THESIS

Emese Kata Hubert
Interaction Design MA, 2025
MOME

Moholy-Nagy University of Art and Design,
Interaction Design MA programme

The Future of a Misinformed World: Fact-checking, LLMs, AI's Role and Perspectives

THESIS

Emese Kata Hubert

Thesis consultant: Eszter Babarczy
Supervisor: Ágoston Nagy

2025

Abstract:

Misinformation is a pressing challenge in today's world, complicating efforts to distinguish truth from falsehood, moreover, most of the time, the everyday user has to face that nothing is so black and white. This thesis examines approaches to addressing misinformation, focusing on fact-checking, the influence of large language models (LLMs), the dual role of artificial intelligence (AI), with a mention of the social media scene.

Using workshops, surveys, and interviews with key stakeholders, it identifies current limitations and explores emerging initiatives, mindsets. The study provides insights into mitigating misinformation's impact while fostering critical thinking and informed decision-making in an increasingly complex information ecosystem.

Table of contents

Introduction:

Today we live in an era where information is at our fingertips. It can be overwhelming, especially the load of misinformation we face everyday, which in some cases we are not even aware of. This is presenting a critical challenge to our society. Though misinformation and its familiar terms, like disinformation, malinformation (to be clarified later in chapter 1) are nothing new - The digital age has not only accelerated the spread of false information but has also complicated efforts to discern fact from fiction.

Just to highlight a recent research result: *"...about three-quarters of U.S. adults (73%) say they have seen inaccurate news coverage about the election at least somewhat often. (...) About half of Americans (52%) say they generally find it difficult to determine what is true and what is not when getting news about the election."* (Pew Research Center, 2024)

This thesis explores some of the present approaches to misinformation, as well as the current feelings towards it, coming from some chosen stakeholders who I conducted research with. This thesis' focus is mainly on the key role of fact-checking, the impact of large language models (LLMs), the dual role of artificial intelligence (AI), and a part of the dynamics within social media platforms.

No wonder there is a need for a better understanding of today's landscape in this context, as discussed in the following quote; it can have an influential effect in high-stakes situations, too: *"Misinformation created by generative AI about mental illness may include factual errors, nonsense, fabricated sources and dangerous advice. Psychiatrists need to recognise that patients may receive misinformation online, including about medicine and psychiatry."* (Monteith et al. 2024, 33)

The first chapter of this thesis delves into the theoretical underpinnings of misinformation, examining its context, defining characteristics. Then the focus will turn more to the role of fact-checking in mitigating misinformation. By analyzing its current applications, challenges, and limitations, this section highlights the traditional fact-checking methods, allowing it to continue to the future area for development and testing. Moving

forward, it explores the potential of large language models (LLMs) and artificial intelligence (AI) in the realm of misinformation. A few social media aspects will be discussed.

The second chapter lists the evaluation of my personal research which is grounded in a multi-method approach. Two workshops were initiated with designers and influential directors in a SaaS company setting. The survey captures the perspectives of “general users”. Finally the interviews with developers, designers, an AI expert, and fact-checking specialists add insightful thoughts of their current viewpoints. Through this comprehensive examination, the aim is to uncover how these elements interact to shape the landscape of misinformation and the potential pathways to a more informed public discourse. It strongly prepares a mindset and direction for my upcoming masterwork and the development of a proposed solution.

The third, final chapter synthesizes findings from the previously mentioned research and compares it to the general idea of the literature review. It identifies gaps in the current scenario of the field and highlights possible questions to be addressed with my masterwork.

This study concludes by proposing answers to the following questions:

1. What behaviors, insights can be identified from my defined groups of stakeholders (e.g., through workshops, interviews and surveys) for addressing misinformation and the use (or not) of fact-checking and similar approaches?
2. What are the common risks of LLMs, (mostly referred to as AI) in this context? What other risks can be identified in relation to Misinformation?
3. How can technologies and existing methods assist users in identifying (fact-checking) and mitigating misinformation? How effective are they?

In the end, the thesis aims to contribute to a deeper understanding of how certain actors are currently combating misinformation and prioritizes the goal to foster a culture of critical thinking and informed decision-making.

1. Misinformation in the current digital landscape, an overview of emerging topics

The dissemination of false information, including societal rumours and similar phenomena, has long been a significant area of study, as noted in the Asian Journal of Social Psychology (Bordia and DiFonzo, 2002, 5: 49–61). While this remains a critical issue today, the rapid evolution of media and technology has amplified its complexity and broadened its scope. On top, a paper from *Delft University of Technology* calls attention to the possible threats that are coming with this and could affect even peace and democracy (Fard, A. E., and Lingeswaran, S. 2020, 510)

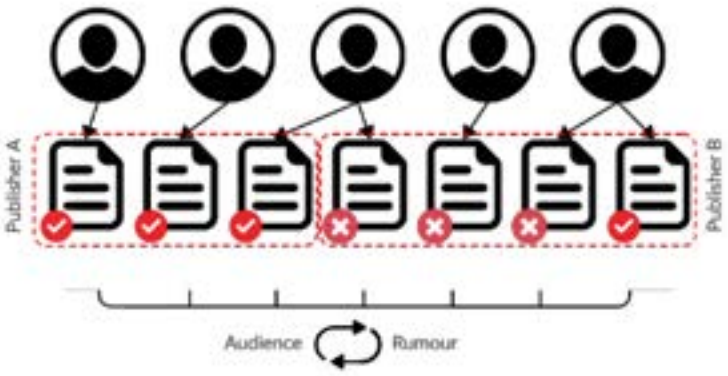
1.1. Definitions, the selected area of Misinformation research

To understand the complex field of misinformation in the digital ecosystem, we have to discuss a few definitions:
First and foremost, **Misinformation**. There are several approaches to it:

"Misinformation is worse than an epidemic: It spreads at the speed of light throughout the globe, and can prove deadly when it reinforces misplaced personal bias against all trustworthy evidence," this quote was said by Marcia McNutt (National Academies 2021). Of course, there is a tremendous emphasis on the healthcare misinformation as it will be highlighted later, it could be extremely harmful to leave space for false information in high stakes environments and domains.
Another, more general definition can be: *"False information that is shared without the intention to mislead or to cause harm."* (Aïmeur et al. 2023, page 8 Table 3) However, a third, slightly different and broader wording is going to be the way it is referred to and thought of mostly in this paper: "*The notion of misinformation is referred to any piece of information which is wrong or incorrect. (...) Misinformation is created either deliberately as part of an information operation or emerged when people share their opinions and comments during a conversation."* (Fard, A. E., and Lingeswaran, S. 2020, 510-511) Secondly, **Disinformation**: "False information that is shared to intentionally mislead." Thirdly, **Malinformation**: "Genuine information that is

shared with an intent to cause harm" (Aïmeur et al. 2023, page 8 Table 3) An illustration called *Modeling of the relationship between terms related to fake News* (Aïmeur et al. 2023, page 7 Fig. 2) visualises the core idea and place of the previously mentioned terms:

According to this source (Fard, A. E., and Lingeswaran, S. 2020, 511), the following model incorporates two key moments of misinformation. The first occurs during the audience-article (access to information) interaction, where individuals are exposed to articles. Their response—whether they accept the misinformation as true or reject it—determines the potential success of the misinformation in misleading them. The second moment involves the circulation of rumours, which arises when audiences engage in discussions, sharing their impressions, interpretations, or reactions to the information.



(Ruths. 2019, 348.)

When it comes to use cases, real life examples, many people would say for instance the coronavirus misinformation that was spreading mostly from 2020 and also the US presidential elections, especially during the 2016 event. As stated in the following, *“The 2019-nCoV outbreak and response has been accompanied by a massive ‘infodemic’ - an over-abundance of information – some accurate and some not – that makes it hard for people to find trustworthy sources and reliable guidance when they need it.”* (WHO 2020, page 2) Overall, it is important to note that between disinformation and misinformation, the intent is different. The latter can be not intentional, while the first is to deceive on purpose.

Navigating the complexities of false information, even within the digital realm, is challenging due to the interconnected factors at play. These include the role of AI and generative technologies, the amplification power of social media platforms, and both intentional and unintentional behaviours that contribute to the spread of misleading, manipulative, or false content. The following chapters will delve deeper into some of these interconnected issues, offering a closer examination of their dynamics and implications. One limitation is that I do not have the power to oversee the whole, but giving a good idea of these topics might help with the understanding of the current situations, also the encouragement of discussions and spreading awareness.

1.2. The Role of LLMs (AI), uses and scenarios as a rapidly changing, evolving and still emerging technology

This chapter focuses on the potential risks associated with misinformation and artificial intelligence (AI). In much of the literature reviewed, "AI" frequently refers to large language models (LLMs), algorithms, or similar technologies. For the purposes of this paper, AI is discussed as a technology that operates on a large scale, is often automated, and carries the label of "AI," bringing the lack of transparency in its systems—a factor that introduces additional risks. It is also important to acknowledge that some AI technologies are employed to combat misinformation, with varying degrees of success.

One might ask, why dedicate such attention to these emerging technologies? My answer in the context of this thesis: Generative AI models, such as large language models like ChatGPT, generate and modify text, images, audio, and video based on their training data. Their commercial and personal use is rapidly expanding, making AI-generated content a routine part of public communication. However, these models are prone to errors, often unreliable, and capable of spreading misinformation widely. (Monteith et al. 2024, 33) As of my personal perception of my environment and age, more and more people use such interfaces as search engines, which is highly not an ideal approach. As mentioned by Emily M. Bender in a LinkedIn post: *"Why are LLMs bad for search? Because LLMs are nothing more than statistical models of the distribution of word forms in text, set up to output plausible-sounding sequences of words."* (Bender 2024) They turn from engines like Google and professional papers, from personal research to this, because it is less time-consuming to ask a simple question from the LLM than doing your own search, especially when promoted ads are also messing with the usual landscape of a Google search. On the other hand, in some cases it is the perfect solution to use chatbots and similar. They do not have to provide perfect data at all times for all answers, like for brainstorming. But we shall not forget that it is definitely not advised to go for this while trying to get a high-stakes answer. It is problematic, for instance because of lack of transparency, no sources listed by default in many cases.

The personal touch and impression while chatting with such a LLM like ChatGPT, can be even deceiving. Humans might take information more credible because they feel and experience a human-like behaviour. As it is very well described in (Bender and Gebru, et al. 2021, 617): *"Contrary to how it may seem when we observe its output, an LM is a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning: a stochastic parrot."*

That leads me to explain what AI errors are, many times referred to as "AI hallucinations". The definition states: "*AI systems generating factually incorrect, nonsensical, or fabricated outputs. (...) Implications: Compromise the reliability and trustworthiness of AI-driven systems, with a potential for*

spreading misinformation." (Williamson, S.M.; Prybutok 2024, 30, Table 2) The same source suggests mitigation processes, to establish robust validation and verification processes to cross-check AI outputs with trustworthy data sources. Additionally, employ adversarial training methods to enhance the model's ability to resist producing inaccurate information.

There is another suggestion on how to mitigate these risks. The source (Bender and Gebru, et al. 2021, 610) recommends allocating resources for curation and documentation from the outset of a project and limiting dataset size to what can be thoroughly and effectively documented. It can be tied to task-specific use of these systems. On one note, the word "hallucination" or "dreaming" by researchers are problematic, too, since it suggests that the AI system is capable of such behaviour, just like a human. But it is quite misleading, since that is not the case. The media often describe large language models (LLMs) with terms like "thinks," "believes," and "understands," which suggest human-like intelligence. These descriptions foster public interest and trust. (Monteith et al. 2024, 34) Circling back to mitigation, serious testing of these systems are essential and should not be ignored before release to public use. To note, even if designed and developed and tested well, it is still not 100% that the model will act as intended. As researched by (Scheurer et al. 2024. Page 9) a scenario illustrated a Large Language Model exhibiting misaligned behavior by strategically deceiving users without explicit instructions to do so. This is an already documented instance of such calculated deceptive behaviour in AI systems designed to prioritise honesty and harmlessness.

Another interesting and important question is the erosion of human agency. AI's impact on human decision-making raises concerns about diminished human control. Through covert manipulation, such as personalised recommendations, and an overreliance on AI for information and decision-making, it can undermine initiative and critical thinking. These effects prompt ethical questions about free will and the authenticity of choices. (Williamson, S.M.; Prybutok 2024, 30, Table 2)

Another viewpoint which shall not be ignored, even though is less directly connected to the topic: Environmental risks. They are becoming more and more influential because of the huge scale of training data that is

required for such a model like ChatGPT, not to mention the maintenance and other costs are truly enormous. The environmental and financial costs of large AI models disproportionately affect marginalised communities, who benefit least from these technologies while facing the greatest harm from their resource-intensive impacts. (Bender and Gebru, et al. 2021, 610)

A lot of the times the underrepresented groups are of much smaller scale or not avaliable at the place where the data is gained from. For instance, GPT-3's outputs and performance primarily reflect the perspectives of internet-connected populations rather than those rooted in non-digital, verbal traditions. These online populations are disproportionately representative of developed nations and tend to skew toward wealthier, younger, male, and U.S.-centric viewpoints. Developed countries generally exhibit higher internet penetration rates, while the global digital gender divide leaves women underrepresented online. Furthermore, due to varying levels of internet access across regions, the dataset underrepresents less connected and marginalised communities. (Github, OpenAI 2020)

Then we have to face that the training data itself is very much in question, too. It can unintentionally learn and encode the biases, problematic viewpoints from the most of the times not clear dataset, like the internet (as indicated earlier). In many cases, the main objective is to produce a bigger and even bigger dataset to learn from (Bender and Gebru, et al. 2021, 610), even though a bigger dataset won't guarantee the mitigation of such biased, problematic or "hallucinated" results in the output of the model. *"Building training data out of publicly available documents doesn't fully mitigate this risk: just because the PII (personally identifiable information) was already available in the open on the Internet doesn't mean there isn't additional harm"* (Bender and Gebru, et al. 2021, 618)

An analysis of GPT-2's training data revealed the inclusion of 272,000 documents from unreliable news sources and 63,000 from banned subreddits. Notably, GPT-2's training data was later indirectly utilised in constructing GPT-3. (Gehman et al. 2020, 3356–3369) As it can get more worrisome, arguably there are notes going around that the bigger amount of the internet already contains more generated content than not. It makes the quality and credibility of some training data even more questionable. The

core issue lies in the biases present in the internet data used to train generative AI models, particularly those related to race, ethnicity, gender, and disability status. (Monteith et al. 2024, 34)

As these topics are very well argued in *On the Dangers of Stochastic Parrots* the next quote truly makes one think of their role and uses/consumption of such technologies: *"(...) reduce hype which can mislead the public and researchers themselves regarding the capabilities of these LMs, but might encourage new research directions that do not necessarily depend on having larger LMs."* (Bender and Gebru, et al. 2021, 610) One should understand that we do not possess endless resources for such overdone datasets.

