Investigating User Behavior Towards Fake News on Social Media Using Gaze and Mouse Movements

Yasmeen Abdrabou*, Elisaveta Karypidou†, Florian Alt*, and Mariam Hassib*
*University of the Bundeswehr Munich, Germany
Email: {yasmeen.essam, florian.alt, and mariam.hassib}@unibw.de
†LMU Munich, Germany
Email: elena.karypidou.ek@gmail.com

***Abstract*—We propose an approach to identify users' exposure to fake news from users' gaze and mouse movement behavior. Our approach is meant as an enabler for interventions that make users aware of engaging with fake news while not being consciously aware of this. Our work is motivated by the rapid spread of fake news on the web (in particular, social media) and the difficulty and effort required to identify fake content, either technically or by means of a human fact checker. To this end, we set out with conducting a remote online study ($N = 54$) in which participants were exposed to real and fake social media posts while their mouse and gaze movements were recorded. We identify the most predictive gaze and mouse movement features and show that fake news can be predicted with 68.4% accuracy from users' gaze and mouse movement behavior. Our work is complemented by discussing the implications of using behavioral features for mitigating the spread of fake news on social media.**

I. Introduction

Fake news became an omnipresent term in our daily life, in particular in the context of major (political) events or topics subject to public debate. Presidential elections, climate change, the COVID-19 pandemic, and the war in Ukraine are but a few examples. People's beliefs regarding what is "truth" has much changed in past years so that even the Oxford dictionary has dedicated a term for the era we are currently living in: "post-truth", defined as "circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief"[1]. There are many reasons behind creating and distributing fake news, including the promotion of agendas, swaying public opinion or merely entertaining people.

Approaches to detect and stop fake news range from relying solely on humans to purely technical solutions. These efforts include crowdsourcing and human fact-checking [1], utilizing machine learning [2]–[4]) and natural language processing [5], as well as hybrid techniques combining the aforementioned methods [6]. Technical approaches relying on machine learning are promising, yet not reliable enough on their own due to the lack of data sets covering all aspects and dimensions of fake news content [7]. Recent surveys of fake news detection methods are provided by Collins et al. [8] and Zhou et al. [9].

[1]https://www.oxfordlearnersdictionaries.com/definition/english/post-truth

Symposium on Usable Security and Privacy (USEC) 2023
27 February 2023, San Diego, CA, USA
ISBN 1-891562-91-6
https://dx.doi.org/10.14722/usec.2023.232041
www.ndss-symposium.org, https://www.usablesecurity.net/USEC/

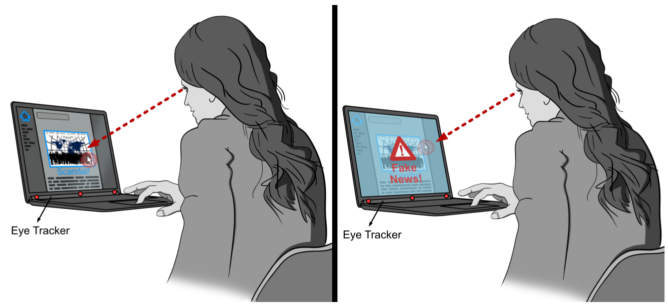


Fig. 1: We explore the concept of predicting exposure to fake news from gaze and mouse movement data. We achieve a prediction accuracy of about 70%. Our work enables novel mitigation strategies to warn users from fake news.

Prior research on fake news consumption looked into the physiological and behavioral responses of humans while consuming fake news as another means to detect the truthfulness of content. For example, sensors such as electroencephalography [10] or eye tracking [7] have been employed in controlled lab studies to explore differences in reading behavior of different types of content. In this work, we investigate if there is a difference in human behavior during the consumption of fake vs. real news online in a naturalistic and ecologically valid setup. In particular, we are interested in investigating users' gaze and mouse movement behavior while scrolling social media posts as predictors of the veracity of the presented information. The ability to detect the veracity of news, while users are reading/scrolling, may enable the displays of instantaneous warnings to end users that the news might be fake, or be used for automatically flagging the posts for sanity checks by fact-checkers. We expecte both cases to help reducing the spread of online fake news.

To this end, we designed and implemented an online remote study (N=54), tracking user behavior (mouse movements, eye gaze) in a natural setup during the consumption of real and fake news on Facebook. We report on the design of the platform and explore the resulting data set regarding user behavior while consuming real and fake news of different types (text, images, articles) and on different news topics. We present the results of the behavior analysis and machine learning classification of the mouse and gaze behavior and discuss the potential of using human behavior to detect fake news online.

II. Background and Related Work

Multiple strands of prior literature guide our work. First, we provide an overview of work investigating fake news consumption on social media and its effects. Second, we discuss prior work, leveraging users' behavior and physiological states in the context of building novel user interfaces. Finally, we discuss efforts toward mitigating the spread of fake news.

A. Fake News, Its Consumption, and Its Effects

Initially, the term fake news referred to satire and entertainment TV, showing parody news. As creating and spreading fake news transitioned from being purely humorous to large-scale hoaxing and even the deliberate production of fake news, the term required new definitions. The Cambridge dictionary defines Fake News as *'False stories that appear to be news, spread on the internet or using other media, usually created to influence political views or as a joke'*[2]. Tandoc et al. [11] identified different types of fake news: news satire, news parody, fabrication, manipulation, advertising, and propaganda. Fake news may not only comprise text, but also other media that can change the perception of a post intentionally [11], such as photos or videos. Lazer et al. [12] state that 'fake news is fabricated information that mimics news media content in form but not in organizational process or intent. Fake news outlets, in turn, lack the news media's editorial norms and processes for ensuring the accuracy and credibility of information'.

There are many reasons for sharing fake news on social media. People might publish false content unintentionally by not doing proper source investigations beforehand. However, fake news is often published on purpose. Talwar et al. [13] investigated the reasons for publishing fake news online. The results included online trust, self-disclosure, fear of missing out, and social media fatigue [13].

In the past years, people have switched from mainstream publishers to social media when reading news online so as to get information about current events. In 2019, Bently et al. investigated how American readers consume news online. They found that 20% of all news sessions started with web search while 16% started from social media. Furthermore, 61% of news sessions only involved a single news domain [14]. Unlike traditional media, such as newspapers, there is no structured review process for content published on social media. Thus, the publication and distribution of fake news can go unnoticed. As it is more likely that people share and circulate contents that are false or controversial, fake news spreads rapidly [15], resulting in an exponential increase in misinformation online.

Flintham et al. [16], conducted a survey on news consumption online. The survey revealed that two-thirds of respondents regularly consumed news via Facebook, and one-third had come across fake news they initially believed to be true [16]. A survey by Edson et al. [17] states that 73% of people ignore fake news on social media when encountering it. The minority would either comment, report, or message the publisher to get rid of the content or to correct it. 12.1% would unfollow or block the person [17]. Ignoring fake news on social media as a behavioral trait is also mentioned by Panwar et al. who investigated the impact of fake news on readers' usage behavior on social media [18]. People only moderately investigate the truthfulness of seen posts [19], [20]. Kim et al. [21] conducted a study on an imitated Facebook interface. The study aimed to find a correlation between preexisting beliefs and biases about a topic and engagement with a post. Results showed that the preexisting beliefs had an impact on the story in the post to be perceived as true. Also, reactions to a social media story in the form of likes, shares, comments, and clicks, are influenced by how much readers believe in it [21]. These studies show the importance of investigating the consumption of fake news via social media and its mitigation strategies.

The spread of fake news online has a strong impact on the individuals and communities. By being exposed to fake news the political view can be manipulated [22]. Studies show that people often believe online content because of their laziness to pursue more content to create an opinion about the topic [23]. Furthermore, the illusory truth effect, which is the belief in misinformation after being frequently exposed to it, can be a result of the massive spread of fake news [24]. Fazio et al. have discovered that this still happens, even though a person has knowledge about the topic before reading it [24]. Finally, the consumption and belief in fake news has been shown to have an impact on humans' physical and psychological health [25].

