RESEARCH PUBLICATION

CHAPTER 9

AI-powered Fact-checking: Strategic Framing of AI Use for Information Verification

Lasha Kavtaradze Department of Communication, Kristiania University College; Department of Information Science and Media Studies at the University of Bergen
Bente Kalsnes Department of Communication, Kristiania University College

Abstract: As the scale of mis/disinformation grows across media and social media platforms, public and professional discussion about the use of artificial intelligence (AI) for information verification is becoming increasingly common. This chapter explores how six companies working on AI-powered services strategically frame mis/disinformation issues and what sort of moral judgments they use when making diagnostic inferences to find solutions for "information disorder". Informed by Entman's framing theory, this study qualitatively analyzes textual data from the websites of AI-powered services for information verification. We find that the companies studied promote services that we identify here as: automated fact-checking, automated credibility assessment, and automated authenticity assessment. Hence, this chapter focuses on the strategic framing of the mis/disinformation problem, along with the solutions promoted by AI-powered services, laying the groundwork for further explorations of how the offered technologies might tackle the problem of spreading fake news.

Keywords: strategic communication, fact-checking, artificial intelligence, framing, information verification

Citation: Kavtaradze, L. & Kalsnes, B. (2024). AI-powered fact-checking: Strategic framing of AI use for information verification. In S. Alghasi, E. C. Vanvik, J. Barland & J. Falkheimer (Eds.), *Strategic communication – contemporary perspectives* (Ch. 9, pp. 177–198). Cappelen Damm Akademisk.
https://doi.org/10.23865/noasp.208.ch9

Introduction

Technological development and accelerated digitalization contribute to the malicious process of spreading mis/disinformation "farther, faster, deeper, and more broadly" (Di Pietro et al., 2021; Vosoughi et al., 2018). As the scale of mis/disinformation grows across media and social media platforms, public and professional discussion about the use of Artificial Intelligence (AI) for information verification is becoming increasingly common (Graves, 2018; Mishra & Setty, 2019; Moreland & Doerrfeld, 2016). The media industry, academia, tech, and civil society are collaborating to come up with AI solutions to deal with fake news defined as false information "packaged to look like real news to deceive readers either for financial or ideological gain" (Tandoc et al., 2019, p. 674). Many hope that remedies for this "informational disorder" (Wardle, 2018) will be found within the technological realms. The current trend of injecting AI into the process of information verification within the media industry is also comfortably situated within such technology-oriented logic. AI-powered services for information verification have been called the "holy grail of fact-checking" (Hassan et al., 2015), and this type of framing is a type of strategic communication that can influence how people view AI-based fact-checking.

Despite rapid developments in the field, AI-powered services for fact-checking are still in their embryonic phase (Jimenez & Li, 2018). Yet, they still need to prove the viability of their products and motivate how AI may offer better solutions for media ecosystems distorted by "fake news". In this chapter, we observe how six leading companies, all of which work on tackling mis/disinformation, strategically frame their AI-powered services on their websites. Companies engage in various kinds of strategic communication to convince users and potential customers about the credibility and efficiency of their solutions. Due to the novelty of such services, it is still too early to talk decisively about their short-term effects or long-lasting implications on information ecosystems or societies. Instead, with the help of framing theory (Entman, 1993), we examine their intentions and strategic positioning within the context of information verification. Thus, this chapter explores how organizations working on AI-powered services strategically frame issues related to mis/disinformation and what sort of ethical judgments they use while making diagnostic inferences to find solutions for information disorder. Accordingly, we answer the following research questions:

RQ1: How do AI-powered fact-checking services frame the problem of information disorder?
RQ2: Which diagnoses and moral judgments do the AI-powered fact-checking services express on their websites?
RQ3: What solutions do the AI-powered fact-checking services recommend for dealing with mis/disinformation?

To address these questions, we selected six leading companies already associated with the use of AI for information verification, designing AI-based systems, and attracting attention from the fact-checking community (Burgess, 2016; Fray, 2016). We will analyze textual data obtained from their websites that demonstrates how they define the problem of mis/disinformation, evaluate causal agents, make judgments, and suggest remedies. First, we discuss the premise of using AI for information verification in media production, which is followed by a short theoretical discussion about Entman's framing theory and a description of the collected data. Finally, the chapter discusses the results of our analysis and the broader implications of strategic framing of AI solutions for information verification.

Literature review

The use of AI in media organizations has attracted considerable scientific and industry attention lately (Diakopoulos, 2019; Whittaker, 2019). The rise of AI technologies has already significantly affected different stages of media work, including "newsgathering, production, and distribution" (Marconi, 2020, p. 22). As Chan-Olmsted (2019) notes, algorithmic technologies have been applied to at least eight main areas in news media: audience content recommendations/discovery, audience engagement, augmented audience experience, message optimization, content management, content creation, audience insights, and operational automation. Checking the veracity of facts is part of the content creation and management process in newsrooms. In this regard, AI is used for tasks like "information search, retrieval, classification, and treatment" (Torrijos, 2021). Media and tech industries are collaborating to create algorithms that would spare human fact-checkers or reporters from the burden of monitoring media spaces for breaking news (Liu et al., 2016), cleaning and analyzing data (Stray, 2019), writing the headlines (van Dalen, 2012) and news stories (LeCompte, 2020), and distributing credible information to the audiences. Accordingly, terms like "automated journalism", "robotic journalism", and "algorithmic journalism" are becoming ubiquitous (Torrijos, 2021).