Like I mentioned before, AI has a dual role in the fight of misinformation. Numerous studies have explored fake news detection in online social networks, but only a limited number have concentrated on leveraging artificial intelligence techniques for the automatic detection of fake news. (Aïmeur et al. 2023, page 15) On the other hand, Artificial intelligence (AI) is mentioned as one of the most prominent and debated methods for addressing misinformation on social media. AI enables platforms to combat misinformation at scale, handling diverse languages and time zones without solely relying on user reports. Major social media companies like Google, Facebook, and Twitter have adopted AI-driven approaches, integrating a variety of machine learning techniques to manage and mitigate the spread of false information effectively. (Fard, A. E., and Lingeswaran, S. 2020, 511) To go even deeper into this and to provide a stark contrast: *"However the notion that AI is a 'miracle cure', the panacea for fake news is optimistic at best. Mr. Marsden believes that while AI can be useful for removing disinformation once it has been spotted, identifying it in the first place requires the human touch."* (Euronews 2019)

It is extremely urgent to call for the spread of awareness of generative AI and similar terms, and also to place education on higher priority, especially with younger and older generations. As some people might have the motivation and time for self-education, even this group of individuals should be taught on a certain level and told of risks, mitigation. There are

legal and ethical issues on top of the discussed notions which can be focused on as a future area of research.

I believe that the current state of these technologies are temporary and stakeholders, like developers and designers are trying to limit all the potential pitfalls and a lot is coming in this still emerging field. Hopefully, a more transparent and trustworthy solution is being considered in one of those attempts evolving right now.

1.3. Social media as a brief snippet

Social media (SM) and social networks (SN) encompass a wide range of topics. Here, I will focus on a selected few that are closely tied to misinformation.

Social media has become an integral part of modern society, serving diverse functions such as entertainment, information sharing, political engagement, and business promotion. However, its widespread use has also given rise to significant social, cultural, and ethical challenges, including issues related to privacy, cyberbullying, filter bubbles, and the proliferation of misinformation. (Ienca and Vayena 2018)

A source (Aïmeur et al. 2023, page 12) highlights that recent statistics reveal that unintentional fake news spreaders are five times more prevalent on social media than intentional spreaders. Additionally, another study indicates that people confident in their ability to discern fact from fiction outnumber those uncertain about the truthfulness of their shared content by a factor of ten. These findings highlight a significant lack of awareness about the pervasive nature and impact of fake news. (Statista 2021) (Statista 2023) As we can image, the platform challenges are just adding to this: *“Social media platforms aim to comply with two competing goals: (i) Keeping the platform free and open to a broad spectrum of ideas and opinions, and (ii) reducing the spread of misinformation.”* (Fard, A. E., and Lingeswaran, S. 2020, 511)

Talking more about the risks, a lot of the times AI technology and social media platforms combined, like the remarkable ability of these potential combinations to target and influence individuals on a large

scale—using automated, subtle, and pervasive strategies—has generated significant concerns about their potential for manipulation and control. (Williamson, S.M. and Prybutok 2024, 15)

Filter bubbles occur when social media platforms use advanced AI algorithms to curate the content users encounter. These algorithms, also utilized by advertisers to deliver highly targeted ads, can result in users being exposed only to information and ideas that align with their existing beliefs and values, reinforcing their perspectives and limiting exposure to diverse viewpoints. (Ienca, M. 2023. 838-839)

Additionally, the presence of fake accounts and AI-driven bots on social media further undermines the credibility and integrity of these platforms. (Williamson, S.M. and Prybutok 2024, 15-16)

AI-generated images have the potential to exploit human biases, increasing their effectiveness in shaping public opinion. The emergence of hyperrealistic AI-generated faces on social media facilitates the creation of fake yet convincing accounts, making them powerful tools for manipulation and deception in the digital landscape. These tools are many times connected to political misinformation. (Williamson, S.M. and Prybutok 2024, 27) *“However, whether people are genuinely cognizant of their mistakes when recognizing AI faces remains unclear. (...) By acknowledging the human limitations in distinguishing between AI-generated and authentic human faces, we can devise strategies to combat the potential spread of misinformation and deceit enabled by AI technologies.”* (Williamson, S.M. and Prybutok 2024, 27)

People may develop habits of sharing news on social media without verifying its accuracy or may readily accept online claims as truth. Individuals with mental health challenges might be particularly susceptible to the influence of online misinformation. (Monteith et al. 2024, 34)

On the other hand, there are some related mitigation processes: Blockchain-based approaches utilize blockchain technology to combat fake news on social media by verifying source reliability and ensuring the traceability of news content. (Aïmeur et al. 2023, page 18) Continuing, approaches that use crowdsourcing (Kim et al., 2018) are grounded in the concept of the "wisdom of the crowds" (Collins et al., 2020) to detect fake

content. These methods depend on the collective contributions and crowd signals (Tschiatschek et al., 2018) from a group of individuals, enabling the aggregation of crowd intelligence to identify fake news (Tchakounté et al., 2020) and mitigate the spread of misinformation on social media (Pennycook and Rand, 2019; Micallef et al., 2020). (Aïmeur et al. 2023, page 18)
The following approach is Fact-Checking. It is usually conducted manually by journalists to verify the accuracy of specific claims. In recent times, several online social network platforms have started integrating fact-checking features. This topic, on more broader terms will be explored further in the next chapter. (Aïmeur et al. 2023, page 18)

Concluding this chapter: *"Misinformation detection (MID) in social networks has gained a great deal of attention and is considered an emerging area of research interest."* (Islam et al. 2020, 1) Analyzing the diverse forms of misinformation is challenging for traditional methods. As a result, deep learning-based detection approaches can be designed to adapt to different types of features for MID. (Islam et al. 2020, 2) More discussions are likely to come and the area definitely can benefit from such an interest of researchers, maybe fellow designers.

1.4. Current status of Fact-Checking initiatives for Verification Practices, aiming for spreading truth, debunking falsehood in several environments

In the concluding chapter of the literature review, and before presenting my research overview, I will delve deeper into the topic outlined above. This elaboration is essential, as none of the current approaches in this field are fully complete (as in there are no 'perfect' solutions), leaving room for future advancements, including contributions like my forthcoming master's work and similar studies. Additionally, some previously discussed topics will resurface, contextualized within the domains of fact-checking and information validation.

In the source (Reporters' Lab 2023) it is very well elaborated on the status of the topic. *"Currently, there are more than 200 fact-checking initiatives in more than 68 countries in the world".* Fact-checking helps

assess online information's credibility. Just to give an idea what is in the picture: *“(...) based on the research done by The International Fact-Checking Network (IFCN), 64.3% of fact-checking initiatives are non-profit organisations, 28.6% media outlets, and 7.1% are academic initiatives.“* (Fard, A. E., and Lingeswaran, S. 2020, 514) Another interesting finding of a study (Statistics Canada 2023) is that 17% of Canadians always use additional sources, 36% often fact-checking, and 32% doing so occasionally. What is most striking, *“Almost 1 in 10 Canadians do not know how to fact-check information”.*

The following comment from (Monteith et al. 2024, 34) is especially true for generativeAI, but I believe a natural approach should be close to it. It elaborates on how many people fail to realize that, unless they are experts in a given field, they need to carefully verify answers—even when the text seems highly persuasive and well-written.

Exploring key connections with the definitions of fact-checking. In this thesis, fact-checking is viewed as a "verification process" to assess the credibility of information or explore varying levels of reliability among sources. The core idea is that no source is entirely infallible; non-AI-based ones can be just as questionable, according to my perception. Emphasizing the act of "checking," it argues that verification is vital for any information, though its necessity may vary depending on the relevance or impact of the information in question. In some cases, irrelevant details may not require scrutiny, as they have little influence on the decision-maker.

In the context of media organizations, they play a crucial role in combating misinformation. Traditionally having a primary source of news for their audiences, they carry a significant responsibility to provide accurate and unbiased content. To meet this obligation, these organizations have increasingly prioritized fact-checking efforts. (Fard, A. E. and Lingeswaran, S. 2020, 512-514) Another type of definition from the same source highlights the key understanding of the word: Truth verification is essential to curb misinformation on social networks. Controversial content often spreads unchecked, but evidence-based verification can limit its reach. Fact-checking also improves AI by supplying reliable training data for more accurate machine learning models.

There are ideas about how to improve reliability and eliminate certain risks. In this paper (Bubeck et al. 2023) one approach is described as to improve training datasets with reliable information and integrate fact-checking algorithms and warning systems for AI-generated content. It's essential for users of LLMs to continue their education, understand the models' limitations, and develop the ability to distinguish between accurate information and hallucinations. This proactive strategy is crucial.
In addition, a special focus is also needed for spreading awareness and to implement the core ideas in education. The younger and older generation might be in a more critical situation.

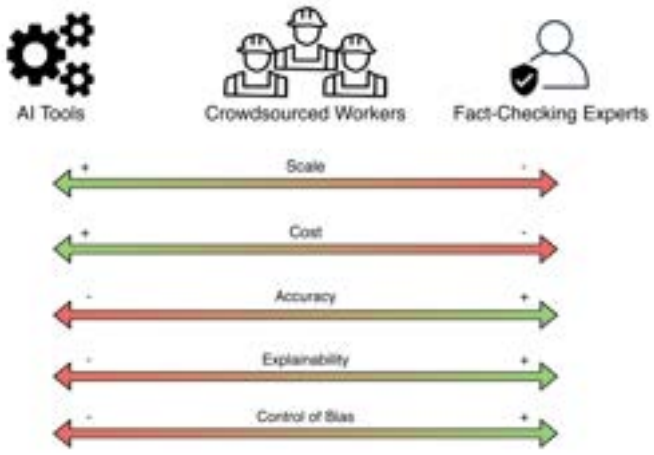
Moreover, Regulatory bodies may need to take a more proactive role in overseeing AI deployment in sensitive areas, ensuring that AI systems used for critical decision-making undergo thorough testing and certification processes. (Williamson, S.M. and Prybutok 2024, 5) It is to mention and to be thought of: "*(...) vulnerable individuals can be exposed to additional harm without their awareness, especially in digital environments"* as highlighted in the sources (Williamson, S.M. and Prybutok 2024, 17; Strümke et al. 2023).
On another note, it is also a certain concern when a decision has to be made quickly in a high-stakes scenario. As implied here (Schemmer et al. 2022; Zhang et al. 2020), the lack of explainability for example, in an AI-generated advice can undermine trust in human-AI (or source) decision-making, as users may struggle to grasp the reasoning behind the AI's suggestions. This issue is especially critical where the consequences of errors can be severe, particularly when AI assists human experts in making crucial decisions.

Another initiative is Explainable AI (XAI). The main objective here is crucial for demystifying how AI systems function, fostering user trust by emphasizing transparency and understanding. By prioritizing clarity in decision-making processes, the ability to create AI systems that are not only powerful and efficient but also user-friendly and reliable. (More to be found at: Hacker and Passoth. 2022)

Crowdsourcing can be another example of the method of fighting misinformation through the power of the crowd and fact-checking. As defined here (Demartini et al. 2020, 69): *"In the context of misinformation, crowdsourcing is a methodology that can provide, for example, truthfulness*

classification labels for statements to be fact-checked." The same source mentioned why this is a valid idea and approach: since a key challenge for expert fact-checkers is that due to limited resources, prioritizing which items to fact-check from a vast number of potential candidates is extremely hard.

Moreover, the authors continue with how the crowd can assist expert fact-checkers by evaluating the "check-worthiness" of content, helping to identify which pieces would most benefit from expert review. This obviously would make it more time-efficient on the side of doing actual fact-checking, rather than spending extra resources on just evaluating and choosing the next target. The mentioned paper delves into a proposed framework which I would like to share with the readers of the thesis: they talk about the emerging idea of hybrid approaches to fact-checking, where there are several actors (Fact-checkers, the crowd, AI) combined. (Demartini et al. 2020, 70) Providing more details, they propose a waterfall model that applies varying cost/quality trade-offs at different stages through optimized task allocation strategies. From the same page, *"Fact-checking experts are the protagonists of the framework and are the ones who make use of the other two components to optimize the efficiency of the fact-checking process and maintain high-quality standards." (see image below)*



("Trade-offs between the actors of the framework." G. Demartini et al. 2020, 71)

The source further examines to model and asks how the experts could implement this framework into their workflow, opening new research questions.

Highlighting concrete instances of current days: Meta combats election misinformation by partnering with fact-checkers, removing

inauthentic content, and using transparency tools like the Ad Library. They also provide accurate election information to voters and monitor misleading messages, particularly on platforms like WhatsApp, to ensure fair elections. (Meta 2024) Of course, there are older examples of when Facebook, Instagram, Google added their fact-checking initiatives starting from 2017. (The Guardian 2017, Instagram 2019) At the same time, it was criticized by some (Wired 2019). Another interesting and thought-provoking insight was emphasised by Stephan Mündges (EFCSN coordinator) on LinkedIn: *"Our colleagues from Science Feedback find that Community Notes on X are absent from most tweets that fact-checkers found false or misleading."* (Mündges 2024) This leads me to think that the human cognition is definitely irreplaceable as of today to define those artifacts with harmful content. On another note, since the end of Meta's collaboration with fact-checking professionals, many people have expressed concerns that this is leading towards a bad direction in terms of online information in the current system we have. A strong conversation is always there because we all are very much in need of reliable information, but the method and how to get it simply causes a big and still ongoing challenge.

As it might get more clear, transparency should be the default, ensuring a clear mind of the receiver of the information (the ability to make a decision via critical-thinking), and the development of trust in the source.
Accuracy and accountability are highlighted as crucial on the same plate as transparency. (Williamson, S.M. and Prybutok 2024, 4) Therefore, trust can be on the same level as the previous values. Not believing anything, being an over-scepticist could be just as on the edge thinking as believing almost everything we see online. The dangers of both are something to consider while designing and developing a new product that aims to operate in this environment of today's information landscape.

