Technovation 151 (2026) 103392

Contents lists available at ScienceDirect

Technovation

journal homepage: www.elsevier.com/locate/technovation

Digital divide and artificial intelligence for health

Jean Clara [a,*], Bussotti Jean-Flavien [b], Cecere Grazia [c], Omrani Nessrine [d], Papotti Paolo [b]

[a] Grenoble Ecole de Management, 12 Rue Pierre Semard, 38000, Grenoble, France
[b] EURECOM, Biot, 06410, France
[c] Institut Mines-Telecom, Business School, Evry-Courcouronnes, 91000, France
[d] Paris School of Business, Paris, 75013, France

ARTICLE INFO

Keywords:
Digital divide
Health
Inequality
Fact-checking
SDGs

ABSTRACT

Social media platforms have become key intermediaries for ad campaigns, but concerns persist regarding the veracity of information presented in ads. In the health sector, false or unsupported claims in ad content can have real-world public health consequences. On these platforms, the display of ads is managed by recommendation systems that match the content of the ad to the interests of the user. This paper investigates whether the use of AI algorithms to recommend ads on social media platforms may help progress toward the Sustainable Development Goals (SDGs). We collected ads across all US states on Meta and Instagram during a period marked by increased public health concerns. Using a fine-tuned deep learning model, we fact-checked the content of these ads. The results of the fact-check show that only 0.2 % of the ads were classified as misinformation, and 15.41 % of the ads were classified as ambiguous. Both types of ads are less likely to be recommended to users located in wealthier states especially when health-related. Also, health-related ads classified as misinformation are more likely to be recommended to users in states with high percentage of people without health insurance. We argue that the use of recommendation systems contributes to widening the digital divide, which can hinder the achievement of SDGs.

1. Introduction

Artificial intelligence (AI), defined as machines, software and algorithms that act by recognizing and responding to their environment (Daron and Pascual, 2020) has caused significant transformations in a variety of industries and sectors (La Torre et al., 2023; Liu et al., 2020; Bahoo et al., 2023; Brynjolfsson et al., 2019, 2023; Abrardi et al., 2022). From manufacturing (Patalas-Maliszewska et al., 2024) and healthcare (Rajpurkar et al., 2022) to marketing (Davenport et al., 2020) and beyond (OECD, 2019), AI has triggered innovations that have fundamentally changed how humans engage with technology (Cockburn et al., 2018). For example, AI is now used to help decision makers with data-driven insights (Duan et al., 2019) or even to automate tasks and processes (Acemoglu and Restrepo, 2018) for supply chain optimization. Rapid adoption of AI in organizations has led to a global AI market valued at more than €130 billion in 2023, highlighting its strategic importance across sectors and organizational functions.[1]

One sector in which the use of AI is particularly pronounced is advertising (Ford et al., 2023). Advertising is a sector characterized by scarce consumer attention (Gentzkow, 2014) and intense advertiser competition (Evans, 2009), in which AI plays a strategic role. Unlike traditional algorithms, AI algorithms can dynamically adapt, enabling real-time resource optimization (Zhang et al., 2021) and highly personalized targeting at scale (Agrawal et al., 2022) that demonstrate tangible benefits. Among them, enhanced ad effectiveness (Davenport et al., 2020; Shumanov et al., 2022), improved user experience, and cost-effective transactions (Goldfarb and Tucker, 2019) stand out.

In this context, social media platforms are interesting strategic intermediaries for organizations' advertising practices. In particular, they facilitate business-to-customer interactions by connecting advertisers on one side with users on the other side. The added value of social media platforms lies in their use of AI algorithms that leverage granular user

This article is part of a special issue entitled: AI x SDGs published in Technovation.

* Corresponding author.
E-mail address: clara.jean@grenoble-em.com (J. Clara).

[1] https://www.europarl.europa.eu/RegData/etudes/ATAG/2024/760392/EPRS_ATA(2024)760392_EN.pdf, https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai, May 2025.

https://doi.org/10.1016/j.technovation.2025.103392
Received 29 September 2024; Received in revised form 1 October 2025; Accepted 13 October 2025
Available online 8 November 2025

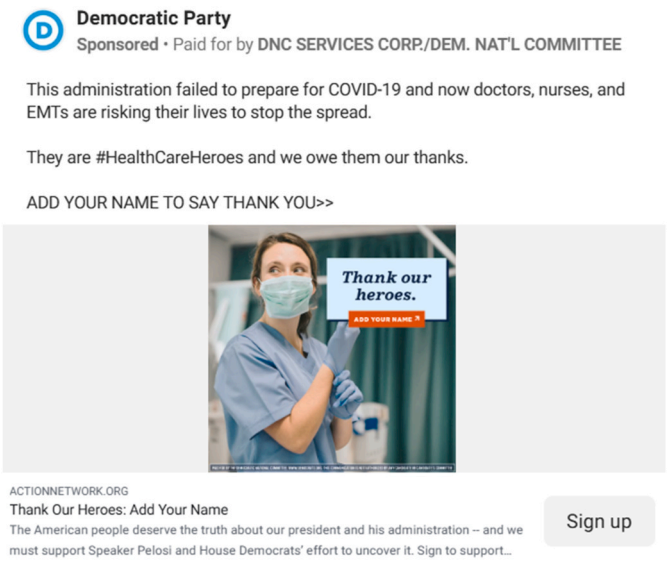

Fig. 1. Health Ads Related to the Pandemic.

Table 1
Examples of labeled ad content used in CT-BERT training.

Misinformation	"china lied about the coronavirus and put american lives at risk! as your state representative, i will join president trump in his efforts to hold china accountable!"
Not Misinformation	"covid-19 has dramatically affected travel. what comes next? what measures should be taken to ensure the safety of passengers? share your opinions! join the conversation."
Ambiguous	"you think covid-19 is bad?"

Table 2
ML model evaluation on 500 manually labeled ads.

Label	Precision	Recall	F1 Score	Support
Not misinformation	0.8795	0.9733	0.9241	375
Ambiguous	0.3333	0.0385	0.0690	52
Misinformation	0.8861	0.9589	0.9211	73

data to tailor advertising content to individual preferences or traits, thus substantially reducing user search costs (Santos et al., 2012) and optimizing advertiser budget allocation. This algorithmic use of data to make suggestions about existing products is often referred to as recommendation systems (Kretschmer and Peukert, 2020; Hosanagar et al., 2014). Low search costs enabled by recommendation systems are essential in an environment where customer attention is limited. This not only benefits advertisers, but also helps explain why social media platforms are increasingly used as primary sources of information, with Meta emerging as the most widely used among them.[2]

However, concerns are raised about the veracity of information and the potential creation of deceptive material (Sandrini and Somogyi, 2023) available on social media platforms (Moravec et al., 2019). Previous studies show that social media platforms are often subject to fake news (Allcott and Gentzkow, 2017), polarization (Azzimonti and Fernandes, 2023; Arora et al., 2022), echo chambers (Kitchens et al., 2020), misinformation (Domenico et al., 2021), and potential algorithmic bias in the diffusion of information on job offers and education (Lambrecht and Tucker, 2019; Sapiezynski et al., 2022). Additional growing concerns are raised about the misinformation contained in ads (Hattori and Higashida, 2014; Rao, 2022), the architectural characteristics of digital platforms (Allcott et al., 2019), and the role of AI algorithms in fostering filter bubbles on social media platforms (Acemoglu et al., 2024).

Table 3
Summary statistics at state level.

Variable	Mean	Std. Dev.	Min	Max
Non-health-related ads				
# Impressions - non-misinformation ads	923.653	959.826	93.779	5013.228
# Impressions - ambiguous ads	942.829	930.913	96.863	4325.648
# Impressions - misinformation ads	469.249	463.901	51.075	2256.227
Health-related ads				
# Impressions - non-misinformation ads	1034.148	1190.013	74.253	6281.785
# Impressions - ambiguous ads	932.576	948.202	90.95	3700.076
# Impressions - misinformation ads	2351.356	2252.841	249.882	10949.464
State characteristics				
Population	6,652,315	7,451,008	581,381	39,029,342
GDP per capita	58,981	10,975	39,157	88,467
% of People w/o Health Insurance	9.722	3.541	3.2	19.9
Tot. COVID-19 cases	10,748,450	11,639,839	500,900	53,548,352
Tot. COVID-19 deaths	144,485	161,668	3263	703,532
Observations		50		

Notes: On average an advertiser page has 379,214 followers and 375,031 likes. We have also the following distribution of advertisers category: 56.39 % are pages that fit into Business & Services, 11.5 % into Politics and Government, 9.31 % unclassified, 8.39 % in Media & Information, 3.38 % in Civil & Society, 3.19 % in Education, 2.74 % in Arts & Culture, 2.19 % in Public Figure, 1.73 % in Health & Science, 1.19 % in Technology & Digital.

The main problem with the spread of inaccurate information is that its consumption generates negative externalities beyond the online space (Carrieri et al., 2019) and harms consumers (Allcott et al., 2020). For instance, its consumption has shown to cause detrimental societal effects ranging from interference in political election (e.g. 2020 US Capitol riots) to health impacts including increased mental distress (Verma et al., 2022), vaccine hesitancy (Do Nascimento et al., 2022; Lee et al., 2022), and promotion of unproven treatments (Suarez-Lledo and Alvarez-Galvez, 2021).[3] During the COVID-19 pandemic, false claims about the dangers of mRNA vaccines, which spread on digital platforms, contributed to hospital overload[4] and increased death rates among unvaccinated individuals.[5] The increased diffusion of misinformation related to health was defined by the World Health Organization (WHO) as "infodemic" (WHO, 2020). Health misinformation, which refers to information that contradicts the established scientific consensus on a given phenomenon, is different from disinformation, as it does not incorporate the notion of intentionality (Swire-Thompson and Lazer, 2020). Unlike political misinformation, health misinformation has strong public health implications with immediate life-threatening consequences. This issue not only threatens individual well-being and equality in information access, but could also undermine broader efforts to achieve the Sustainable Development Goals (SDG) in a context where investments are being made to tackle current socioeconomic development challenges around the world (Johnson and Acemoglu, 2023;

[2] https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/, April 2024.

[3] https://www.cancer.gov/news-events/cancer-currents-blog/2021/cancer-misinformation-social-media, https://www.who.int/europe/news/item/01-09-2022-infodemics-and-misinformation-negatively-affect-people-s-health-behaviours–new-who-review-finds, May 2025.

[4] https://www.bloomberg.com/graphics/2021-covid-surge-shows-overwhelming-cost-of-being-unvaccinated-america/, May 2025.

[5] https://ourworldindata.org/grapher/united-states-rates-of-covid-19-deaths-by-vaccination-status, May 2025.

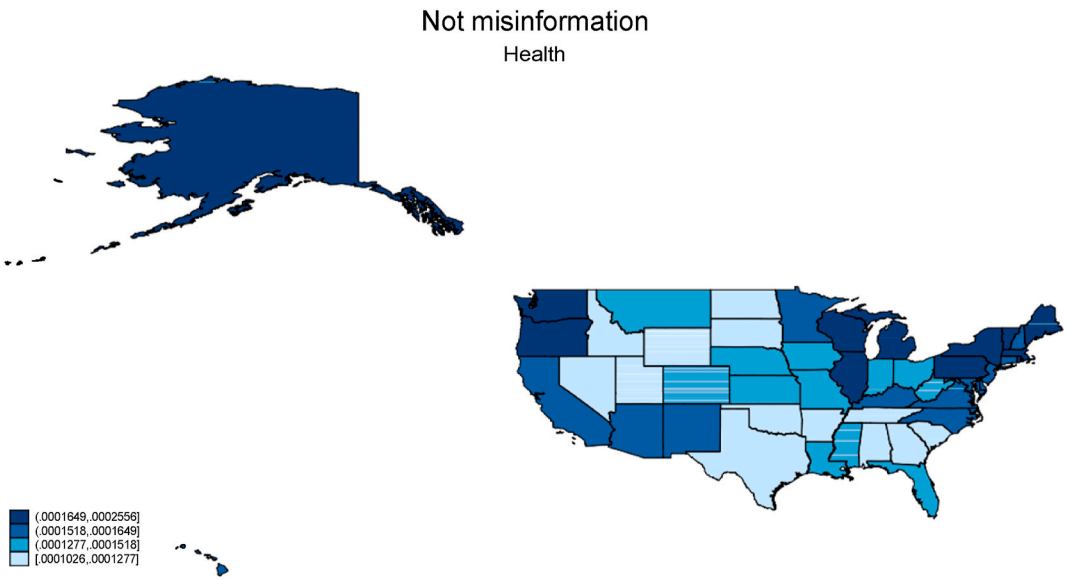

Fig. 2. Impressions per capita - Not misinformation & Health-related ads.