Moreover, one of the cornerstones of journalistic practice – fact-checking – has also turned into a separate genre of media production (Juneström, 2020) and is attracting scientific and industry attention in the context of using AI. The utilization of AI for corroborating or refuting the veracity of information, be it textual claims or audio-visual artifacts, has thus far mostly been studied from the computer and data science perspective (Zeng et al., 2021). The necessity of finding technological solutions for information verification has become especially acute after recognizing the dangerous scale of fake news spreading worldwide. Lately, researchers have been focusing on the use of AI technologies in determining the credibility of sources of information (Choudhary et al., 2020) as well as breaking down the logic behind manual fact-checking and the automatization of its various steps. This process was epitomized by the emergence of terms like, automatic detection of fake news, automated fact-checking, or computational fact-checking (Ciampaglia et al., 2015; Lahlou et al., 2019; Zeng et al., 2021). However, finding the "holy grail" (Hassan et al., 2015) for identifying fake news is still in the distant future. Nevertheless, tech and media enterprises are investing time and resources to design and exploit AI-based tools for harvesting, repurposing, and delivering information backed by verified facts.

Before AI-powered solutions of information verification materialize on a large scale, it is worth looking at the reasoning behind putting so much industry and professional effort into a phenomenon that is often described as a solution for mis/disinformation-related societal problems. How do AI-powered organizations frame the problem of fake news? What do they see as the solution to this problem? How do they position themselves in this process? What is the role of AI technologies in attempts to deal with the problem of information disorders? At present, the scientific literature is unable to answer such questions sufficiently. Here, with the help of Entman's framing theory, we examine the strategic communication of the AI-powered services as expressed on their websites. This chapter examines examples where AI is strategically used to curb the fake news problem. Further, we demonstrate how this process is communicated strategically, relying on framing theory as discussed in the next section of this chapter.

Entman's analytical framework

Though framing as an analytical tool is often used to study how certain topics are covered by news media, Hallahan (1999) notes that framing analysis also proves to be useful in analyzing how public actors communicate to wider audiences. Accordingly, public relations, including strategic communication can be examined via this framing lens. AI-powered fact-checking services communicate their mission statements on their websites, like most other companies. This allows us to examine how they frame their purpose as well as solutions to specific problems through information selection and salience – what is called framing. As described by Entman (1993, p. 52), to frame is to "select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described". So, framing can be a way to conduct strategic communication on a company's website.

In recent decades, there has been considerable interest in intangible aspects and assets such as brands, organizational identity, reputation, image, and legitimacy (Falkheimer & Heide, 2014). Here, we understand strategic communication as the targeted and formal communication processes planned and activated as a means for organizations to reach overall goals (Falkheimer & Heide, 2014; Holtzhausen & Zerfass, 2015). Strategic communication has also been defined as "[t]he purposeful use of communication by an organization to fulfill its mission […]. The concept further implies that people will be engaged in deliberate communication practice on behalf of organizations, causes, and social movements" (Hallahan et al., 2007, p. 7). An organization, in the context of this chapter, refers to private companies, public authorities, formal networks, associations, and interest groups currently developing AI-powered tools within the fact-checking industry.

As Hallahan (2008) highlights, even though framing has received disproportionate attention in the media since the 1970s, the concept has the potential to illuminate broader communicative processes, in spheres such as organizational behavior, economy, politics, and sociology. More importantly, he notes that "framing is not merely useful but is essential to public relations" (Hallahan, 1999, p. 229). Elsewhere, Hallahan (2008) offers "strategic framing" as a concept to delineate the purposeful use of communicative techniques to convey a desired interpretation of reality to audiences or to promote certain products by directing audience attention to the desired aspects of reality. Eventually, to quote Hallahan, "the goals of strategic framing are to telegraph meaning and to focus audience attention on particular portions of a message or aspects of a topic to gain favorable response" (2008, p. 1).

As stated by Entman, the idea of framing offers a way to describe the power of a communicating text, and analysis of frames illuminates the precise way in which the influence of human consciousness is exerted by the transfer (or communication) of information from one location – such as speech, utterance, news report, novel – to that consciousness. Thus, Entman's framework is divided into four steps that we will apply to our analysis of the six websites: 1) define the problem – delineating what a causal agent is doing with what costs and benefits, usually defined with common culture values; 2) diagnose causes – identifying the forces causing the problem; 3) making moral judgments – evaluating causal agents and their effects; 4) suggest remedies – offering treatment of the problems and predicting their likely effect.

Data material, methods and the study limitations

We have selected a specific public communication channel, the websites of companies currently developing AI-powered services, to qualitatively study how they frame and diagnose mis/disinformation-related issues. As Holtzhausen and Zerfass (2015) note, since internet technologies and Web 2.0 have become universally accessible, stakeholder-owned media channels like websites, blogs, and social media can be used for maintaining coherent strategic communication. Accordingly, nowadays so-called "participatory websites" (Walther & Jang, 2012) represent one of the primary means for companies to communicate their mission, strategies, or products to a broader audience.