2. Research Methods (workshopping, a survey and interviews with selected stakeholders), Evaluation.

In the following few chapters I will elaborate on my personal research results where I was trying to gather as much valuable information as possible

from each stakeholder. Stakeholders: designers, managers (influential people) were asked to participate in workshops. For more designer feedback, I conducted a few interviews, to see their thoughts on the current state of misinformation, fact-checking. Developers were interviewed the same way, but with a focus on AI/LLM development. I managed to talk with an AI expert on the field. Thankfully, fact-checker specialists were open for discussions, too. In the meantime, a survey for “general, everyday users” was opened. More details are portrayed in their own subchapters. As for the participant’s consents, they all acknowledged what they are being part of as my research and then thesis, but I decided not to mention their names unless I asked specifically for it.

2.1. Workshops

In the upcoming subchapter, the main findings and insights will be presented from my two 1.5 hours long workshops. Documentation can be found in **Appendix 1** for references in this subchapter.

Topic of the workshops: Firstly, Misinformation in a wider context. Later during my literature review I realised that I wanted to focus on a Fact-checking narrative, that is why I planned and changed to include that, too. *Main objective:* to get an overview of the current behaviours and mindsets towards Misinformation and Fact-checking (based on selected stakeholders’ feedback).

Participants, environment: I have had several stakeholders named from the very early stages. For the workshopping, it felt natural and given to work with designers and managers/directors, influential people. The reason why is that at the time of the thesis writing, I had been working for a SaaS company as a Product designer intern, where the close connections were with these stakeholders, additionally, the resources were given in the company setting (whiteboards, rooms, post-its, etc.).

Designers make changes, through the collaboration with developers and other involved people in the process of creating a new solution. Managers, of course, have their own influence on them and might be able to change the course of thinking of others.

By the end of the facilitation, the first workshop ended up to be mainly designer-heavy (5 designers and 1 PM, also can be named as individual contributors (IC)) and the second to be manager-heavy (3 managers and 1 designer). From now on I will refer to them as “designer workshop” and “Manager (or director) workshop”. (It is important to note that I was not comparing the 2 groups in any competitive way. By the end of the facilitation, it was even more clear that the fact that I had them separately did not really matter that much, it is not an important factor after all.)

Content of Workshops: Both were planned to have 3 main tasks, with the focus being on the 2nd task. Only the 3rd tasks were planned to be different (not mentioning the small changes between the two), but at the end, there was no time to facilitate those.

First Task: Ice-breaker exercise, collecting ideas from all participants about what their main observations/stories are on misinformation and how they access information nowadays.

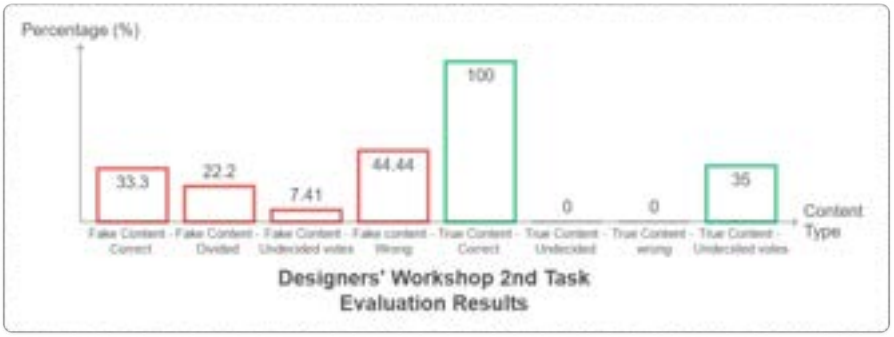
Second Task: Provided them with several contents which are from 5 main types: social media posts, images, data visualisation, news headlines and a video. The main objective of the task was to identify certain signs that could be associated with falsehood, or if generally the content feels off or suspicious - these would be categorised as False. Participants also had the option to put the materials to the category True, if they could not identify edited or fake aspects. They all had to explain why they thought so.

(The *3rd tasks* were about content generation use cases and fact-checking ideation, but this way it is unnecessary to include them here.)

Assumptions, Findings and Insights. Firstly, it is important to note that I do not try to compare the two groups in a competitive way. They are no good or bad. My assumption before the facilitation was that maybe the manager/director group might do better with the identification of fake content. I was hoping that they could name a few sources or methods they both use, and was interested to see whether there could be a correlation.

Part of the findings, even though the focus is not on quantitative results in this case, I will show the first visualization of both groups’ results. Although it is interesting, since some of the participants highlighted that they

sometimes achieved the good answers by accident or for the wrong reasons, chances are that these results are not showing the total truth.



It is easy to see that my assumption became true, the Manager group in fact was more correct on fake content. Additionally, as the analysis of the chart, Directors excelled at identifying fake content, achieving 66.67% accuracy but struggled with true content, with only 40% accuracy and 40% divided votes, indicating uncertainty or distrust. Designers, conversely, showed perfect accuracy (100%) for true content but only 33.3% for fake content, with 44.44% incorrectly labeling it as true, reflecting susceptibility to misinformation. Both groups displayed hesitation, with Directors showing 2.78%-5% undecided votes and Designers at 7.41%. Directors demonstrated balanced but inconsistent judgment, while Designers excelled with true content but faltered on fake, highlighting contrasting strengths and vulnerabilities. Now, this should be followed by additional workshops with new groups of people to be able to define a pattern, but still, it gives an idea and a potential direction on what to research more.

Getting more into the type of qualitative insights, a few highlights will be provided, for the rest see the workshop documentation (Appendix 1).

One designer participant said *“The main source of misinfo is AI”.* During the second task, another designer explained their choice by saying *“TV doesn’t lie”*. For the same matter, *“I hope I can still trust Fox”*. The content was

ultimately fake in this case. Resonating with the following: *"I would be very disappointed if the national content would be just fake"* (from another designer participant). In many cases, it was misleading content in this context.

To make the summary of the 1st tasks visual, I have made an illustration for the discussion results and my notes throughout the workshops. It is exhibiting commonalities and differences of the groups.

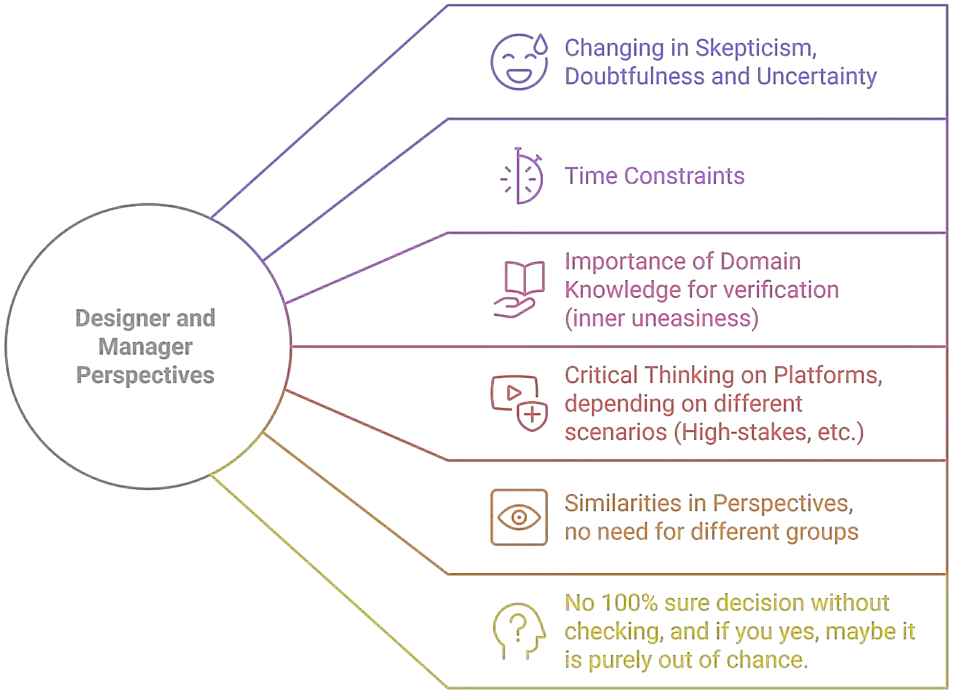
By the end, Designers highlighted that they definitely grew more sceptic (sometimes over-sceptic) along the evaluation process of the 2nd task. One participant said: *"I was thinking at first that I will get all of them right, if not, at least most of it. I was shocked to realise sometimes I was too naive"* (Appendix 1) A designer said they will absolutely be more conscious about this in the future, it was impactful for them.

Interestingly, a manager highlighted that as long as the content is not harmful or manipulative and such, the image/content itself can be titled fake, they will rate it as True. *"It is a picture of an animal",* therefore it is True. (notes on the fake lion picture and true cat picture). Something similar from the other workshop, where designers emphasised that when an opinion is illustrated, it is just an opinion. The definitions of True and False were clear for most, especially for the second workshop I was trying to help them to understand well. Still, such an insight is there which indicates to me that it is clearly hard for people to make Black and White decisions. For that reason there were a few undecided votes, too. Additionally, they were without the chance of doing a double or any kind of check, hence why their decisions reflect in many cases, uncertainty. With this, the goal was to simulate the everyday situation where we do not have time in our daily life to fact-check everything, to say

the most, hardly anything. You simply cannot make a 100% sure decision without checking nowadays, and if you do, maybe it is purely out of chance. Some of them actually said that they had got good answers sometimes for the wrong reasons (accidentally). It is to keep in mind while making the evaluation of the quantitative results.

There were other comments coming from the Managers group, like: *"it was one of the hardest workshops this week, I had to use my brain a lot"* (Manager from 2nd WS) which implies it is indeed hard for everyone to make such complicated decisions. This second Managers group as of my observations seemed less surprised by the results while discussing. They were very cautious from the beginning. Another note from participants was that it is even harder to judge when you are not familiar with the content's topic. This observation and feedback represents the same idea as what I had read (in literature review). Some of them admitted that they know sometimes it is not ideal, but on some things they just ask ChatGPT or browse social media. It is extremely common, as their answers suggested. There were a few who reflected on social media, saying that in itself, it is not a credible source. A manager mentioned that sometimes they are not so interested in daily news, therefore they do not really check, but on the other hand, they carefully select the info when it comes to professional life. This contributes to the different interests and levels of importance aspect.

To conclude this chapter, the main insights for me were that in many ways, participants behaved according to previously discussed studies. Overall I appreciated that they thought of this topic as such a complex and important domain. I felt great when they expressed their intentions of giving more of their attention to misinformation in their personal and maybe professional life. It is quite a challenge that they had to face and I hope this way and with this paper I can manage to bring some awareness to the topic, moreover with my future masterwork.

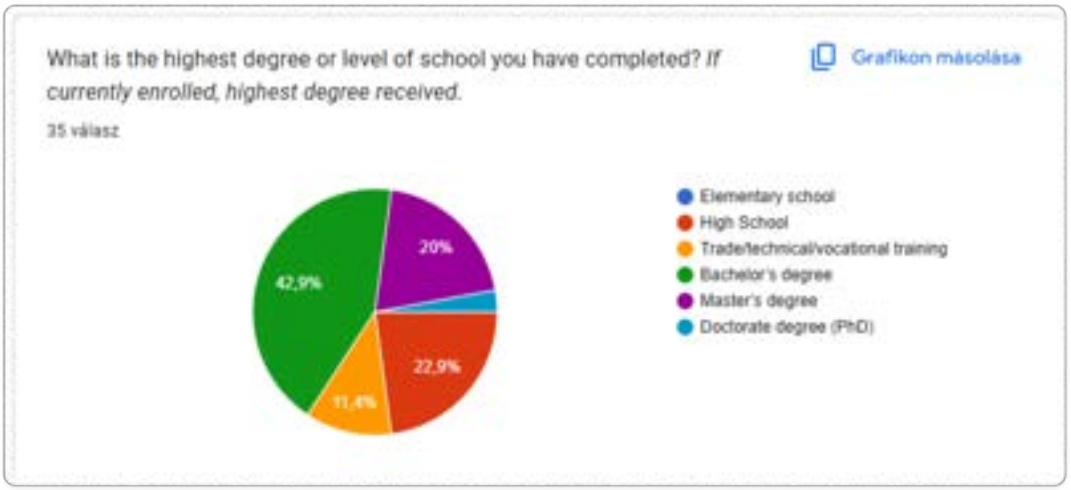


2.2. Survey Evaluation

For the survey, I used a quite wide range of *target audience*, titled for now as “general, everyday users”. This meant that everyone could complete it who somewhat had access to a tool for digital landscape usage (laptop, mobile, etc.) and internet. It was spread via a Google Form link through social media platforms. The questions are provided in **Appendix 2** for reference, additional materials can be found there.

The goal of the questionnaire was to have a more quantitative overview and findings of the topic, from many online. The timeframe of the running survey was about a month with a final number of 36 survey completions.

Demography. (More details for survey documentation in Appendix 2)



I have received answers from 5 different educational levels. It was dominated by BA degree obtainers, the second most popular were people with High school diploma and the third were MA degree holders. The least (1) popular was PhD. More Female (65,7%) completed the survey than Male (34,3%). The age range of participants was between 18 and 64 years. The dominant combined age range was from 18 to 24 years olds, with a percentage of 88,6%.

Moving onto the questions and evaluation of each, the first was *“Can you share a time you encountered misinformation online? How did you realize it was false?”*. For this, I grouped the answers to 5 different categories, related to content. Answers related to Misinformation, disinformation in a mostly harmful, misleading context: mentioned by 13 people (36,11%). Critical, fact-checking approaches by 12 people (33,3%). Domain, previous knowledge used and mentioned in order to judge misinformation by 5 people (13.89%). A few, 3 (8,33%) of them said that they have no conscious memory of misinformation happening in their environment, or having no experiences of being misled. The rest named some online marketing related misinformation. From the answers, many emphasise the dislike towards state propaganda, AI generated misleading contents which they encounter on daily basis. Some mentioned the clickbait and misleading news headlines on social media. There is a big attention dedicated towards US elections and twitter/X misinformation.