B. Physiological Responses to Fake News

A strand of research looked at physiological responses and physical behaviors of users, collected via sensors when being exposed to fake news online. Moravec et al. investigated cognitive processes during reading news on social media using Electroencephalography (EEG) [10]. They found that while articles flagged as fake increased readers' cognitive activity, it did not impact their beliefs in the reliability and truthfulness of these articles [10]. Hansen et al. [26] conducted a deception-based lab study where participants were asked to read headlines and rate them in order of recency while their eye gaze was being tracked. This served as the basis for an ensemble learner using eye gaze data and showed that there is a significant difference in visual perception of false news headlines than true ones in the form of visual attention. They concluded that false news headlines are being read with less attention compared to headlines with truthful content [26]. Simko et al. [7] looked into the behavior of reading and labeling social media posts as real or fake on a simulated social media feed with article-based posts. In a second run, after being informed about the purpose of the study, the time spent while evaluating the truthfulness of the posts was collected. The results show that people who spent more time investigating the headline than the article had a higher failure rate when judging the post's veracity [7]. Lutz et al. [27] conducted a lab experiment where they collected eye tracking and heart rate data while users were reading news displayed on a screen. They found lower Heart Rate Variability and a larger number of eye fixations per second to be related to a higher probability of fake classification [27].

C. Summary

There is no doubt that the spread of fake news on social media is a challenge. Utilizing human physiological behavior collected from eye gaze or other sensors as an indicator of the factuality of news is a promising research direction.

[2]https://dictionary.cambridge.org/dictionary/english/fake-news

Prior work investigated the relationship between time and reading types of news online [7], eye gaze and heart rate [27] or eye gaze behavior while reading headlines [26]. We build on top of this by (1) adding another source of behavioral data, that is mouse movement, which is a viable indicator of cognitive load and attention [28], [29] and (2) designing and conducting an experiment within the context of social media websites.

III. Concept and Research Questions

We believe human behavior to be a predictor of a person's belief regarding news posts. This knowledge can be leveraged in many ways, including approaches that raise user awareness about the possibility of currently being exposed to fake news.

We set out to explore user behavior when consuming news on social media. In particular, we chose to investigate eye gaze and mouse movement behavior for multiple reasons. Prior work has found differences in eye fixations and saccades while reading news that may be true or false [21], [26], [27], [30]. In addition, eye gaze behavior has been shown to be an indicator of cognitive load in security-related contexts [31]. Mouse behavioral features, such as slower and longer transitions, have been shown to indicate high cognitive load while browsing [29], mind wandering [28] and susceptibility to phishing [32].

In addition, an important objective of our study was to conduct it in an ecologically valid remote setting, mimicking real-life interaction with social media. To this end, we also chose sensors that can be employed in the users' vicinity rather than requiring users to be augmented with on-body sensors.

Prior work used eye trackers to analyze where users look while reading fake news [21]. Hence, we hypothesize users' gaze and mouse behavior to indicate if users are being exposed to fake news. Our hypothesis is based on the belief that fake news triggers unconscious user behavior which can be captured using sensing technologies. Hence, the first driving research question is: How well can we predict fake news from gaze behavior, mouse movements, or both (*RQ1*)? We investigate the best gaze and mouse features reflecting the consumption of fake news. Second, we expect the content type and news domain to affect users' perception and, hence, their consumption behavior, resulting in the second research question: Does the content and post type affects user behavior (*RQ2*)? We compare users' behavior while reading different news types prevalent on social media: text, images, and articles.

IV. Data Collection

To explore our research questions, we designed and developed a platform mimicking a social media website. Users eye gaze and mouse behavior were being collected while interacting and scrolling. We designed and conducted a remote field study. In the following, we explain the developed platform and study instruments, the study design, and the procedure.

A. Study Design: Social Media Platform and News Content

We chose Facebook as our social media of choice as it is currently one of the most popular social media platforms. In a recent survey, around 30% of participants from 12 different countries stated that they use it as a source of news [33]. It is also one of the platforms most prone to sharing (fake) news.

Since the Facebook API does not offer the possibility to get the user feed and manipulate it, we created our own Facebook profile page, imitating a real feed with publicly available user information (for example, name, profile photo). The layout of the feed resembles the actual Facebook feed in structure, style, and icons. It only differs in the content of the posts. The user interface can be seen in Figure 3.

News Categories: We identified five different news categories which we include in our study design, covering a wide range of current topics to be distributed equally through the posts shown. The topics were chosen based on the popularity of news types in Germany. The picked categories were: Health, Environment, Entertainment, and European and US Politics.

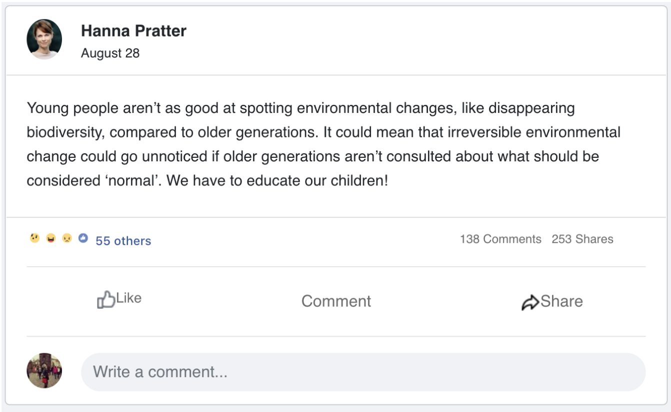
1) *Health:* Since the coronavirus outbreak 2020 and the resulting pandemic, health news is consistently published on a daily basis as one of the main topics.
2) *Environment:* Climate change is a global problem, discussed repeatedly by news media and politics. The debate on the consequences of global warming as well as the large amount of misinformation online about the topic makes it a relevant one for our study.
3) *Entertainment:* This category covers a broad spectrum of topics often concerning media and celebrities. Fake news covering movie actors and other celebrities often spreads through social media. Hence we considered entertainment a news category in our study.
4) *Politics:* Political news is one of the most popular topics to inform people about current events in the world and notify citizens about political decisions. Since the study was performed in Germany with mostly German and other European participants, we decided to use German/European politics as a category. Since American politics are also popular among German citizens we decided to include American politics as a post category as well.

Post Types: There are three types of posts we investigate in our work that normally appear on a Facebook news feed: text-based posts, image-based posts, and article-based posts. We chose not to explore video-based posts on social media currently as videos are a complex media type and may need to be separately investigated (for example, deep fakes).

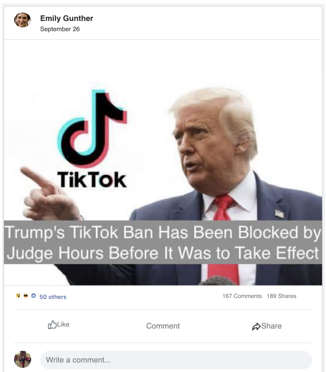
- *Text-based posts:* contain solely text without media.
- *Image-based posts:* contain a photograph or illustration which conveys the main message. They can also include text as a description.
- *Article-based posts:* contain a link to an external website, which provides an article in newspaper form. The post contains an image of the linked website, the headline, and the first few lines of the article.

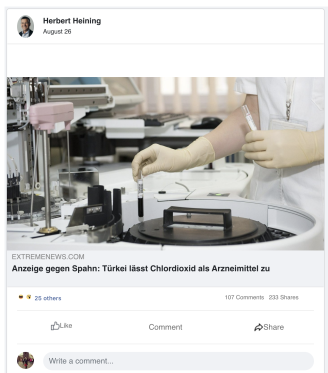
All three types of posts have a post header and post footer in common (Figure 2). The post header consists of the publisher/author image, the name of the publisher/author, and the timestamp of the publishing. The footer contains the number of likes/reactions/comments and shares, as well as the possibility to react by liking, commenting, or sharing the post.