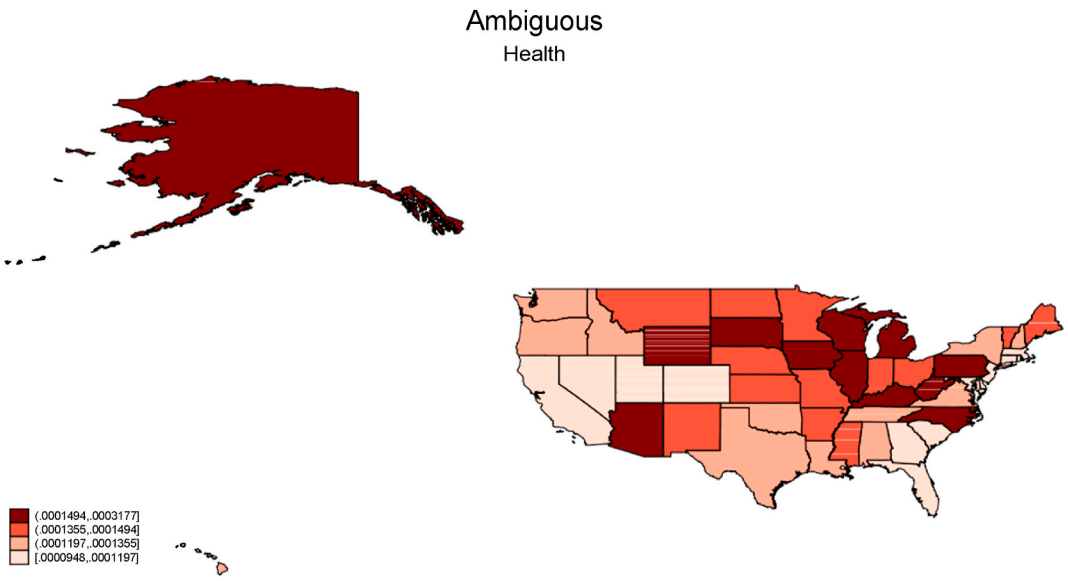

Fig. 3. Impressions per capita - Ambiguous & Health-related ads.

Deaton, 2024; Banerjee and Duflo, 2012).

In this paper, we explore whether recommendation systems on social media platforms, used as a primary source of information,[6] could threaten progress toward the SDGs related to reducing inequalities (SDG10) and increasing health and well-being (SDG3).[7] While most existing research highlights the positive impacts of AI use in various sectors (Tomašev et al., 2020; Chatterjee et al., 2022; Liu et al., 2020), its dark side remains underexplored (Mikalef et al., 2022; Papagiannidis et al., 2023) and its effects on the achievement of the SDGs remain unclear (Gupta et al., 2021; Vinuesa et al., 2020; Sætra, 2021; Di Vaio et al., 2020; Camodeca and Almici, 2021). We study in particular how AI algorithms, used to recommend ad content, might contribute to widen economic and social differences, and thus the digital divide. The underlined research questions being addressed are: *Do AI algorithms efficiently control the veracity of information contained in ad content? Do AI algorithms recommend ads containing misinformation based on socioeconomic indicators and influence the digital divide?*

To address these research questions, we employed a method that involved the use of an innovative fact-checking algorithm that takes advantage of a variety of different technologies to detect misinformation related to health issues. The fact-checking algorithm allowed an evaluation of the likelihood that an ad contained false or unsupported claims designed as misinformation (Nyhan, 2020). Our approach combined a broad understanding of complex textual claims enabled by pre-trained language models (LM) with domain-specific knowledge allowed by the fine-tuning of these models with curated datasets.[8] To train the model,

[6] https://www.bbc.com/news/articles/c93lzyxkklpo, September 2025.

[7] https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/, April 2024.

[8] Fine-tuning involved taking pre-trained LMs and training them further on smaller, specific datasets to refine their capabilities and improve their performance on a particular task or in a specific domain.

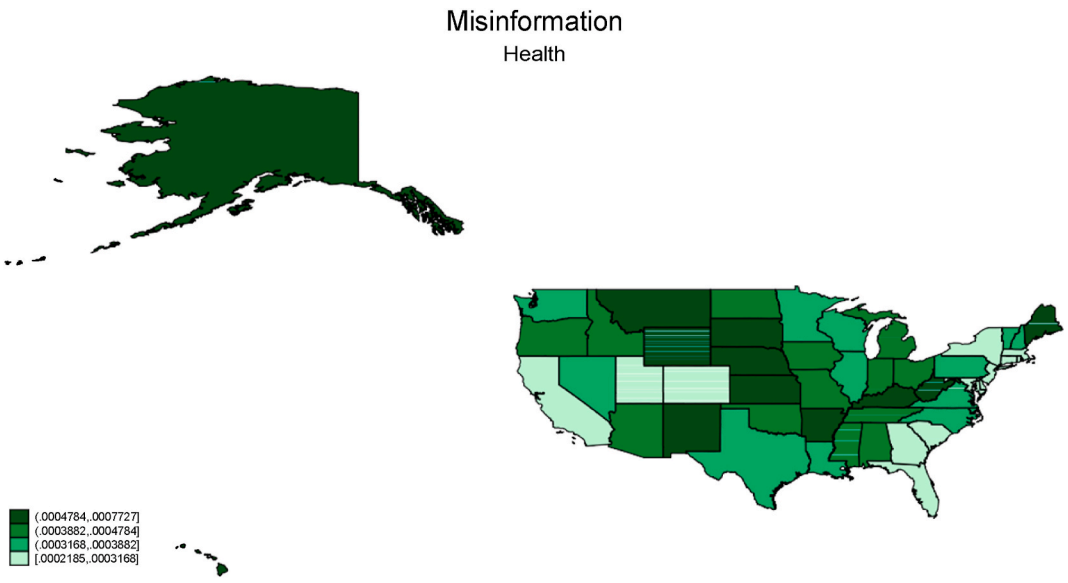

Fig. 4. Impressions per capita - Misinformation & Health-related ads.

Table 4
Ad recommendation and GDP per capita.

	Non-Health-related Ads			Health-related Ads		
	Not Misinformation	Ambiguous	Misinformation	Not Misinformation	Ambiguous	Misinformation
	(1)	(2)	(3)	(4)	(5)	(6)
High GDP States	−8.142	−110.643**	−53.780***	92.066	−132.244	−274.715**
	(34.206)	(54.802)	(17.15)	(77.852)	(116.175)	(127.756)
Constant	79.327**	178.589***	83.201***	−27.223	222.892**	510.539***
	(33.150)	(52.355)	(19.378)	(74.503)	(98.378)	(121.746)
R-squared	0.979	0.954	0.975	0.929	0.820	0.948
Observations	50	50	50	50	50	50
Population	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS estimates. The dependent variable is the average number of impressions displayed in a given state. Standard errors are clustered at the state level. We include the state population. Significance at 1 %; 5 % and 10 % levels indicated respectively by ***,** and *.

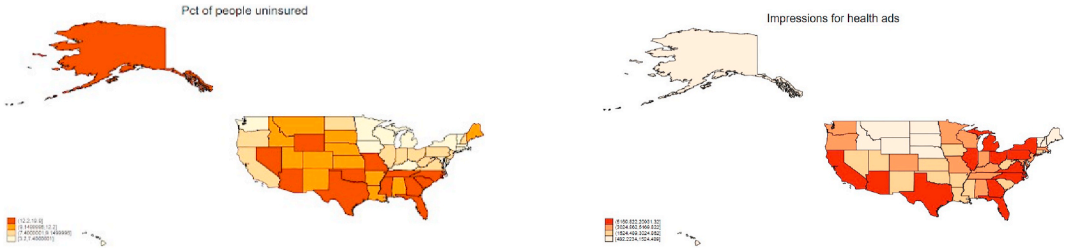

Fig. 5. Percentage of People without Health Insurance per State and Health-related ad Impressions.

we used human annotations and relied on synthetic data produced by a generative AI tool to increase the training sample size. Generating synthetic data using a large language model (LLM) is a novel approach to supplement our training data set, particularly for the misinformation class, which is inherently rare in moderated platform data from the real world. This method allows us to overcome the limitations of relying solely on scarce human-annotated examples, providing a more robust and balanced training set for the fact-checking task.

We collected ads displayed in all US states on Meta and Instagram from the Meta Ad Library between January and June 2020, a period marked by the COVID-19 pandemic, during which many ads were related to health.[9] For each ad, we obtained information on the text of the ad, total impressions, and their distribution by age, sex, and state. We combined these data with the results of the fact-checking algorithm, identifying whether a given ad in our data set included misinformation or not, and US state-level administrative data. Our aim is twofold. First, to assess the ability of AI algorithms to evaluate the veracity of information contained in health-related ads. Second, to assess whether algorithmic recommendation of health-related ads – especially those classified as misinformation on social media platforms – is correlated

[9] See Section 3 for more details.

Table 5
Disparities in health ad recommendation linked to health insurance coverage.

	Non-Health-related Ads			Health-related Ads		
	Not Misinformation	Ambiguous	Misinformation	Not Misinformation	Ambiguous	Misinformation
	(1)	(2)	(3)	(4)	(5)	(6)
% People w/o Health Insurance	−18.720**	−0.784	8.927***	−56.161***	−9.641	50.466**
	(7.763)	(8.884)	(3.145)	(15.658)	(17.075)	(24.024)
Constant	246.654***	139.874	−20.497	523.068***	255.986	−63.536
	(86.760)	(85.449)	(28.835)	(160.876)	(160.709)	(255.128)
R-squared	0.983	0.950	0.977	0.954	0.817	0.951
Observations	50	50	50	50	50	50
Population	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS estimates. The dependent variable is the average number of impressions displayed in a given state. Standard errors are clustered at the state level. We include the state population. Significance at 1 %; 5 % and 10 % levels indicated respectively by ***,** and *.

Table 6
Information divide: More ambiguous ad impressions for highly COVID-19 impacted states.

	Health-related Ads		
	Not Misinformation	Ambiguous	Misinformation
	(1)	(2)	(3)
States with High COVID-19 Cases	72.280	366.583**	242.737
	(94.723)	(137.793)	(185.897)
Constant	−6.452	79.851**	338.241***
	(37.929)	(30.057)	(79.718)
R-squared	0.928	0.841	0.946
Observations	50	50	50
Population	Yes	Yes	Yes

Notes: OLS estimations. The dependent variable is the average number of impressions displayed in a given state. Standard errors are clustered at state level. We include state population. Significance at 1 %; 5 % and 10 % levels indicated respectively by ***,** and *.

with socioeconomic indicators of US states.

Our findings related to the fact-check indicate that very few ads displayed on Meta and Instagram were classified as misinformation. Although the fact-checking algorithm classified 15.41 % of the ads as ambiguous, only 0.2 % were considered to provide misinformation. However, we found evidence of significant relationships between algorithmic recommendation of ads and the socioeconomic characteristics of US states, which has implications for the digital divide. Algorithmic recommendation of ads classified as misinformation or ambiguous is negatively associated with US states with high GDP per capita, especially when health-related. This suggests that users in wealthier states are less likely to be recommended low-quality information i.e., misinformation and ambiguous ads. We also observe a positive correlation between the algorithmic recommendation of health-related ads classified as misinformation and US states with a high percentage of uninsured individuals. However, we do not observe any correlation between the algorithmic recommendation of health-related ads and the number of COVID-19 cases at the state level.

This paper has several managerial and policy implications.

From a managerial perspective, problems related to the management of misinformation were considered in 2024 by the World Economic Forum to be the most prominent risk in the coming two years.[10] The huge volume of ad content offered on social media platforms underscores the need for scalable fact-checking. Combining human oversight with robust AI algorithms to curate and manage this content could be a step in that direction. In health-related contexts, social media platforms could implement tools that assess the likelihood of misinformation in ad content to help users navigate information more safely. Our fact-checking algorithm directly addresses this need by offering a timely automated ad content moderation tool. Although most fact-checking tools focus on organic posts rather than ad content (Barrera et al., 2020), our study bridges this gap and offers practical solutions – especially relevant in light of algorithmic platform transparency obligations under the European Digital Services Act (DSA).

From a policy-maker perspective and especially given a recent 2025 context in which fact-checking programs are dropped on social media platforms,[11] there is a pressing need for an effective solution to verify the validity of claims and data online (Nakov et al., 2021). Scalable solutions could support regulators, policymakers, and journalists by significantly reducing the pressure on human fact-checkers and promoting more timely debunking of misleading claims. This aligns with European Union recommendations on the use of technology to combat misinformation (Martens et al., 2018). Policymakers might encourage platforms to adopt such tools or require the creation of dedicated recommendation systems for public service campaigns, in order to ensure their reach across socioeconomically diverse audiences.

Our paper is organized as follows. Section 2 reviews the current literature on online misinformation and fact-checking, recommendation systems, and the role of internet technologies in socioeconomic inequality. Section 3 introduces the context of the study and describes the collection of data through the Meta Ad Library API and external data sources. Section 4 details the construction and training of our fact-checking algorithm. Section 5 offers a preliminary analysis through descriptive evidence. Section 6 presents our main results. Section 7 discusses the potential mechanisms that explain our results and the underlying implications. Section 8 highlights some limitations and concludes.

2. Literature review

This article relies on three streams of literature. First, we refer to the literature on online misinformation and fact-checking. Second, we contribute to the literature related to recommendation systems. Third, we build on the literature that investigates how internet technology affects socioeconomic inequality.

2.1. Online misinformation and fact-checking

Numerous studies highlight social media platforms, including Meta (Guess et al., 2021) and Twitter, as widespread channels for disseminating misinformation online (Allcott and Gentzkow, 2017). For example, Vosoughi et al. (2018) find that false news spreads faster and

[10] https://www.weforum.org/publications/global-risks-report-2024/, January 2024.