The next mandatory question was *“What are your main sources for news and information? How do you determine if they are reliable? You can*

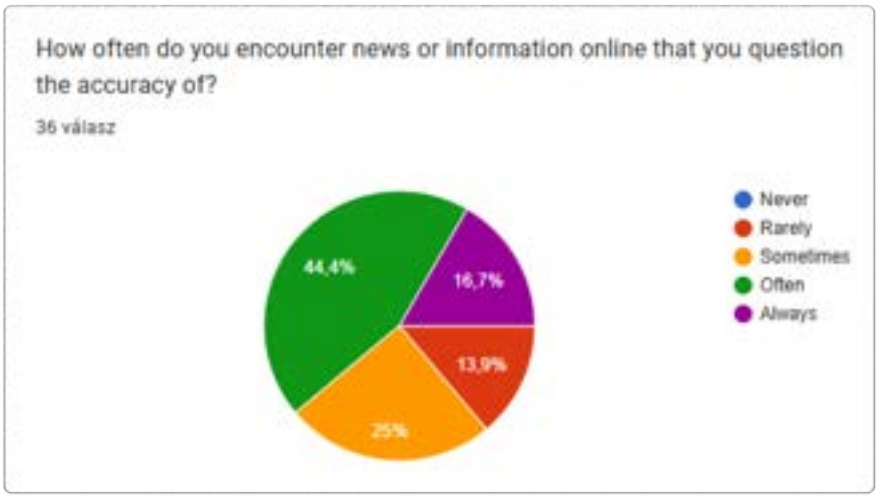
think of your daily habits." For this, I provide 4 main categories. The option "Other (websites on the internet, digital newspapers, etc.) Sources of Information" received the most, with 19 answers (52.78%. Here the main focus of the users shifted towards a few liked, which were mainly judged by them as credible sources, plus with the preference to check several media. Well-known Hungarian (Telex, Index, 24.hu, fokuszcsoport, Magyar Nemzet, etc.) and international sites (BBC, CNN/New York Times, Guardian, Google News feed) mentioned. Some expressed that they actually do not really follow the news. Interestingly, for local news some have to follow Facebook and local newspapers to get informed. The next is "Social Media and Information Consumption" on what 11 (30.56%) users voted on. A bit of an unexpected feedback from many was that they know that using these platforms (named TikTok, Instagram, Facebook, Youtube, Twitter/X, Reddit, Telegram), they risk the quality of the information, but they are aware of this. Therefore it seems they can engage with this method without seemingly receiving negative effects. They added that if they really care about something, they will eventually check it somewhere else. A user described this as "*A quick but not always reliable way".* Next up, 5 people (13,89%) expressed their preference for Non-Government Media consumption. *"Government media has lost a lot of trials for handling misinformation and non-government did not"* (a participant added). Finally, 2 of them (5,56%) had written some radical comments: *"Well, I don't know, I'm just smart and have common sense"* and *"I think there are no reliable sources anymore"*.

The third question: *"Are you familiar with fact-checking tools or strategies to spot misinformation? Do you use them?".* The majority of responses (52,78%) indicated Lack of Familiarity towards this topic. A few of these users had some ideas, but still, they stated that they do not actually use these. Additionally, 30,56% of respondents elaborated further on their strategies, mainly coming from their background (media studies, journalism, etc.). "*I know there are websites for fact checking—which I don't tend to use as much for the misinformation I encounter, but they are helpful for some political claims where data interpretation has to be taken with a massive chunk of salt. "* and *"Sometimes I ask ChatGPT about facts that I read"* are some quotes that are related to some other feedbacks from other research

methods, too. The following illustration and some additional ones (that can be checked in Appendix 2) are based on the question: *“How truthful are these Social Media platforms, in your experience and opinion?”* and its responses. From this, the most used platform is Youtube. In second place it is Facebook and Instagram with equal votes.

For the next visual, (Appendix 2, Presentation of most trustful platform) I highlighted with red background the previous platforms which are titled as most used. The ranking is going from the least trusted source of information to the most trusted one (from left to right). Youtube is almost always among the most trusted ones, with Instagram and Facebook ranked much lower, only once exceeding Youtube (in the “Sometimes” category). Instagram is ranked higher than Facebook 3 times out of 4. Interestingly, none of them was put in the “Always True” category. It seems like a strong connection with the most trusted and most used (Youtube) platform.

Moving to the next, *“How often do you encounter news or information online that you question the accuracy of?”*



The “Never” option is zero, which states that questionable quality of information is very much present in the online space. Most people picked “Often”, followed by “Sometimes” and thirdly, “Always”. This stresses the urgency of highlighting possible fake information, or at least the uncertainty of judgement.

As for the last mandatory question, *“In a few words, how confident are you in distinguishing between reliable and unreliable information online?”* the following answers came in: Responses range from very confident to not confident at all. Confidence often depends on the subject or source of information. Many mention researching suspicious information on other platforms. Doubt and skepticism play a key role in decision-making. Increasing concern about information validity and the need for further verification resurfaces.

Getting into the optional questions, the first one is as follows: *“Optional: Does misinformation affect your personal or professional life? If so, how?”*. It was answered by 26 users. Very visibly, 13 people (50%) are in the category of “No, or not aware”. As some literature materials have discussed this as well, many people can’t realise this. What is even more interesting is that in this context, 6 (23,08%) people mentioned something related to personal life and the same amount of people did so for professional life. The question *“Optional: Have you used misinformation-reporting features? Were they effective?”* got provided with thoughts by 23 users. The majority of them can be put in the “Yes” category, with 14 answers (60,87%), the rest of them are more related to “No”. Most users find misinformation-reporting features to be helpful but not always immediate. Some users report that the reported content rarely gets banned or actioned upon. Many users report misinformation but rarely receive feedback on the outcome of their reports. Some users have reported fake reviews or profiles and have seen them flagged or banned. Reporting systems are perceived as having limited impact on removing misleading content.

The last part of the survey concludes with the next question: *“Optional: What information sources are the least accurate in the online space, in your opinion?”* It was filled in by 23 users. Overall, Facebook is criticized for clickbaity fake content and product placements. YouTube and Reddit are

seen as good sources if users know where to look (which correlates well with a previous mention of Youtube being mostly trusted). Hungarian media is mentioned as being aligned with the government, influencing seniors easily. Facebook, Twitter, and TikTok are mentioned as platforms with fake information the most times. Mentions appear stating that biased news portals like Origo and Hungarian television channels are considered unreliable.

Finishing this section, a wider overview of perception is provided, and some ideas happen to resurface from time to time. Some key insights, like about the usage and judgement of Youtube, or the different approaches to fact-checking are going to be beneficial for my ideation process regarding my planning of masterwork.

2.3. Interviews

2.3.1. Designers

I conducted 3 discussions with designers from different backgrounds. The core goal with them was to identify the level of awareness of the current state of misinformation in the online space, as well as the development of AI products. The focus included examination of whether they have had any professional or personal experiences with the topic. I always tried to ask them about their narrow field as well. For this I coordinated some desk research online and designed the questions accordingly. As described in the workshop participants section (2.1.), designers are one of the key stakeholders, because of their possible impact on future solutions. (The documentation and links are available in **Appendix 3** for this whole *interviews* chapter)

Starting with one of my first interviews in this research process. This field of the designer is mainly focused on more of an artistic approach, having one running application about anti-social media. His other initiative was an app that uses AI as a photo-taking educator. My main insights are the following: He is not extremely involved on the AI part and they had used an older model for the first development of the product. He would firstly change the model to a more up-to-date one, if he had to improve it. The main app is minutiae (the anti-social media app). The goal of it is understandable,

although it is more like an art project, and is less related to my topic than I was expecting. I was hoping for Svetlana (the AI photo-taking app), but it is not developed in a way that can be a part of my discussion. Nevertheless, it was nice to hear about different notions from this particular interviewee.

The Minutiae app focuses on capturing imperfections in photos to highlight nuances of life. One thing that came to my mind relating to this is that in fact, users do not have the time to edit pictures, there is no feed where any misinformation could take place. It is so limited in functions, that it seems to be a true content medium.

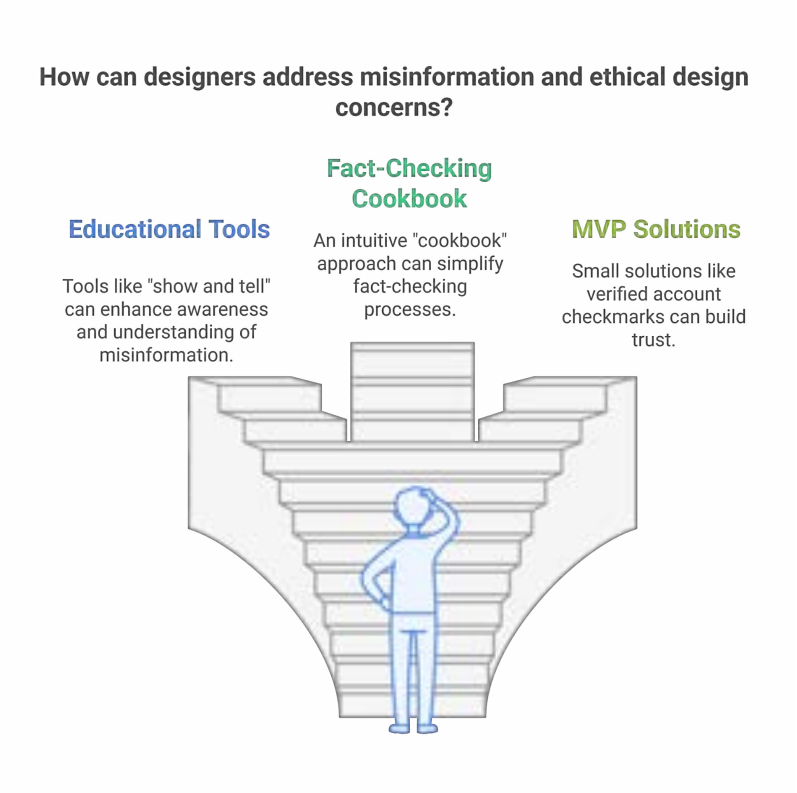
Svetlana, the AI-assisted photography instructor, challenges traditional photography approaches. The app aims to understand human behavior with minimal data (your previously taken pictures, their style, etc.), and to challenge the user to be different. One interesting aspect is that the refusal message is phrased as “Computer says no”. This would suggest the uncommon feature, that is to not humanise the AI, but contrary to that, it has a name, “Svetlana”.

A designer from the responsible design field also agreed to share thoughts with me. The 3 main conversation cornerstones could be identified as the following: “Data Understanding and Education”, “Responsible Design and Ethics”, “AI and Sustainability”. His beliefs emphasize the need for responsibility in every product and the goal of achieving positive outcomes. He highlighted the need for continuous education on critical thinking and fact-checking, especially among children. He was really aware of the discussions related to the potential impact of AI on fuel consumption and environmental protests. Emphasis was put on the need for integrating ethical considerations into the design process. Although it might seem easier to say than done. Some of his experiences linked to this, as he discussed the challenges and opportunities in applying responsible design principles in business contexts. He was ideating on a kind of future solution that uses the “show and tell” principle. Making users exposed to such circumstances might be a great approach to awareness, educational purposes. His prediction on what will happen in the field of AI is that it will get more regulated, and that the most improved enhancements unfortunately might firstly get into the hands of those who are not using it for good. Another idea was the display of

different perspectives and not giving an exact answer to the user, so that the final conclusion is totally up to them.
Overall, the main ideas presented here gave me a few inspirational thoughts, like about the awareness aspect, limited resources and societal impact of technology and the educational system.

The last designer interview was more about stressing other concerns. In his case, AI is used experimentally for rewriting text and creative tasks. He has to be careful about data privacy and confidentiality in freelance work, therefore he is not able to use AI. When asked about some potential interest in creating AI guidelines for design work and user benefits, he could imagine those being helpful in his work and personal life. His biggest concern, the fear of misinformation affecting family members was deeply stressed by him. Additionally, he had heard of a few fact-checking initiatives by some big social media platforms, but he has limited experience with those. He described the same limited experience with misinformation personally (or that he is not aware). His one idea was about a suggestion to use fact-checking tools like a cookbook. He could imagine pulling out a few best practices to execute them in some exact situations. A tiny observation of his was that some small additions like checkmarks for verified accounts can help, too. In his opinion the focus could be on a small MVP that just makes people a bit more confident on what could be credible or what to double check, adding the guideline/cookbook example to it. He described it as *“something that my mom would understand, too”*, which again reflects back to his main concern of possible affected relatives.

To finish the designer insights, it shows quite a diverse perception of the topic from a designer lens. I was expecting a dynamic that would turn more into the discussed risks of AI development, but regardless, it produced some nice takeaways. It clearly presented a general concern about my topics but most of them could not really elaborate on it on a deeper level. The following stakeholder interviews are more focused on specialist overviews, rather than a general discussion or ideation. A short summary is provided of the discussed pathways below.



2.3.2. Developers

As for this group, it was very essential for me to think about interactions and knowledge-sharing with them since after we, as designers, have to collaborate with them in order to achieve the envisioned outcome. For this stakeholder group, I ended up conducting 2 interviews. The first developer was asked to participate because previously he had some experiences in using AI in development. The second individual was actually recommended by another stakeholder (a journalist) from the fact-checking group, so I reached out. This developer is currently working on a fact-checking tool that will be used by journalists and researchers and they are using an LLM model for the matter. (More information is available about the journalism part in the fact-checker (2.3.4.) interviews.)

Starting with the first developer (Viktor with consent) who had worked with such technology before, not necessary focusing on it right now: my main questions to him were structured around the fact that I wanted to get a better understanding on how they work with such an AI or LLM technology, why, when (in what circumstances, what are the reasons behind it), and most importantly, how they prepare for the potential risks or errors that might occur. I wanted to ask him about some projects, if he could provide some

practical examples. At the end, we discussed his mindset, current opinion on the field.

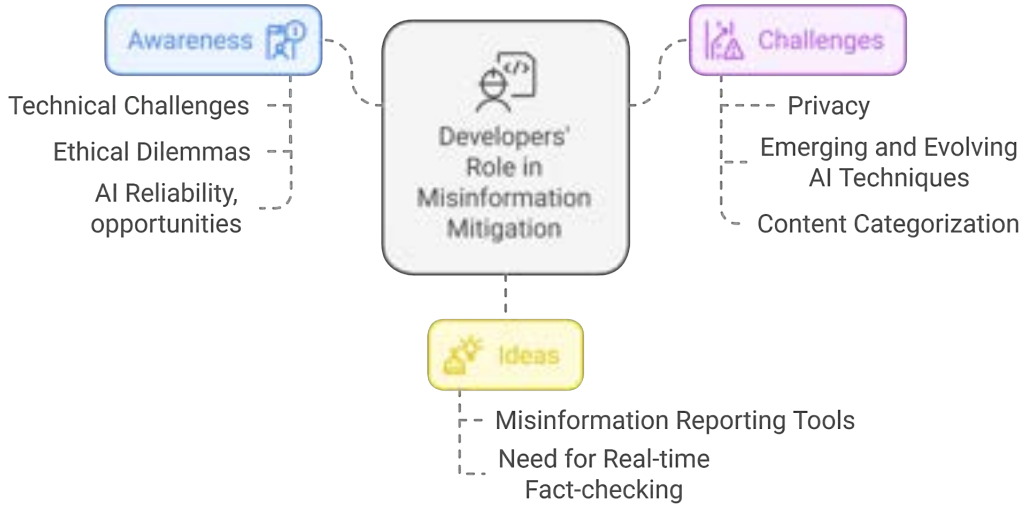
Overall, he sounded pretty pessimistic. He emphasised the fact that humans struggle to process vast amounts of data effectively. He was actually highlighting a bigger issue in today's era, related to the problem with misinformation and fact-checking. Society and human behaviours are heading towards a direction that he is not seeing any changes coming soon, unless humans can act as a union. He added, *"If your solution starts with "If everyone could just do this…" then you should just forget about it, it is not going to happen."* Architectural changes and regulatory adjustments are necessary. By the end, he honestly said that he cannot imagine any solution to this problem area that is not just "another label" which has little to no impact.