To collect qualitative data on the AI-powered organizations that closely associate themselves with the mis/disinformation ecology, we have purposefully selected six leading companies, both in terms of public recognition, adoption by media companies, and innovative solutions: a UK-based organization, Company 1; a Norwegian startup, Company 2; services from three US-based organizations, Company 3, Company 4 and Company 5; and Company 6, which currently operates in three countries (the US, the UK and India). The data was collected from the companies' six websites between February and May 2022.

The companies differ from each other when it comes to their organizational and institutional characteristics. Some are more business/profit oriented (Company 2, Company 5 and Company 6) aiming to monetize their tools to earn revenues. Other services were established in more academic (Company 3 and Company 4) or non-governmental/charity (Company 1) contexts. Several of the selected organizations are already established within the fact-checking and information production landscape (Company 1, Company 5 and Company 6), while others are still in their embryonic phase, and in the process of developing their products (Company 2 and Company 3). Accordingly, we have included a variety of organizations that allow us to look at different framing strategies for their solutions to mis/disinformation-related issues.

After selecting the organizations, we created a database of texts from their websites. Within the database we have collected data about the type of organization (business/profit-oriented, charity/non-governmental, academic); organizational slogans/mottos, associations with company names, texts from their mission statements and "about us" sections, information about funding, existing and imagined user profile, description of the products and the role of humans in the functioning of the products. Key information on the websites for this study includes organizational mission statements and so-called "about us" sections, as well as descriptions of the products, users, and funding sources, which represent how organizations identify and analyze the problems they are working on and reveal their solutions to the identified problems. Besides, associations with company names, as well as their slogans and mottos allowed us to interpret their moral stance when it comes to operating within the mis/disinformation ecology. In the case of two organizations (Company 2 and Company 3), the information on the website was relatively short, making it difficult to do a proper analysis. Hence, we complemented the website information with information from papers published by representatives from the organizations or their founders and promoted on their websites as a part of a public communication practice (in total two conference proceedings and two academic journal articles).

After populating the database with information from the websites, we thematically analyzed the data with the help of "in-vivo coding" (Manning, 2017). As Saldaña notes, in-vivo codes are parts of the text (often words or phrases) that "seem to stand out as significant or summative of what is being said" (2014, p. 17). As in-vivo codes, we have taken snippets from the texts collected on the AI-powered service websites. The length of in-vivo codes varied depending on the meaning snippets conveyed. To illustrate, a few examples of the codes with the respective frame analysis steps are – "unprecedented amount of falsehoods, hyperboles and half-truths" (problem), "claims made by politicians, public institutions and journalists" (cause), "harm to people's lives, health, finances and to democracy" (impact), "scalable, robust, automated fact-checking" (solution). According to Entman "the text contains frames, which are manifested by the presence or absence of certain keywords, stock phrases, stereotyped images, sources of information, and sentences that provide thematically reinforcing clusters of facts or judgments" (1993, p. 52). This approach has provided us with a theoretical framework to analyze the data, connecting in-vivo codes to the four main steps of the frame analysis: description of the problem, diagnosis, moral judgment, and treatment recommendation. In-vivo codes have been grouped under analytical categories within the respective steps of the framing analysis, as shown in Figure 1.

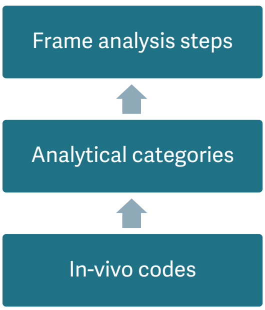

Figure 1 Analysis of framing ai-powered fact-checking solutions.

Accordingly, the frame analysis steps and analytical categories under each of them represent the logical backbone of the study results presented in the next section of this chapter and described in Figure 2.

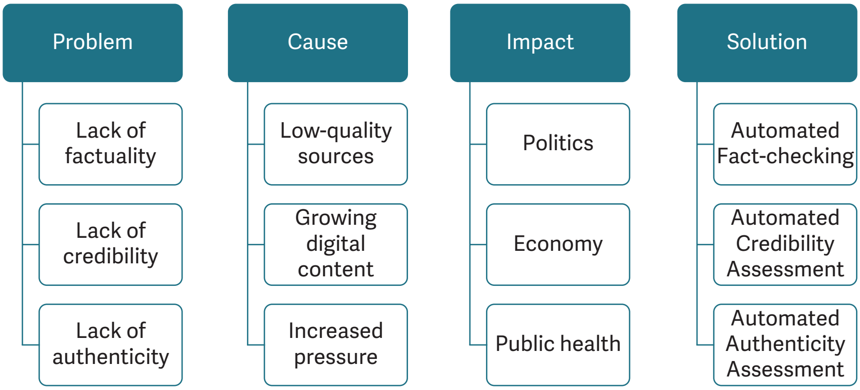

Figure 2 Analytical categories under each frame analysis step.

Before discussing the results, it is worth noting that the companies included in the analysis were strategically selected. The study does not aspire to get representative results, generalizable to all the organizations working within the fact-checking industry that use AI in some form. Our selection of companies prioritized the services that address information verification of texts, hence largely omitting visual and audio media content. Apart from that, the goal of this chapter is not to discuss in depth the root causes, scale, and general effects of or remedies for information disorder. Instead, we focus on how AI-powered services reconstruct the social reality of spreading mis/disinformation and imagined solutions by making their technological offerings salient. Despite these limitations, we believe our research makes a valuable contribution to understanding the mis/disinformation problem from an overlooked perspective. To our best knowledge, there are no other studies in the media industry that have analyzed the strategic framing of AI solutions to information verification.