[11] https://www.nytimes.com/live/2025/01/07/business/meta-fact-checking, May 2025.

reaches more people on Twitter than true news. Similarly, Chiou and Tucker (2018) highlight how Facebook groups, especially in anti-vaccine communities, amplify false information dissemination, a pattern also confirmed by Törnberg (2018), who show that echo chambers facilitate the spread of misinformation and rumors. An explanation is that people who often share misinformation focus their attention on factors other than the accuracy of the news they read (Pennycook et al., 2021).

Although misinformation might seem insignificant in terms of impacts on individuals, the reality contradicts this assumption. Its effects have been observed both at the societal and individual levels (Adams et al., 2023) and have been shown to have strong implications for social welfare (Glaeser and Ujhelyi, 2010). The main issue with the consumption of misinformation is the development of harmful social behaviors (Imhoff et al., 2022; Pennycook and Rand, 2021), which are particularly pronounced in the context of health, where viruses are often taken as examples (Ghenai and Mejova, 2017; Valecha et al., 2020; Ho et al., 2023). During the COVID-19 pandemic, a large amount of information has circulated, occasionally causing adverse effects on public health (Van Der Linden, 2022) with a direct effect on vaccination intent (Loomba et al., 2021), resulting in an increase in the total number of cases and deaths in the initial stages of the pandemic (Bursztyn et al., 2020).

To address the challenges raised by online misinformation that was initially concentrated in the political sphere (Amazeen, 2016; Graves, 2016), fact-checking has emerged as an interesting solution (Graves, 2018). Described as an internal process used to verify facts before publication (Graves and Amazeen, 2019), fact-checking has gained popularity over time. Key contributing factors include weak democratic institutions (Amazeen, 2020), the decline of traditional journalism, rapid technological change, and periods of social or political instability (Amazeen, 2019). Although some reluctance to use online fact-checking services has been observed (Brandtzaeg et al., 2018), their proven effectiveness (Walter et al., 2020) makes them a powerful tool for journalists to expose political spin and increasingly sophisticated media manipulation techniques targeting users (Dobbs, 2012). However, fact-checking has also faced criticism (Graves, 2017), ranging from selection bias – particularly in the choice of claims being debunked – to concerns about the accuracy of statement verifications (Amazeen, 2013).

Considering the vast amount of information available online, it is necessary to understand the sociotechnical context of fact-checking (Micallef et al., 2022). Saeed et al. (2022) explore the use of crowdsourcing as a potential approach to improve fact-checking practices. Although the authors present crowdsourcing as an effective fact-checking strategy in specific settings, the inconsistency and lack of actionable results prompt the exploration of alternative solutions for more effective fact-checking of online misinformation. In this context, AI algorithms are interesting support for fact-checkers by enhancing scalability. As an illustration, Peskine et al. (2023b) apply transformer-based models[12] (CT-BERT) and node embedding techniques[13] (node2vec) to address COVID-19-related conspiracy theories using tweet text and user interaction graphs, and show that this is a viable approach to address this challenge. Just (2024) advocates Natural Language Processing (NLP) as a non-human innovation intermediary, enhancing decision-making by expanding information analysis and reducing costs through automation. In the same vein, Song et al. (2017) underscore the efficiency of NLP tools, particularly the F-term approach, in classifying patent documents based on technical attributes.[14]

[12] A transformer model is a type of neural network that learns context and consequently, meaning, by discerning connections within sequential data.

[13] Node Embeddings are vectors that reflect properties of nodes in a network.

[14] The F-term (Schellner, 2002) approach is a method used to classify patent documents using the k-nearest neighborhood method.

Until now, there has been ongoing discussions about the potential of AI algorithms to address the issue of detecting online misinformation (Martens et al., 2018). Our paper contributes to the academic literature that believes that AI should be part of the solution when it comes to managing misinformation (Zhou and Zafarani, 2020; Thorne and Vlachos, 2018). Although traditional emphasis on posts or tweets was commonly targeted in previous efforts, our article diverges from previous studies with the development of a fact-checking algorithm specifically tailored to advertising content, where recent and growing concerns about misinformation spread are raised (Rao, 2022; Fong et al., 2024).

2.2. Recommendation systems

Given the large volume of content available online, digital platforms increasingly rely on AI algorithms to process content and data. AI, defined as general-purpose technology, is of great interest to an increasing number of private and public organizations (Agrawal et al., 2022) due to its scalability. Among AI algorithms, systems that match content with users based on personal traits or preferences – commonly referred to as recommendation systems (Resnick and Varian, 1997) – are central to online personalization (Koren et al., 2009; Ricci et al., 2010). Applications range from web search in e-commerce (Yuan et al., 2025), music industry (Kretschmer and Peukert, 2020; Datta et al., 2018; Hosanagar et al., 2014; Aguiar and Waldfogel, 2018), to streaming services and video platforms (Gomez-Uribe and Hunt, 2015; Qian and Jain, 2024).

Social media platforms that act as an information intermediary (DeNardis and Hackl, 2015) are a major application of recommendation systems (Narayanan, 2023; Stray et al., 2024). In particular, personalized algorithmic recommendations are essential to increase consumer product consumption and significantly shape the dynamics of market competition (Chen and Tsai, 2024). For example, Senecal and Nantel (2004), using an online experiment, show that online recommendations made by recommender systems significantly influence consumer choice where products were selected nearly twice as often when they were recommended. In a similar vein, Bakshy et al. (2015), using a large-scale dataset, demonstrate how algorithmic news rankings, in combination with peer sharing, shape patterns of news consumption on social media platforms. Going a step further, Aridor et al. (2022) emphasize that the increase in consumption enabled by recommendation systems is driven by the informational value these systems provide to users.

However, given the need for personalization to ensure optimal content-user matches (Agarwal and Dhar, 2014), recommendation systems can lead to unintended side effects. This includes algorithmic bias (Kordzadeh and Ghasemaghaei, 2022; Cowgill and Tucker, 2019; Obermeyer et al., 2019), filter bubbles (Pariser, 2011; Berman and Katona, 2020) and the potential spread of misinformation. For example, Liu et al. (2021) find that content-based and collaborative filter recommendation algorithms could contribute to the formation of filter bubbles. In the same vein, Bourreau and Gaudin (2022) demonstrate that recommendation systems might introduce recommendation bias if the platform recommends a content type that differs from the optimal mix of consumers. These side effects are largely driven by feedback loops on social media platforms, where over-personalization by recommendation systems restricts users' exposure to diverse information (Acemoglu et al., 2024).

Although prior work has examined how recommendation systems can contribute to the amplification of misinformation (Pathak et al., 2023), the effect of existing AI algorithms on the recommendation of false and misleading information remains not sufficiently studied (Fernandez and Bellogín, 2020). In this paper, we aim to fill this gap. Building on this literature, our article shifts focus from misinformation amplification to assess whether recommender systems promote equal exposure to ad content through the lens of Sustainable Development Goals (SDGs).

2.3. Role of internet technologies in socioeconomic inequality

The third strand of literature focuses on examining the role of internet technologies on socioeconomic inequality. The early diffusion of the Internet has been associated with a wider gap in the access and use of information and communication technologies (ICT), creating the so-called digital divide (Van Dijk, 2005). This has contributed to an increase of existing inequalities (DiMaggio and Hargittai, 2001), where sociodemographic and socioeconomic characteristics are well-known drivers (Robinson et al., 2015). The digital divide[15] – referring to differences in the gains from the use of the Internet (Lutz, 2019) – has been previously documented in the literature (Goldfarb and Prince, 2008; Cecere and Corrocher, 2012). As an illustration, Forman et al. (2002) underline the geographical spread of Internet diffusion, increasing inequality between urban and rural areas.

Recently, a new form of digital divide has been observed by Gran et al. (2021) through the ability of individuals to navigate on the Internet while being aware of the role of AI algorithms. Although the use of these algorithms has shown tangible benefits with fairer decisions for African-American and Hispanic individuals in pretrial bail decisions (Kleinberg et al., 2018), there are also significant risks with strong impact on socioeconomic inequalities (Capraro et al., 2024). AI algorithms can harm vulnerable populations (O'Neil, 2017) by reinforcing established structural inequalities (Noble, 2018). In the context of algorithmic automation, Eubanks (2018) illustrates how such systems can exacerbate socioeconomic disparities. She documents cases where automation errors led to the wrongful denial of food stamps, failures in managing homelessness, and inaccurate risk assessments related to child maltreatment. One possible explanation lies in the sociotechnical nature of AI algorithms, which tend to reproduce structural discrimination and informational inequality (Noble, 2018).

Given the control of online information enabled by AI algorithms (Acemoglu, 2021), Barocas et al. (2023) warns about two dominant legal implications of their use, i.e., disparate impact and disparate treatment (Kleinberg et al., 2019). Although disparate treatment refers to intentional discrimination, disparate impact refers to practices that have a dis-proportionate effect on a protected class. The latter is at play when AI algorithms on social media platforms are deployed, especially towards content access. The fact that not everyone benefits equally from technology requires more research toward establishing clear technology needs that could lead to the development of more coherent frameworks and policies to bridge digital gaps (Lythreatis et al., 2022). We enhance this body of literature by investigating how the nature of information disseminated on social media platforms by AI algorithms may relate to geographical digital divides and inequalities, which are underexplored in ad-specific contexts.

3. Empirical section

3.1. Data collection

We collected data from the Meta ad library to empirically test the diffusion of online (mis)information and its recommendation according to the socioeconomic characteristics of US states. We chose Meta due to its prominence as the main player in news distribution channels (Martens et al., 2018). Given the crucial importance of managing health-related misinformation, we focused on pandemic-related ad content. Fig. 1 shows an example of a health-related ad. We collected data using the Meta API.[16]

Our final sample included 145,272 ads from 1,096 advertisers. We limit our sample to only ads displayed in all US states from January to June 2020. As advertisers can decide to target the whole country or only a certain set of states, we consider only ads published in all states. This aims to mitigate advertiser targeting bias and focus on algorithmic ad recommendations. Our data include information on the average number of impressions per ad, broken down by demographics and US states (see Table 8).

Online advertising is an essential communication tool. Although online advertising may be used for commercial purposes, it is also used by public institutions and non-governmental organizations to inform individuals. Thus, algorithmic detection of information is triggered by the coexistence of different types of information. The urgency surrounding the COVID-19 pandemic meant that the platforms had to be able to discern whether the ads were appropriate or potentially contained inaccurate information (De Alves et al., 2022).

Before publication, all paid ads on Meta platforms - Meta and Instagram - are reviewed by an automated ad-screening system to ensure compliance with Meta's advertising policy. This algorithmic ad screening helps the platform identify ad content that could harm users through what the platform terms 'unacceptable content'. Examples of unacceptable content include ads promoting child sexual exploitation, abuse, and nudity, discriminatory practices, hate speech, inaccurate health information, and anti-vaccine content.[17] Ad that includes any of these types of content violates Meta's advertising policy and is removed from the platform.

3.2. Measuring ad distribution inequality

To assess inequality in health information access, we augmented the Meta ad-related data with three sources of administrative data at the US state level. First, we collected open data from the US Census Bureau for the year 2020 on the percentages of people without health insurance for each US state. Second, we collected open data from the US Centers for Disease Control and Prevention (CDC) on the number of cases of COVID-19 and COVID-19-related deaths per state and per month in 2020. Third, we collected data on GDP per state for the year 2020 from the US Bureau of Economic Analysis (BEA). We aggregated ad-level data to the state level, enabling us to match ad-related data with US state administrative data.

4. Fact-checking using a deep learning model

To analyze the ad content, we use a fact-checking method that relies on a deep learning model. We rely on a fact-checking model to verify and annotate the content of real ads published on Meta's ad library platform. To annotate the whole dataset, we use a fine-tune CT-BERT (Müller et al., 2023) pre-trained model to predict a label for each ad. CT-BERT is a deep learning model trained with text data related to a crisis, including posts on social media and domain-specific vocabulary. The purpose of this model is to better understand and interpret the language used during various types of emergencies. Our choice of CT-BERT is not arbitrary; unlike standard BERT models, CT-BERT is specifically pre-trained on crisis-related communications. This is particularly beneficial for our study that focuses on COVID-19-related (mis)information, as CT-BERT is designed to understand the linguistic patterns and vocabulary prevalent during public health crises, offering a potentially more sensitive and accurate detection of misinfor-mation in this domain than generic models. Although CT-BERT is a relatively small model, we find that its performance is better than much larger models, such as LLaMA 2 7B (Touvron et al., 2023) fine-tuned for classification and ChatGPT (OpenAI, 2024) instructed with prompting.

[15] We are referring to the third level of digital divide.
[16] API is Application Programming Interface.
[17] See https://transparency.fb.com/policies/ad-standards/for a complete list, February 2024.

4.1. Data labelling process

As any encoder model, CT-BERT relies on fine-tuning to be instructed for our dataset and labels. We further train the original CT-BERT model using custom examples. We use human annotation to classify the content of a random sample of 2,600 ads, including health-related ads and non-health-related ads. We follow the same approach used by Cecere et al. (2021) based on the guidelines provided by the statement of the Australian Competition and Consumer Commission (ACCC) to distinguish between false and mis-leading claims.[18] We also rely on the Federal Trade Commission (FTC) statement related to unproven health-related claims.[19] In addition to the general guidelines outlined in the FTC and ACCC documents, the FTC report includes specific examples of misleading COVID-19-related claims, helping annotators interpret what constitutes misinformation in this context. Similarly, the ACCC document provides examples of misleading and deceptive advertising content, further supporting the annotation process.