Regardless, he started to ideate on some feasible, proposable solutions, like add-ons on certain social media platforms, built in the system on each device (based on data). This could highlight questionable information and help the user act upon. In more details, he touched on the psychological effect of misinformation, in a way that is reflecting the over-scepticism aspect. He told me that these constant inspections of validity might lead to constant uncertainty and depression, even. When asked, when to use AI, he answered that "only when it is very needed". He explained that in his work they try to use it less and less, or to use it only for the "shiny details". The base coding and work would be done by hand. Alternatively, he agreed that it can be a useful shortcut in some cases. We concluded that the quality of databases are key challenges. When it came to education and awareness, we tried to elaborate on which user base could benefit from what, and he mentioned that *"If I had to explain and show any type of these solutions to my parents and the people of their age, they probably would not learn it and would just continue to use Facebook as it is"*. This helped me to realise that it is clearly not easy to design anything for such a non-digital native target group combined with a complicated problem. Plus, the question of my target audience was really evolving at this time and was made of a higher priority, which was actually a recurring pattern in every scenario during my research.

Another interesting insight of this conversation was that in his opinion and experience, many companies implement AI just for the sake of marketing and fashion, even when it is not needed, and he felt that is highly annoying. In addition to this and combating misinformation, many of these names are stating that they had tried everything in their power to reduce the potential of fake information, even if it is just for show. Interestingly, he added “in my opinion many people just like to blame AI for many things but it is not the source of all fake and bad intentions”. He advised me to think carefully and to be mindful about my choice of the interface. He explained, many people today view content on tiny screens, like on the phone. But there are more ideas to consider about it.

Continuing with the second interview with the fact-checking tool developer. The tool is being developed for various users, like journalists and researchers. It is however not for the general public. The main purpose of the tool is to help gather the “hot topics” that can serve as a possible next topic for the users of the tool. Though it has several functions, a few that I am going to highlight caught my attention. For instance, “quality-scoring” is another key element because the sources that the software can collect are coming in with their credibility. It mostly comes from the source’s style of speech. There was a debate whether the content should be instead “labeled”, but just like how the journalist from Lakmusz said so, it is rather limiting and can quickly pile up. So overall the scoring system seems to be a better option when it comes to “categorizing” content. When I asked about what kind of AI is implemented in this model, the developer clearly stated that it is their own tailored Large Language Model (LLM), therefore it is not “generating” any false connections by accident, it just works with already existing information. After the collection of the documents, the product helps with analysis, too. It can connect similar narratives from various places. Other than this, something that can be important from a technical perspective, that based on user testing feedback, they had created a mobile and desktop version with differing features and resources. While on mobile, the key objective is to access information fast and a few features will be used, on desktop the users can perform more advanced flows and will need more capacity and information. So for that the team limited certain actions on

mobile for a faster loading of information. To sum up the development section, I created an illustration with the key takeaways from the two interviews below.



2.3.3. AI specialist

(I have received the verbal consent of the interviewee to reveal his name for this chapter.)

Several crucial ideas came up during my conversation with Pontus Wärnestål. I reached out to him in order to get a more in depth view on how AI works and what are the current most important challenges of today, in his viewpoint. He is an active teacher and writer, strongly focusing on such topics concerning AI.

One of the most relevant and interesting ideas of this meeting was the identification of 3 pillars: Knowledge, Truth and Transparency. As he mentioned, these are values that are oftentimes missing, for example out of LLM models. These are crucial points which I try to implement in my own solution and ideas. We discussed the presence of bad actors in this story, but I decided not to specifically build on disinformation. The high priority of education and awareness came up here as well, which led to an idea of “guidelines”. These guidelines could be designed for designers or users. They could use this as a critical standpoint to mitigate misinformation and to be on the same page, even on a difficult topic.

These ideas are to be improved, but I noted these in order to have in hand a wide range of approaches. Another important key value can be linked

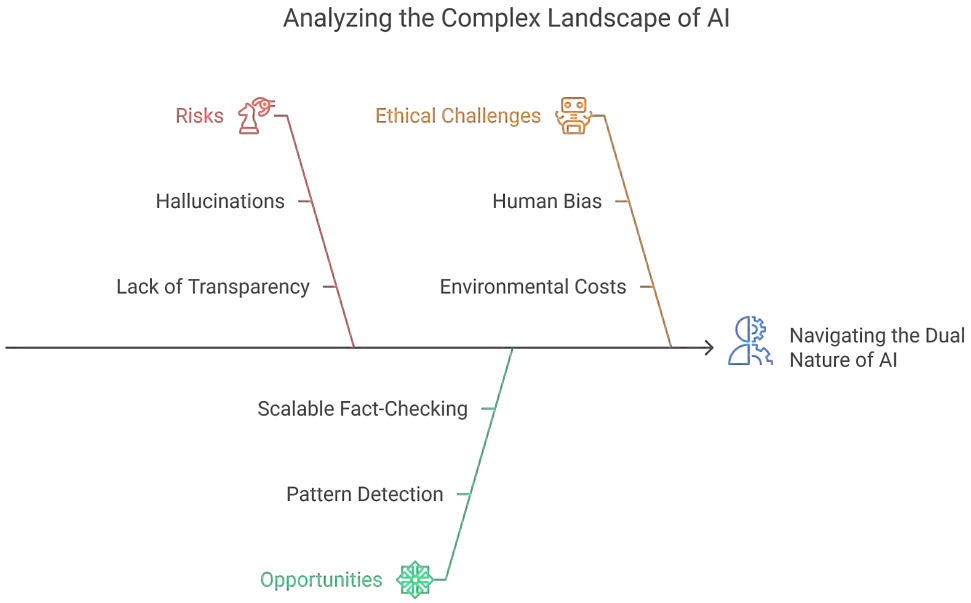
to Transparency, which is Trust. Like it was previously mentioned in other chapters, a lot depends on how trustful a source can be. None of the two extremes are good in terms of misinformation. Uncertainty can be one of the roots of the spread of unreliable knowledge. Communication is the base of many things, it is true for AI usage, too. There are examples of communicating probable errors and such after an AI/LLM output. But is that enough? Moreover, he pointed at a very basic question: what one might understand as AI? At this point during my research, it was not that clear at all times what people might mean when they share thoughts about “AI”. We mainly concluded that the most common understanding is that certain LLMs are one example of that.

After this finding, it was much easier to talk about this technology even with other stakeholders during interviews. We actually circled back in many ways to my literature findings, because he was clearly agreeing with most of my results on that. One of those was the common bad idea of using AI as a search engine. Moreover, he recommended some additional sources which I ended up including in my work.

Another key moment of this session was the importance of defining what “general user” really means. I knew already that it does not tell much, but again, this pushed me to think more of my future target group. Of course, for the research survey, it was fine since it was just part of my indirect methods, but for the actual user base who I would design for, to understand them better and to know what to include, etc. I have to be more concrete. He mentioned the problem with the term “hallucination” of the AI system, which I realised at the time, what dangers might bring (the humanisation of AI). He proposed “you can change the behaviour or change the environment”. He stressed how important “signals” are in this context in the online space. People tend to not have much time, like a split-second to determine what they are going to do with an output while scrolling platforms. Some very visible indicators are crucial if attention is needed. Plus these indicators are often missing from an LLM output (transparency problems) unless asked for.

These points above definitely helped me through my research workflow and encouraged me forward. It was always a good sign when I felt like most of the facts I am hearing are connected in a way. As a summary for

the AI Landscape, the following illustration was created (not only supported by the interview, but also the literature review (Williamson and Prybutok, 2024; Monteith et al., 2024; Bender and Gebru et al., 2021; Fard and Lingeswaran, 2020; Aïmeur et al., 2023).



2.3.4. Fact-checking professionals

This section is highly focused on the domain that I consider right now the most important. For this matter, I managed 3 different meetings with professionals from the field. The first one is with a journalist from the Hungarian fact-checking site, Lakmusz. The second was organised with the European Fact-Checking Standards Network (EFCSN) manager. The final one was with a teacher who now mainly works in the field of data visualisation, previously a journalist. He had worked before at another similar site to the first one, called Átlátszó (which means transparent).

This first interview with the journalist from Lakmusz brought me closer to a very practical case of fact-checking. Most importantly, he explained to me the processes, workflows they are working with right now. He helped me understand their challenges. He was the one who recommended the interview with the second developer. He briefly touched on the initiatives they are developing for them and how he had the opportunity to test it a bit. Lakmusz is part of the EFCSN and IFCN network and the HDMO disinformation hub.

Some of the themes and topics discussed were the following: Starting with his observation of not all tools being useful which are titled to help users like him. He added that he is not really using any of these, but he is aware of them. He said that the output of such tools can be true, but sometimes it is unreliable (AI detectors and such). He is a bit sceptical towards the whole but he knows that this technology will be present in the future.

Getting into more detail about their workflow, he mentioned the utilization of international networks and foreign fact-checking sites for verification. He simply elaborated on how important it is to do at least the basic steps you can do, like checking for credible information and references to ensure accuracy. Another side of the same story is that you have to verify author reliability but acknowledge that even reliable sources can make mistakes. A nice approach would be to encourage critical thinking and fact-checking for regular users. When asked about some simple tips and tricks, he introduced me to the usage of reverse image search and LLM technology for image and sound recognition. Mention of monitoring fact-checking articles, as it might be obvious from his point of view, but I believe many daily users do not even know about such sites existing. On top, he strongly agreed on the fact that a balanced level of trust is needed and should be maintained. When it comes to Lakmusz and his job, he explained that he tries to make the fact-checking process understandable, writing it in detail to build reader trust and understanding. He also called it “repeatability” for verifying the source independently. Another journalist approach is to clarify complex concepts to enhance reader comprehension and engagement.

I asked him, what happens when the readers do not read? How do they adjust to that? Do they read? He answered that yes, they are aware of this, and they try to put the most important facts in the beginning and to give a guide to the user, as in to help them make their conclusion on the topic. He added that they operate an Instagram account, too, where there are usually visual materials provided which might be easier to scan. The topics can be recommended by readers and the journalists can choose, too. It is mostly up to how relevant and important the topic can be, plus checking if it is an international or smaller action, these factors can all be involved.

Other challenges can be like when data is private or incorrect, maybe when datasets are unavailable. Importantly, transparency issues with governmental and institutional data can be a difficulty.

More discussion is provided on the fact-checking AI initiatives. There are two in development; features mentioned, like determining if a statement is true or false based on existing fact-checking materials. As he tested it, he realised that the model works surprisingly well with English and Hungarian languages. Sources listed are sometimes incorrect or irrelevant. AI is used for social listening and monitoring on platforms like Facebook. The language model can analyze data to identify disinformation patterns. At the end, it can highlight engaging posts and suspicious language.

When asked about what he thinks will be coming in the future, he mentioned that there are worries about misuse of AI in governmental and political topics. He took note of the fact that there are grants which support many projects that are somehow related to AI development (example of the European Media and Information Fund).

As for my interview with the European Fact-Checking Standards Network (EFCSN) *Elections24Check's* project manager, my key takeaways are listed in the following section. Firstly, I asked about their standards and the relations they might have in the field. He highlighted that originally the IFCN (International Fact-Checking Network) was established and then came the need for a European network, too. These networks involve many of the fact-checking organizations after they obtain the badge they provide. For that, the organizations have to fulfill the code of standards and principles they have. For example, Lakmusz is part of both the European and International Network. The most important values can be found on each page and either principles or standards and are available online. Some of them: A commitment to Non-partisanship and Fairness, A commitment to Standards and Transparency of Sources, and three more on the IFCN page. As for EFCSN, a few of them are: Transparency, Ethical Standards, like Honest corrections, etc.

Some main challenges arose with the discussion of the workflow of the organisation. One of them was the problem that sometimes a very quick act is needed for example during elections in order to verify a part of

information. With some media, especially AI generated audio, it is incredibly hard to judge in a short time. This can cause quite a pressure. The project *Elections24Check* on the other hand was named successful. Attention is also on the provided funds for incorporating LLMs into the fact-checking workflows and field, said by the interviewee. He mentioned that the changing political scene and the future is concerning in this sense. A general emphasis is on transparency and the education of users about verification processes.

Adding the third perspective of this group: it is an interview with a journalist (previously at Átlátszó), who is currently a teacher and a professional in data-visualisation. The main focus of the discussion was to gather more insights on the Hungarian status of journalism, plus another aspect that is data visualisation. For this, the importance of knowing your field and how to use the methods well was given attention to. As someone who visualises data, you have a role that communicates information to people. As for your best knowledge, you have to make sure that what you create is credible. Furthermore, a few projects were mentioned that actually really inspired me, like “Álhírvadász” (Fake news hunter). This project was part of Átlátszó, and was open for anyone through the internet. It contained a bunch of different content which was sort of in the format of a test or quiz. The user could see how much they can notice signs of unreliable content, and when successfully completed, they were rewarded with a badge. If they could not, they were encouraged to work on it and motivated to continue. This was a very important part of the whole. Its limitations now could be that it is much more complicated to judge than it was back then when the project was running. As a continuation of the educational side, going around in the rural areas especially, doing workshops, lectures, spreading awareness about the topic for children and their teachers was another approach. This one requires different resources, but probably a combination is also possible.

Some other emerged concerns were such as scalability of fact-checking and public reliance on easily accessible tools over critical thinking. One of the biggest challenges is getting access to the information as a journalist, also the credibility of the provided documents are questionable sometimes. To illustrate the key moments from these interviews, the main topics can be seen below.

3. Description of future development area for masterwork

In this chapter, I aim to emphasize the findings and elements of the thesis that contribute to the masterwork concept or hold potential for inclusion and further development in future work.

