Social Media Misinformation: Trust, Perception, & Public Awareness in the Age of AI

Kayla Council
Computer Science Department
Hampton University
Hampton, VA, USA
kayla.council@my.hamtponu.edu

Chutima Boonthum-Denecke
Computer Science Department
Hampton University
Hampton, VA, USA
chutima.boonthum@hamptonu.edu

ABSTRACT

Misinformation that spreads through social media platforms is becoming very important since artificial intelligence (AI) is helping the spread of misinformation. A lot of misinformation spreads through manipulated social media posts, and this can cause people to reduce their trust in what they see online, and some people may even be easily persuaded and believe what they see in these manipulated posts. Since artificial intelligence is growing and becoming the newest big thing, it's being used to create posts that are manipulated, and this makes it hard for people to distinguish between real and manipulated posts. This research will analyze how well people can assess social media posts and what factors play a role in their ability to determine what posts are real from the ones that are manipulated. A survey was conducted where people had the opportunity to choose what post they believe is manipulated, and then they were asked what made them choose that choice, and then they rated their confidence level. After gathering all of the results from the survey, the results will be compared with tools that are already created for being able to detect misinformation that's generated by AI. Comparing the human results from the survey with the detection tools will help determine if humans' ability to spot misinformation is just as good as the detection tools. This research will highlight the difficulties of spotting misinformation in AI-manipulated posts, and it will also show how the cybersecurity side of things can help people continue to trust what they see online with the help of detection tools.

KEYWORDS

Misinformation, Social Media, Artificial Intelligence, Online Trust, Cybersecurity, Public Awareness.

I. Introduction

Social media is one of the best and main ways that information can spread since a lot of people use it. Social media can spread information really fast, but this also means that misinformation can spread just as fast, and this can cause a lot of problems. This can cause problems because posts that are inaccurate and appear to be real can influence people's opinions, and it can reduce their trust in sources that are actually credible and authentic. Due to the rise of AI, it's becoming very hard to determine what's real from what's fake since AI is being used to create manipulated photos, videos, and most importantly, manipulated social media posts.

This research will mainly focus on how people perceive and react to social media posts in situations where it can sometimes be hard to determine whether the post is manipulated or authentic. There are three research questions that will be addressed throughout this research: how accurately people can distinguish authentic from manipulated social media posts, how their awareness of AI-generated content affects their confidence in their ability to make decisions, and how certain factors, such as age, technical background, and their social media use, affect their ability to spot misinformation. As mentioned before, a survey will be conducted to see how well people can identify AI-manipulated posts and what influences their decisions.

Also, as previously mentioned, this research will take into account the role AI detection tools play, and this will help bridge the two perspectives. One perspective is the technical solutions that are being created to address AI-generated misinformation, and the other perspective is the human difficulties of identifying misinformation in social media posts.

II. Methodology

A. Target Audience

The targeted audience for this survey is people who are at least 15 and up to 65+; having such a wide range of ages will help gather different types of social media users, and this will also help reduce biased results. Also, having such a diverse age group will allow for high school and college students, working adults, and senior citizens who use social media sites like Facebook, Instagram, LinkedIn, and Reddit to participate. Having a wide range of participants is very beneficial when it comes to analyzing how people's confidence in assessing social media posts and how their ability to identify misinformation can be influenced by their age, their social media usage, and their background experiences.

B. Research Approach

This research will use a survey-based methodology that will be used to investigate how people perceive and react to misinformation that they see on social media. In the survey, there's a section where participants will review a variety of real and AI-manipulated social media posts, and they will select which post they think is the manipulated one. Then, after they make their decision, they will rate their confidence level (1-5) and select which

factor influenced their decision (wording, image quality, prior knowledge, or other).

Google Forms was used to help develop the survey, which helps make it simple to gather and arrange all of the different responses. The analysis of this research will derive from these three main areas: the accuracy of the detection, the confidence in the responses, and the impact of factors such as age, technical background, and social media use on participants' trust in what they see online.

After all of the responses from the survey are collected, they will be compared with current AI-based misinformation detection systems to determine where human judgment and machine accuracy vary. Comparing the two will help connect this study to the cybersecurity side of things.

In addition to the social media platforms previously mentioned, the survey will also be shared with the help of personal connections outside of Hampton University. By collecting responses from peers from other universities, such as North Carolina A&T, Winston-Salem State University, Norfolk State University, and others, this will help gather a more diverse range of responses. Also, reaching people outside of my network who are from different age groups, occupations, states, and even different countries will be easier with the help of social media, specifically Reddit. Having multiple different ways of gathering responses will help ensure that the data collected from the survey is represented by a range of different viewpoints and experiences. This also helps reduce data being gathered from a single community, since this will reach a lot of different people.

C. Literature Review

This section examines five important sources that discuss how people trust what they see online, the different AI detection techniques that are currently being used, and misinformation on social media and its impacts. All of these sources are very important because they explain how misinformation can spread, how people view and recognize it, and how there are a lot of different tools that are used to help identify and reduce the spread of misinformation. Together, all of these sources support the human and the cybersecurity side of this study.

Characterizing AI-Generated Misinformation on Social Media. *Drolsbach, C., & Pröllochs, N. (2025).*
This article explains how artificial intelligence is used to produce and distribute manipulated posts on social media [5]. It also goes into detail about why posts that are created by AI usually look realistic and why these posts can be hard for people to recognize. Also, the concept of my research is similar to this article's discussion of how challenging it's becoming to distinguish authentic information from manipulated information that's produced by AI. This article is also helpful because it discusses how people perceive misinformation and how AI systems are used to detect it. Overall, this source strengthens the relationship between cybersecurity and the human components of this research.