Three individuals participated in the annotation task. Two of the three annotators were experienced researchers familiar with fact-checking practices. The third annotator worked under the supervision of one of the authors of this paper. Each was instructed to carefully review the FTC and ACCC guidelines before independently evaluating the data set of 2,600 ads. To avoid bias or mutual influence, the annotators worked independently, without access to each other's assessments. For each ad, they were asked to refer to the FTC and ACCC documents and classify the content as misinformation, not misinformation, or "cannot say" – the latter being used to denote ambiguous cases. Therefore, annotators have three classifications available: "Not misinformation", "Ambiguous", and "Misinformation". The "Ambiguous" class represents a conservative approach to fact-checking. Rather than forcing a binary label "Misinformation" or "Not misinformation" on every ad, a third class acknowledges uncertainty and avoids potentially misclassifying ads that, even for human annotators, are genuinely ambiguous or lack sufficient context for definitive classification. This cautious approach prioritizes avoiding false positives (labeling something as misinformation when it is not), which we deemed crucial in the sensitive domain of health information.

The final label for each ad was determined by taking the statistical mode of the three independent evaluations. For example, if two annotators classified an ad as "Not misinformation" and one as "Misinformation", the ad was labeled as "Not misinformation" based on the majority decision. In cases where the three annotators assigned different labels, the ad was marked as "N/A" and excluded from the CT-BERT training data set. This occurred in only five out of the 2,600 ads.

Examples of ads labeled as misinformation, not misinformation, and ambiguous are presented in Table 1 below.

4.2. CT-BERT algorithm training

Following the labeling process described in Section 4.1, we observed a significantly skewed distribution of the labels, with "Not misinformation" being the most represented label. As a result, the data was not directly fed into the model in its original form. The class unbalance is an important problem in the training set, as this may lead the model to predict only the popular class. We therefore address this issue with two combined approaches.

First, we augment the data in the "Misinformation" class by generating 400 synthetic examples (Meng et al., 2022; Bussotti et al., 2023). We generate such examples using ChatGPT (OpenAI, 2024) with two methods based on "in context" learning. The first exploits the availability of misinformation definitions discussed above (from ACCC and FTC). We include in the LLM prompt the definition of misinforming text to steer the generation (Peskine et al., 2023a). A second method is also based on prompting LLMs, but instead of descriptions, we use human-labeled misinformation examples following the "few shots" approach (Wang et al., 2020). The newly generated data set is then evaluated by humans. Ultimately, only 250 examples are considered high quality and added to the training data. All details are available in Appendix A.

Second, at training time, we also optimize the process by giving more importance to human data and taking into account class imbalance with a custom loss function. A loss function is a mathematical method used to measure how well the model's predictions match the actual known values, helping to guide the improvement of the model during training. Examples of an under-represented class, such as the "Misinformation" class, are given more importance individually during the training step, so that the model is more likely to recognize them, even if they are rare. The higher importance given to ads of such a class permits obtaining better predictions for the underrepresented classes. Additional information on the process is reported in Appendix B.

4.3. Results of the fine-tuning for CT-BERT

The model we use for the final annotation of the unlabeled texts is trained on 2,100 manually annotated examples. To assess the quality of the model, we keep 500 ads of the original set of 2,600 human-annotated examples out of the training for testing. We also include the 250 synthetic ads for the "Misinformation" class in the training set.

Table 2 reports the results of the model evaluation. In general, the model can effectively predict both "Misinformation" and "Not misinformation" classes, with an overall precision of 0.923. This number is obtained by computing the average of the F1 scores available in Table 2 between the "Not misinformation" and "Misinformation" classes. At inference time, 76.3 % of the original *Misinformation* claims are either labeled as *Misinformation* or *Ambiguous*, confirming the quality of the predictions of the model.

Our model achieves robust F1 scores for both the *Not misinformation* (0.9241) and *Misinformation* (0.9211) labels. This indicates that when the model expresses a definitive classification, it does so with a high degree of accuracy. The *Ambiguous* label primarily reflects cases where the model identifies uncertainty, rather than widespread misclassification across classes. This label, by definition, encompasses ads with nuanced, unclear, or context-dependent claims. Identifying and consistently classifying such ambiguous content is a known challenge in NLP, even for human annotators, and is particularly difficult for automated models that rely on explicit textual features.

Overall, the analysis of the deep learning model indicates that our dataset includes 0.2 % of ads that contain misinformation and 15.41 % of ads that are ambiguous. It should be noted that the 0.2 % of ads classified as "Misinformation" correspond to content that was clearly assessed as being false or misleading. We deliberately avoid referring to this content as "disinformation", as establishing intentionality is extremely difficult in our setting. Additionally, the 15.41 % of ads classified as "Ambiguous" do not necessarily constitute misinformation; rather, they reflect cases where insufficient context or uncertainty prevented a definitive classification. This conservative labeling strategy was adopted to reduce the risk of misclassification and ensure robustness in our analysis.

This low percentage suggests that AI algorithms are likely to filter ads containing misinformation on social media platforms. Results of the classification are available in Table 7 in Appendix C.

5. Descriptive evidence

Our objective is to investigate the correlation between algorithmic recommendation of ads and socioeconomic characteristics of US states.

[18] See the definition: https://www.accc.gov.au/consumers/advertising-and-promotions/false-or-misleading-claims, February 2024.

[19] https://www.ftc.gov/business-guidance/blog/2021/04/advertisers-stop-unproven-covid-claims-or-face-penalties-under-new-law, February 2024.

Algorithmic recommendation of ads is proxied by the number of impressions, defined as the number of times an ad is displayed to users. We compute the average number of impressions by state and by ad category, i.e., health-related or not, and classification made by the fact-checking algorithm, i.e., not misinformation, ambiguous, or misinformation. Therefore, we end up with 50 observations corresponding to the 50 US states, with the District of Columbia not being available in the data. Table 3 presents the descriptive statistics of the main variables used in the empirical analysis. In addition to the average number of impressions per ad types and state characteristics, the data includes for each state, the population, GDP per capita, percentage of people without health insurance, total cases of COVID-19, and total number of COVID-19-related deaths.

Overall, we observe a significantly higher number of impressions for health-related compared to non-health-related ads. While we observe a higher number of impressions for ads classified as not containing misinformation or as ambiguous among non-health-related ads, the recommendation pattern is different for health-related ads. More specifically, health-related ads classified as misinformation are more likely to be recommended than other types of ads. This is in line with previous work that shows that misinformation proliferates faster and broadly on social media platforms (Vosoughi et al., 2018). This first evidence raises concerns about the equality of user access to health information in online ads.

The data, summarized in Table 8 in Appendix D, shows that California has the highest impressions share, while Wyoming has the lowest, which aligns with the population size of these states.[20] To account for population differences, we normalize the data by dividing the number of impressions in each state by its population, giving us a ratio of impressions per capita. Since our primary interest is in the algorithmic recommendation of health-related ads, we focus on this subsample.

Figs. 2–4 show the ratio of impressions per capita for health-related ads classified respectively as "Not misinformation", "Ambiguous", and "Misinformation". Fig. 2 shows that the ratio of impressions per capita is higher in states located on the East and West coasts of the US when health-related ads are classified as "not misinformation". Figs. 3 and 4 show that the ratio of impressions per capita for health-related ads classified as "ambiguous" and "misinformation" are higher for states in the Mideast and in the Midwest where the GDP per capita is also known to be the lowest.[21]

6. Empirical analysis

The empirical evidence shown in Figs. 2–4 shed light on the disparities in terms of the ratio of impressions per capita by ad categories. Rather than establishing causal relationships, we aim to uncover consistent patterns in ad recommendation and key covariates. In this section, we examine how algorithmic recommendation of ads correlates with the socioeconomic characteristics of US states.

Our approach relies primarily on ordinary least squares (OLS) estimates. We use as a dependent variable the variable $Impressions_i$ that measures the average number of times the ad is displayed in a given state i. The equation we estimate is as follows:

$$Impressions_i = \beta_0 + \beta_1 X_i + \epsilon_i. \quad (1)$$

We have three main X_i explanatory variables: 1) *High GDP States* which is a dummy variable and takes the value 1 if the GDP per capita in a given state is above the median value equals to 58,007.85 dollars and 0 otherwise, 2) *% People w/o Health Insurance* which is a continuous variable indicating the percentage of people in a given state without health insurance, 3) *States with High COVID-19 Cases* which takes the value 1 if the total number of COVID-19 cases in a given state is above the median value equals to 7,447,673 and 0 otherwise. We study the correlation between the average number of impressions and each explanatory variable separately due to the small sample size.[22] ϵ_i indicates the error term.

6.1. Is there a link between misinformation ad recommendation and GDP per capita?

Table 4 investigates the correlation between the average number of impressions and the GDP per capita at the US state level. As mentioned previously, the dummy variable *High GDP States* takes the value of 1 if the GDP per capita in a given state is above the median value of 58,007.85 dollars and 0 otherwise. We split the sample between non-health-related ads (columns (1) to (3)) and health-related ads (columns (4) to (6)).

Columns (1) and (4) of Table 4 show that, whether health-related or not, there is no significant correlation between the average number of impressions and states with high GDP per capita for ads classified as not misinformation. However, columns (2) and (3) show a different pattern. There is a negative and significant correlation between the average number of impressions and states with high GDP per capita for non-health-related ads classified as ambiguous or as misinformation, respectively. This pattern holds true for the subset of health-related ads where column (6) shows a negative and significant correlation between the average number of impressions and states with high GDP per capita for ads classified as misinformation, although no correlation is observed for ads classified as ambiguous (column (5)). We find that the magnitude of the coefficient is larger for health-related ads (column (6)) compared to non-health-related ads (column (3)). Therefore, users living in states with high GDP per capita are less likely to be shown ambiguous or misinformation ads, particularly in health-related contexts, pointing to disparities in information access based on socioeconomic characteristics across US states.

Our results are robust regardless of the GDP measure chosen (see Section G.1 in Appendix G). The coefficients available in Table 11 show consistent results with Table 4, both in terms of signs and significance levels, when a continuous measure of GDP per capita (log-transformed) is considered. We further find a positive and statistically significant correlation between the average number of impressions and continuous GDP per capita for health-related ads classified as not misinformation. As state GDP per capita increases, users are more likely to be shown health-related ads not containing misinformation. Table 12 provides information on the monotonicity and direction of the above associations using GDP per capita quartiles. The results suggest a negative and significant correlation between US states in the top quartile of GDP per capita and algorithmic recommendation of ads classified as misinformation, especially when ads are health-related. Thus, users living in the wealthiest 25 % of states are less likely to be recommended health-related ads classified as misinformation compared to those in the poorest 25 % of states, suggesting the existence of a digital divide.

6.2. Are health-related ads with misleading claims displayed more in States with higher rates of uninsured individuals?

This section aims to investigate whether algorithmic ad recommendation, especially ads classified as misinformation or ambiguous, is likely to be correlated with the percentage of people without health insurance at the state level. Fig. 5 shows the percentage of people without health insurance by state on the left and health-related ad impressions by state on the right.

Overall, US states with a high percentage of people without health

[20] https://www.statsamerica.org/sip/rank_list.aspx?rank_label=pop1, February 2024.

[21] https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_GDP#/media/File:GDP_by_U.S._state.svg, March 2024.

[22] A correlation matrix between variables is reported in Section F.

insurance, which are represented by darker shades of orange in Fig. 5, are less exposed to health-related ads.

We used an empirical analysis to study the potential correlation between state health insurance coverage and algorithmic recommendation of ads across the different ad categories. Table 5 shows the results. The dependent variable is the average number of impressions at the state level. Columns (1) to (3) provide the estimates for the subsample of non-health-related ads, and columns (4) to (6) provide estimates for the subsample of health-related ads.

We observe for both non-health and health-related ads a negative and significant correlation between the average number of impressions and the percentage of people without health insurance for ads classified as not misinformation (columns (1) and (4)). The coefficient is larger for the subsample of health-related ads in column (4). In contrast, columns (3) and (6) show a positive and significant correlation between the average number of impressions and the percentage of people without health insurance for ads classified as misinformation. The coefficient is also larger for the health subsample in column (6). Therefore, as the percentage of uninsured individuals in a state increases, users living in this state are more likely to be shown misinformation ads, particularly in a health context. This suggests that the level of health insurance coverage in a state shapes the type of information accessible to users, thereby perpetuating social inequalities among individuals.

6.3. *Are advertising recommendations influenced by the scale of the health crisis?*

Given our objective to highlight the correlation between the socio-economic characteris-tics of US states and algorithmic recommendation of ads, we took into consideration the number of COVID-19 cases per state. We assume that this metric could affect the algorithmic recommendation of ads, as states more severely impacted by the pandemic may show a higher volume of health-related ads. We create a dummy variable *States with High COVID-19 Cases* which takes the value of 1 if the total number of cases of COVID-19 in a given state is greater than the median national value, equal to 7,447,673 cases, and 0 otherwise. Table 6 presents the results for the subsample of health-related ads. The dependent variable is the average number of impressions at the state level. Estimates are provided in column (1) for ads classified as not misinformation, in column (2) for ads classified as ambiguous, and in column (3) for ads classified as misinformation. We observe no correlation between states with a high number of COVID-19 cases and the algorithmic recommendation of health-related ads – whether classified as misinformation (column (3)) or not (column (1)). Only column (2) reveals a positive correlation between states with high COVID-19 case counts and the recommendation of health-related ads classified as ambiguous. This suggests that, apart from ambiguous ads, there is no correlation between algorithmic recommendation of health-related ads and COVID-19 cases at the US state level.