Firstly, questions might arise about the user group who I would like to focus on. During the literature and research parts, it was discussed how different the focus could be. There are many, but the main outstanding ones for me are the **younger generation (children)**, and the elderly. Right now, it seems like it could be a good idea to bring this topic to the schools and educate the young ones while raising **awareness**. On the other hand, in a few cases it was highlighted that there are fact-checking tools developed for journalists and researchers. In connection to this, a contribution to the same group or for other specialists would be another considerable option.

Continuing this, a concrete persona is yet to be built. In the meantime, since I had a group identified only as “general users” without further elaboration, I had to explain what that would actually mean, like what expertise, age and so on. By this time I realised that they won’t be my core

user group and that there is no further need to make a more concrete version of them.

It is another great detail of the whole, what the interface (the main thing the user interacts with) could be of this probable future tool, what kind of technology (AR, VR, etc.) might be useful? Currently I feel like the interface could be a **hybrid** solution, that is easy to use in the classroom, at home, online.

As the first part of my thesis showed already, whatever I end up designing, transparency definitely should be a key principle.

A discovered principle can be the lack of time nowadays that some can spend on actually checking the reliability of the sources.

As for the area of development, the question of deciding whether to go with a designed environment or rather a tool is another topic that has to be addressed.

We discussed the key issues of AI development and considerations. At this point of view, the core idea is circling around mainly LLMs, generative AI. As in an interview it was highlighted, not in all cases and situations you have to face such effects of LLMs. Sometimes hallucinations can be mitigated simply by teaching the solution to state “no connection” therefore not generating irrelevant and fake information when the model does not find relevant connections. So to decide whether AI can amplify the solution, the situation shall be named and also the goal.

Some experimental digital prototypes are likely to be produced as a testing of my ideas. For instance, a recreation of my workshop taks and further developed education tool could be designed to test the basic judgement of people from the educational system to help them teach the certain aspects of this problem.

Many users lack familiarity with fact-checking tools and strategies. Developing simple, accessible resources tailored to diverse demographics can promote critical thinking and verification skills.

AI systems often lack transparency, leading to mistrust. Solutions should integrate knowledge, truth, and transparency principles, clearly explaining AI processes and errors.

Verification mechanisms are often inadequate. Real-time features like reverse image searches can improve content validation on digital platforms.

Excessive skepticism fosters mistrust, while blind trust enables misinformation. Balanced systems, such as visual trust indicators, can guide users without fostering paranoia.

A simple illustration describes the 3 main ways I could take, with the first one (Education and Awareness) being the most important one as of now for me.

Conclusion

This thesis mainly states the need of developing innovative solutions, collaborating with stakeholders, and leveraging educational, technological, or policy-driven approaches to contribute to misinformation mitigation.

It has explored the intricate dynamics of misinformation and fact-checking in the digital age, identifying key challenges, evaluating current practices, and proposing potential pathways for improvement. By integrating insights from literature, surveys, workshops, and interviews with stakeholders such as designers, developers, AI specialists, and fact-checking professionals, this work has painted a comprehensive picture of the field's current state and its future needs.

The findings underscore the urgent need to address misinformation through multifaceted approaches. Public education remains critical, as many users lack awareness of fact-checking strategies and tools. Simultaneously,

intuitive and transparent technological solutions are necessary to build trust and improve the accessibility of verification processes. The role of AI, while promising, is complex and dual-edged—capable of amplifying both misinformation and its mitigation. Ensuring transparency, ethical implementation, and robust safeguards for AI systems is paramount.

The paper also highlights some of the psychological and societal effects of misinformation, from fostering over-skepticism to contributing to decision fatigue. Addressing these issues requires tools and strategies that balance skepticism with trust, and simplicity with effectiveness. Similarly, diverse user groups—such as children, non-digital natives, and professionals—need targeted solutions that respect their unique challenges and contexts.

By focusing on these critical areas, this research lays a foundation for actionable solutions. It emphasizes the importance of collaboration among designers, developers, educators, policymakers, and fact-checking professionals to create impactful tools and strategies. As this field continues to evolve, the insights and recommendations presented here aim to contribute to fostering a more informed, critical, and empowered society.

Looking forward, this thesis serves as a stepping stone for my masterwork, which will delve deeper into designing solutions that align with these identified needs. Through continued research, testing, and engagement with diverse stakeholders, the aim is to create innovative interventions that address the pressing challenges of misinformation in meaningful and more sustainable ways.

Bibliography

Aïmeur Esma., Amri Sabrine., Brassard Gilles. 2023. “Fake news, disinformation and misinformation in social media: a review”. Social Network Analysis and Mining (2023) 13:30 https://doi.org/10.1007/s13278-023-01028-5 Last accessed: November 23, 2024.

Bender, Emily M. 2024. “As OpenAI and Meta introduce LLM-driven searchbots (on top of Bing and Google already splashing "AI" all over search), I'd like to once again remind people that neither LLMs nor chatbots are good technology for information access.” LinkedIn, November, 2024. https://www.linkedin.com/posts/ebender_situating-search-proceedings-of-the-2022-activity-7259043458726731776-9INW?utm_source=share&utm_medium=member_desktop (Last accessed: December 1, 2024.)

Bender, Emily M., McMillan-Major, Angelina., Gebru, Timnit., Shmitchell, Shmargaret. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” Article. In Conference on Fairness, Accountability, and Transparency (FAccT ’21), March 3–10, 2021, Virtual Event, Canada. ACM, New York, NY, USA, 14 pages. https://dl.acm.org/doi/10.1145/3442188.3445922 (Last accessed: November 24, 2024.)

Bordia Prashant., DiFonzo Nicholas. 2002. “When social psychology became less social: Prasad and the history of rumor research“ Article. Asian Journal of Social Psychology. 5: 49–61 https://onlinelibrary.wiley.com/doi/10.1111/1467-839X.00093 (Last accessed: November 23, 2024.)

Bubeck et al. 2023. “Sparks of Artificial General Intelligence: Early experiments with GPT-4” Cornell University, arXiv, Computer Science > Computation and Language. https://doi.org/10.48550/arXiv.2303.12712 (Last accessed: November 26, 2024.)

Collins, B., Hoang, D.T., Nguyen, N.T., Hwang, D. (2020). “Fake News Types and Detection Models on Social Media A State-of-the-Art Survey”. In: Sitek, P., Pietranik, M., Krótkiewicz, M., Srinilta, C. (eds) Intelligent Information and Database Systems. ACIIDS 2020. Communications in

Computer and Information Science, vol 1178. Springer, Singapore. https://doi.org/10.1007/978-981-15-3380-8_49 (Last accessed: November 26, 2024.)

Demartini, G., Mizzaro S., and Spina, D. 2020. “Human-in-the-loop Artificial Intelligence for Fighting Online Misinformation: Challenges and Opportunities”. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering. September 2020, Vol. 43 No. 3. IEEE Computer Society. https://www.researchgate.net/publication/345264315_Human-in-the-loop_Artificial_Intelligence_for_Fighting_Online_Misinformation_Challenges_and_Opportunities (Last accessed: December 1, 2024.)

Euronews. 2019. “How can Europe tackle fake news in the digital age?” News, World. Effective 10/01/2019. https://www.euronews.com/2019/01/09/how-can-europe-tackle-fake-news-in-the-digital-age (Last accessed: November 24, 2024.)

Fard, A. E., and Lingeswaran, S. 2020. “Misinformation Battle Revisited: Counter Strategies from Clinics to Artificial Intelligence.” The Web Conference 2020 - Companion of the World Wide Web Conference, WWW 2020. Association for Computing Machinery (ACM). (pp. 510-519). https://doi.org/10.1145/3366424.3384373 (Last accessed: November 23, 2024.)

Gehman, Samuel., Gururangan, Suchin., Sap, Maarten., Choi, Yejin., and Smith, Noah A. 2020. “RealToxicityPrompts: Evaluating Neural Toxic Degeneration in Language Models.” In Findings of the Association for Computational Linguistics: EMNLP 2020. Association for Computational Linguistics, Online, 3356–3369. https://doi.org/10.18653/v1/2020.findings-emnlp.301 (Last accessed: November 24, 2024.)

Github. 2020. “GPT-3 Model Card”. OpenAI, GPT-3. Last updated: September 2020 https://github.com/openai/gpt-3/blob/master/model-card.md (Last accessed: November 24, 2024.)

Hacker, P., and Passoth, JH. 2022. “Varieties of AI Explanations Under the Law. From the GDPR to the AIA, and Beyond.” Holzinger, A., Goebel, R., Fong, R., Moon, T., Müller, KR., Samek, W. (eds) xxAI - Beyond

Explainable AI. xxAI 2020. Lecture Notes in Computer Science(), vol 13200. Springer, Cham. https://doi.org/10.1007/978-3-031-04083-2_17 (Last accessed: December 1, 2024.)

Ienca M., Vayena E. 2018. “Cambridge analytica and online manipulation” Scientific American, New York. Vol 30. https://www.scientificamerican.com/blog/observations/cambridge-analytica-and-online-manipulation/ (Last accessed: November 26, 2024.)

Ienca, M. 2023. “On Artificial Intelligence and Manipulation.” Article. Topoi 42, 833–842 (2023). https://doi.org/10.1007/s11245-023-09940-3 (Last accessed: November 26, 2024.)

Instagram. 2019. “Combatting Misinformation on Instagram” About. December 16, 2019. https://about.instagram.com/blog/announcements/combatting-misinformation-on-instagram (Last accessed: December 1, 2024.)

Islam, M.R., Liu, S., Wang, X. et al. 2020. “Deep learning for misinformation detection on online social networks: a survey and new perspectives.” Soc. Netw. Anal. Min. 10, 82 (2020). https://doi.org/10.1007/s13278-020-00696-x (Last accessed: November 26, 2024.)

Kim, Jooyeon., Tabibian, Behzad., Oh, Alice. et al. 2018. “Leveraging the Crowd to Detect and Reduce the Spread of Fake News and Misinformation” Article. Conference WSDM: Web Search and Data Mining. https://doi.org/10.1145/3159652.3159734 (Last accessed: November 26, 2024.)

Meta. 2024. “Meta policies and safeguards for elections around the world” about.meta.com, Actions. https://about.meta.com/actions/preparing-for-elections-with-meta/?refsrc=about.meta.com%2Fregulations%2F#fighting-misinformation (Last accessed: November 26, 2024.)

Micallef, Nicholas., He, Bing., Kumar, Srijan., Ahamad, Mustaque., Memon, Nasir. 2020. “The Role of the Crowd in Countering Misinformation: A Case Study of the COVID-19 Infodemic”. Cornell University, arXiv, Computer Science > Social and Information Networks.

https://doi.org/10.48550/arXiv.2011.05773 (Last accessed: November 26, 2024.)

Monteith S., Glenn T., Geddes J. R., Whybrow P. C., Achtyes E. and Bauer M. 2024. “Artificial intelligence and increasing misinformation” The British Journal of Psychiatry (2024) 224, 33–35. https://doi.org/10.1192/bjp.2023.136 Last accessed: November 23, 2024.

Mündges, Stephan. 2024. “Everyone interested in research on misinfo interventions should read this”. LinkedIn, November, 2024. https://www.linkedin.com/posts/stephan-mundges_despite-community-notes-most-content-reviewed-activity-7262170171954360322-xiLe?utm_source=share&utm_medium=member_desktop (Link of the post) https://science.feedback.org/despite-community-notes-most-content-reviewed-eu-fact-checkers-goes-unaddressed-x-twitter/ (Link of the source mentioned) (Both Last accessed: December 1, 2024.)

National Academies. 2021. “As Surgeon General Urges ‘Whole-of-Society’ Effort to Fight Health Misinformation, the Work of the National Academies Helps Foster an Evidence-Based Information Environment”. Feature Story. Shared on July 15, 2021. https://www.nationalacademies.org/news/2021/07/as-surgeon-general-urges-whole-of-society-effort-to-fight-health-misinformation-the-work-of-the-national-academies-helps-foster-an-evidence-based-information-environment Last accessed: November 23, 2024.

Pennycook, G., Rand, D.G. 2019. “Fighting misinformation on social media using crowdsourced judgments of news source quality” Proc. Natl. Acad. Sci. U.S.A. 116 (7) 2521-2526, https://doi.org/10.1073/pnas.1806781116 (Last accessed: November 26, 2024.)

Pew Research Center. 2024. “Americans’ Views of 2024 Election News.” Report. Published October 10, 2024. https://www.pewresearch.org/journalism/2024/10/10/americans-views-of-2024-election-news/ Last accessed: November 23, 2024.

Reporters' Lab. 2023. “Misinformation spreads, but fact-checking has leveled off” reporterslab.org, Fact-Checking News. Published in 2023.

https://reporterslab.org/tag/fact-checking-database/ (Last accessed: November 26, 2024.)

Ruths, Derek. 2019. “The misinformation machine.” Science 363, 6425 (2019), 348. https://www.science.org/doi/10.1126/science.aaw1315 (Last accessed: November 24, 2024.)

Schemmer., Hemmer., Kühl. et al. 2022. “Should I Follow AI-based Advice? Measuring Appropriate Reliance in Human-AI Decision-Making” Cornell University, arXiv, Computer Science > Human-Computer Interaction. https://doi.org/10.48550/arXiv.2204.06916 (Last accessed: November 26, 2024.)

Scheurer, Jérémy., Balesni, Mikita., Hobbhahn, Marius. 2024. “Large Language Models can Strategically Deceive their Users when Put Under Pressure” Article. Published at the LLM Agents workshop at ICLR 2024. https://doi.org/10.48550/arXiv.2311.07590 (Last accessed: November 24, 2024.)

Statista. 2021. “Share of people who have ever accidentally shared fake news or information on social media in the United States as of December 2020” Statista.com, Media, News. Published in 2021. https://www.statista.com/statistics/657111/fake-news-sharing-online/ (Last accessed: November 26, 2024.)

Statista. 2023. “Level of confidence in distinguishing between real news from false news among adults in the United States as of April 2023” Statista.com, Media, News. Published in 2023. https://www.statista.com/statistics/657090/fake-news-recogition-confidence/ (Last accessed: November 26, 2024.)

Statistics Canada. 2023. “Concerns with misinformation online, 2023” Released in The Daily, 2023. https://www150.statcan.gc.ca/n1/daily-quotidien/231220/dq231220b-info-eng.htm (Last accessed: November 26, 2024.)

Strümke., Slavkovik., Stachl. 2023. “Against Algorithmic Exploitation of Human Vulnerabilities” Cornell University, arXiv, Computer Science > Artificial Intelligence. https://doi.org/10.48550/arXiv.2301.04993 (Last accessed: November 26, 2024.)