Frame analysis of AI-powered services

In order to frame the entities that are in charge of the communication, first, we must single out the problem they are trying to fix (Entman, 1993). This is followed by identification of the causal agents of the problem and the impact it has on the social world. Lastly, we examine how the companies present their solutions to the problem. Accordingly, the AI-powered information verification services introduce the problem they are attempting to solve on their websites, before presenting their evaluation of the situation and suggesting remedies for restoring the information ecology.

Identifying the problem of distorted information landscapes

In describing mis/disinformation issues and the harm “information disorder” brings to societies, AI-powered companies identify various sets of problems. They use different terminology to illustrate the situation, such as “fake news”, “bad information”, “misinformation and disinformation”, “falsehoods, hyperboles, and half-truths”, “harmful claims”, and “damaging and misleading information”, Company 6, offering different kinds of services, including AI-assisted fact-checking, intelligence reporting, and countering extremist online content, offers to assist in separating “the facts from the fakery in … the news diet”. The UK-based fact-checking organization Company 1 indicates that they work “to tackle misinformation and disinformation without harming free speech”. Some of the organizations put particular emphasis on the increased risks of spreading malicious information during politically or socially critical moments, such as “elections, public health emergencies, and natural disasters”.

Terminological ambiguities and diversity around mis/disinformation issues have already drawn major scholarly attention (Fallis, 2014; Wardle & Derakhshan, 2017). Despite divergent approaches to classifying misleading or deceiving information, based on our analysis, AI-powered services problematize mis/disinformation issues in terms of three main aspects: lack of factuality, lack of credibility, and lack of authenticity.

Lack of factuality

A core problem under the scrutiny of AI-powered services is the factual quality of information bits, such as claims, multimedia artifacts, news stories, and information campaigns. Selected organizations emphasize the value of factuality and take for granted that facts represent the vital building blocks of quality media experiences. This belief is even manifested in the names of some of the companies (Company 1, Company 2 and Company 5). Facts are the main “instruments” that journalists use to depict reality and reconstruct the truth about the topic they are covering. As Zelizer (2004) puts it, “facts”, along with “truth” and “reality” are journalism’s “god-terms” (Godler & Reich, 2013). Hence, AI-powered services that tackle information verification issues emphasize their devotion to determining the factuality of the information units.

Lack of credibility

Selected organizations also highlight the problem of credibility in media content. Company 5 and Company 6 in particular emphasize the need to evaluate the credibility of online media products. According to Company 5's website, their tool evaluates the information value of articles published in the media to help information consumers determine if they can be trusted and perceived as unbiased. Discussing the issue of media trust, Mrazek notes that "credibility becomes an essential asset of future journalism" (2019, p. 132). In the same fashion, the company 5 highlights that "the unbiased movement" they are aspiring to, should "help people trust the news again". According to the organization, the credibility of media content is inherently connected to another ongoing, yet already well-documented problem in several Western democracies – a decline in media trust (Hanitzsch et al., 2018).

Lack of authenticity

With the rise of social media and its paramount role in information production, one of the key objectives of information verification is to determine if the information derives from an authentic source or has been manipulated by a malicious actor. For example, Company 4, which was launched in 2014, uses machine learning to determine if a specific Twitter user is authentic or is a "social bot" i.e., an account pretending to be a genuine human user that is actually controlled by computational means (Ferrara, 2017). As Company 4's website mentions, "social bots can be used to manipulate social media users by amplifying misinformation". Accordingly, tech services like Company 4, which specializes in social media account analysis, emphasize the problem of source authenticity online.

Before we discuss the causal agents of information disorder according to AI-powered services, we should underline that the organizations analyzed in this study focus on three main problematic areas of information production in the digitized world: authenticity and credibility of online information sources and the factuality of the media content. To deal with these problems they offer particular AI-based solutions.

Causes of information disorder

As our analysis further demonstrates, AI-powered fact-checking organizations identify at least three different causes of the distorted online media landscape: 1) low-quality information sources; 2) growing digitally mediated content; 3) increased pressure on media professionals to verify the information. In the following subsections, we provide examples of how the causal agents of information disorder are described by the six AI-powered fact-checking initiatives.

Low-quality sources

One of the key drivers of information disorder, according to the organizations, is ill-intended or erroneous sources, especially within the digital realm. Such sources might differ by origin, medium, intensity, or intentions, but they all contribute to the construction of an aberrated social reality. Our analysis shows that some organizations emphasize the involvement of real-life public figures and institutions in the process of producing and amplifying mis/disinformation. As Company 1's website claims, they "fact-check claims made by politicians, public institutions and journalists, as well as viral content online". Such actors often make erroneous claims regularly, which enhances the growing scale of dis/misinformation. As Hassan et al. underline in their paper on aspects of the Company 3 initiative, "politicians repeatedly make the same false claims. Fake news floods cyberspace and even allegedly influenced the 2016 [US presidential] election" (2017, p. 1803). Thus, there is a dire need to hold politicians or institutions accountable for the information they share with the help of fact-checking, which resonates with the normative function of media professionals as watchdogs to hold public actors responsible for their actions.