Decoding Misinformation: Why We Fall for Fake News. *Ispos. (2025).*
This source talks about how there are different behavioral and psychological factors that can increase the possibility of people believing and trusting misinformation that they see online [7]. Specifically, this source talks about how when people view social media posts online, they're influenced by a lot of factors such as emotions, biases, and social norms. This connects with this research because the overall purpose of this research is to determine how people's perceptions of social media posts are influenced by factors such as their confidence levels and knowledge of AI-manipulated social media posts. This source is a good source because it also discusses that it's important for people to educate themselves more on AI-generated content, because this can help them recognize false information more successfully. Essentially, this source provides a broader understanding of why people typically trust misinformation rather than trusted and confirmed facts.

AI in Disinformation Detection. *Djenouri, Y., & Puczyńska, J. (2024).*
This source is very important because it talks about how artificial intelligence can be used to identify as well as stop the spread of misinformation online [4]. It goes into more detail about how we can use AI models and algorithms to identify certain patterns of manipulation in text, photos, and videos. This is a very helpful source because it can be used to compare the human responses from the survey conducted in this paper with the AI detection results that are discussed in this source. This source also supports the concept that both humans and machines are crucial when it comes to properly handling misinformation.

(Why) Is Misinformation a Problem? *Adams, Z., Osman, M., Bechilvanidis, C., & Meder, B. (2023).*
This is another good source because it explains the reasons why misinformation is a constant issue in today's society [1]. This source discusses how fast misleading information can spread and circulate on social media and how this affects people's behavior and confidence. It also emphasizes how hard it is to stop misinformation from spreading online after it has already reached a significant number of people. All of the concepts that are discussed in this source explain the significant effect that misinformation has on society, and this is a very relevant and beneficial source for this paper.

How Misinformation on Social Media Has Changed News. *Micich, A. (2025).*
This source highlights the impact of getting news from social media platforms and how people trust what they see on social media [8]. More specifically, this source explains how there are engagement-driven designs and algorithms on social media platforms like Facebook and X (previously known as Twitter) that help the circulation and spread of misinformation. This paper's research emphasizes how social media exposure and usage impact how we trust what we see online, and this source reinforces that. This source also discusses how user activity influences what people think or question online, and this is also directly related to a section that's in the survey. Overall, this source provides a lot of important background knowledge on how social media affects the accuracy of information in today's society.

III. Survey Results

This section focuses on the results that were gathered for this research to answer the three research questions that are outlined in the methodology section. Also, this survey was broken up into 4 sections: Background and Demographics, Awareness and Trust, Post Evaluation, and then Reflection.

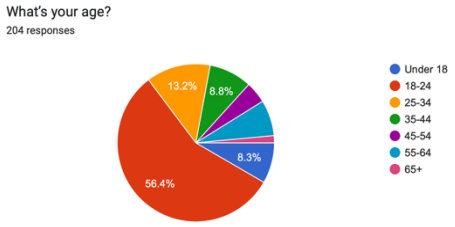

Figure 1.1
After distributing the survey over the course of 2 weeks, I was able to get over 200 responses from a diverse group of participants. Starting with the background demographics section, this was the first question in the survey, and about 50 percent of the participants who took the survey were between the ages of 18 and 24; however, the rest of the age groups were evenly divided, except for the 65-plus age group, which was only about 2 percent of participants.

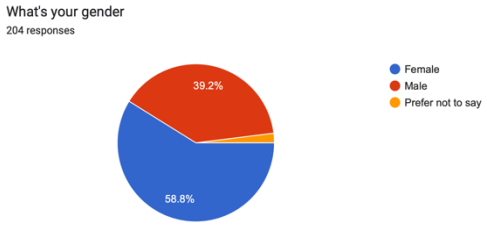

Figure 1.2
This chart shows the difference in gender of the participants; more females than males took the survey, it was about a 20 percent difference.

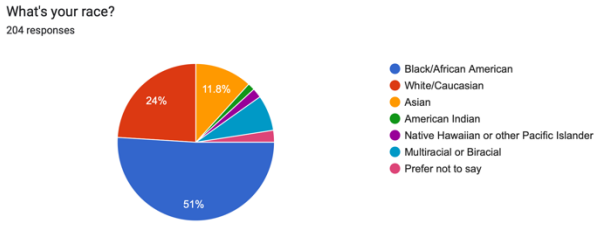

Figure 1.3

For the race percentages in this chart, 51 percent African Americans took the survey, then it was followed by 24 percent white, and then other races were evenly distributed.

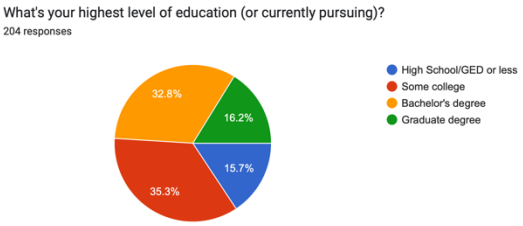

Figure 1.4
This chart shows the highest level of education that the participants have achieved, and as shown in the chart, the data was pretty evenly distributed. There wasn't an education level that was significantly higher or lower than the others, and this shows that this survey has a diverse group of participants from different backgrounds.