We run a series of robustness checks on the correlation between the average number of impressions and COVID-19 incidence at the state level. Results are available in Section G.2 of Appendix G. Table 13 replicates the specification of Table 6 using a continuous measure of the number of COVID-19 cases (log-transformed). Consistent with previous results, we find no association between the average number of impressions and the number of COVID-19 cases. The previously positive and significant correlation for the subset of ads classified as ambiguous no longer holds. Table 14 provides additional information on direction and monotonicity of the relationship (or lack thereof) using quartiles of COVID-19 cases. Column (1) shows consistent results for ads classified as not misinformation. Column (2) shows no difference in algorithmic ad recommendations between US states in the top and bottom quartiles of COVID-19 cases for the subsample of ambiguous ads, even if states in the second quartile of COVID-19 cases, designated as states with moderated COVID incidence, are more likely to be shown ambiguous ads than states with low COVID incidence. Column (3) similarly shows no difference in algorithmic ad recommendations between US states in the top and bottom quartiles of COVID-19 cases for the subsample of misinformation ads, but states in the third quartile, designated as elevated COVID incidence were shown more ads containing misinformation. Alternatively, because the number of COVID-19 cases may imperfectly capture COVID-19 incidence, we also use COVID-19-related deaths as an alternative proxy. Results, presented in Table 15, are aligned with previous findings where no association can be established between the average number of impressions and COVID-19 incidence at the state level.

7. Discussion and implication

Social media platforms are widely used worldwide as a primary source of information. Given the proliferation of misinformation on these platforms, policymakers and regulators urge social media to improve the management of content available on their platforms. The widespread dissemination of online misinformation poses challenges to the integrity of various markets, including media, cybersecurity, and social media. In particular, promoting equality in accessing reliable health-related information is crucial from a public health perspective. However, given the increased role of recommendation systems in matching ad content to users, little is known about how recommendation systems contribute to the digital divide when it comes to health-related information on social media platforms.

Our paper aims to bridge the gap between the use of AI algorithms, especially recommendation systems, and progress towards achieving the SDGs related to reducing inequality (SDG 10) and improving health and well-being (SDG 3). Through the construction of an innovative fact-checking algorithm and the analysis of ad-related data collected from the Meta ad library, we provide evidence that the use of AI algorithms can contribute to expanding the digital divide, which conflicts with the achievement of the SDGs (SDG 10). Even if our findings show that very few ads published on Meta's platform include misinformation and around 15 % have been classified as ambiguous, we provide evidence that exposure to ads managed by recommendation systems is uneven. This disparity, linked to the socioeconomic characteristics of US states, prevents equal access to information, especially when health-related, as stated in SDG3.

These findings have significant implications, both nationally and internationally. Countries with high percentages of people without health insurance can experience exacerbated inequalities, given that the uninsured population often comprises working-age adults with lower education and income levels. In addition, such implications extend to countries with high income and wealth inequality indices. For example, India has seen a surge in misinformation queries during the pandemic, ranked highest in the risk of disinformation and misinformation dissemination.[23] This has contributed to a widening global digital divide between countries, with a new challenge emerging in the form of the use of health misinformation for geopolitical purposes.[24] Therefore, companies, especially social media platforms, need to strengthen their efforts when it comes to health-related ads recommended by AI algorithms, where the relationship between their use and inequality to information access has been observed.

Currently, understanding the role of social media platform structures in the diffusion of misinformation remains imperative (Zhuravskaya et al., 2020; Acemoglu et al., 2024). Our research reinforces this necessity, emphasizing the importance of algorithmic transparency and accountability. This also aligned with the increasing regulatory pressure on social media platforms to regulate the ad content they distribute.

[23] https://www.who.int/images/default-source/digital-health/google-data-insights.jpg?sfvrsn=16b5b112_5, https://www.statista.com/chart/31605/rank-of-misinformation-disinformation-among-selected-countries/, April 2024.

[24] https://www.europarl.europa.eu/RegData/etudes/ATAG/2020/649369/EPRS_ATA(2020)649369_EN.pdf, April 2024.

While the Digital Service Act represents an initial step, regulations such as the AI Act further enhance algorithmic transparency. From a practical perspective, regulatory bodies could help social media platforms by providing an ad-focused fact-checking tool to scrutinize algorithmic behavior on social media. Our fact-checking tool directly addresses this market need, as it allows ad claims to be checked for validity related to health issues. The main difference from previously built tools is that it is tailored to social media ads, which was not the case before, and for which growing concerns are expressed (Rao, 2022; Fong et al., 2024). Additionally, social media platforms could implement strategies such as detecting and flagging disproportionate ad impressions given a set of predetermined attributes, using health-related ad content run by reputable organizations to assess the veracity of other contents on a similar topic, standardizing geotargeting practices in crisis context, implementing a public service advertising mode that equally reach users when it comes to health, and expanding user feedback mechanisms to combat misinformation effectively. These measures collectively aim to improve transparency, combat misinformation, and foster a safer online environment. Finally, the generalizability of our findings beyond the United States should be interpreted with caution. Different countries have varying regulatory environments, health communication policies, and platform governance standards, which can significantly affect both the creation and the dissemination of misinformation. As such, the patterns observed in the United States may not directly translate to other national contexts. Future research should replicate and extend this analysis in different countries to better understand how local regulations and socio-political factors shape algorithmic ad recommendation and exposure to misinformation.

8. Conclusion and limitation

We collected data from the Meta ad library on ads displayed in all US states. To ensure the reliability of our analysis, we collected data over six months from January to June 2020. Online advertising on social media platforms is highly dynamic and sensitive to external events such as major holidays (e.g., Black Friday) or political campaigns, which can cause significant shifts in advertising strategies and user engagement. By selecting a shorter data collection window, we aimed to capture a more stable and representative period of typical platform activity, minimizing the risk of co-occurring events introducing bias into the analysis. Although this choice limits the long-term generalizability of our findings, it allows for a more controlled examination of recommendation systems behavior. Future research could extend the time frame to confirm the persistence of the observed patterns over longer periods. We augmented ad-related data with US state level administrative data, including GDP per capita, percentage of individuals without health insurance, and numbers of COVID-19 cases and deaths in a given state in 2020. To evaluate the platform's ability to detect misinformation in ads, we developed an innovative, fine-tuned fact-checking algorithm based on a deep learning model trained on human-annotated data and synthetic data generated by means of LLM, which we applied to textual content available in the ads.

Our fact-checking tool showed that only a small fraction of the ads (0.2 %) in the sample were classified as misinformation, even if around 15 % were classified as ambiguous. This suggests that the platform's curation is effective in terms of assessing the content available on the platform. However, we found evidence of a digital divide in the algorithmic ad recommendation. We find a negative and significant correlation between the average number of impressions and states with high GDP per capita for ads classified as ambiguous or as misinformation, particularly those related to health. Conversely, there is a positive and significant correlation between the average number of impressions and the percentage of individuals without health insurance for health-related misinformation ads. Although no strong correlation is identified between COVID-19 incidence and algorithmic recommendation of health-related ads, these findings overall suggest disparities in information access based on socioeconomic characteristics of US states. In general, users in wealthier states and in states with lower rates of uninsured individuals are more likely to be exposed to higher-quality information, as algorithmic recommendations tend to not recommend misinformation ads in an health context.

Our study has some limitations. First, the six-month time interval may not be enough to account for potential learning and adjustments by AI algorithms over time, even though it helps control for co-occurring events. Second, despite our efforts to rectify class imbalances in our fact-checking tool, we acknowledge the potential for false positives. As is the case when studying any form of AI algorithms, distortions can occur due in particular to unrepresentative data. Training the algorithm on synthetic data could lead to bias amplification, which we initially used to address the class imbalance in the training data, being one of the major sources of bias replication. Bias amplification occurs if the synthetic data inherits biases in the original training data. However, the use of high-quality synthetic data can produce economies of scale by reducing the volume of human-generated data needed for training, with direct cost implications for businesses. Another limitation concerns cross-sectional data aggregated by states which do not take into account psychological or behavioral factors of individuals. Individuals in lower GDP per capita states may exhibit higher levels of skepticism toward expert information sources, making them more prone to engage with alternative, potentially misleading content. As such, our results may reflect not only algorithmic recommendation patterns, but also the algorithmic feedback loop. Future research would be valuable to disentangle these effects. Furthermore, while our analysis focuses on socioeconomic characteristics of US states, we did not control for political orientation at the state level, which previous studies (Guess et al., 2019; Grinberg et al., 2019) have shown to be strongly associated with consumption of misinformation. Future research should integrate political orientation measures to provide a more nuanced understanding of algorithmic ad recommendations and user interactions.

Despite these limitations, our work should be helpful to policymakers and help to debunk misleading claims. It also highlights the need for social media platforms to actively address and mitigate the inequalities that are generated by recommendation systems in the advertising context.

CRediT authorship contribution statement

Jean Clara: Writing – review & editing, Writing – original draft, Validation, Software, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Bussotti Jean-Flavien:** Writing – review & editing, Writing – original draft, Validation, Software, Data curation. **Cecere Grazia:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Omrani Nessrine:** Writing – original draft, Project administration, Conceptualization. **Papotti Paolo:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Methodology, Investigation, Data curation, Conceptualization.

Declaration of interest statement

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Acknowledgments

We thank the participants of the Digital Transformation Society 2024 conference, held in Naples, for their valuable comments and suggestions. We also thank the anonymous reviewers and the guest editorial team for their constructive comments and guidance during the revision of this manuscript, which helped improve its quality.

I'm building a fake news detection model. For this, I need misinformation claims. Given the following rules, can you write me some claims that would be classi-fied as misinformation? The claims of product and service must be true, sub-stantiated, and include accurate information on price, quality, and benefits. - Misrepresentations, withholding key information, or making false origin claims are illegal. -Exaggerations are often permissible, but objectively false claims, especially about prices and comparative advantages, are not. -COVID-19-related advertising must avoid false or unproven treatment claims; comparative claims require accurate information. -Claims about product quality, safety, and environmental impact must be factual and specific. -All claims, including those about COVID-19 prevention or treatment, need credible scientific support and transparent pricing. -The FTC enforces against deceptive COVID-19 claims; platforms monitor and penalize such misinformation.

Here is a list of covid misinformation:

['whoa … we just ignore that aliens are real?! watch the first episode of 'real america' with graham allen!in turning point usa's brand new series, graham cuts through the fake news & tackles the biggest stories of the week. join us on youtube & facebook every thursday at 4 est! #realamerica', 'selinexor killed the virus in a Petri dish. next steps?', "the false chinese government propaganda against our president has gone too far. as usual, china is trying to brainwash the rest of the world while throw-ing america under the bus.it's time to put an end to their communist propaganda once and for all, but we can't do it without you. president trump is calling on every american to step up and defend their country against the chinese communist party lies!please sign our official petition demanding an end to the chinese government propaganda against our president!", "some claim it could cure covid-19. here's the story of the controversy surrounding the antiviral drug avigan in italy.", […]]

Can you add other items to the list. They should be written in a similar style.

Your task is to generate short sentences that contain misinformaion techniques in order to train a fact-checking model. The definition of a misinformation technique is the following: The claims of products and services must be true, substantiated, and include accurate information on price, quality, and benefits. Misrepresentations, withholding key information, or making false origin claims are illegal. Exagger-ations are often permissible, but objectively false claims, especially about prices and comparative advantages, are not. COVID-19-related advertising must avoid false or unproven treatment claims; comparative claims require accurate informa-tion. Claims about product quality, safety, and environmental impact must be factual and specific. All claims, including those about the prevention or treatment of COVID-19, need credible scientific support and transparent pricing. The FTC enforces against deceptive COVID-19 claims; platforms monitor and penalize such misinformation.

Here are some examples: 'have you seen more people wearing face masks in re-sponse to the coronavirus? turns out, they might not be as effective as they think.', […]

Please generate sentences that contain the misinformation technique, as detailed above, similar to the examples on similar topics.

This work has been supported in part by the ANR project ATTENTION (ANR-21-CE23-0037), CARNOT TSN, Maison des Sciences de l'Homme Paris-Saclay under grant 22-EM-06 and the Digital Organizations & Society chair of Grenoble Ecole de Management.

Appendix A. Prompts for Misinformation Example Generation

We report the three prompts that we used to generate additional training examples (labeled as "Misinformation") to fine-tune the fact-checking model.

In the first prompt, we give the model guidelines for creating claims. These guidelines are sourced from the FTC and ACCC. Any examples are given to the model. Therefore, we ask the model to generate claims that violate at least one of the guidelines provided.

Each claim should break one or multiple rules above. The claim should not be too obvious. For example, avoid writing in the claim that there is no evidence that this is true. Also, the claim should not contain the evidence that proves it is misinformation.

In the second prompt, we only provide the model with a small list of ads from our dataset and an instruction on what to do. The task is to supplement the list with additional items.

The third and last prompt is a combination of the two previous prompts: We provide both models and rules to generate claims.