Posts Data Set Creation: After defining the categories and post types, we set out to collect and create a data set of real and fake posts that we describe in the following.



(a) Text-based post





(b) Image-based post (c) Article-based post

Fig. 2: Sample of the three types of posts used in our study: text, image, and article-based posts

First, we determined the time frame for scrolling through posts in our study. As the focus of our study is exploring the overall behavior of consuming news on a Facebook feed, which includes scrolling, seeing, reacting, and reading the posts (e.g., by opening an article in a new tab), we used 7 minutes as a task duration, inspired by Simko et al. [7] ($Mean = 5; SD = 2$).

We calculated the number of posts depending on how long it takes for the social media user to consume media feed content. Based on Facebook studies the average time spent on each post in order to consume the content is 2.5 seconds for desktop users[3]. For article-based posts We decided to double the time that may be spent reading them, given that the article would be opened in a new tab. We intended to distribute the post types equally. Therefore, based on the average time spent per each post, 2.5 s and 5 s, we created 120 posts, 24 in each news category (health, entertainment, environment, European and US politics) and 40 in each news type (article, image, and text) half of which are real and the other half is fake.

We followed two different approaches for creating the real and fake posts for the study. First, fake news were taken from fact-checking web sites as follows:

Health Feedback[4] and *Climate Feedback*[5] were both created to help readers of online content verify which news to trust in the area of health and environment. They are sub-pages of *Science Feedback*[6], a non-profit organization verifying the credibility of influential claims and media coverage that claims to be scientific. Scientists comment on media articles that are science-based and rate the credibility of them.

Regarding entertainment topics, *Gossip Cop* fact-checks celebrity gossip to reduce misinformation about entertainment, celebrities, and royal families. Next to Gossip Cop, there is *Snopes*[7], another fact-checking resource. It also contains a lot of fact checks about misinformation regarding entertainment but also other topics like technology, health, and politics.

One of the most popular fact-checking websites regarding politics is *Politifact*[8], which is cooperating with Facebook and Tiktok to slow down the spread of misinformation. The reporters aim to fact-check statements made by political officials or about online content concerning US politics.

Regarding German politics the fact-checking website *Correctiv* was used. It provides fact-checking also for other topics, such as health, environment and foreign countries[9].

Real posts were collected from popular real news websites in Germany as a basis for our simulated feed. After collecting all 120 post contents – 60 valid and 60 fake – we designed and formatted the posts according to the previously decided post types. Note that for all post categories, we chose posts that are marked by the websites as entirely false/true.

As explained earlier, we created the header and footer of the posts. For the publisher/author, we either used the image and name of the real publisher of the content, as mentioned on the fact-checking website, or we used a fictive name and profile picture, meant to be from friends on the news feed. As timestamps, we used the publication date of the used content. As the number of likes / shares / comments likely influence the credibility of the news and might change their users' behavior and physiological state, we chose a similar number of comments shares for all types and categories of posts. The number was a chosen randomly between 200 and 300.

B. Participants and Recruitment

We recruited participants through University mailing lists, social media groups, and personal contacts. The study requirements were to have a social media profile on Facebook and own a laptop with a camera. Each participant accomplishing the study was reimbursed with a 5 Euro voucher or a reimbursement point for university-specific studies. In total, 54 participants completed the study, out of which 28 were female, and 26 were male. The average age was 25.8 (SD=7.7, range=18-58). The participants had a wide range of educational backgrounds: currently completing undergraduate studies (25 participants), bachelor degree (13), masters degree (13), one participant finished an apprenticeship and two did not have a degree. The majority of the participants (37) were students.

[3]https://www.facebook.com/business/news/insights/capturing-attention-feed-video-creative

[4]https://healthfeedback.org/

[5]https://climatefeedback.org/

[6]https://sciencefeedback.co/

[7]https://www.snopes.com/

[8]https://www.politifact.com/

[9]https://correctiv.org/faktencheck/

C. Apparatus

Gaze data estimation as performed using webcam data. We used the GazeCloudAPI.js tool[10] to enable gaze estimation on our interface. Since we depend on webcams, data collection was conducted with a rate of 30 frames per second. We saved users' raw gaze data (XY position) associated with the post ID they are scrolling over. For capturing the mouse movements, we used native HTML to save the mouse position whenever an event was fired. We saved mouse clicks, double clicks, hovers, and selections, as well as the ID of the post it is inside. Both mouse and gaze data were synced using timestamps.

D. Procedure

The participants visited a URL that was presented to them in the study invitation. The linked website included the study instructions and procedure shown in Figure 4. The study started with an introduction and an explanation of the task as follows:

> *We will ask you to sign in with your Facebook account. Only information publicly available will be used (i.e. your name, and profile picture). After you have signed in with your Facebook account you will be forwarded to a Facebook feed. Your task will be to casually scroll through your Facebook feed. Just behave normally and read, react as you would normally do. This part of the study will end after seeing around 80 posts or after spending around 7 minutes. A button on the right panel will then be activated, enabling you to move to the next stage.*

Participants were then shown a consent form that informed them that only their public Facebook profile will be accessed and that their webcam feed and mouse movements will be recorded for the study purposes. After the participants agreed to the consent form, they were instructed to ensure that they are sitting in a well-lit room at an adequate distance from the webcam. We explained that they will then perform a webcam calibration. For this, the Gaze Recorder API, employing a standard 9-point calibration[11], was used. Participants were then directed to the Facebook feed showing their public profile information and the real/fake posts from our created data set. The order of the posts was randomized per participant. After the 7 minutes or at least 80 posts a button on the right panel appeared through which they could move to the next stage or continue scrolling for seven minutes.

Participants were then directed to the eye-tracking accuracy test. Here, participants had to follow a red circle with their eyes on the screen which appears in nine positions once at a time, three per row, similar to the calibration phase. The main aim of the accuracy test was to check gaze data accuracy at the end of the study to be able to eliminate users with a strong gaze shift (cf. section VI-A1). During the process, the circle number, the circle position, the eye position, and the check if it was a valid gaze entry were collected and saved. After the accuracy test was completed, participants were directed to a post-labeling task, in which they were asked to label each post they had seen as real or fake. Finally, they were asked to fill out a post-study questionnaire.

[10]https://gazerecorder.com/gazecloudapi/

[11]Gaze Recorder Calibration: https://gazerecorder.com/faq/

The post-study questionnaire first collected demographic data: age and gender, the highest level of education, and the current employment status. We asked participants how much they use Facebook: never, less than once a month, once or twice a month, once a week, once or twice a week, once a day, and more than once a day. We asked about their intentions of using Facebook and their interest in the five chosen news category topics. We used a five-point Likert scale (1=not interested at all; 5=very interested).

We alaso asked based on which aspects participants considered posts to be true or fake. Possible options for answering were the publisher of the post, quality of the image, the content of the article (read by opening tab), text content of the post, number of likes, number of shares, and the possibility to write an own answer. Multiple choices were possible.

E. Limitations

Since our main goal was to collect behavioral data in an ecologically valid setup through a remote field study, we acknowledge some limitations to our study. We clearly stated that the participants should be seated in a well lit quiet room and not interrupted. However, we could not ensure this. In our data cleaning step, we accounted for such uncontrollable issues by assessing eye tracking accuracy and looking at the time spent for the entire study. We removed any data sets with obvious discrepancies or incomplete results.

V. Feature Extraction & Classification Approach

In the following, we describe our step-by-step process to extract eye gaze and mouse movement features from the collected data for fake news detection. In our data set, for every timestamp of behavioral input a data point consists of the current fixation position, fixation duration, mouse position, mouse duration, and whether users looked or pointed at a post. Based on these features, we further extracted other gaze and mouse features per post. We highlighted three main areas of interest (AOI) which we believe are important while reading fake and real news. These AOIs are post header, post content, and post footer. The post header includes the name of the person who posted it and the time. Poster content is the actual content (consisting of different types, as explained earlier), Finally, the post footer includes the number of likes, comments, and shares as well as the space for likes and comments.