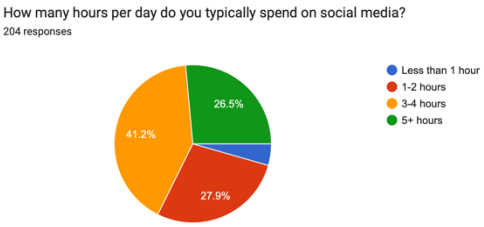

Figure 1.5
For the set of data within this chart, this is very important because how many of hours a day a participant typically spends on social media can play a big role in how well they can spot posts that are manipulated by artificial intelligence. Most participants, 41.2 percent, said that they spend 3-4 hours a day on social media. Then that was followed by 27.9 percent of participants who said they spend 1-2 hours on social media a day. Participants who spent more than 5 hours a day on social media were pretty close, and that was 26.5 percent. Only 4.4 percent of participants said they spend less than an hour a day on social media.

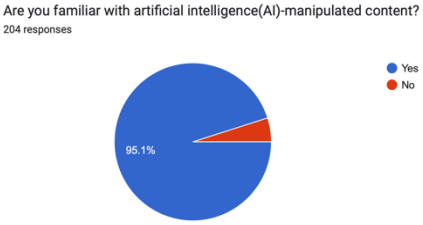

Figure 2.1
This is the first question within the Awareness and Trust section of the survey, and this question provides insight into how familiar participants are with AI-manipulated content. About 95 percent of the participants said they are familiar with it, and only 5 percent said no.

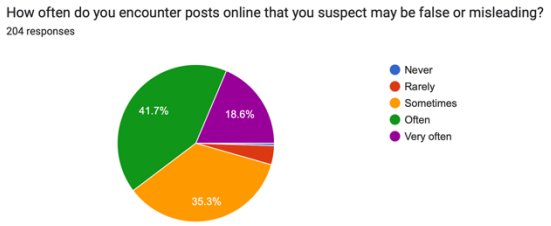

Figure 2.2
This second question in this section asks how often participants encounter social media posts that they think are false or misleading, and the options they were given were: never, rarely, sometimes, often, and very often. There were 41.7 percent of participants who said "often", 35.3 percent said "sometimes", 18.6 percent said "very often", 3.9 percent said "rarely", and only 0.5 percent, which is only one participant, said "never".

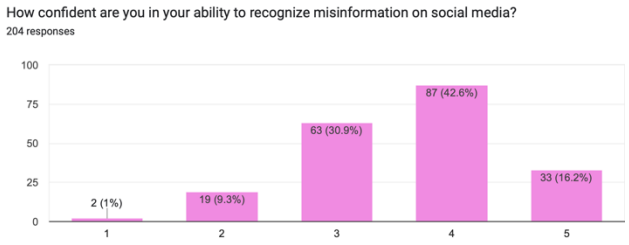

Figure 2.3
This is the last question in this section. Participants were asked to rate their confidence levels in being able to recognize misinformation on social media, and this is very important because the same question will be asked again after participants evaluate the AI-manipulated social media posts. The participants were given a 1-5 rating with 1 being not confident and 5 being very confident, and 2 participants chose 1, 19 participants chose 2, 63 participants chose 3, 87 participants chose 4, and 33 participants chose 5.

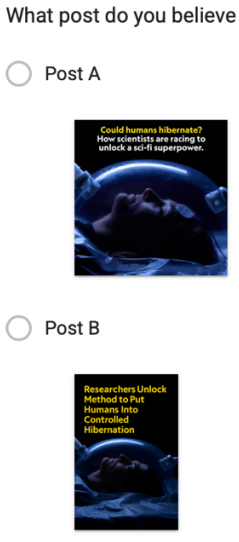

Figure 3.1.1

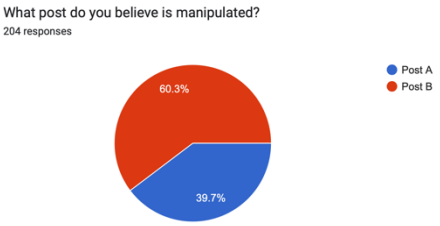

Figure 3.1.2
These two sets of figures are for the first question in the Post Evaluation section, and in this section, the participants are evaluating six sets of posts and figuring out which post in the pair is manipulated. For the set of posts in the figures above, Post B is the manipulated post. As you can see in the results data, 60.3 percent of participants got the answer correct, while 39.7 percent got the question wrong.

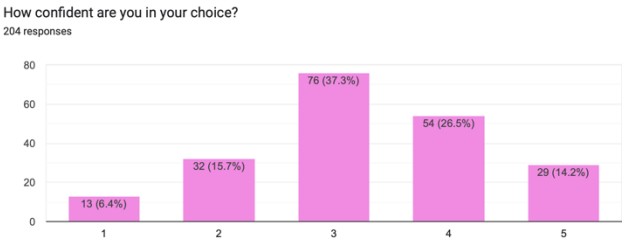

Figure 3.1.3

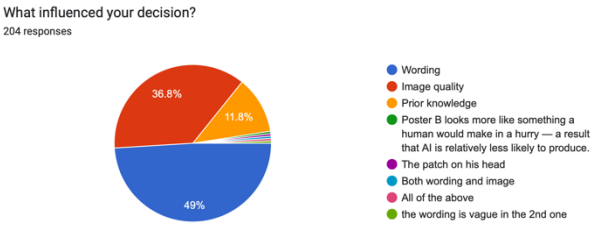

Figure 3.1.4
The two figures above are questions that are a part of the first set of posts. These two questions ask the participants to rate their confidence in their choice, as well as select what influenced their decision, with a set of choices being: wording, image quality, prior knowledge, or other, where users are able to leave an open-ended response. For the first set of posts, 13 participants rated their confidence as 1, 32 participants rated their confidence as 2, 76 participants rated their confidence as 3, 54 participants rated their confidence as 4, and 29 participants rated their confidence as 5. Also, for choosing what factors influenced their decision, 49 percent chose "wording", 36.8 percent chose "image quality", 11.8 percent chose "prior knowledge", and 2.4 percent chose other.