Other services focus on more non-human entities within online spaces, such as bots and inauthentic social media accounts. For instance, Company 6 is involved in the credibility assessment of information sources both from human and non-human origin. Company 4 is also exploiting AI technologies to automatize the detection of inauthentic Twitter accounts. As noted on their website, "there are many kinds of social bots. Some are harmless or even useful or amusing. But malicious bots can be used to manipulate social media users by amplifying misinformation". In some cases, such sources are involved in "coordinated disinformation campaigns" or "political astroturfing, a centrally coordinated disinformation campaign in which participants pretend to be ordinary citizens acting independently" (Keller et al., 2020, p. 256). They usually attempt to exploit the vulnerability of media professionals and the readership to digest large amounts of information coming from every corner of the digital infrastructure, hence catering to certain political, economic, or ideological interests.

Growing digital content

Thus, the campaigns mentioned above are marked by an abundance of information that also increases the risk of spreading mis/disinformation. By simple logic, the more information produced, the higher the chance of sources that make mistakes or manipulate information. AI-powered fact-checking services highlight the overload of digitally produced content and consequently the growing pressure on media professionals to verify astronomical amounts of data as one of the key drivers of information disorder. The Norwegian startup, Company 2, points out that there are "500 million new tweets, 29 million new blog posts, 5,642,511,302 Google searches, and 720,000 hours of uploads on YouTube" daily, which affects the fact-checking process because it becomes harder to find a way through this ocean of information. Human fact-checkers need to operate efficiently on at least three different levels of fact-checking: claim identification, claim verification, and distribution of fact-checks (Graves, 2018). Meanwhile, the informational overload clogs the pipeline of fact-checking on each level. It becomes harder to make decisions about which claim to choose from a myriad of claims, sort through information, and make sure that all aspects of a claim have been thoroughly checked. "Deep research takes time, and the increasing amount of misinformation is not making it any easier", reads the website, underlining the hardships of keeping up with the scale of misinformation fact-checkers have to deal with.

Pressure to verify information

Besides, information verification as an epistemic activity is a mentally and materially demanding task that requires significant financial and human resources. As Company 3's creators note, "the human fact-checkers cannot keep up with the amount of misinformation and the speed at which they spread" (Hassan et al., 2017, p. 1803). Fact-checking is a relatively young branch of journalism, often characterized by low interest from the general public and low chances of getting advertisement revenues. This makes it unrealistic to expect such resourcefulness from fact-checking organizations and to ask human fact-checkers to deal with all the malicious content in the public discourse. To ease this pressure, AI-powered services offer different solutions "to accelerate research and fact-checking" (Company 2) and to "ensure the accuracy of their news stories" (Company 3), which we will discuss in detail after presenting the companies' take on the implications of spreading mis/disinformation.

The moral judgment: Impact of information disorder

Distorted information ecosystems come at a cost. AI-powered fact-checking organizations provide rich descriptions of the societal consequences of spreading mis/disinformation. As the data analysis show, the selected organizations focus on three major aspects of human life: politics, public health, and economic issues. Fake news as a harmful socio-technical practice certainly goes beyond these aspects and could be considered equally as, for example, an issue of national security (Belova & Georgieva, 2018) education and literacy (Higdon, 2020), climate crises (Allen & McAleer, 2018), etc. Here, we do not claim that AI-powered services provide a full account of the impact information disorder is having on human lives. Instead, we focus on the areas of social life the organizations make salient while positioning strategically within a broader public discourse on information verification practices.

Politics

First and foremost, AI-powered services emphasize the effects of information disorder on politics, especially on the quality of democratic governing, political polarization, and specific political events such as elections. Company 1 repeatedly mentions that mis/disinformation "hurts our democracy, by damaging trust in politicians and political processes" as well as "leads to bad decisions, by disrupting public debate […]". Similarly, scientists working on Company 3's technologies claim that "unprecedented amounts of falsehoods, hyperboles, and half-truths […] do harm to wealth, democracy, health, and national security". Meanwhile, Company 5 and Company 6 emphasize the negative consequences of the information crisis in connection with political, media and social polarization.

Public health

Another recurrent topic among the analyzed websites is how fake news affects public health. Such an emphasis on health-related topics is exemplified by the fact that the COVID-19 pandemic shifted from a health issue to an informational issue, manifested in the emergence of the term infodemic (Solomon et al., 2020). In the beta version of the AI editor created by Company 2, the organization allows users to check claims related to the pandemic and especially vaccines, such as the claim that "vaccines don't cause autism". Considering the scale of the resistance towards vaccines across the globe, it is logical why AI-powered services try to demonstrate the efficiency of their tools by examining public health-related claims.

Economy

The selected organizations also highlight the negative consequences of the information disorder on economy-related issues, underlining the exceptional damage that the spread of fake news has had on businesses, economies, and public wealth in general. Company 4 highlights instances involving inauthentic social media accounts and bots in "committing financial fraud, suppressing or disrupting speech, spreading malware or spam, trolling/attacking victims, and other types of abuse".