Appendix B. Technical Details for the Classification Model

To use the CT-BERT classification model in our setting, we need to adjust some of its functionalities and parameters. In the training process, we give more importance to human data with respect to ChatGPT-generated data, and we also take into consideration class unbalance. For this goal, we use a custom loss function to train the model. The custom loss takes into consideration the percentage of each label to compute the loss. In our case, it gives higher importance to texts with the "Misinformation" label, as they are rare in the dataset. The custom loss also gives human training examples a coefficient two times superior to the one for generated examples. This choice reflects the intuition that the original examples are more representative

of reality, thus it is preferable to privilege them in the training.

We run the training and inference of our CT-BERT model on a cluster of OS Linux workstations.[25] The experiments were run on a single graphics processing unit (GPU), a NVIDIA TITAN Xp that has a Video Random Access Memory (VRAM) capacity of 12 GB. This is a low-resource setting that allows most people to run this model without needing a high-end GPU. The fact-checking models rely on PyTorch (Paszke et al., 2019) and on the HuggingFace libraries (Wolf et al., 2019). To train the CT-BERT model, we set the following hyperparameters: 10 epochs, a batch size for training and evaluation of 8. We define the learning rate to be $5e^{-5}$.

Appendix C. Predictions of the Fact-Checking Algorithm

Table 7 shows the predictions of the fact-checking algorithm. Overall, 84.4 % of ads were classified as not misinformation, suggesting that social media platforms are capable of ensuring a largely safe online environment. However, 15.41 % of ads remain ambiguous, either due to a lack of context or a lack of information, and nearly 2 % were classified as containing misinformation. Although the percentage of ads containing misinformation is low, it should be noted that these ads have a significantly higher number of impressions, particularly when it comes to health, and can reach and spread more widely than truthful information, raising concerns from the perspective of access to information by users of social media platforms.

Table 7
Results of the CT-BERT prediction at ad level

Prediction	Percent	Cum. Percent
Not misinformation	84.40	84.40
Ambiguous	15.41	99.81
Misinformation	0.19	100

Appendix D. Impressions Share by US States

As mentioned previously, we find a positive correlation between impressions share and population size of US states. For example, 10.5 % of impressions were shown in California, while only 2 % were shown in Wyoming. According to the population estimates from 2024, Wyoming has the lowest population, while California has the highest, which justifies our strategy of dividing the average number of impressions by the state's population to obtain a ratio of impressions per capita.[26]

Table 8
Impressions share by US states

Variable	Mean	Std. Dev.	N
Wyoming	0.002	0.001	145,272
North Dakota	0.002	0.003	145,272
Vermont	0.003	0.002	145,272
South Dakota	0.003	0.002	145,272
Delaware	0.003	0.004	145,272
Alaska	0.003	0.007	145,272
Rhode Island	0.003	0.002	145,272
Hawaii	0.004	0.003	145,272
Montana	0.004	0.002	145,272
New Hampshire	0.005	0.008	145,272
Idaho	0.006	0.003	145,272
Maine	0.006	0.012	145,272
Nebraska	0.006	0.006	145,272
West Virginia	0.007	0.006	145,272
Utah	0.007	0.007	145,272
New Mexico	0.007	0.01	145,272
Nevada	0.008	0.012	145,272
Mississippi	0.009	0.011	145,272
Kansas	0.009	0.007	145,272
Arkansas	0.01	0.007	145,272
Connecticut	0.01	0.005	145,272
Iowa	0.011	0.019	145,272
Oklahoma	0.012	0.01	145,272
Louisiana	0.013	0.008	145,272
South Carolina	0.014	0.015	145,272
Alabama	0.015	0.014	145,272
Kentucky	0.016	0.012	145,272
Oregon	0.017	0.011	145,272
Maryland	0.017	0.012	145,272
Minnesota	0.018	0.018	145,272
Colorado	0.018	0.02	145,272
Wisconsin	0.019	0.031	145,272

(*continued on next page*)

[25] OS stands for Operating Systems.

[26] https://www.statsamerica.org/sip/rank_list.aspx?rank_label=pop1, April 2021.

Table 8 (*continued*)

Variable	Mean	Std. Dev.	N
Missouri	0.02	0.012	145,272
Indiana	0.021	0.009	145,272
Tennessee	0.021	0.011	145,272
Arizona	0.022	0.036	145,272
Massachusetts	0.023	0.015	145,272
New Jersey	0.024	0.015	145,272
Virginia	0.026	0.016	145270
Washington	0.028	0.026	145268
Georgia	0.028	0.015	145,272
North Carolina	0.031	0.033	145,272
Michigan	0.037	0.04	145,272
Ohio	0.038	0.018	145,272
Illinois	0.038	0.02	145,272
Pennsylvania	0.042	0.036	145,272
Florida	0.064	0.035	145,272
New York	0.065	0.038	145,272
Texas	0.073	0.036	145,272
California	0.105	0.065	145,272

Appendix E. Further Empirical Analysis

In this section, we adopt an econometric approach at the ad level to study whether and how ambiguous and misinformation ads have been recommended according to individual demographic characteristics. Our approach relies on ordinary least squares (OLS) estimates. We use as a dependent variable the proportion of impressions displayed to a given age cohort for an ad i on a month t. Equation (2) captures our main econometric specification. Standard errors are clustered at the advertiser page level. All regressions include month-fixed effects. The equation we estimate is as follows:

$$Prop.Impressions_{it} = \beta_0 + \beta_1 Misinformation + \beta_2 Ambiguous + \lambda_t + \epsilon_i. \quad (2)$$

The binary variable *Misinformation* takes the value 1 if the ad i was classified as misinformation by the fact-checking algorithm. The variable *Ambiguous* takes the value 1 if the ad was classified as ambiguous, and ϵ is the error term. *Not misinformation* is used as the reference variable. Table 9 presents empirical estimates that measure the proportions of impressions by age groups, as indicated. Contrary to the other age groups, the result in column (1) suggests that users below the majority threshold are likely to see a higher proportion of impressions for ads classified as misinformation. This is a result that can only be observed for this age cohort.

Table 9
Proportion of ad display by age cohorts

	Display 13–17 (1)	Display 18–24 (2)	Display25-34 (3)	Display 35–44 (4)	Display 45–54 (5)	Display 55–64 (6)	Display over 64 (7)
Ambiguous	0.011	0.014	0.008	0.014*	0.004	−0.010	−0.042***
	(0.007)	(0.009)	(0.013)	(0.008)	(0.006)	(0.011)	(0.013)
Misinformation	0.112**	0.026	0.015	0.069	−0.053	−0.066	−0.104
	(0.047)	(0.032)	(0.035)	(0.058)	(0.043)	(0.060)	(0.065)
Constant	0.003*	0.100***	0.143***	0.118***	0.138***	0.214***	0.283***
	(0.002)	(0.015)	(0.013)	(0.006)	(0.005)	(0.010)	(0.022)
R-squared	0.016	0.006	0.007	0.011	0.006	0.007	0.015
Observations	145,272	145,272	145,272	145,272	145,272	145,272	145,272

Notes: OLS estimations. "Not misinformation" is used as the reference variable. Errors are clustered at the advertiser page level. Significance at 1 %; 5 % and 10 % levels indicated respectively by ***,** and *.

Appendix F. Correlation Matrix

Table 10 presents the pairwise correlations between three key state-level variables: whether a state has a high GDP, the percentage of people without health insurance, and whether the state had a high number of COVID-19 cases. We observe a negative correlation between being a high-GDP state and the percentage of people without health insurance. This suggests that wealthier states tend to have slightly lower percentages of uninsured individuals. Although there is no meaningful correlation between being a high-GDP state and having a high number of COVID-19 cases, there is a very weak positive correlation between the percentage of people without health insurance and states with high numbers of COVID-19 cases.

Table 10
Cross-correlation table

Variables	High GDP States	% of People w/o Health Insurance	States with High COVID-19 Cases
High GDP States	1.000		
% of People w/o Health Insurance	−0.266 (0.061)	1.000	
States with High COVID-19 Cases	−0.040 (0.783)	0.108 (0.456)	1.000

Appendix G. Robustness Checks

G.1 Continuous Measure of GDP Per Capita

Our main specification in Table 4, which examines the correlation between the average number of impressions and GDP per capita, uses a binary variable as a proxy, namely *High GDP States*. As a complementary analysis, we include in this section a similar specification using a continuous variable of GDP per capita. Table 11 below replicates Table 4 using the continuous variable *GDP per Capita*. Since the distribution of this variable is skewed, we use a logarithmic transformation. In order to control for the scale effect whereby states with larger populations might receive more impressions, we include the variable *Population* as a control variable. To ensure that there were no multicollinearity issues between the two independent variables, a VIF estimation was performed, yielding a score of 1.09. The dependent variable is the average number of impressions. We divided the sample between non-health-related ads (columns (1) to (3)) and health-related ads (columns (4) to (6)).

In line with results from Table 4, column (1) shows that there is no significant correlation between the average number of impressions and GDP per capita for non-health ads classified as not misinformation. Coefficients available in columns (2), (3), (5), and (6) are aligned with what was previously shown in Table 4 both in terms of significance and direction. However, unlike before, column (4), where no correlation had previously been established, shows a positive and significant correlation at 5 % between GDP per capita and the average number of impressions of health-related ads classified as not mis-information. This reinforces our previous findings, as a state's GDP per capita increases, users are more likely to be shown health-related ads that do not contain misinformation, widening the digital divide.

Overall, the use of the continuous variable *GDP per Capita* corroborates our previous findings and even reinforces the fact that users in wealthier states are less likely to be shown misinformation ads, particularly in the context of health.

Table 11
Ad recommendation and continuous measure of GDP per capita

	Non-Health-related Ads			Health-related Ads		
	Not Misinformation	Ambiguous	Misinformation	Not Misinformation	Ambiguous	Misinformation
	(1)	(2)	(3)	(4)	(5)	(6)
Log (GDP per Capita)	94.143	−304.739**	−242.262***	666.473**	−297.776	−1453.349***
	(114.724)	(116.349)	(49.724)	(293.117)	(226.881)	(362.415)
Constant	−952.352	3461.315***	2707.081***	−7268.784**	3420.605	16271.234***
	(1246.717)	(1290.066)	(546.734)	(3200.896)	(2514.174)	(3956.644)
R-squared	0.979	0.953	0.980	0.937	0.819	0.957
Observations	50	50	50	50	50	50
Population	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS estimates. The dependent variable is the average number of impressions displayed in a given state. Standard errors are clustered at the state level. We include the state population. Significance at 1 %; 5 % and 10 % levels indicated respectively by ***,** and *. VIF score = 1.09.

To provide more details on the direction and monotonicity of these associations, Table 12 presents the same specification as Table 11, with the variable *GDP* per capita (in logarithms) divided into quartiles. The reference category corresponds to users living in the poorest 25 % of states, which includes, for example, users in the state of Alabama.

First, the results from columns (1), (2), and (5) indicate that the sign of the coefficients changes, suggesting a non-monotonic relationship between *GDP per Capita* and the average number of impressions; the coefficients decrease as GDP per capita increases for ads classified as ambiguous, whether health-related or not.

Second, users living in the wealthiest 25 % states (e.g., California) are less likely to be shown health-related ads classified as ambiguous or as misinformation (columns (2) and (3)). A similar pattern is identified for health-related ads classified as misinformation (column (6)). In contrast, column (4) indicates that users living in the wealthiest 25 % states are more often exposed to health-related ads that do not contain misinformation, compared to those in the poorest 25 % states. Overall, these findings align with previous results and reinforce our main argument.

Table 12
Ad recommendation and GDP per capita quartiles

	Non-Health-related Ads			Health-related Ads		
	Not Misinformation	Ambiguous	Misinformation	Not Misinformation	Ambiguous	Misinformation
	(1)	(2)	(3)	(4)	(5)	(6)
Moderated GDP per Capita	36.235	123.563	−11.719	161.760	365.433*	−0.059
	(46.525)	(87.911)	(22.742)	(106.303)	(182.509)	(165.315)
Elevated GDP per Capita	−13.035	4.665	−15.201	48.212	127.440	−17.380
	(46.004)	(54.993)	(23.693)	(84.164)	(100.101)	(182.967)
High GDP per Capita	35.422	−113.045***	−110.183***	311.724***	−46.324	−569.093***
	(44.716)	(36.252)	(24.437)	(112.565)	(59.715)	(188.518)
Constant	63.765**	118.333***	86.219***	−95.494	49.120	496.548***
	(27.328)	(34.562)	(17.757)	(58.565)	(55.240)	(115.825)
R-squared	0.979	0.958	0.981	0.937	0.844	0.956
Observations	50	50	50	50	50	50
Population	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS estimates. The dependent variable is the average number of impressions displayed in a given state. Standard errors are clustered at the state level. We include the state population. Significance at 1 %; 5 % and 10 % levels indicated respectively by ***,** and *.

G.2 Continuous Measures of COVID-19 Cases and Deaths

Our specification in Table 6 examines the correlation between the average number of impressions and the number of COVID-19 cases at the US state level, using a binary variable to indicate COVID-19 incidence in a given state. As a complementary analysis, we estimate a specification similar to Table 6 using a continuous measure of COVID-19 cases. Specifically, we calculate the COVID-19 incidence per 100,000 inhabitants, defined as the number of COVID-19 cases in a given state divided by the state population and multiplied by 100,000. Because the distribution of this variable is skewed, we apply a logarithmic transformation. The variable *Population* is included as a control. The results, presented in Table 13, are overall consistent with previous findings where no correlation between COVID-19 incidence and the algorithmic recommendation of health-related ads can be observed.