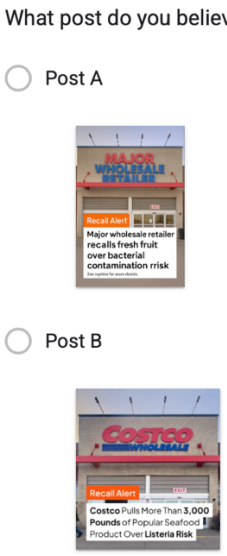

Figure 3.2.1

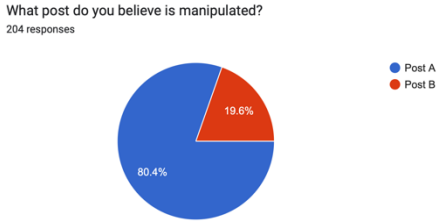

Figure 3.2.2

The second set of posts is shown in the two figures above, and for this set of posts, Post A is the manipulated post. There were 80.4 percent of participants who got the answer correct, and 19.6 percent got the answer wrong.

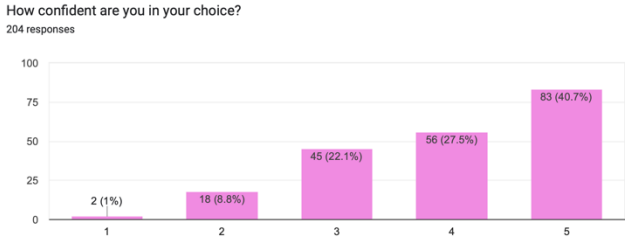

Figure 3.2.3

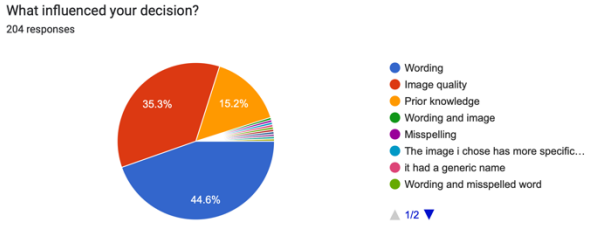

Figure 3.2.4

For this set of posts, a lot of participants felt more confident. Only 2 participants rated their confidence as 1, 18 participants rated their confidence as 2, 45 participants rated their confidence as 3, 56 participants rated their confidence as 4, and 83 participants rated their confidence as 5. Also, for the factors that influenced their decisions, 44.6 percent chose "wording", 35.3 percent chose "image quality", 15.2 percent chose "prior knowledge", and only 4.9 percent chose "other".

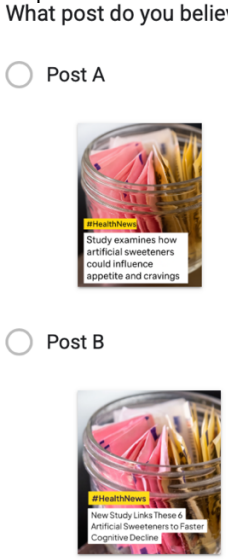

Figure 3.3.1

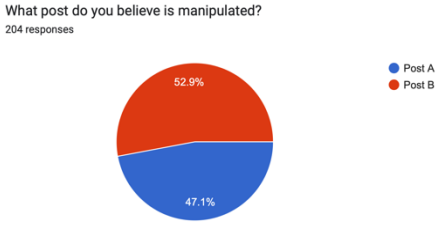

Figure 3.3.2

The third set of posts is shown in the two figures above, and the manipulated post is Post A. The majority of the participants, which was 52.9 percent, thought that Post B was the manipulated post. However, 47.1 percent of the participants selected the correct answer.

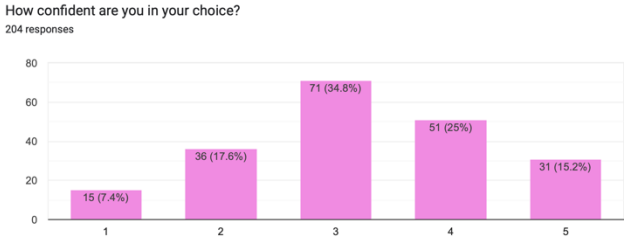

Figure 3.3.3

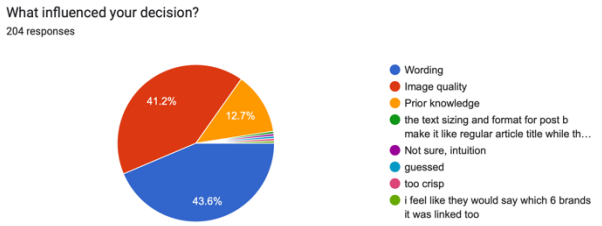

Figure 3.3.4

For the confidence levels, 15 participants were unsure of their selection, and their confidence rating was 1; 36 participants rated their confidence as 2; most of the participants, which was 71, rated their confidence as 3, 51 participants rated their confidence as 4, and 31 participants rated their confidence as 5. There were 43.6 percent of participants who chose "wording" as the factor that influenced their decision, 41.2 percent chose "image quality", 12.7 percent chose "prior knowledge", and 2.5 percent chose "other".