AI-powered solutions for distorted information landscapes

Apart from discussing the negative impact of mis/disinformation, AI-powered organizations also demonstrate the importance of verified information and how their solutions can contribute to it. "Good quality information is fundamental to our daily lives", as stated on Company 1's website; "we want to ensure that the right information is available to the right people at the right time". Company 6 notes that they are designing solutions "to protect democratic debate and process and provide access to trustworthy information". For this, the selected services are offering socio-technical solutions that recommend using AI to tackle the mis/disinformation issue on a scale that we will discuss in the following sections of the chapter.

While communicating to broader audiences primarily via their websites, the companies responsible for these services highlight the value of AI-driven solutions in the process of tackling the mis/disinformation problem. They attempt to make the case that automating different processes of information verification is a preferred response to the increased pressure and impact of mis/disinformation on media ecosystems, even though the level and understanding of automation differs from organization to organization. The studied organizations are developing various types of AI-based services related to information verification, including but not limited to what we here identify as automated (or semi-automated) fact-checking for determining the factual value of claims made by relevant public actors; automated credibility assessment of media content; and automated authenticity assessment of information sources.

Automated fact-checking

Some AI-powered services, such as Company 1 and Company 3, aim to create technology capable of autonomously conducting fact-checking. This includes identifying checkworthy claims, comparing them to existing fact-checks or authoritative databases, and making judgments about their factual value. As Company 1's website notes, the organization is "building scalable, robust, automated fact-checking tools to be used in newsrooms…" to make "fact-checking dramatically more effective using existing technology". However, creating automated fact-checking technology is not the primary goal of the organization, and represents only one part of their effort to tackle mis/disinformation issues. Importantly, Company 1 makes a disclaimer that they are not trying to substitute fact-checkers with machinery. Instead, they want to empower fact-checkers with the technologies. Thus, the AI solutions Company 1 designs are part of the bigger process and are based on the prior efforts of human-led fact-checking.

Company 3 also highlights the importance of human contributions to ensuring the functioning of its service. The website declares that the goal of Company 3 is to create technological solutions to online falsehoods. They even hint at the ambition of creating the so-called "holy grail" or end-to-end system capable of conducting information verification autonomously (Hassan et al., 2017). Accordingly, the promise of creating a fully functional automated fact-checking system is something the organization sees as a solution to information disorder.

Other organizations, like Company 6, have a slightly different understanding of what automated fact-checking means. In a report published on their website, Company 6 describes how they used AI for live fact-checking of the 2020 US presidential election. According to the report, "when any of the [US presidential] candidates or moderators spoke, their speech was automatically transcribed into text, and from this transcription, our AI extracted claims and compared them with our existing fact check library".

Automated credibility assessment

Services like Company 5 use AI to assess media content based on the criteria of credibility. The company website states that their algorithm is capable of evaluating articles by considering the quality of the article's website, the expertise of the author, the quality and diversity of sources used, and the tone of the article. Thus, the goal of Company 5 is to determine the information value of the media content based on these metrics and to assist users in deciding whether certain content can be considered credible or not. Credibility evaluation is also one of the goals of Company 6, assessing whether the media source by itself seems credible. Company 6 offers a browser extension to its users to help them navigate websites with confidence that information sources are credible.

Automated authenticity assessment

Another aspect the companies try to address is finding inauthentic information sources and identifying coordinated mis/disinformation campaigns these sources are involved in. For instance, Company 4 analyzes accounts on Twitter with the help of a machine learning algorithm capable of calculating a score where "likely human accounts" get low scores and "likely bot accounts" get high scores. In this context, Company 4 defines a social bot "as a social media account controlled at least in part through software" and more notably, "deceptive bots take on inauthentic personas and are controlled by unknown entities". Such social media accounts are often involved in coordinated mis/disinformation campaigns, an inauthentic process of information production and dissemination. Coordinated mis/disinformation campaigns are also targeted by other AI-powered services such as Company 1 and Company 6.

Conclusion

Though there are a number of initiatives that have been launched to tackle the problem of mis/disinformation, in this chapter we decided to focus on one particular type of endeavor – companies using AI-powered services to address this issue. Informed by Entman's (1993) framing theory as an analytical tool, we studied the strategic communication of six websites presenting AI solutions for various aspects of information verification. This chapter has captured the way AI-powered services frame the problem of information disorder (RQ1), how they view the cause and impact of the spread of mis/disinformation (RQ2), and how they frame their AI-based solutions as a way to deal with the situation (RQ3).

As our analysis shows, AI-powered services problematize current information ecosystems primarily in terms of a lack of factuality and credibility in media content, as well as a deficiency of credible and authentic information sources engaged in creating and circulating information online. The explanation for this state of affairs can certainly not be diminished to the few reasons we have identified in this chapter. From current geopolitics to economic struggles or various kinds of crises (public health, environment, migration, to name a few) the reasons behind this distorted informational reality can be found in many areas of the social-political realm. According to the websites of the selected AI-powered companies, the root causes of information disorder are the amplification of the information sources and media content, along with decreasing quality. Additionally, the websites highlight the increased pressure on people working within the information verification industry. Given the scale of harm, AI-based information verification services highlight the implications of spreading mis/disinformation in politics, especially on the state of democracy, on elections, etc. As the websites emphasize, the information ecosystem is saturated with fake news and other types of falsehoods that negatively affect other aspects of human life, such as the economy and public health.