A. Gaze Features

We extracted saccades and fixations from the raw data. Fixations are defined as maintaining the gaze on a single location [34]. Saccades are defined as the rapid eye movement that shifts the center of gaze from one fixation to another [35]. We identified fixations using the Dispersion-Threshold Identification algorithm [36]. Then we extracted a set of six main gaze features, inspired by prior literature [26], [30], [31].

- **Fixation Count**: This feature provides the number of total fixations inside one post.
- **Average Fixation Duration**: The fixation durations were grouped by post and summed up to obtain the average fixation duration.

Fig. 3: Imitated Feed with participant data

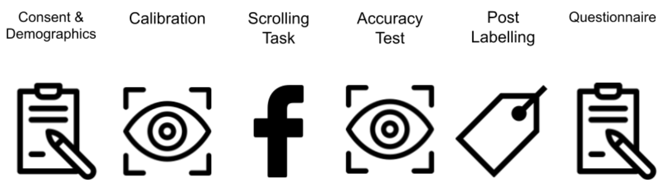


Fig. 4: Study Procedure

- **Average Fixation Distance**: The distances of fixations inside a post were calculated by using the start and end position for each fixation. We then summed them up to obtain the average fixation distance.
- **Average Saccadic Duration**: The saccadic duration is calculated by subtracting the timestamps of two consecutive fixations. The average saccadic duration per post was determined.
- **Average Saccadic Length**: Similarly, the average distance between two fixations per post was calculated.
- **Gaze Duration inside Post**: For each post, the total duration of gaze movements was calculated. We calculated the duration per AOI (header, content, footer).

B. *Mouse Features*

Eight mouse features were calculated from the mouse position and actions, inspired from prior work [29], [32].

- **Number of Mouse Actions**: This features includes the total number of actions per post (hovers, clicks, double clicks, or selections).
- **Mouse Hover**: We extracted the number of mouse hovers inside a post. A mouse hover was defined as the mouse position not changing between 100 ms and 3000 ms [32].
- **Mouse Hover Ratio**: The mouse hover ratio was calculated by dividing the number of mouse hovers by the non-mouse hovers inside a post.
- **Slow Mouse Movement**: We calculated slow mouse movements and saved the number of occurrences in a post. We divided every mouse movement speed by the average velocity of the mouse movements in the post. If the value was below 0.25 we considered it a slow mouse movement, as suggested by Yu et al. [32].
- **Slow Mouse Movement Ratio** Same as the mouse hover ratio we calculated the slow mouse ratio by dividing the number of slow mouse movements by the number of non slow mouse movements per post.
- **Average Mouse Distance** The distances of mouse movements were summed up and divided by the total number to get the average.
- **Mouse Duration inside a Post**: We calculated the overall duration of the mouse inside a post. Again, we separately looked at header, content, and footer.
- **Number of Mouse Click**: We extracted the mouse clicks inside a post. We distinguished between a simple mouse click, a double click, and a mouse selection. All were calculated separately and the number of clicks and selections was used as own features per post. We used the features for the header, content, and footer area of the posts as their own features as well.

C. *Other Features: Time*

Finally, we calculated the time spent on a post. For that, we combined the mouse and gaze actions per post. We used the end time of the last entry for a post and subtracted it from the starting time of the first entry of that post.

D. *Classification Approach*

We present a classification approach to map a feature vector of behavioral data (gaze and mouse features), computed from a time window of data to one of the classes corresponding to the news type (fake vs. real). As behavioral data differs between users, we test user-dependent classifiers, to understand the personalized behavior on an individual basis by using a leave-one-out approach. We additionally explore user-independent classifiers. We test three different feature sets: 1) gaze only features, 2) mouse only features, and 3) combined features.

We compared the performance of three classifiers that are mostly used in the literature [31], [37]: Support Vector Machines (SVM), Logistic Regression (LR), and Random Forest. We chose different types of classifiers categories. To optimize performance, hyper parameters for each classifier were empirically optimized on a small set of values.

1) User-Dependent Classifier: The user-dependent classifier is trained on the data from each user separately. First, we clean the data by removing the data outside our areas of interest (i.e., post area). After that, we calculate the features for both gaze and mouse movements. The collected data is synchronized using the timestamp for the analysis. This is followed by splitting the data into training and test data (leave-one-out approach). Then, we applied 10-fold cross-validation. We then report the average accuracy of all participants.

2) User-Independent Classifier: We split the data into training and test data, using the leave-one-out approach. This time, we left a whole participant data set out and then applied 10-fold cross-validation. Finally, we averaged the accuracies.

VI. RESULTS

We present and analyze the collected data from our remote field study. We first start with data cleaning and pre-processing, present insights from the labeling tasks, present the statistical analysis of the computed gaze and mouse features, and finally the machine learning classification outcomes.

A. Data Cleaning

We pre-processed the collected data based on 1) the eye tracking accuracy test, and 2) the study completion time. In addition, we removed gaze and mouse data outside our three areas of interest (post header, post content, and post footer).

1) Accuracy Test: To check the results of the accuracy-test data collected, we followed the guide for accuracy testing provided by Tobii [38]. The accuracy value should not exceed 1 degree, that is the gaze point should not have a distance larger than 55 px from the target circle. The vertical and horizontal accuracy values had to be calculated separately. The formulas for calculating the accuracy values were:

$$Acc_{vertical} = \left(\frac{\frac{\sqrt{(xCircle - xGaze)^2 + (yCircle - yGaze)^2)}}{2}}{\frac{distanceToScreen * displayResolutionHeight}{displayDimensionHeight(inmm)}}\right) * 2$$

$$Acc_{horizontal} = \left(\frac{\frac{\sqrt{(xCircle - xGaze)^2 + (yCircle - yGaze)^2)}}{2}}{\frac{distanceToScreen * displayResolutionWidth}{displayDimensionWidth(inmm)}}\right) * 2$$

We calculated the accuracy value of each viewpoint. We computed the screen width and height from the dots shown in the accuracy test. The ninth dot was at the position of 100 pixels less than the screen width and 100 pixels less than the screen height. Due to the nature of our study being conducted in naturalistic settings, we could not discern clearly the distance to the screen. Since all of our participants completed the study on a laptop we used a distance of 40 cm which is the minimum preferred viewing distance[12]. The accuracy test resulted in having all accuracy values for all participants under 1 degree. Thus, the data sets for all users finishing the study could be used for further analysis.

2) Short Study Duration Period: The average duration of the study was six minutes. While 42 participants were interacting with the feed for a period longer than six minutes, four participants spent less than four minutes. Hence, we chose to eliminate these four participants from the data analysis as we assumed they scrolled too fast without consuming the content.

[12]https://lookafteryoureyes.org/eye-care/screen-use/

TABLE I: Participants' post labeling (percentage, and count) presented by post type (text, image, article) and post veracity (fake and real). Correctly labeled posts are highlighted in bold.

Post Type		labeled as Real	labeled as Fake
Text	Real	**29.2% (141)**	19.5% (94)
	Fake	14.7% (71)	**36.6% (177)**
Image	Real	**29.9% (175)**	17.9% (105)
	Fake	17.1% (100)	**35.1% (206)**
Article	Real	**34.6% (185)**	15.0% (80)
	Fake	12.0% (64)	**42.4% (205)**

B. Data Overview

For both gaze and mouse data, we sampled data at 30 Hz from the eye tracker and from mouse events. This led to an average of 12600 samples per user, resulting in overall 630 K samples for all participants. Due to different scrolling and reading behavior, participants saw an overall of 779 real posts and 823 fake posts. The seen posts contained 483 text-based posts, 586 image-based posts, and 534 article-based posts.