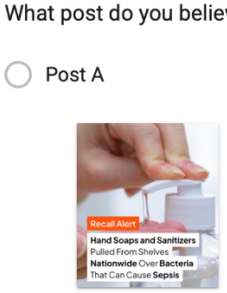

Figure 3.4.1

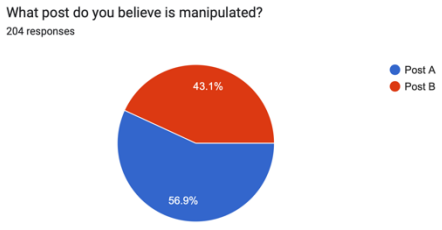

Figure 3.4.2

The manipulated post for the figure shown above is Post B; however, most of the participants, which was 56.9 percent, chose Post A. Only 43.1 percent of participants got the answer correct.

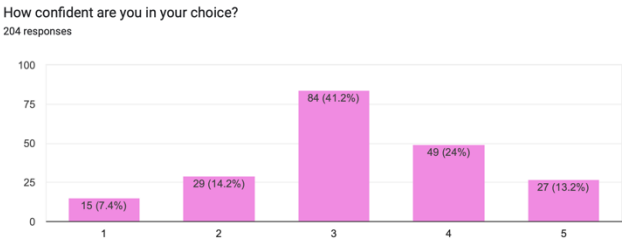

Figure 3.4.3

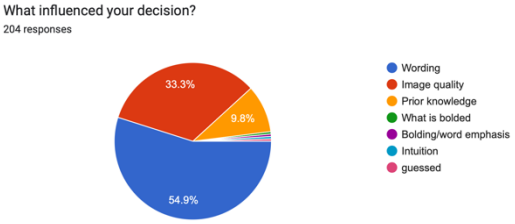

Figure 3.4.4

The majority of participants were split in the middle and were unsure about their selection. There were 15 participants who rated their confidence as 1, 29 participants rated their confidence as 2, 84 participants rated their confidence as 3, 49 participants rated their confidence as 4, and 27 participants were 100 percent certain and rated their confidence as 5.

Figure 3.5.1

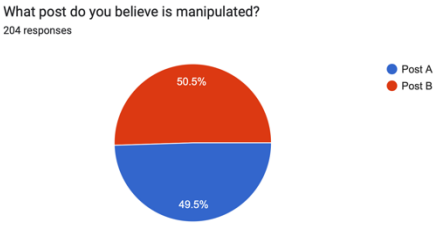

Figure 3.5.2

The manipulated post in the set shown above was Post B, and for this post, it was very close, and there was only a 1 percent difference in accuracy. There were 50.5 percent of participants who chose the correct post, and 49.5 percent chose the incorrect post.

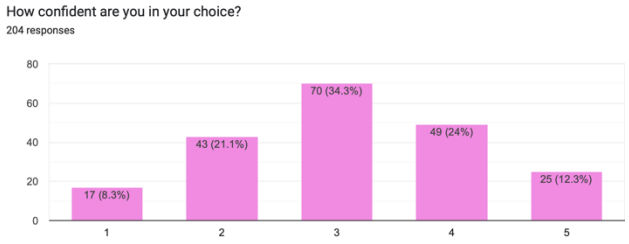

Figure 3.5.3

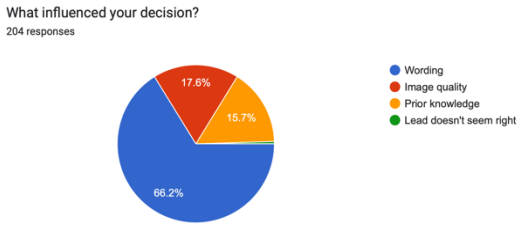

Figure 3.5.4

For this set of posts, the confidence rating was split in the middle as well. There were 17 participants who rated their confidence as 1, 43 participants rated their confidence as 2, 70 participants rated their confidence as 3, 49 participants rated their confidence as 4, and only 25 participants rated their confidence as 5. Also, the majority of the participants, 66.2 percent, said the wording influenced their decision, 17.6 percent chose "image quality", 15.7 percent chose "prior knowledge", and less than 1 percent chose "other".

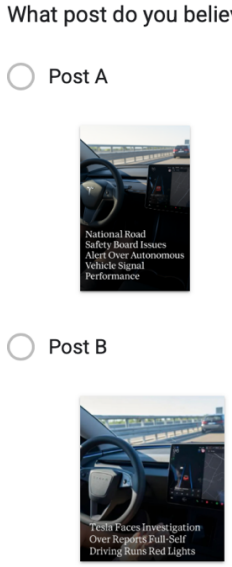

Figure 3.6.1

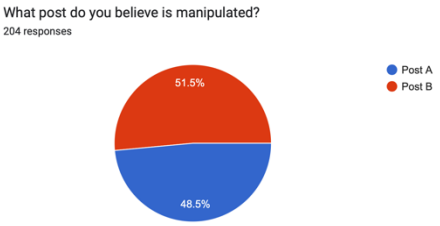

Figure 3.6.2

The sixth set of posts, which is the final posts in the survey, is shown above, and the manipulated post was Post A. However, 51.5 percent of participants chose Post B, and 48.5 percent chose the correct answer.