To respond to this challenge, the companies offer various kinds of automated services. Even though these technologies are far from performing without mistakes, AI-powered organizations and initiatives emphasize that the key to solving the mis/disinformation problem might be in lifting the burden of manually verifying information from the shoulders of human fact-checkers. The selected companies suggest using services that we have identified as automated fact-checking, automated credibility assessment, and automated authenticity assessment.

Using the term “automated” in the presentation of these solutions creates an expectation that these services should be capable of conducting information verification, credibility assessment, or authenticity evaluation without humans. The limitations of this chapter prevent us from analyzing the performance of each AI solution within each category, but we can confidently say that this is not the case. Human effort is very much required even in the most basic steps of operating such technologies, either for producing and sorting the data, labeling it, taking final decisions regarding the correctness of facts, or simply initiating the information verification process. Some of the companies even express cautious optimism about the functioning of their tools, emphasizing that using AI might not be a “silver bullet” that will solve the problem of distorted information ecosystems. Company 1 expects “most fact-checks to be completed by a highly trained human, but we want to use technology to help”. Nevertheless, some initiatives (Company 3) still keep pushing the idea that creating the “holy grail” or an end-to-end information verification system is possible, while others (Company 5) emphasize that their algorithm works without human input, and that they “can offer a consistent assessment of news articles in just seconds”.

Different AI-powered services choose different approaches to communicate their strategic goals depending on their intended audience, which may vary from media professionals to laymen, depending on the simplicity of the tool or the goals the AI-powered companies are trying to achieve. Though all of these companies emphasize the importance of automation and express optimism about AI-based solutions for mis/disinformation, analyzing their websites illuminates that the involvement of humans in information verification remains immutable. Thus, the role of human effort in autonomous information verification systems, as well as the actual functioning of AI-powered services could be a topic of further exploration in order to observe the strategic positioning of the companies and the actual results their services yield.

References

Allen, D. E. & McAleer, M. (2018). Fake news and indifference to scientific fact: President Trump’s confused tweets on global warming, climate change and weather. *Scientometrics*, *117*(1), 625–629. https://doi.org/10.1007/s11192-018-2847-y

Belova, G. & Georgieva, G. (2018). Fake news as a threat to national security. *International Conference Knowledge-based organization*, *24*(1), 19–22. https://doi.org/10.1515/kbo-2018-0002

Burgess, M. (n.d.). Google is helping Full Fact create an automated, real-time fact-checker. *Wired UK*. https://www.wired.co.uk/article/automated-fact-checking-full-fact-google-funding

Chan-Olmsted, S. M. (2019). A review of artificial intelligence adoptions in the media industry. *International Journal on Media Management*, *21*(3–4), 193–215. https://doi.org/10.1080/14241277.2019.1695619

Choudhary, N., Singh, R., Bindlish, I. & Shrivastava, M. (2020). Neural network architecture for credibility assessment of textual claims. *ArXiv:1803.10547 [Cs]*. http://arxiv.org/abs/1803.10547

Ciampaglia, G. L., Shiralkar, P., Rocha, L. M., Bollen, J., Menczer, F. & Flammini, A. (2015). Computational fact checking from knowledge networks. *Plos One*, *10*(6),. https://doi.org/10.1371/journal.pone.0128193

Dalen, A. van (2012). The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists. *Journalism Practice*, *6*(5–6), 648–658. https://doi.org/10.1080/17512786.2012.667268

Di Pietro, R., Raponi, S., Caprolu, M. & Cresci, S. (2021). *New dimensions of information warfare* (Vol. 84). Springer International Publishing. https://doi.org/10.1007/978-3-030-60618-3

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, *43*(4), 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Falkheimer, J. & Heide, M. (2014). From public relations to strategic communication in Sweden: The emergence of a transboundary field of knowledge. *Nordicom Review*, *35*(2), 123–138. https://doi.org/10.2478/nor-2014-0019

Fallis, D. (2014). The varieties of disinformation. In L. Floridi & P. Illari (Eds.), *The philosophy of information quality* (Vol. 358, pp. 135–161). Springer International Publishing. https://doi.org/10.1007/978-3-319-07121-3_8

Ferrara, E. (2017). Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday*. https://doi.org/10.5210/fm.v22i8.8005

Fray, P. (2016, 19 April). Is that a fact? Checking politicians' statements just got a whole lot easier. *The Guardian*. https://www.theguardian.com/commentisfree/2016/apr/19/is-that-a-fact-checking-politicians-statements-just-got-a-whole-lot-easier

Godler, Y. & Reich, Z. (2013). How journalists think about facts: Theorizing the social conditions behind epistemological beliefs. *Journalism Studies*, *14*(1), 94–112. https://doi.org/10.1080/1461670X.2012.689489

Graves, L. (2018). *Understanding the promise and limits of automated fact-checking* [Factsheet]. Reuters Institute for the Study of Journalism; University of Oxford. https://reutersinstitute.politics.ox.ac.uk/our-research/understanding-promise-and-limits-automated-fact-checking

Hallahan, K. (1999). Seven models of framing: Implications for public relations. *Journal of Public Relations Research*, *11*(3), 205–242. https://doi.org/10.1207/s1532754xjprr1103_02

Hallahan, K. (2008). Strategic framing. In W. Donsbach (Ed.), *The International Encyclopedia of Communication*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781405186407.wbiecs107