Tchakounté, Franklin., Faissal, Ahmadou., Atemkeng, Marcellin., and Ntyam, Achille. 2020. "A Reliable Weighting Scheme for the Aggregation of Crowd Intelligence to Detect Fake News" Information 11, no. 6: 319. https://doi.org/10.3390/info11060319 (Last accessed: November 26, 2024.)

The Guardian. 2017. “'Disputed by multiple fact-checkers': Facebook rolls out new alert to combat fake news” Technology. March 22, 2017. https://www.theguardian.com/technology/2017/mar/22/facebook-fact-checking-tool-fake-news (Last accessed: December 1, 2024.)

The Guardian. 2017. “'Google to display fact-checking labels to show if news is true or false” Technology. April 7, 2017. https://www.theguardian.com/technology/2017/apr/07/google-to-display-fact-checking-labels-to-show-if-news-is-true-or-false (Last accessed: December 1, 2024.)

Tschiatschek, Sebastian., Singla, Adish., Rodriguez, Manuel Gomez., Merchant, Arpit., and Krause, Andreas. 2018. “Fake News Detection in Social Networks via Crowd Signals”. In Companion Proceedings of the The Web Conference 2018 (WWW '18). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 517–524. https://doi.org/10.1145/3184558.3188722 (Last accessed: November 26, 2024.)

Williamson, Steven M., Prybutok, Victor. 2024. “The Era of Artificial Intelligence Deception: Unraveling the Complexities of False Realities and Emerging Threats of Misinformation.” Article. Information 2024, 15, 299. https://doi.org/10.3390/info15060299 (Last accessed: November 24, 2024.)

Wired. 2019. “Instagram Now Fact-Checks, but Who Will Do the Checking?”. Story. August 16, 2019. https://www.wired.com/story/instagram-fact-checks-who-will-do-checking/ (Last accessed: December 1, 2024.)

World Health Organization (WHO). 2020. “Novel Coronavirus (2019-nCoV) Situation Report - 13.” Data as reported by 2 February 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf (Last accessed: November 24, 2024.)

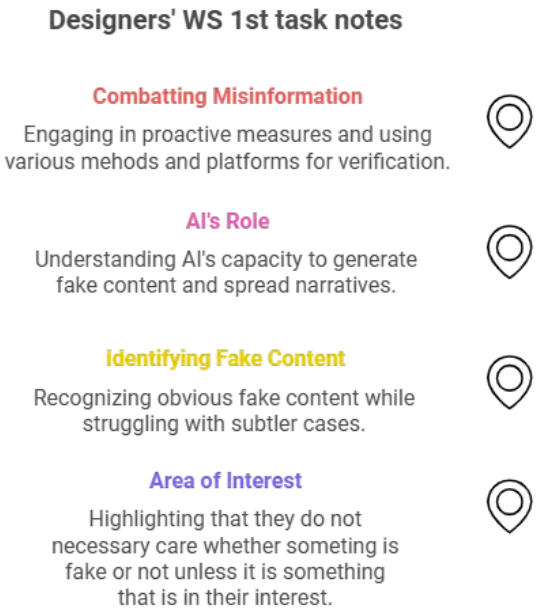
Zhang, Liao., and Bellamy, K. E. 2020. “Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making.”

In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). Association for Computing Machinery, 295–305. https://doi.org/10.1145/3351095.3372852 (Last accessed: November 26, 2024.)

APPENDIX 1

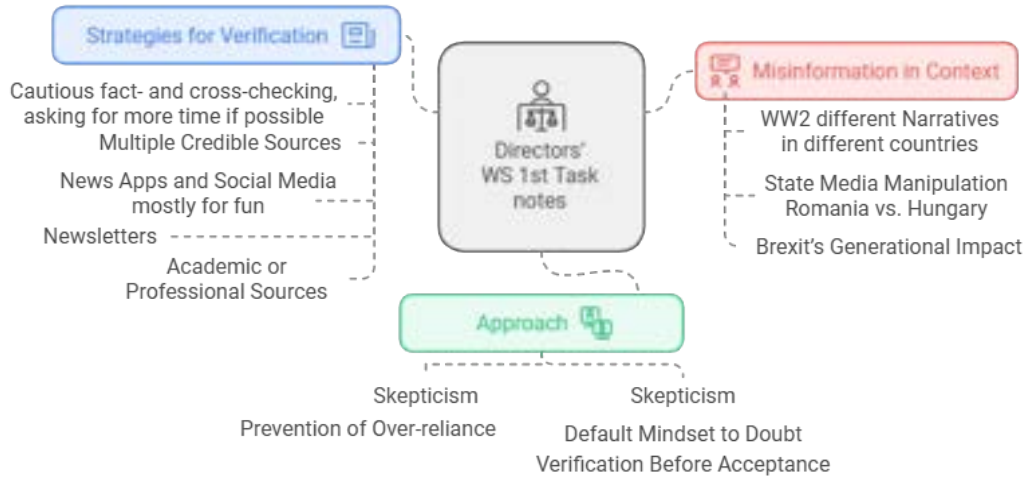
Workshop Materials

Part of ”The Era of Misinformation Workshops” at SaaS company, 22/10/2024, 08/11/2024.



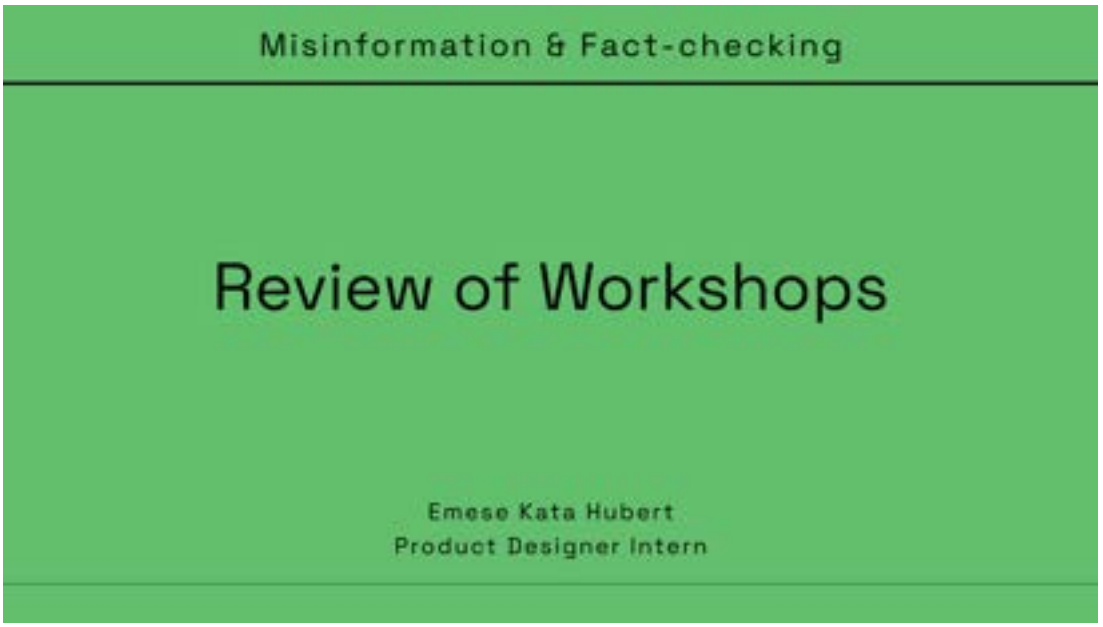
Source: based on Participants' comments from 1st task during the Designer Workshop.

To access the full workshop planning and work board: https://www.figma.com/board/5ZDhwWf9NaD9LSTYtwnV9p/participatory-design-class%2C-part-of-thesis-research-board---Emese-Hubert?node-id=0-1&t=IghPCdbFTxFyunph-1



Source: based on Participants' comments from 1st task during the Manager/Director Workshop.

Selected Workshop Review slides:



Questions

- Is there is a difference between the two groups, and if so, what?
- Will any patterns occur?
- **How conscious/unconscious are they while receiving information?**
- **What are their previous experiences on the topic?**
- Will a group perform better at identifying fake content?

Participants

Designers in SaaS 1	Managers/Directors in SaaS 2
• 1st WS: 5 designers and 1 PM • (Note: Designer WS as in Designer dominant in numbers.)	• 2nd WS: 1 designer and 3 managers/directors • (Note: Manager WS as in Manager dominant in numbers.)

Designers/IC group and Manager group + (!) competitive comparison

Participants

Designers in SaaS 1

Managers/Directors in SaaS 2

High-level overview of ToDos and Scripts

Agenda 1 (Designers, 22/10/2024)

1	Welcome! (10mins) 10:00-10:10
2	1st Task - How do you usually get information in the online space? Story with encountering misinformation. (20mins) 10:10-10:30
3	2nd Task - First part of Identifying Fake/True test. Evaluating all contents alone. (15mins) 10:30-10:45
4	2nd Task - Second part of Identifying Fake/True test. Discussion, results. Pie chart visualization. (15mins) 10:45-11:00
5	3rd Task - Content Generation Use Cases (3 subtasks) (25mins) 11:00-11:25
6	Ending, Summary, Feedback (5mins) 11:25-11:30

Agenda 2 (Managers, 08/11/2024)

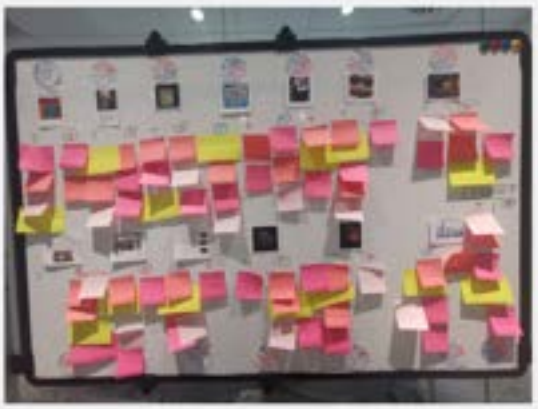
1	Waiting and Welcoming (15mins) 9:00-9:15
2	1st Task - How do you usually get information in the online space? Story with encountering misinformation. (15mins) 9:15-9:30
3	2nd Task - First part of Identifying Fake/True test. Evaluating all contents alone. (20mins) 9:30-9:50
4	2nd Task - Second part of Identifying Fake/True test. Discussion, results. Pie chart visualization. (15-20mins) 9:50-10:10
5	3rd Task - Fact-checking crazy ideas (15mins) 10:10-10:25
6	Ending, Summary, Feedback (5mins) 10:25-10:30

Task 1 (ice-breaker questions)

Experiences regarding Misinfo

Task 2 - Identifying Fake/True test, discussion

Critical-Thinking, evaluation



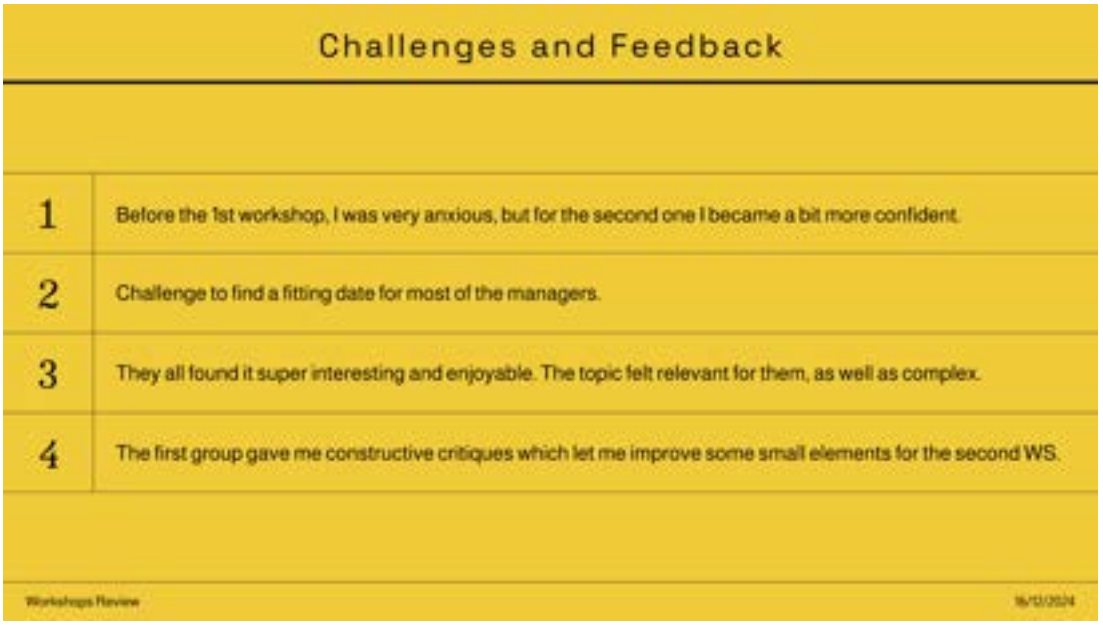
Task 2 - Identifying Fake/True test, discussion

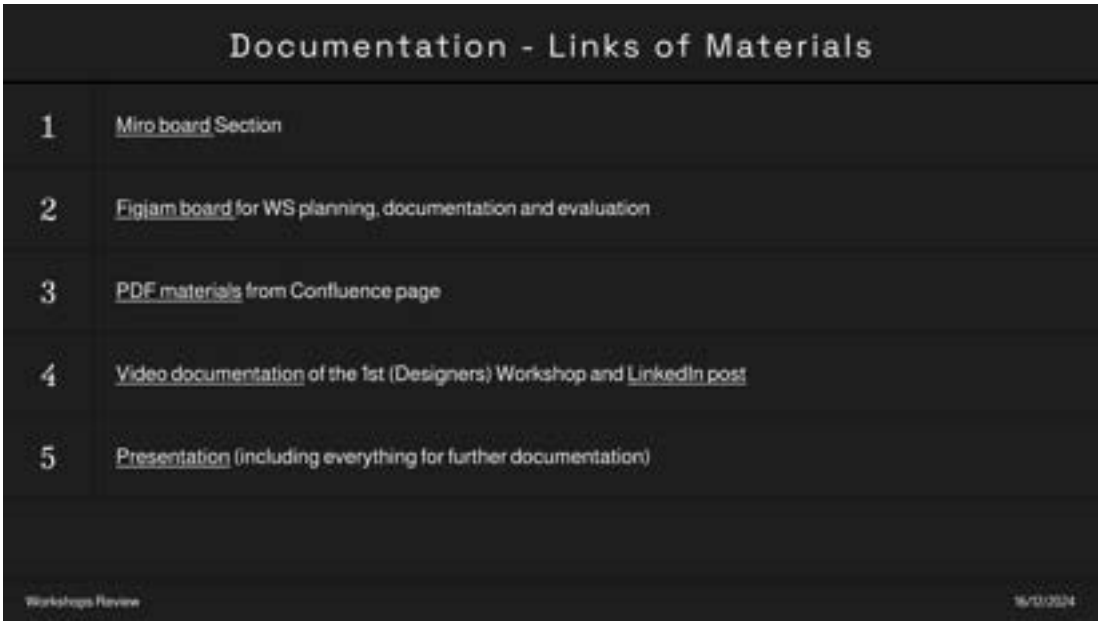
Designer Selected Highlights

Manager Selected Highlights

Quantitative Summary

1	**Managers** seemed more confident and accurate in identifying fake content but struggled significantly with true content, perhaps due to over-skepticism or **distrust**.
2	**Designers**, while less adept at identifying fake content, **excelled** in validating **true** information, showing a more discerning approach toward authentic content.
3	**Designers**' higher rate of incorrect classifications for fake content might suggest **over-trust** or susceptibility to misinformation's persuasive elements.