C. Labeling Task Results

1) Labeling Task Analysis per News Type and Category: We first analyze participants' perceptions of the seen posts by type and category. Table I shows participants' labeling performance. We found that 54.1% of the judgments classified a post as fake. With a small difference, text-based posts have a higher rate of being judged as fake (56.1%) than image-based (53.4%) or article-based posts (53.1%). However, the results show that image-based and text-based posts are misjudged (i.e. judged wrongly) more often than article-based posts. 35% of image-based and 34.2% of text-based posts are misjudged while only 27% of article-based posts were misjudged. This answers to *RQ2*.

Furthermore, two-thirds of the posts seen in total were judged correctly. Our findings here indicate that text-based and image-based posts are more likely to be wrongly interpreted, compared to article-based posts where participants can click on the links and further read about the content.

For labeling per news category, we found that overall, participants saw 298 environment news posts, 327 health, 332 environment, 333 European politics, and 313 American politics posts. Table II reflects participants' labeling per post category. While posts about entertainment, health, environment, and German/European politics are more often labeled as fake than real, more fake news are labeled as real in the American politics category (35.1% of posts in the category are misjudged). This might be due to our sample from Europe, with participants being less familiar with global political news. The environment news category achieved the highest score with 72.6% of correctly labeled posts. The rest of the categories all had similar scores with 64% to 68% of posts correctly labeled and around 31% to 35% of wrongly labeled posts. This confirms, in line with the literature [10], that participants struggle with accurately detecting fake news.

2) Labeling Response Time: Analysing the time spent with each post, we found that participants took longer to label fake posts compared to real posts. Participants required on average 9.6 s to label post as fake – which is longer than

TABLE II: Participants post labeling (percentage and count) presented by post category (entertainment, health, environment, European and American politics) and post veracity (fake and real). Correctly labeled posts are highlighted in bold.

Post Category		**labeled as** Real	Fake
Entertainment	Real	**29.2% (87)**	15.8% (47)
	Fake	16.1% (48)	**38.9% (116)**
Health	Real	**30.9% (101)**	22.3% (73)
	Fake	11.6% (38)	**35.2% (115)**
Environment	Real	**33.1% (110)**	15.4% (51)
	Fake	12.5% (40)	**39.5% (131)**
European Politics	Real	**30.6% (102)**	18.9% (63)
	Fake	13.2% (44)	**37.2% (124)**
American Politics	Real	**32.3% (101)**	14.4% (45)
	Fake	20.8% (65)	**32.6% (102)**

TABLE III: Overall duration of news labeling per post type. Correctly labeled posts are highlighted in bold.

Post Type		**labeled as** Real	Fake
Text	Real	**6.1 sec**	8.2 sec
	Fake	8.8 sec	**12.7 sec**
Image	Real	**8.8 sec**	8.7 sec
	Fake	11.0 sec	**8.4 sec**
Article	Real	**9.3 sec**	7.7 sec
	Fake	10.0 sec	**9.9 sec**

when labeling post as real (8.8 s). There is a difference between comparing the duration by post type or by category. We also found that participants required less time labeling text-based posts compared to image and article-based ones.

The participants tend to spend more time inspecting a text-based post before classifying it as fake (Mean=11.1 s) than classifying it as valid (Mean=7.0 s). Table III shows the amount of time participants spent on a post with real content and fake content based on the post type. It also shows the average time of consuming actual real or fake posts and the duration of the consumption with a resulting right or wrong judgment. The time spent on text-based posts which are fake is on average 11.6 s – almost twice the time spent on real content (6.9 s). Also, the time spent on a post to then label it incorrectly as fake real is more for all three post types than incorrectly labeling it as true. Overall, participants spent an average of 9.3 s on text-based, 9.0 s on image-based, and 9.4 s on article-based posts.

Regarding the news categories, Table IV shows the labeling duration per news category. As seen, participants spent most time reading health posts before judging them (9.7 s), whereas the least time was spent on entertainment posts (7.8 s). Overall, participants tend to focus longer on posts they later classify as fake. Our findings suggest that reading and labeling fake news induces cognitive load, as indicated in literature showing that task response time can be used as a metric of cognitive load when participants are engaged in a primary task [39].

D. Behavioral Analysis

Below we reflect on the statistical analysis of both mouse and gaze behavior while reading real and fake news. Unless otherwise stated, data were non-normally distributed (confirmed by Shapiro-Wilk and Anderson-Darling tests). We performed non-parametric tests. We report mean values (M).

TABLE IV: Overall duration of news labeling per post category. Correctly labeled posts are highlighted in bold.

Post Category		**labeled as** Real	Fake
Entertainment	Real	**7.1 sec**	7.5 sec
	Fake	9.2 sec	**8.3 sec**
Health	Real	**9.2 sec**	9.3 sec
	Fake	10.9 sec	**11.6 sec**
Environment	Real	**9.4 sec**	7.6 sec
	Fake	7.3 sec	**12.4 sec**
European Politics	Real	**8.2 sec**	7.7 sec
	Fake	11.8 sec	**9.7 sec**
American Politics	Real	**6.8 sec**	8.9 sec
	Fake	10.7 sec	**8.8 sec**

TABLE V: Wilcoxon signed-rank tests for the mouse features. Reading and classifying real and fake news have an effect on most mouse features (Significant results in bold, $P < .05$)

Mouse Features	**Real** Mean	SD	**Fake** Mean	SD	**Wilcoxon** Z	P
Avg. Speed	**.0034**	**.0063**	**.1030**	**.1285**	**-6.093**	**<.001**
Hover Ratio	**.3575**	**.4088**	**.2692**	**.3475**	**-2.857**	**.004**
Hover Counts	.9198	.7241	1.0	1.4	-.224	.823
Action Counts	**2.6**	**2.3**	**7.0**	**7.8**	**-5.585**	**<.001**
Slow Mouse Mov. Rat.	**.0061**	**.0172**	**.0768**	**.1143**	**-5.556**	**<.001**
Slow Mouse Mov.	**.0761**	**.2192**	**1.8**	**2.4**	**-5.840**	**<.001**
Avg. Distance	91.9	86.9	98.5	92.9	-1.278	.201
Post Mouse Dur.	**3860.9**	**3402.2**	**26761.2**	**28804.6**	**-5.973**	**<.001**
Click Count	**.1941**	**.3585**	**.4869**	**1.2**	**-2.212**	**.027**
Double Click Count	.0071	.0291	.0229	.0851	-1.260	.208
Selection Count	.0298	.1278	.0949	.3208	-1.680	.093
Content Mouse Dur.	3296.4	3011.1	4309.3	4539.3	-1.507	.132
Footer Mouse Dur.	456.4	616.9	428.0	639.4	-.025	.980
Header Mouse Dur.	**108.2**	**216.3**	**57.8**	**105.4**	**-2.562**	**.010**
Other Mouse Dur.	1964.4	1953.3	2430.4	3731.6	-.035	.972
# Content Click	**.0971**	**.1659**	**.0407**	**.2296**	**-2.709**	**.007**
# Footer Click	**.0951**	**.2319**	**.0321**	**.2752**	**-5.059**	**<.001**
# Header Click	**.0019**	**.0136**	**.0839**	**.0829**	**-6.032**	**<.001**
# Other Click	**.0807**	**.1561**	**.0116**	**.2021**	**-3.163**	**.002**
# Content Double Click	**.0007**	**.0051**	**.0752**	**.0565**	**-5.970**	**<.001**
# Footer Double Click	**.0063**	**.0278**	**.0815**	**.0785**	**-6.094**	**<.001**
# Header Double Click	**.0000**	**.0000**	**.0866**	**.0811**	**-6.094**	**<.001**
# Other Double Click	**.0000**	**.0000**	**.0752**	**.0565**	**-5.970**	**<.001**
# Content Selection	**.0208**	**.1084**	**.0613**	**.1222**	**-4.835**	**<.001**
# Footer Selection	**.0090**	**.0422**	**.0759**	**.0914**	**-5.632**	**<.001**
# Header Selection	**.0000**	**.0000**	**.0866**	**.0811**	**-6.094**	**<.001**
# Other Selection	**.0131**	**.0660**	**.0771**	**.0825**	**-5.737**	**<.001**

1) Mouse Features Analysis: We analyzed mouse features and click behavior. Overall, we found a statistically significant effect of news type on mouse features using a Friedman test (χ^2(53)= 2081.9, P<.001). Table V summarizes findings for mouse features. Overall, using a Wilcoxon test for pairwise comparisons, we found that reading fake news significantly affects most mouse movements. For example, while reading fake news, participants spent more time hovering over the post, performed more mouse actions such as clicks and double clicks, and performed more slow mouse movements. This indicated that reading fake news induces cognitive load, as confirmed by literature showing that slower and longer mouse movements are indicators of high cognitive load [29]. Moreover, more slow mouse movements show that users were attentive while reading the posts and, hence, correctly identified them [28], [32].