Hallahan, K., Holtzhausen, D., van Ruler, B., Verčič, D. & Sriramesh, K. (2007). Defining strategic communication. *International Journal of Strategic Communication*, *1*(1), 3–35. https://doi.org/10.1080/15531180701285244

Hanitzsch, T., Van Dalen, A. & Steindl, N. (2018). Caught in the nexus: A comparative and longitudinal analysis of public trust in the press. *The International Journal of Press/Politics*, *23*(1), 3–23. https://doi.org/10.1177/1940161217740695

Hassan, N., Adair, B., Hamilton, J. T., Li, C., Tremayne, M., Yang, J. & Yu, C. (2015). The quest to automate fact-checking. *Proceedings of the 2015 computation+ journalism symposium.*

Hassan, N., Arslan, F., Li, C. & Tremayne, M. (2017). Toward automated fact-checking: Detecting check-worthy factual claims by ClaimBuster. *Proceedings of the 23rd ACM SIGKDD*

International Conference on Knowledge Discovery and Data Mining, 1803–1812. https://doi.org/10.1145/3097983.3098131

Higdon, N. (2020). *The anatomy of fake news: A critical news literacy education*. University of California Press.

Holtzhausen, D. & Zerfaß, A. (Eds.). (2019). *The Routledge handbook of strategic communication*. Routledge.

Jimenez, D. & Li, C. (2018). An empirical study on identifying sentences with salient factual statements. *2018 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN.2018.8489601

Juneström, A. (2020). An emerging genre of contemporary fact-checking. *Journal of Documentation*, *77*(2), 501–517. https://doi.org/10.1108/JD-05-2020-0083

Keller, F. B., Schoch, D., Stier, S. & Yang, J. (2020). Political astroturfing on Twitter: How to coordinate a disinformation campaign. *Political Communication*, *37*(2), 256–280. https://doi.org/10.1080/10584609.2019.1661888

Lahlou, Y., El Fkihi, S. & Faizi, R. (2019). Automatic detection of fake news on online platforms: A survey. *2019 1st International Conference on Smart Systems and Data Science (ICSSD)*, 1–4. https://doi.org/10.1109/ICSSD47982.2019.9002823

Liu, X., Li, Q., Nourbakhsh, A., Fang, R., Thomas, M., Anderson, K., Kociuba, R., Vedder, M., Pomerville, S., Wudali, R., Martin, R., Duprey, J., Vachher, A., Keenan, W. & Shah, S. (2016). Reuters tracer: A large scale system of detecting & verifying real-time news events from Twitter. *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 207–216. https://doi.org/10.1145/2983323.2983363

Manning, J. (2017). In Vivo coding. In J. Matthes, C. S. Davis & R. F. Potter (Eds.), *The international encyclopedia of communication research methods* (1st ed., pp. 1–2). Wiley. https://doi.org/10.1002/9781118901731.iecrm0270

Marconi, F. (2020). *Newsmakers: Artificial intelligence and the future of journalism*. Columbia University Press.

Mishra, R. & Setty, V. (2019). SADHAN: Hierarchical attention networks to learn latent aspect embeddings for fake news detection. *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*, 197–204. https://doi.org/10.1145/3341981.3344229

Moreland, K. & Doerrfeld, B. (2016, 25 February). Automated fact checking: The holy grail of political communication. *Nordic APIs*. https://nordicapis.com/automated-fact-checking-the-holy-grail-of-political-communication/

Mrazek, T. (2019). Truth and trust: Credibility secures the sustainability of journalism. In T. Osburg & S. Heinecke (Eds.), *Media trust in a digital world* (pp. 127–133). Springer International Publishing. https://doi.org/10.1007/978-3-030-30774-5_9

Saldaña, J. (2014). Coding and analysis strategies. In P. Leavy (Ed.), *The Oxford handbook of qualitative research* (pp. 580–598). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199811755.013.001

Solomon, D. H., Bucala, R., Kaplan, M. J. & Nigrovic, P. A. (2020). The "infodemic" of COVID-19. *Arthritis & Rheumatology*, *72*(11), 1806–1808. https://doi.org/10.1002/art.41468

Stray, J. (2019). Making artificial intelligence work for investigative journalism. *Digital Journalism*, *7*(8), 1076–1097. https://doi.org/10.1080/21670811.2019.1630289

Tandoc, E. C., Jenkins, J. & Craft, S. (2019). Fake news as a critical incident in journalism. *Journalism Practice*, *13*(6), 673–689. https://doi.org/10.1080/17512786.2018.1562958

Vosoughi, S., Roy, D. & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Walther, J. B. & Jang, J. (2012). Communication processes in participatory websites. *Journal of Computer-Mediated Communication*, *18*(1), 2–15. https://doi.org/10.1111/j.1083-6101.2012.01592.x

Wardle, C. (2018). The need for smarter definitions and practical, timely empirical research on information disorder. *Digital Journalism*, *6*(8), 951–963. https://doi.org/10.1080/21670811.2018.1502047

Wardle, C. & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking*. Council of Europe. https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

Whittaker, J. (2019). *Tech giants, artificial intelligence, and the future of journalism* (1st ed.). Routledge.

Zeng, X., Abumansour, A. S. & Zubiaga, A. (2021). Automated fact-checking: A survey. *Language and Linguistics Compass*, *15*(10). https://doi.org/10.1111/lnc3.12438