Table 13
COVID-19 incidence does not shape algorithmic ad display

	Health-related Ads		
	Not Misinformation	Ambiguous	Misinformation
	(1)	(2)	(3)
Log (COVID-19 incidence per 100,000)	−96.934	128.575	152.719
	(59.953)	(104.527)	(177.225)
Constant	1167.497	−1365.993	−1425.485
	(730.071)	(1212.705)	(2096.859)
R-squared	0.928	0.817	0.945
Observations	50	50	50
Population	Yes	Yes	Yes

Notes: OLS estimates. The dependent variable is the average number of impressions displayed in a given state. Standard errors are clustered at the state level. We include the state population. Significance at 1 %; 5 % and 10 % levels indicated respectively by ***,** and *.

As in the previous section, Table 14 provides more details on the direction and monotonicity of these (or lack thereof) associations, using the COVID-19 cases variable (transformed in logarithms) divided into quartiles. The reference category corresponds to the 25 % of states with the fewest COVID-19 cases, designated as low COVID incidence states, such as the state of Colorado. The results of column (2) show no difference in algorithmic ad recommendations between US states in the top and bottom quartiles of COVID-19 cases for the subsample of ambiguous ads, even if states in the second quartile of COVID-19 cases, designated as states with moderated COVID incidence, are more likely to be shown ambiguous ads than states with low COVID incidence. Column (3) similarly shows no difference in algorithmic ad recommendations between US states in the top and bottom quartiles of COVID-19 cases for the subsample of misinformation ads, but states in the third quartile, designated as elevated COVID incidence, were shown more ads containing misinformation. Therefore, we retrieve our results from Table 6.

Table 14
Quartile analysis of COVID-19 cases on algorithmic ad recommendations

	Health-related Ads		
	Not Misinformation	Ambiguous	Misinformation
	(1)	(2)	(3)
Moderated COVID incidence	152.097	418.163**	254.414
	(136.688)	(194.128)	(204.045)
Elevated COVID incidence	−154.210	246.747*	421.738**
	(93.672)	(129.470)	(186.135)
High COVID incidence	−49.585	121.660	136.265
	(52.715)	(78.247)	(182.377)
Constant	31.522	−0.750	215.680*
	(64.133)	(30.578)	(126.093)
R-squared	0.936	0.842	0.949
Observations	50	50	50
Population	Yes	Yes	Yes

Notes: OLS estimates. The dependent variable is the average number of impressions displayed in a given state. Standard errors are clustered at the state level. We include the state population. Significance at 1 %; 5 % and 10 % levels indicated respectively by ***,** and *.

Since COVID-19 cases may be an imperfect proxy for COVID incidence, we also consider the number of COVID-19-related deaths per month and per state as an alternative measure. Table 15 presents the results of the same specification as Table 13, but using this new variable. Overall, the results are consistent with those in Table 13, showing no correlation between COVID-19-related deaths and the algorithmic recommendation of health-related ads.

Table 15
COVID-19 deaths do not shape algorithmic ad recommendations

	Health-related Ads		
	Not Misinformation	Ambiguous	Misinformation
	(1)	(2)	(3)
Log (Deaths per Capita)	3777.708	11144.305	4042.316
	(5333.535)	(7490.792)	(10083.455)
Constant	−61.275	−44.973	319.380*
	(115.876)	(119.973)	(181.086)
R-squared	0.928	0.823	0.945
Observations	50	50	50
Population	Yes	Yes	Yes

Notes: OLS estimates. The dependent variable is the average number of impressions displayed in a given state. Standard errors are clustered at the state level. We include the state population. Significance at 1 %; 5 % and 10 % levels indicated respectively by ***,** and *.

Data availability

Data will be made available on request.

References

Abrardi, L., Cambini, C., Rondi, L., 2022. Artificial intelligence, firms and consumer behavior: a survey. J. Econ. Surv. 36 (4), 969–991.

Acemoglu, D., 2021. Harms of AI. National Bureau of Economic Research. Technical report.

Acemoglu, D., Ozdaglar, A., Siderius, J., 2024. A model of online misinformation. Rev. Econ. Stud. 91 (6), 3117–3150.

Acemoglu, D., Restrepo, P., 2018. The race between man and machine: implications of technology for growth, factor shares, and employment. Am. Econ. Rev. 108 (6), 1488–1542.

Adams, Z., Osman, M., Bechlivanidis, C., Meder, B., 2023. (Why) is misinformation a problem? Perspect. Psychol. Sci. 18 (6), 1436–1463.

Agarwal, R., Dhar, V., 2014. Editorial—big data, data science, and analytics: the opportunity and challenge for IS research. Inf. Syst. Res. 25 (3), 443–448.

Agrawal, A., Gans, J., Goldfarb, A., 2022. Prediction Machines, Updated and Expanded: The Simple Economics of Artificial Intelligence. Harvard Business Press.

Aguiar, L., Waldfogel, J., 2018. Platforms, Promotion, and Product Discovery: Evidence from Spotify Playlists. National Bureau of Economic Research. Technical report.

Allcott, H., Braghieri, L., Eichmeyer, S., Gentzkow, M., 2020. The welfare effects of social media. Am. Econ. Rev. 110 (3), 629–676.

Allcott, H., Gentzkow, M., 2017. Social media and fake news in the 2016 election. J. Econ. Perspect. 31 (2), 211–236.

Allcott, H., Gentzkow, M., Yu, C., 2019. Trends in the diffusion of misinformation on social media. Res. Politics 6 (2), 2053168019848554.

Amazeen, M.A., 2013. A Critical Assessment of fact-checking in 2012. New America Foundation, pp. 1–40.

Amazeen, M.A., 2016. Checking the fact-checkers in 2008: predicting political Ad scrutiny and assessing consistency. J. Polit. Market. 15 (4), 433–464.

Amazeen, M.A., 2019. Practitioner perceptions: critical junctures and the global emergence and challenges of fact-checking. Int. Commun. Gaz. 81 (6–8), 541–561.

Amazeen, M.A., 2020. Journalistic interventions: the structural factors affecting the global emergence of fact-checking. Journalism 21 (1), 95–111.

Aridor, G., Gonçalves, D., Kluver, D., Kong, R., Konstan, J., 2022. The informational role of online recommendations: evidence from a field experiment. arXiv e-prints, arXiv–2211.

Arora, S.D., Singh, G.P., Chakraborty, A., Maity, M., 2022. Polarization and social media: a systematic review and research agenda. Technol. Forecast. Soc. Change 183, 121942.

Azzimonti, M., Fernandes, M., 2023. Social media networks, fake news, and polarization. Eur. J. Polit. Econ. 76, 102256.

Bahoo, S., Cucculelli, M., Qamar, D., 2023. Artificial intelligence and corporate innovation: a review and research agenda. Technol. Forecast. Soc. Change 188, 122264.

Bakshy, E., Messing, S., Adamic, L.A., 2015. Exposure to ideologically diverse news and opinion on Facebook. Science 348 (6239), 1130–1132.

Banerjee, A., Duflo, E., 2012. Poor Economics: a Radical Rethinking of the Way to Fight Global Poverty. PublicAffairs. ISBN 9781610391603. Retrievable at. https://books.google.fr/books?id=2dlnBoX4licC.

Barocas, S., Hardt, M., Narayanan, A., 2023. Fairness and Machine Learning: Limitations and Opportunities. MIT press.

Barrera, O., Guriev, S., Henry, E., Zhuravskaya, E., 2020. Facts, alternative facts, and fact checking in times of post-truth politics. J. Publ. Econ. 182, 104123.

Berman, R., Katona, Z., 2020. Curation algorithms and filter bubbles in social networks. Mark. Sci. 39 (2), 296–316.

Bourreau, M., Gaudin, G., 2022. Streaming platform and strategic recommendation bias. J. Econ. Manag. Strat. 31 (1), 25–47.

Brandtzaeg, P.B., Følstad, A., Chaparro Domínguez, M.Á., 2018. How Journalists and social media users perceive online fact-checking and verification services. Journal. Pract. 12 (9), 1109–1129.

Brynjolfsson, E., Li, D., Raymond, L.R., 2023. Generative AI at Work. National Bureau of Economic Research. Technical report.

Brynjolfsson, E., Rock, D., Syverson, C., 2019. Artificial intelligence and the modern productivity paradox. In: The Economics of Artificial Intelligence: an Agenda, 23, pp. 23–57.

Bursztyn, L., Rao, A., Roth, C.P., Yanagizawa-Drott, D.H., 2020. Misinformation During a Pandemic. National Bureau of Economic Research. Technical report.

Bussotti, J.-F., Veltri, E., Santoro, D., Papotti, P., 2023. Generation of training examples for tabular natural language inference. Proceed. ACM Manage. Data 1 (4), 1–27.

Camodeca, R., Almici, A., 2021. Digital transformation and convergence toward the 2030 agenda's sustainability development goals: evidence from Italian listed firms. Sustainability 13 (21), 11831.

Capraro, V., Lentsch, A., Acemoglu, D., Akgun, S., Akhmedova, A., Bilancini, E., Bonnefon, J.-F., Brañas-Garza, P., Butera, L., Douglas, K.M., et al., 2024. The impact of generative artificial intelligence on socioeconomic inequalities and policy making. PNAS Nexus 3 (6), 191.

Carrieri, V., Madio, L., Principe, F., 2019. Vaccine hesitancy and (Fake) news: quasi-experimental evidence from Italy. Health Econ. 28 (11), 1377–1382.

Cecere, G., Corrocher, N., 2012. The usage of VoIP services and other communication services: an empirical analysis of Italian consumers. Technol. Forecast. Soc. Change 79 (3), 570–578.

Cecere, G., Jean, C., Lefrere, V., Tucker, C.E., 2021. Trade-offs in automating platform regulatory compliance by algorithm: evidence from the COVID-19 pandemic. Available at SSRN, 3603341.

Chatterjee, S., Chaudhuri, R., Vrontis, D., 2022. AI and digitalization in relationship management: impact of adopting AI-embedded CRM system. J. Bus. Res. 150, 437–450.

Chen, N., Tsai, H.-T., 2024. Steering via algorithmic recommendations. Rand J. Econ. 55 (4), 501–518.

Chiou, L., Tucker, C., 2018. Fake News and Advertising on Social Media: a Study of the Anti-vaccination Movement. National Bureau of Economic Research. Working paper.

Cockburn, I.M., Henderson, R., Stern, S., et al., 2018. The Impact of Artificial Intelligence on Innovation, 24449. National Bureau of Economic Research, Cambridge, MA, USA.

Cowgill, B., Tucker, C.E., 2019. Economics, fairness and algorithmic bias. Preparation for: J. Econ. Perspect.

Daron, A., Pascual, R., 2020. The wrong kind of AI? Artificial intelligence and the future of labour demand. Camb. J. Reg. Econ. Soc. 13 (1), 25–35.

Datta, H., Knox, G., Bronnenberg, B.J., 2018. Changing their tune: how consumers' adoption of online streaming affects music consumption and discovery. Mark. Sci. 37 (1), 5–21.

Davenport, T., Guha, A., Grewal, D., Bressgott, T., 2020. How artificial intelligence will change the future of marketing. J. Acad. Market. Sci. 48, 24–42.

De Alves, C., Salge, C., Karahanna, E., Thatcher, J., 2022. Algorithmic processes of social alertness and social transmission: how bots disseminate information on Twitter. MIS Q. 46 (1), 229–259.

Deaton, A., 2024. The Great Escape: Health, Wealth, and the Origins of Inequality. Princeton University Press.

DeNardis, L., Hackl, A.M., 2015. Internet governance by social media platforms. Telecommun. Policy 39 (9), 761–770.

Di Vaio, A., Palladino, R., Hassan, R., Escobar, O., 2020. Artificial intelligence and business models in the sustainable development goals perspective: a systematic literature review. J. Bus. Res. 121, 283–314.

DiMaggio, P., Hargittai, E., et al., 2001. From the 'Digital Divide' to 'Digital Inequality': studying internet use as penetration increases. Princeton: Center for Arts and Cultural Policy Studies, Woodrow Wilson School, Princeton University 4 (1), 4–2.

Do Nascimento, I.J.B., Pizarro, A.B., Almeida, J.M., Azzopardi-Muscat, N., Gonçalves, M. A., Björklund, M., Novillo-Ortiz, D., 2022. Infodemics and health misinformation: a systematic review of reviews. Bull. World Health Organ. 100 (9), 544.