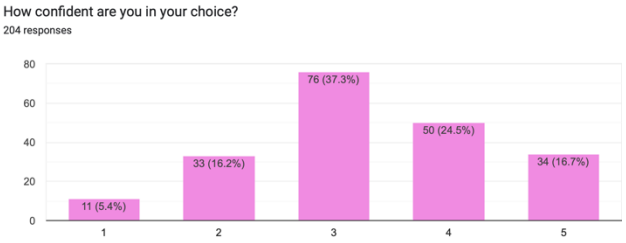

Figure 3.6.3

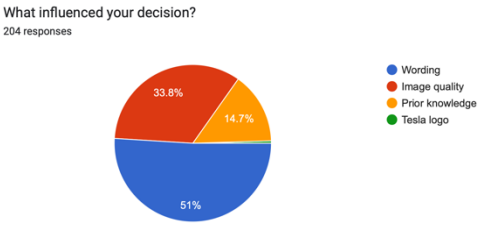

Figure 3.6.4

A lot of participants were unsure about this set of posts as well. There were 11 participants who rated their confidence level as 1, 33 participants rated their confidence as 2, 76 participants rated their confidence as 3, 50 participants rated their confidence as 4, and 34 participants rated their confidence as 5. There were 51 percent of participants who said the wording influenced their decision, 33.8 percent said "image quality", 14.7 percent said "prior knowledge", and less than 1 percent chose "other".

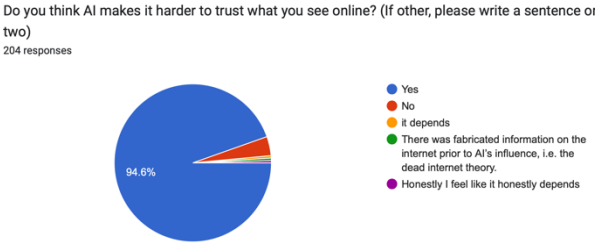

Figure 4.1

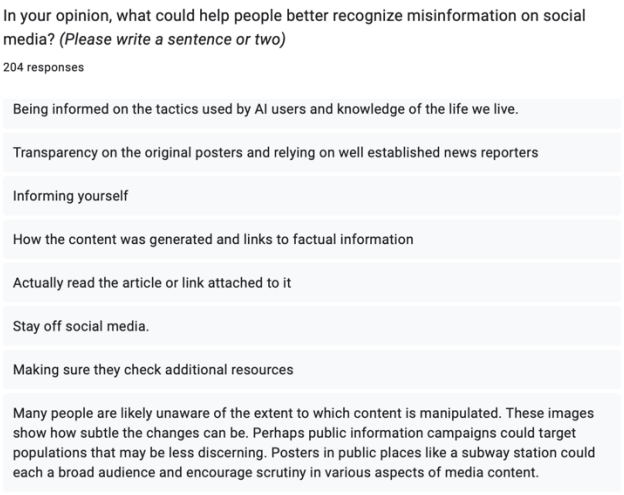

Figure 4.2

The last section of the survey is the Reflection section, and it only has three questions. The first two questions are shown in the two figures above, and the first question asks the participants if they think AI makes it harder to trust what they see online. 94.6 percent of participants said "yes", 3.9 percent said "no", and 1.5 percent said "it depends". The second question is an open-ended question, and it asks the participants what they think could help people recognize misinformation on social media. Most participants said "being aware", "educating yourself", and others went into more detail.

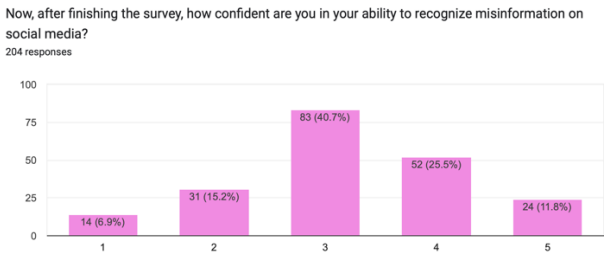

Figure 4.3

The last question of the survey is very important because it asks users to rate how confident they are in recognizing misinformation on social media, and this is important because this question was also asked before evaluating the posts. There were 14 participants who rated their confidence as 1, 31 rated it as 2, 83 rated it as 3, 52 rated it as 4, and only 24 rated it as 5.

IV. Analysis of Results

A. The Overall Accuracy of Results

When it came down to picking the correct post that was manipulated by AI, the accuracy percentages were practically split down the middle across all 6 sets of posts. The first two posts were the only ones where most participants were correct when it came to identifying the correct post. However, for the last 4 sets of posts, you can tell that people were starting to become unsure because half of the participants would select the correct post, while the other half would select the incorrect post. The post that was the easiest for the participants was the second set of posts *(refer to Figure 3.2.1),* and the post that was the hardest was the fifth set of posts *(refer to Figure 3.5.1)*. The set of posts that were more heavily worded was the one that seemed the hardest for participants to decipher. Also, for the set of posts where the background images were practically the same was also hard for participants to decipher because then they had to rely on the captions in the post to help them determine which post was the manipulated one. Overall, these findings helped me address my first research question: how accurately can people distinguish manipulated from authentic social media posts?

B. How Participant Awareness Affects Their Confidence

My second research question that I wanted to address was: how do confidence levels affect participants' ability to make decisions? For the overall confidence within the survey, the majority of the participants were not confident in any of the posts except for the second set of posts *(refer to Figure 3.2.2)*. For all of the other posts, most of the participants had ranked their confidence as 3, and this shows that they were split down the middle when it came to how certain they were about their selection. It was very interesting to see how many people are unsure of what post was correct because this shows that a lot of people don't know what's manipulated by AI, and this is a very concerning issue. It's also important to note that there was a question in the survey that was asked twice, before evaluating the posts, and after evaluating the posts, which asked the participants to rank their confidence levels in being able to recognize misinformation on social media. It was interesting to see that before evaluating the posts, participants had felt that they were very confident because the majority of the participants ranked their confidence as a 4, and only 2 participants ranked their confidence as a 1 *(refer to Figure 2.3)*. However, at the end of the survey, the confidence levels changed, and the majority of the participants ranked their confidence as 3, and then 14 participants ranked their confidence as 1 *(refer to Figure 4.3)*. This shows that before actually seeing misinformation online, people were overly confident, but after being presented with sets of posts, it made a lot of people question their confidence and realize that they might not be as confident as they thought they were.