2) Gaze Features Analysis: Similar to the mouse movements, we analysed users' gaze movements and features. Overall, we found statistically significant effects of news type on gaze features using a Friedman test ($\chi^2(17) = 686.9, P <$

TABLE VI: Wilcoxon signed-rank tests for gaze features. Reading and classifying real and fake news have an effect on some gaze features (significant results in bold, $P < .05$).

Gaze Features	Real Mean	SD	Fake Mean	SD	Wilcoxon Z	P
Avg. Fix. Dur.	3549.5	14685	3970.2	15578.2	-1.457	.145
Fixation Count	**2.4**	**3.1**	**4.2**	**5.4**	**-5.287**	**<.001**
Avg. Fixa. Dist.	141.9	109.7	131.1	73.5	.542	.588
Avg. Sacc. Dur.	7109	5974.7	8713.3	12197.7	.174	.862
Avg. Sacc. Len.	217.9	98.3	222.6	79.2	.174	.862
Post Gaze Dur.	**1465.9**	**2281.2**	**3381**	**5787.3**	**-5.227**	**<.001**
Content Gaze Dur.	1313.6	2107.7	1398.1	2681.8	.124	.901
Footer Gaze Dur.	129.9	279.5	401.8	2177.6	-1.248	.212
Header Gaze Dur.	22.3	50.2	15.9	26.9	-1.378	.168

.001). Pairwise comparisons using a Wilcoxon test was not significant for many features (see Table VI). We only found a statistically significant effect of news type on number of fixation and overall gaze duration on the posts. Our analysis shows that while reading fake news, participants had significantly more fixation and spent more time on the post. As the duration per post and fixation duration is longer, this indicaties that users were struggling while reading fake news due to higher cognitive load [40], [41]. Generally, we found that more fixations and longer fixation durations occur when reading news headlines, which is again in line with the literature [26].

E. Qualitative Analysis of Post Questionnaire

In our post questionnaire, we had asked participants due to which aspects they classified posts as real or fake. 42 participants most often mentioned the text content of the post, 35 people mentioned the publisher of the post is important for the decision, 33 people said mentioned reading the article posted by opening it in a new tab, and 29 see the quality of the image as an indicator for the classification. Three people said that they look at the number of likes and shares. Only one person mentioned looking for the source of the information.

The interest levels differed for each category. The highest interest levels were found for health and environment, providing a possible explanation for why the veracity ratings in the environment category was highest. The least interesting category for participants was entertainment.

F. Machine Learning Classification Results

We compared the performance of three different models: SVM, random forest, and logistic regression. We conducted two classifications: user-independent and user-dependent classifiers. Below, we reflect on each of them.

1) User-Dependent Classifier: As users' reading behavior is unique, we built a user-dependent classifier. Table VII shows the overall performance for each classifier across the different features. Although SVM and LR yielded similar accuracy, the SVM resulted in better accuracy in most cases. Hence, we will focus on and report the SVM results. As seen, our classifiers were able to predict fake news by 64.2% from users' gaze features only and 63.9% from mouse features only. Finally, by combing both features, our classifier was able to predict fake news with 64.2%. Although the difference in accuracy is not substantial between mouse and gaze features, gaze features are slightly more accurate and the classifier with both features together provide a higher accuracy than each on its own.

TABLE VII: User-Dependent Classification Accuracy: standard deviation for the different classifiers (random forest, logistic regression and support vector machines) across different feature sets with best accuracy in bold.

Classifier		Gaze Features	Mouse Features	Both Features
RF	**Accuracy:**	62.6% ± 17%	61.3% ± 19.8%	61.7% ± 22%
	AUC:	54.2%	51.8%	53%
LR	**Accuracy:**	63.2% ± 19.6%	61.9% ± 20%	64.3% ± 19.7%
	AUC:	50.6%	50.1%	53.3%
SVM	**Accuracy:**	**64.2% ± 18.3%**	**63.9% ± 18.3%**	**64.2% ± 18.3%**
	AUC:	50%	50%	50%

TABLE VIII: User-Independent Classification Accuracy: standard deviation for the different classifiers (random forest, logistic regression and support vector machines) across different feature sets with best accuracy in bold.

Classifier		Gaze Features	Mouse Features	Both Features
RF	**Accuracy:**	57.9% ± 3.5 %	53.3% ± 6.5%	49.2% ± 6.5%
	AUC:	48.9%	48.4%	48.3%
LR	**Accuracy:**	66.7% ± 2.7%	63.3% ± 6.4%	62.5% ± 8.7%
	AUC:	49.8%	50.6%	49.8%
SVM	**Accuracy:**	**67% ± 3%**	**65.9% ± 2.2%**	**68.4% ± 2.9%**
	AUC:	50%	50%	50%

2) User-Independent Classifier: To understand the generalizability of our approach, we created a user-independent classifier. Table VIII shows the overall performance for each classifier across the different features. Similar to the user-dependent results, we also found that SVM gave better results for all features. Hence, we will only report the SVM results. Our classifiers can predict fake news with 67% accuracy from users' gaze features only and 66% from mouse features only. By combing both features, our classifier was able to predict fake news with 68.4% accuracy, answering *RQ1*). Similar to the user-dependent classifier, we also found slight differences between mouse and gaze features classification accuracies. However, gaze features are slightly more accurate.

The user-independent classifier provided higher classification accuracy than the user-dependent one (around 4% higher). We hypothesize that this is due to differences in users' behavior which is reflected in the classifier's standard deviation, reaching 18%. This means that the classifier can provide an accuracy of 46% to 82%, depending on the user.

3) Feature Importance: We investigated which features contribute to the accuracy of the classifiers. As user-independent classifiers gave better accuracy, here we show the features for user-independent classifiers only. We used SHAP [42], a tool that explains the output of a machine learning model by computing the contribution of each feature to its prediction. Figure 5 shows the feature importance. We observed that for the gaze features, average saccadic and fixation duration strongly affect the classification accuracy. For the mouse features, we observed that mouse duration in the header, mouse clicks on the content, and mouse selection count for the header are the most important features for classification. For both features, we found that mouse features have a stronger influence on the model's accuracy than gaze features.

