in *2022 International Conference on Computer Communication and Informatics (ICCCI)*, Jan. 2022, pp. 1–7. doi: 10.1109/ICCCI54379.2022.9740886.

[15] D. Choudhury and T. Acharjee, ‘A novel approach to fake news detection in social networks using genetic algorithm applying machine learning classifiers’, *Multimed Tools Appl*, vol. 82, no. 6, pp. 9029–9045, Mar. 2023, doi: 10.1007/s11042-022-12788-1.

[16] S. S. Nikam and R. Dalvi, ‘Machine Learning Algorithm based model for classification of fake news on Twitter’, in *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Oct. 2020, pp. 1–4. doi: 10.1109/I-SMAC49090.2020.9243385.

[17] Y.-F. Huang and P.-H. Chen, ‘Fake news detection using an ensemble learning model based on Self-Adaptive Harmony Search algorithms’, *Expert Systems with Applications*, vol. 159, p. 113584, Nov. 2020, doi: 10.1016/j.eswa.2020.113584.

[18] S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru, and S. Satoh, ‘SpotFake: A Multi-modal Framework for Fake News Detection’, in *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, Sep. 2019, pp. 39–47. doi: 10.1109/BigMM.2019.00-44.

[19] K. Somandepalli, T. Guha, V. R. Martinez, N. Kumar, H. Adam, and S. Narayanan, ‘Computational media intelligence: human-centered machine analysis of media’, *Proceedings of the IEEE*, vol. 109, no. 5, pp. 891–910, 2021.

[20] C. Vaccari and A. Chadwick, ‘Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news’, *Social Media+ Society*, vol. 6, no. 1, p. 2056305120903408, 2020.