C. How Demographic Factors Influence Decision Making

Another research question that I wanted to address was whether there were any particular factors, such as age, education levels, or other demographic factors would play a role in how well people are able to spot AI-manipulated posts. After conducting the survey, I noticed that all of the age groups were pretty mixed when it came to deciphering which post was manipulated; there wasn't a particular age group that was able to get more correct answers compared to other age groups. Also, when it came to education levels and awareness of AI-manipulated content, that didn't affect any results either. There wasn't a particular group or community that was able to 100 percent accurate when it came to determinging which post was manipulated.

D. Factors That Influence Decisions

When the participants were selecting a factor that influenced their decision, for each set of posts, the majority of the participants chose the "wording" option, and "image quality" was a close second. However, for the "other" option, which was open-ended, some participants left some interesting comments. Some participants pointed out the "bolding of the words", others said "the image is too crisp", and some even mentioned "the text sizing and format". For the participants who left specific comments like that, they were able to accurately choose the manipulated posts, and this shows that even when you take the time to pay attention to minor details, this could help you possibly more accurately detect which post is manipulated.

E. Key Takeaways of Survey Results

Overall, after analyzing all of the results, it's clear to see that it can be hard for people to accurately determine which post is manipulated by AI, and this can be a very concerning problem within our society. This shows that a lot of people can be fooled by minor details or even just the way certain things or worded to make manipulated posts seem more convincing. After gathering all of the results, only 3 participants out of 204 were able to accurately identify all of the posts that were manipulated, and those participants all came from different backgrounds and were in different age groups.

F. Analysis of AI Detection Tools

Even though artificial intelligence is used to produce manipulated social media posts, it can also be used as a detection tool. AI has built-in systems that use natural language processing (NLP) and machine learning [6]. NLP is a form of AI, and it helps computer systems to be able to understand and even generate human language [6]. Also, machine learning is very similar, and it essentially does the same thing, but it also recognizes patterns within data so that it can make decisions without the help of humans [6]. Both of these AI systems are very helpful when it comes to detecting misinformation that's generated by AI because they can pick up on things that humans may not easily notice or see. NLP and machine learning are able to analyze text, which helps verify if posts are authentic, and it's able to detect misinformation within posts [6]. On the other hand, machine learning uses algorithms that learns from data so that their performance can improve over time, and this helps so that algorithms can be trained to detect misinformation due to the patterns that they learn [6]. Along with NLP and machine learning, there are also detection tools for deepfakes, and they can see if social media posts have been manipulated [8].

Since there are a lot of different AI tools and features that are currently in place to detect misinformation in social media posts, this is very beneficial, and it brings a lot of strengths. One of the biggest benefits of these tools is that they can scan large amounts of data to detect misinformation [2]. Also, once it takes in those large amounts of datasets, the tools are also being trained so that they can improve their accuracy when it comes to the detection process [6]. As mentioned earlier, a big concern with misinformation in social media is that it can spread very quickly; however, AI detection tools can help mitigate the rapid spread of misinformation [2]. Having these AI detection tools also improves the public's trust since these tools can offer an extra hand when it comes to detecting misinformation on social media.

Even though AI tools can be beneficial and bring ease, they also have many weaknesses. Sometimes it can be very hard for the AI detection tools to be able to understand sarcasm and context, and this is a very big issue if we were to rely heavily on AI detection systems [6]. Another concern that these tools can bring is that they can sometimes have false positives because they misidentify real posts as manipulated ones [6]. Also, with deepfakes making things look extremely real, they're getting harder to detect by AI tools, and in order to properly detect them, the detection systems need to have more training [10]. Since these AI detection tools are being trained with large sets of data, some of the data that they're taking in may have some biases, and this is an ethical concern because these models could potentially be biased due to learning from biased data [9].

G. Comparing Human Judgement & AI Detection Tools

When it comes to humans being able to detect misinformation within social media posts, there are a few factors in which humans could possibly perform better than AI. For starters, humans are able to understand tone, sarcasm, context, and they have common sense, which are things that AI struggles with [7]. Along with those factors, humans can also relate to what they see online and tie what they see back to their own personal experiences. This is also something that AI lacks since AI doesn't have emotions or feelings [7]. Humans can also combine their critical thinking skills with their emotions, and this can help them better identify misinformation in social media posts [10].

Even though humans could perform better due to having emotions, there are also some situations where AI could outperform humans. Artificial intelligence is able to detect AI at a higher rate compared to humans since it has a 97% accuracy rate [6]. It has such a high accuracy rate since AI is able to take in large amounts of data at a time, and it can see small details in posts that humans aren't able to see [6]. Also, since AI doesn't have emotions, it won't fall for posts that are emotional or persuasive [10].

Humans and AI detection tools have their own strengths of when they would be better to use in certain situations; however, they both have common struggles. For example, both humans and the AI detection tools can struggle when it comes to determining if a post has misinformation, as shown in the results of my survey, as well as research that shows where AI has failed [5]. Also, when it comes to social media algorithms, they often push out more misinformation due to a lot of people engaging with it, and this can cause humans to be more likely to fall for misinformation since a lot of people are engaging with the post, so it may seem as if the post is true or real [2]. It's important to remember that humans or AI can't fully detect misinformation by themselves, and it's best when combining the two together [1].