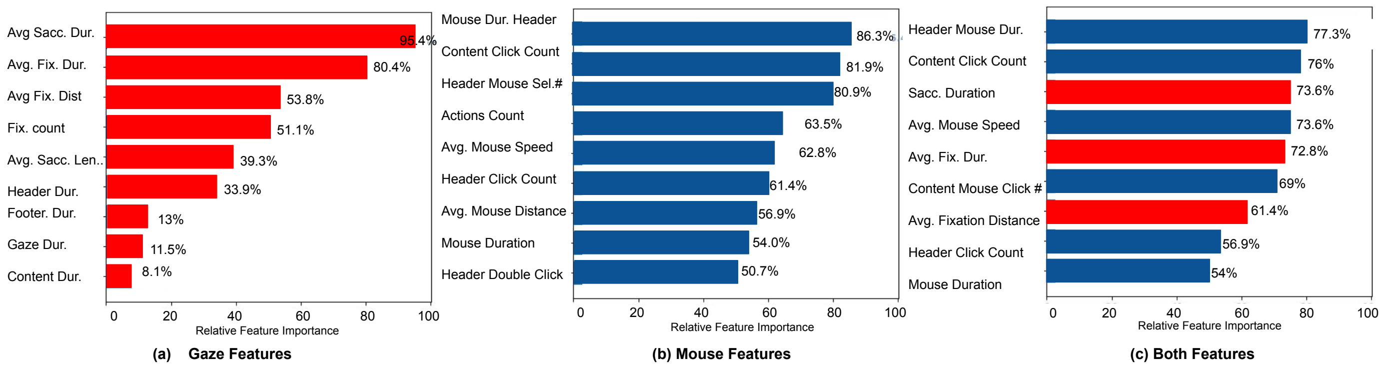


Fig. 5: Results of the feature importance analysis across the tested feature groups for the user-independent classifier

G. Summary

From our results, we can see that reading and classifying real and fake news affect users' task response time, and mouse and gaze behavior. We conclude that reading fake news induces cognitive load. In addition, it induced cognitive load, reflected in users' fixation count and slow mouse movements. Finally, our classification models showed that we can identify fake news from users' behavior with an accuracy of 68% for a user-independent model working across all users.

VII. Discussion and Future Work

A. Classification Accuracy and User Behavior

From our classification results, we found that gaze features provide a better accuracy than mouse features as well as features from both sensors together. Although the differences between gaze and mouse data are not huge, gaze features were shown to reflect unique user behavior which affects the classification accuracy. This unique behavior is reflected in the way users read and skim the posts, highlighting that users' reading behavior has a strong impact on classification accuracy [7]. Another influencing factors is whether users read with their eyes only or use the mouse as an anchor for their readings. We hypothesize that this is the reason for which gaze data yielded better classification accuracies. However, as this is the first attempt to utilize gaze and mouse to judge fake news on social media during a natural task (which is scrolling with no restrictions), there is room for enhancing the overall accuracy of the classifiers. Future work can focus, for example, on only using the best-performing features from our list of eye gaze and mouse features, instead of using all of them. Future work can also introduce more features, such as the ratio of fixations on the AOIs, which can reflect different reading behaviors and, thus, posted news' trustworthiness. Another approach can be collecting more data over a longer period of time and recruiting more participants.

B. User Perception of Real and Fake News

We have chosen to conduct our study in a real and uncontrolled setup to try and capture users' real-world behavior while consuming real and fake content on social media. Table I) shows that while 36.7% of the posts seen by participants were correctly identified as fake, they misjudged 14.7% by labeling the posts as true whereas they were fake. This means that 29% of all fake posts seen were labeled as true. This validates the challenge of identifying fake news on social media platforms and further motivates the need for action against fake news. Overall, 32% of all posts were wrongly identified as fake or real. However, participants might have been primed to re-think the information they are consuming when specifically asked to rate the veracity of the posts so that they identify content they have never heard of, but that is actually true, as fake.

Our study has shown that participants had the highest rate of judging a post as true while it was fake for image-based posts. 33% of the posts containing misinformation in the form of an image were later labeled as true. This motivates future research on technical approaches to detecting fake news online.

Reflecting on the post categories and types, we found that article-based posts are more often labeled correctly compared to image-and text-based posts. This might be due to the amount of content provided in each post type: text-based and image-based posts tend to be short, which may make it harder to judge from a small amount of text, compared to more text and the ability to check the link for more information in article-based posts. This is also reflected in our finding that reading and labeling fake content induces cognitive load (and longer task duration compared to real news). Future work should investigate the length of the posts and its relationship to users' ability to correctly label the posts. At the same time, the familiarity of the topic might also have an impact on users' behavior. Although we asked participants about their interest in the presented topics, we could not find any correlation between their interest and correctly labeling the posts. Future work should investigate the topic familiarity in correlation to the participant's ability of correctly labeling the posts.

Finally, we used a similar number of likes / comments / shares for all posts. Future work should investigate the effect of different numbers on users' ability to correctly label posts.

C. Mouse and Eye Gaze Behaviors When Consuming News

Participants' mouse and gaze behavior showed interesting tendencies in our study. They performed more mouse hovers on posts they later labeled as true, but also had a higher hover rate on real posts and on posts that they correctly identified as fake or real. Additionally, slower mouse movements occur more often for posts later classified as true. An explanation might be that when trying to identify the post veracity while using the mouse, hovers and slow mouse movements indicate more concentration on the content. Whereas we do not know of similar studies investigating mouse movements and fake news, prior research has investigated mouse behavior, for example, for detecting phishing awareness [32] or for detecting attention [43] and cognitive load [29]. We see this as an opportunity to further investigate mouse behavior features in larger data sets.

A similar behavior is also observed for gaze fixations. Users tend to fixate more on posts they correctly identify as real or fake. Also, the average fixation duration on a post labeled correctly was found to be higher than on misjudged posts. This is in line with prior work on identifying real / fake headlines [26]. Participants were able to correctly label posts when their behavior reflected more focus (more fixations, slower mouse movements, high hovers).

D. Implications for Fake News Behavior Detection

As we presented earlier, current efforts to detect fake news and address its quick spread are still far from perfect. A combination of tools and techniques – machine learning, manual fact-checking, and raising awareness through educating users – can lead to better results. With the continuous advancement of sensing technologies (e.g. eye tracking, physiological sensing) and their inclusion in everyday devices, such as smartphones, wearables, and laptops, we expect rich data to be available in many contexts. Hence, behavior sensing can be used to build user-dependent models and help more quickly detect the veracity of a seen post. Subsequently, feedback can be provided to the user. More specifically, the user could receive a warning message not making them aware of fake news, but also nudging them to flag the news or make them reconsider whether or not to share it.

VIII. Conclusion

We investigated human behavior on social media when exposed to fake and real news. For this case, we created a remote online study in which 54 participants had to scroll through a social media interface that resembled Facebook. The feed shown to participants contained text-based, image-based, and article-based posts in the categories of entertainment, environment, health, German/European politics, and American politics. We collected users' gaze and mouse behavior during scrolling through the news feed. The data served as the basis for a predictive model. Our results show that fake news can be detected from users' behavioral data. An explanation is the higher cognitive load while reading fake news. Our work lays the foundations for interventions that flag or report posts for further investigation.

Acknowledgment

The presented work received funding from the German Research Foundation (DFG) under project no. 316457582 and from the Studienstiftung des Deutschen Volkes. This research was also funded by dtec.bw – Digitalization and Technology Research Center of the Bundeswehr (Voice of Wisdom). dtec.bw is funded by the European Union – NextGenerationEU.

References

[1] G. Pennycook and D. G. Rand, "Fighting misinformation on social media using crowdsourced judgments of news source quality," *Proceedings of the National Academy of Sciences*, vol. 116, no. 7, pp. 2521–2526, 2019.

[2] M. Granik and V. Mesyura, "Fake news detection using naive bayes classifier," in *2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON)*. IEEE, 2017, pp. 900–903.

[3] N. Kshetri and J. Voas, "The economics of "fake news"," *IT Professional*, vol. 19, no. 6, pp. 8–12, 2017.

[4] S. Yang, K. Shu, S. Wang, R. Gu, F. Wu, and H. Liu, "Unsupervised fake news detection on social media: A generative approach," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 5644–5651.

[5] V. Klyuev, "Fake news filtering: Semantic approaches," in *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. IEEE, 2018, pp. 9–15.