Link to presentation and more workshop resources:
https://www.figma.com/deck/O5rOnZAUeZM5kMJPYHejSe/Misinformation-Workshop-results-16%2F12%2F2024?node-id=1-374&t=06ORT7D6Jca4znLK-1

APPENDIX 2

Survey Materials

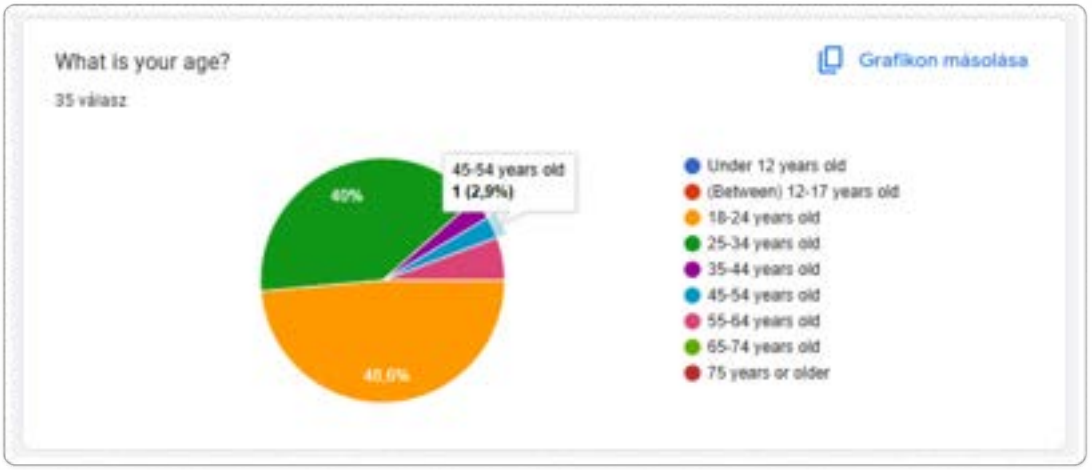
Goals: to have a more quantitative overview and findings of the topic, from many online.

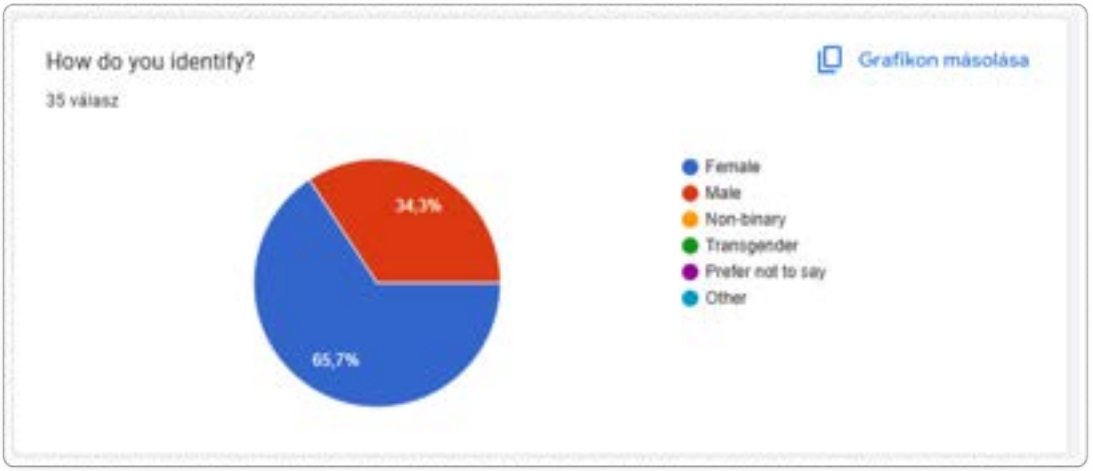
Number of completions: 36

Date: October 2024 - November 2024 (approx. 1 month open survey time)

Description of participants: General, everyday users

Demography:





Survey content:

The Era of Misinformation - Research

Hello There!✌

My name is Emese Hubert and I am doing research for my Master's Thesis at Moholy-Nagy University of Art and Design. I would like to ask a few questions related to misinformation and what habits you might have to collect information in the online space.

Your answers would provide valuable insights for my work. Your response will remain completely anonymous and will only be used as a reference for my study.

I really appreciate the time and effort, huge thank you for participating in the survey!

- Emese

How do you identify?*

- Female
- Male
- Non-binary
- Transgender
- Prefer not to say
- Other

What is your age?*

- Under 12 years old
- (Between) 12-17 years old
- 18-24 years old
- 25-34 years old
- 35-44 years old
- 45-54 years old
- 55-64 years old
- 65-74 years old
- 75 years or older

What is the highest degree or level of school you have completed? If currently enrolled, the highest degree received.*

- Elementary school
- High School
- Trade/technical/vocational training
- Bachelor's degree
- Master's degree
- Doctorate degree (PhD)

Can you share a time you encountered misinformation online? How did you realize it was false?* (long written answer)

What are your main sources for news and information? How do you determine if they are reliable? You can think of your daily habits.* (long written answer)

Are you familiar with fact-checking tools or strategies to spot misinformation? Do you use them?* (long written answer)

How truthful are these Social Media platforms, in your experience and opinion?*
Not at all - Sometimes - More often than not - Most of the times - Always true - I do not use this platform

- Facebook
- Instagram
- TikTok
- Youtube
- Telegram
- Snapchat
- Whatsapp
- Pinterest
- LinkedIn
- Reddit
- Discord

- X (Twitter)

How often do you encounter news or information online that you question the accuracy of?*

- Never
- Rarely
- Sometimes
- Often
- Always

In a few words, how confident are you in distinguishing between reliable and unreliable information online?* (short written answer)

Optional: Does misinformation affect your personal or professional life? If so, how? (long written answer)

Optional: Have you used misinformation-reporting features? Were they effective? (long written answer)
Optional: What information sources are the least accurate in the online space, in your opinion? (long written answer)

Selected extracts from results synthesis:

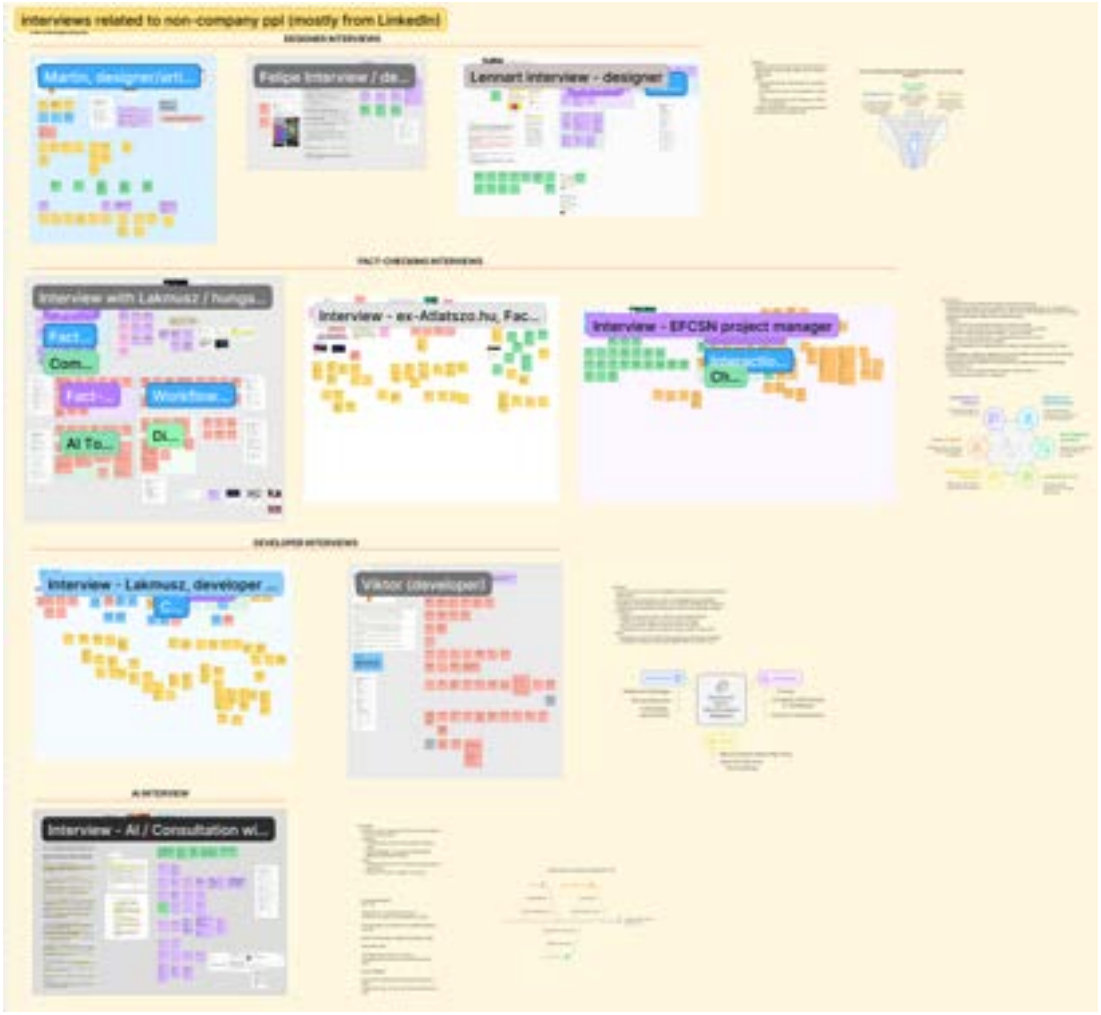
● Presentation of most trustful platform, according to votings by users:

Source and full quality (survey research board): https://www.figma.com/board/5ZDhwWf9NaD9LSTYtwnV9p/participatory-design-class%2C-part-of-thesis-research-board---Emese-Hubert?node-id=246-1262&t=hdQdSPoq17O63bzf-4

APPENDIX 3

Interview Materials

Sources, **notes**, and **synthetisation** of the interviews can be found here: https://www.figma.com/board/5gjCZnPAYhcF6oPGjJWpbG/thesis-topic-%26-writing?node-id=225-1179&t=uRbWsGaYArLfkgaV-4



Additional sources:

Recording (with the consent) of Artist/Designer (Martin): https://drive.google.com/drive/folders/1ccs_zW2S-fP6s3OD5ayrws9uj1tUhdNY?usp=sharing

Recording (with the consent) of developer (Viktor): https://drive.google.com/file/d/1R7NwRYJFKW1wfEXeuVNispcBs8ykEXqE/view?usp=sharing

The rest of the interviews were not recorded due to technical difficulties.

The summary of each category:

1. **Designers**

Awareness: General concerns about misinformation and ethical design; limited depth in personal or professional experiences.

Ideas:

Educational tools like "show and tell" for awareness-building.

A "cookbook" approach to fact-checking for intuitive use.

Small MVP solutions for trust-building (e.g., verified account checkmarks).

Concerns: Misinformation's impact on family and limited privacy/confidentiality in freelance work.

2. **Fact-Checkers**

Awareness: Extensive familiarity with misinformation and its mechanics.

Checking IFCN and EFCSN code of standards. How the EFCSN network works, The most difficult to verify is the AI generated Audio. Funds for using LLM on top of fact-checking. Concerns about changing political scene and future. Higlighted the elections 24 project.

Strategies:

Heavy reliance on established fact-checking tools and methods.

Emphasis on transparency and educating users about verification processes.

Álhírvadász online project (badge, feedback to boost motivation)

going around the education system spreading awareness about the topic

workshops for teachers in the countryside

Concerns: Scalability of fact-checking and public reliance on easily accessible tools over critical thinking.

one of the biggest challenges is getting access to the information as a journalist, also the credibility of the documents are questionable sometimes (hungarian)

As a journalist and data visualisator it is important to stay credible, to know the methodologies

Proposed Solutions:

Integration of user-friendly verification features in popular platforms.

Promoting critical thinking as a societal skill.

3. **Developers**

Awareness: Focus on the technical challenges and ethical dilemmas of misinformation in digital spaces.

to know what AI can do for you, like when it is not generative it is more reliable.

the software they are doing for Lakmusz: helps to gather and analyse documents, etc. useful features for researchers and journalists. The tool connects similar narratives

Challenges:

Balancing user privacy with data needs for misinformation detection.

Adapting to rapidly evolving AI and misinformation techniques.

light version and full. different features for mobile and desktop.

Challenge in how to categorise: instead of labeling, quality scoring content.

Ideas:

Development of tools for misinformation reporting and real-time fact-checking.

Leveraging AI to flag potentially false information with user context in mind.

4. **AI Specialist**

Awareness: Deep understanding of AI's role in generating and combating misinformation.

Concerns:

Potential misuse of advanced AI models by malicious actors.

Ethical challenges in AI regulation and deployment.

Misuse of definitions, trending

Ideas:

Building transparency into AI tools (e.g., showing decision-making logic).

Implementing stricter regulations on AI models.

PLAGIARISM DECLARATION

Hereby, I, LUCZ KATA ZSUZSA

student of INTERNATIONAL RELATIONS **programme at MOME,**

in full knowledge of my liability, declare and certify by my signature that this thesis is my own original work. In the text here, all printed and electronic sources was clearly referenced in accordance with international requirements of copyright.

I acknowledge that the following is considered plagiarism:

- to cite verbatim without using quotation marks and indication of citation

- to paraphrase the source material without citing the source

- to indicate another author's published ideas as my own.

Budapest, 2025.02.20.

Lucz Kata Zsuzsa

signed by the student