H. How We Can Protect What We Believe Online

It's important to protect what we believe online because what we fall for can have a large impact on things that we may not even think of. Misinformation on social media can influence elections, crises, health concerns, and just overall any major topic that may be going on in the world [9]. The results of my survey show how easy it is for people to fall for misinformation because AI can make things look very realistic. In order for humans not to fall for misinformation that's produced by AI, they have to educate themselves in the patterns or details that AI usually uses [7]. It's also important for people to educate themselves on a topic that they may see in a post instead of just automatically following it [10]. Also, social media platforms should integrate more AI control within their systems so that they can improve transparency as well as moderate content that's produced by AI [2]. Social media platforms also need to do a better job at encouraging users to fact-check what they see online, especially before they start sharing posts [10]. Also, social media platforms need to push out more credible news outlets to people's algorithms instead of just pushing what's currently popular or trending at the moment, because this could help reduce the spread of misinformation [10].

V. Conclusion

This research was conducted to prove how hard it can be for people to detect misinformation in social posts. Since a lot of people have social media, this increases the amount of misinformation that's being spread online through social media, and artificial intelligence helps play a role in that. Artificial intelligence is growing every single day, and it's becoming more advanced; so advanced that it can produce AI-generated social media posts that can spread misinformation, and this makes it even harder for people to detect misinformation in social media posts. Throughout this research, there were three research questions that I addressed: how accurately people can spot AI-generated social media posts, how confident people are, and what demographic factors play a role in detection. After gathering and analysing all the responses from my survey, it answered my research questions by proving that people struggle in determining what's authentic vs manipulated, people realised that it was more difficult to decipher posts than what they had originally thought, and demographics didn't really have an effect on how well people could decipher the posts. Also, as mentioned before, only 3 participants out of 204 were able to correctly identify which posts were AI-generated, and this is very concerning because it shows that a lot of people fall for misinformation that's produced by artificial intelligence. Also, within this research, I wanted to compare human judgment and AI-detection tools to see if they align or differ when it comes to detecting AI. I noticed that humans detect AI better when it comes to dealing with emotions and tone within a post, AI-detection tools are better because it can recognize certain patterns that humans aren't able to see, but both humans and AI-detection tools have areas where they lack so it's important to remember that we should combine human judgement and AI-detection tools in order to get the best results. After conducting this research, I learned that misinformation that's produced by AI has a huge negative impact, and since it's starting to look more realistic, it can have an effect on info about elections, health issues, etc., and this weakens the trust that people have in information that's produced online. In order to mitigate the issues that AI-generated social media posts cause, as a community, we need to work together to promote media literacy and to always fact-check what we see online. Also, social media platforms need to improve their AI-detection tools so that they can fact-check better and flag posts that are AI-generated so that users can be aware and not blindsided. So, hopefully this research shows how important it is to take the extra steps to become more aware of what you're seeing online and to educate yourself before you believe what you see, because AI can scarily produce posts with misinformation and make them seem like they're real.

REFERENCES

[1] Adams, Z., Osman, M., Bechilvanidis, C., & Meder, B. (2023, February 16). *(Why) Is Misinformation a Problem?* . Retrieved from National Library of Medicine: https://pmc.ncbi.nlm.nih.gov/articles/PMC10623619/.

[2] American Psychological Association. (2024, March 1). *8 recommendations for countering misinformation*. Retrieved from American Psychological Association Logo: https://www.apa.org/topics/journalism-facts/misinformation-recommendations.

[3] Caceres, M. M., Sosa, J., Lawrence, J., Sestacovschi, C., Johnson, A., Rasool, M., . . . Fernandez, J. (2022, January 12). *The impact of misinformation on the COVID-19 pandemic*. Retrieved from NIH Library of Medicine: https://pmc.ncbi.nlm.nih.gov/articles/PMC9114791/#s3.

[4] Djenouri, Y., & Puczyńska, J. (2024, December 31). *AI in Disinformation Detection*. Retrieved from ACIG journal: https://www.acigjournal.com/AI-in-Disinformation-Detection,200200,0,2.html#S5.

[5] Drolsbach, C., & Pröllochs, N. (2025, May 15). *Characterizing AI-Generated Misinformation on Social Media*. Retrieved from ARXIV: https://arxiv.org/html/2505.10266v1#:~:text=Contributions:%20To%20the%20best%20of,and%20harmful%20as%20conventional%20misinformation.

[6] Erokhin, D. (2025, February 28). *Artificial Intelligence Tools in Misinformation Management during Natural Disasters*. Retrieved from Springer Nature Link: https://link.springer.com/article/10.1007/s11115-025-00815-2#Sec4.

[7] Iposos. (2025, April 17). *Decoding Misinformation: Why we fall for fake news*. Retrieved from Iposos: https://www.ipsos.com/en-us/decoding-misinformation-why-we-fall-for-fake-news.

[8] Micich, A. (2025, July 30). *How misinformation on social media has changed news*. Retrieved from U.S Pirg: https://pirg.org/edfund/articles/misinformation-on-social-media/.

[9] Palfrey, J. (2025, September 9). *Misinformation and disinformation*. Retrieved from Britannica: https://www.britannica.com/topic/misinformation-and-disinformation.

[10] Unicef. (2025, February 10). *A quick guide to spotting misinformation*. Retrieved from unicef: https://www.unicef.org/eca/stories/quick-guide-spotting-misinformation.