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

[7] J. Simko, M. Hanakova, P. Racsko, M. Tomlein, R. Moro, and M. Bielikova, "Fake news reading on social media: An eye-tracking study," in *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, ser. HT '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 221–230. [Online]. Available: https://doi.org/10.1145/3342220.3343642

[8] B. Collins, D. T. Hoang, N. T. Nguyen, and D. Hwang, "Trends in combating fake news on social media–a survey," *Journal of Information and Telecommunication*, vol. 5, no. 2, pp. 247–266, 2021.

[9] X. Zhou and R. Zafarani, "Fake news: A survey of research, detection methods, and opportunities," *arXiv preprint arXiv:1812.00315*, vol. 2, 2018.

[10] P. Moravec, R. Minas, and A. R. Dennis, "Fake news on social media: People believe what they want to believe when it makes no sense at all," *Kelley School of Business Research Paper*, no. 18-87, 2018.

[11] E. Tandoc, Z. Lim, and R. Ling, "Defining "fake news": A typology of scholarly definitions," *Digital Journalism*, vol. 6, pp. 1–17, 08 2017.

[12] D. M. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild *et al.*, "The science of fake news," *Science*, vol. 359, no. 6380, pp. 1094–1096, 2018.

[13] S. Talwar, A. Dhir, P. Kaur, N. Zafar, and M. Alrasheedy, "Why do people share fake news? associations between the dark side of social media use and fake news sharing behavior," *Journal of Retailing and Consumer Services*, vol. 51, pp. 72 – 82, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0969698919301407

[14] F. Bentley, K. Quehl, J. Wirfs-Brock, and M. Bica, "Understanding online news behaviors," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–11. [Online]. Available: https://doi.org/10.1145/3290605.3300820

[15] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018. [Online]. Available: https://science.sciencemag.org/content/359/6380/1146

[16] M. Flintham, C. Karner, K. Bachour, H. Creswick, N. Gupta, and S. Moran, *Falling for Fake News: Investigating the Consumption of News via Social Media*. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–10. [Online]. Available: https://doi.org/10.1145/3173574.3173950

[17] J. Edson C Tandoc, D. Lim, and R. Ling, "Diffusion of disinformation: How social media users respond to fake news and why," *Journalism*, vol. 21, no. 3, pp. 381–398, 2020. [Online]. Available: https://doi.org/10.1177/1464884919868325

[18] P. Charu and A. Anshu, "Impact of fake news on readers' usage behaviour for news items on facebook and twitter," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, pp. 2564–2568, 2019.

[19] L. Plotnick, S. R. Hiltz, S. A. Grandhi, and J. Dugdale, “Real or fake? user behavior and attitudes related to determining the veracity of social media posts,” *ArXiv*, vol. abs/1904.03989, 2019.

[20] C. Geeng, S. Yee, and F. Roesner, “Fake news on facebook and twitter: Investigating how people (don’t) investigate,” in *Proceedings of the 2020 CHI conference on human factors in computing systems*, 2020, pp. 1–14.

[21] A. Kim and A. Dennis, “Says who?: How news presentation format influences perceived believability and the engagement level of social media users,” 01 2018.

[22] Y. Thijssen, “Breaking the news : the effects of fake news on political attitudes,” 2017.

[23] G. Pennycook, “Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning,” *Cognition*, vol. 188, 06 2018.

[24] L. Fazio, N. Brashier, B. Payne, and E. Marsh, “Knowledge does not protect against illusory truth,” *Journal of experimental psychology. General*, vol. 144, 08 2015.

[25] M. S. Islam, T. Sarkar, S. H. Khan, A.-H. Mostofa Kamal, S. M. M. Hasan, A. Kabir, D. Yeasmin, M. A. Islam, K. I. Amin Chowdhury, K. S. Anwar, A. A. Chughtai, and H. Seale, “Covid-19–related infodemic and its impact on public health: A global social media analysis,” *The American Journal of Tropical Medicine and Hygiene*, vol. 103, no. 4, pp. 1621–1629, 2020. [Online]. Available: https://www.ajtmh.org/content/journals/10.4269/ajtmh.20-0812

[26] C. Hansen, C. Hansen, J. G. Simonsen, B. Larsen, S. Alstrup, and C. Lioma, “Factuality checking in news headlines with eye tracking,” in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR ’20. New York, NY, USA: Association for Computing Machinery, 2020, p. 2013–2016. [Online]. Available: https://doi.org/10.1145/3397271.3401221

[27] B. Lutz, M. T. Adam, S. Feuerriegel, N. Pröllochs, and D. Neumann, “Affective information processing of fake news: Evidence from neurois,” in *Information systems and neuroscience*. Springer, 2020, pp. 121–128.

[28] M. D. da Silva and M. Postma, “Wandering minds, wandering mice: Computer mouse tracking as a method to detect mind wandering,” *Computers in Human Behavior*, vol. 112, p. 106453, 2020.

[29] M. Grimes and J. Valacich, “Mind over mouse: The effect of cognitive load on mouse movement behavior,” 2015.

[30] E. Bozkir, G. Kasneci, S. Utz, and E. Kasneci, “Regressive saccadic eye movements on fake news,” in *2022 Symposium on Eye Tracking Research and Applications*, 2022, pp. 1–7.

[31] Y. Abdrabou, J. Schütte, A. Shams, K. Pfeuffer, D. Buschek, M. Khamis, and F. Alt, “”your eyes tell you have used this password before”: Identifying password reuse from gaze and keystroke dynamics,” in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: https://doi.org/10.1145/3491102.3517531

[32] K. Yu, R. Taib, M. A. Butavicius, K. Parsons, and F. Chen, “Mouse behavior as an index of phishing awareness,” in *Human-Computer Interaction – INTERACT 2019*, D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, and P. Zaphiris, Eds. Cham: Springer International Publishing, 2019, pp. 539–548.

[33] N. Newman, “Overview and key findings of the 2020 digital news report,” 2020.

[34] M. Rucci and M. Poletti, “Control and functions of fixational eye movements,” *Annual review of vision science*, vol. 1, p. 499, 2015.

[35] M. D. Binder, N. Hirokawa, and U. Windhorst, Eds., *Saccade, Saccadic Eye Movement*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 3557–3557. [Online]. Available: https://doi.org/10.1007/978-3-540-29678-2_5190

[36] D. D. Salvucci and J. H. Goldberg, “Identifying fixations and saccades in eye-tracking protocols,” in *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, ser. ETRA ’00. New York, NY, USA: Association for Computing Machinery, 2000, p. 71–78. [Online]. Available: https://doi.org/10.1145/355017.355028

[37] Y. Abdelrahman, A. A. Khan, J. Newn, E. Velloso, S. A. Safwat, J. Bailey, A. Bulling, F. Vetere, and A. Schmidt, “Classifying attention types with thermal imaging and eye tracking,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 3, sep 2019. [Online]. Available: https://doi.org/10.1145/3351227

[38] T. T. AB, “Accuracy and precision test method for remote eye trackers,” 2011.

[39] R. Brunken, J. L. Plass, and D. Leutner, “Direct measurement of cognitive load in multimedia learning,” *Educational psychologist*, vol. 38, no. 1, pp. 53–61, 2003.

[40] V. Navalpakkam, J. Rao, and M. Slaney, “Using gaze patterns to study and predict reading struggles due to distraction,” in *CHI’11 Extended Abstracts on Human Factors in Computing Systems*, 2011, pp. 1705–1710.

[41] Q. Wang, S. Yang, M. Liu, Z. Cao, and Q. Ma, “An eye-tracking study of website complexity from cognitive load perspective,” *Decision Support Systems*, vol. 62, pp. 1–10, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167923614000402

[42] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” *Advances in neural information processing systems*, vol. 30, 2017.

[43] I. Arapakis and L. A. Leiva, “Learning efficient representations of mouse movements to predict user attention,” in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020, pp. 1309–1318.