Dobbs, M., 2012. The Rise of Political fact-checking How Reagan Inspired a Journalistic Movement: a Reporter's Eye View. New America Foundation.
Domenico, G.D., Sit, J., Ishizaka, A., Nunan, D., 2021. Fake news, social media and marketing: a systematic review. J. Bus. Res. 124, 329–341.
Duan, Y., Edwards, J.S., Dwivedi, Y.K., 2019. Artificial intelligence for decision making in the era of big data–evolution, challenges and research agenda. Int. J. Inf. Manag. 48, 63–71.
Eubanks, V., 2018. Automating Inequality: How high-tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.
Evans, D.S., 2009. The online advertising industry: economics, evolution, and privacy. J. Econ. Perspect. 23 (3), 37–60.
Fernandez, M., Bellogín, A., 2020. Recommender Systems and Misinformation: the Problem or the Solution? Working Paper.
Fong, J., Guo, T., Rao, A., 2024. Debunking misinformation about consumer products: effects on beliefs and purchase behavior. J. Mark. Res. 61 (4), 659–681.
Ford, J., Jain, V., Wadhwani, K., Gupta, D.G., 2023. AI advertising: an overview and guidelines. J. Bus. Res. 166, 114124.
Forman, C., Goldfarb, A., Greenstein, S., 2002. Digital Dispersion: an Industrial and Geographic Census of Commerical Internet Use. National Bureau of Economic Research. Working Paper 9287.
Gentzkow, M., 2014. Trading dollars for dollars: the price of attention online and offline. Am. Econ. Rev. 104 (5), 481–488.
Ghenai, A., Mejova, Y., 2017. Catching zika fever: application of crowdsourcing and machine learning for tracking health misinformation on Twitter. In: 2017 IEEE International Conference on Healthcare Informatics (ICHI). August. 518.
Glaeser, E.L., Ujhelyi, G., 2010. Regulating misinformation. J. Publ. Econ. 94 (3–4), 247–257.
Goldfarb, A., Prince, J., 2008. Internet adoption and usage patterns are different: implications for the digital divide. Inf. Econ. Pol. 20 (1), 2–15.
Goldfarb, A., Tucker, C., 2019. Digital economics. J. Econ. Lit. 57 (1), 3–43.
Gomez-Uribe, C.A., Hunt, N., 2015. The netflix recommender system: algorithms, business value, and innovation. ACM Trans. Manage. Inform. Sys (TMIS) 6 (4), 1–19.
Gran, A.-B., Booth, P., Bucher, T., 2021. To be or not to be algorithm aware: a question of a new digital divide? Inf. Commun. Soc. 24 (12), 1779–1796.
Graves, L., 2016. Deciding What's True: the Rise of Political fact-checking in American Journalism. Columbia University Press.
Graves, L., 2017. Anatomy of a fact check: objective practice and the contested epistemology of fact checking. Commun., Cult. Critique 10 (3), 518–537.
Graves, L., 2018. Boundaries not drawn: mapping the institutional roots of the global fact-checking movement. Journal. Stud. 19 (5), 613–631.
Graves, L., Amazeen, M., 2019. Fact-checking as idea and practice in journalism. Tech. Rep.
Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., Lazer, D., 2019. Fake news on Twitter during the 2016 US presidential election. Science 363 (6425), 374–378.
Guess, A., Aslett, K., Tucker, J., Bonneau, R., Nagler, J., 2021. Cracking open the news feed: exploring what us Facebook users see and share with large-scale platform data. J. Quantit. Descrip.: Digit. Media 1.
Guess, A., Nagler, J., Tucker, J., 2019. Less than you think: prevalence and predictors of fake news dissemination on Facebook. Sci. Adv. 5 (1), eaau4586.
Gupta, S., Langhans, S.D., Domisch, S., Fuso-Nerini, F., Felländer, A., Battaglini, M., Tegmark, M., Vinuesa, R., 2021. Assessing whether artificial intelligence is an enabler or an inhibitor of sustainability at indicator level. Transport. Eng. 4, 100064.
Hattori, K., Higashida, K., 2014. Misleading advertising and minimum quality standards. Inf. Econ. Pol. 28, 1–14.
Ho, L., Breza, E., Banerjee, A., Chandrasekhar, A.G., Stanford, F.C., Fior, R., Goldsmith-Pinkham, P., Holland, K., Hoppe, E., Jean, L.-M., et al., 2023. The impact of large-scale social media advertising campaigns on Covid-19 vaccination: evidence from two randomized controlled trials. AEA Papers Proceed. 113 (May), 653–658.
Hosanagar, K., Fleder, D., Lee, D., Buja, A., 2014. Will the global village fracture into tribes? Recommender systems and their effects on consumer fragmentation. Manag. Sci. 60 (4), 805–823.
Imhoff, R., Zimmer, F., Klein, O., António, J.H., Babinska, M., Bangerter, A., Bilewicz, M., Blanuša, N., Bovan, K., Bužarovska, R., et al., 2022. Conspiracy mentality and political orientation across 26 countries. Nat. Hum. Behav. 6 (3), 392–403.
Johnson, S., Acemoglu, D., 2023. Power and Progress: Our thousand-year Struggle over Technology and Prosperity. Hachette UK.
Just, J., 2024. Natural language processing for innovation search – reviewing an emerging non-human innovation intermediary. Technovation 129, 102883.
Kitchens, B., Johnson, S.L., Gray, P., 2020. Understanding echo chambers and filter bubbles: the impact of social media on diversification and partisan shifts in news consumption. MIS Q. 44 (4), 1619–1649.
Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., Mullainathan, S., 2018. Human decisions and machine predictions. Q. J. Econ. 133 (1), 237–293.
Kleinberg, J., Ludwig, J., Mullainathan, S., Sunstein, C.R., 2019. Discrimination in the age of algorithms. J. Legal Anal. 10, 113–174.
Kordzadeh, N., Ghasemaghaei, M., 2022. Algorithmic bias: review, synthesis, and future research directions. Eur. J. Inf. Syst. 31 (3), 388–409.
Koren, Y., Bell, R., Volinsky, C., 2009. Matrix factorization techniques for recommender systems. Computer 42 (8), 30–37.
Kretschmer, T., Peukert, C., 2020. Video killed the radio star? Online music videos and recorded music sales. Inf. Syst. Res. 31 (3), 776–800.
La Torre, D., Appio, F.P., Masri, H., Lazzeri, F., Schiavone, F., 2023. Impact of artificial intelligence in business and society: opportunities and challenges. Routledge.
Lambrecht, A., Tucker, C., 2019. Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of stem career ads. Manag. Sci. 65 (7), 2966–2981.
Lee, S.K., Sun, J., Jang, S., Connelly, S., 2022. Misinformation of COVID-19 vaccines and vaccine hesitancy. Sci. Rep. 12 (1), 13681.
Liu, J., Chang, H., Forrest, J.Y.-L., Yang, B., 2020. Influence of artificial intelligence on technological innovation: evidence from the panel data of China's manufacturing sectors. Technol. Forecast. Soc. Change 158, 120142.
Liu, P., Shivaram, K., Culotta, A., Shapiro, M.A., Bilgic, M., 2021. The interaction between political typology and filter bubbles in news recommendation algorithms. In: Proceedings of the Web Conference 2021, 4, pp. 3791–3801.
Loomba, S., de Figueiredo, A., Piatek, S.J., de Graaf, K., Larson, H.J., 2021. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat. Hum. Behav. 5 (3), 337–348.
Lutz, C., 2019. Digital inequalities in the age of artificial intelligence and big data. Human Behav. Emerg. Technol. 1 (2), 141–148.
Lythreatis, S., Singh, S.K., El-Kassar, A.-N., 2022. The digital divide: a review and future research agenda. Technol. Forecast. Soc. Change 175, 121359.
Martens, B., Aguiar, L., Gomez-Herrera, E., Mueller-Langer, F., 2018. The digital transformation of news media and the rise of disinformation and fake news. Working paper 2018-02. Digital Econom. Joint Research Centre Technical Reports.
Meng, Y., Huang, J., Zhang, Y., Han, J., 2022. Generating training data with language models: towards zero-shot language understanding. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. NIPS '22. November, pp. 462–477.
Micallef, N., Armacost, V., Memon, N., Patil, S., 2022. True or false: studying the work practices of professional fact-checkers. Proceed. ACM Human- Comput. Interact. 6 (CSCW1), 1–44.
Mikalef, P., Conboy, K., Lundström, J.E., Popovič, A., 2022. Thinking responsibly about responsible AI and 'the Dark Side' of AI. Eur. J. Inf. Syst. 31 (3), 257–268.
Moravec, P.L., Minas, R.K., Dennis, A.R., 2019. Fake news on social media. MIS Q. 43 (4), 1343. A13.
Müller, M., Salathé, M., Kummervold, P.E., 2023. Covid-twitter-bert: a natural language processing model to analyse COVID-19 content on Twitter. Front. Artif. Intell. 6, 1023281.
Nakov, P., Corney, D., Hasanain, M., Alam, F., Elsayed, T., Barrón-Cedeño, A., Papotti, P., Shaar, S., Da San Martino, G., 2021. Automated fact-checking for assisting human fact-checkers. In: *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*. 8. International Joint Conferences on Artificial Intelligence Organization, pp. 4551–4558. Survey Track.
Narayanan, A., 2023. Understanding Social Media Recommendation Algorithms. Technical Report.
Noble, S.U., 2018. Algorithms of oppression: how search engines reinforce racism. In: Algorithms of Oppression. New York university press, pp. 117–120.
Nyhan, B., 2020. Facts and myths about misperceptions. J. Econ. Perspect. 34 (3), 220–236.
Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S., 2019. Dissecting racial bias in an algorithm used to manage the health of populations. Science 366 (6464), 447–453.
OECD, 2019. Artificial intelligence in society. Tech. Rep. OECD.
O'Neil, C., 2017. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown.
OpenAI, 2024. Interaction with Chatgpt. Accessed on: 2024-02-05. https://chat.openai.com/chat.
Papagiannidis, E., Mikalef, P., Conboy, K., Van de Wetering, R., 2023. Uncovering the dark side of AI-based decision-making: a case study in a B2B context. Ind. Mark. Manag. 115, 253–265.
Pariser, E., 2011. The Filter Bubble: How the New Personalized Web is Changing what we Read and How we Think. Penguin.
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E.Z., De- Vito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: an imperative style, high-performance deep learning library. In: Wallach, H.M., Larochelle, H., Beygelzimer, A., d'Alché- Buc, F., Fox, E.B., Garnett, R. (Eds.), Advances in Neural Information Pro- Cessing Systems 32: Annual Conference on Neural Information Processing Systems 2019, *Neurips* 2019, pp. 8024–8035. December 8-14, 2019, Vancouver, BC, Canada. December. https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html.
Patalas-Maliszewska, J., Szmołda, M., Łosyk, H., 2024. Integrating artificial intelligence into the supply chain in order to enhance sustainable production – A systematic literature review. Sustainability 16 (16), 7110.
Pathak, R., Spezzano, F., Pera, M.S., 2023. Understanding the contribution of recommendation algorithms on misinformation recommendation and misinformation dissemination on social networks. ACM Trans. Web 17 (4), 1–26.
Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A.A., Eckles, D., Rand, D.G., 2021. Shifting attention to accuracy can reduce misinformation online. Nature 592 (7855), 590–595.
Pennycook, G., Rand, D.G., 2021. The psychology of fake news. Trends Cognit. Sci. 25 (5), 388–402.
Peskine, Y., Korencic, D., Grubisic, I., Papotti, P., Troncy, R., Rosso, P., 2023a. Definitions matter: guiding GPT for multi-label classification. Retrievable at. https://aclanthology.org/2023.findings-emnlp.267.
Peskine, Y., Papotti, P., Troncy, R., 2023b. Detection of COVID-19-related conpiracy theories in tweets using transformer-based models and node embedding techniques. In: Mediaeval 2022, Multimedia Evaluation Workshop, 12-13 January 2023, 1, pp. 12–13. Bergen, Norway.
Qian, K., Jain, S., 2024. Digital content creation: an analysis of the impact of recommendation systems. Manag. Sci. 70 (12), 8668–8684.

Rajpurkar, P., Chen, E., Banerjee, O., Topol, E.J., 2022. AI in health and medicine. Nat. Med. 28 (1), 31–38.
Rao, A., 2022. Deceptive claims using fake news advertising: the impact on consumers. J. Mark. Res. 59 (3), 534–554.
Resnick, P., Varian, H.R., 1997. Recommender systems. Commun. ACM 40 (3), 56–58.
Ricci, F., Rokach, L., Shapira, B., 2010. Introduction to recommender systems handbook. In: Recommender Systems Handbook. Springer, pp. 1–35.
Robinson, L., Cotten, S.R., Ono, H., Quan-Haase, A., Mesch, G., Chen, W., Schulz, J., Hale, T.M., Stern, M.J., 2015. Digital inequalities and why they matter. Inf. Commun. Soc. 18 (5), 569–582.
Saeed, M., Traub, N., Nicolas, M., Demartini, G., Papotti, P., 2022. Crowdsourced fact-checking at Twitter: how does the crowd compare with experts?. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, pp. 1736–1746. October.
Sætra, H.S., 2021. A framework for evaluating and disclosing the ESG related impacts of AI with the SDGs. Sustainability 13 (15), 8503.
Sandrini, L., Somogyi, R., 2023. Generative AI and deceptive news consumption. Econ. Lett. 232, 111317.
Santos, B.D.l., Hortaçsu, A., Wildenbeest, M.R., 2012. Testing models of consumer search using data on web browsing and purchasing behavior. Am. Econ. Rev. 102 (6), 2955–2980.
Sapiezynski, P., Ghosh, A., Kaplan, L., Rieke, A., Mislove, A., 2022. Algorithms that "Don't See Color"' measuring biases in lookalike and special Ad audiences. In: Proceedings of the 2022 AAAI/ACM Conference on AI. *Ethics, and Society*. May, pp. 609–616.
Schellner, I., 2002. Japanese File Index classification and F-terms. World Pat. Inf. 24 (3), 197–201.
Senecal, S., Nantel, J., 2004. The influence of online product recommendations on consumers' online choices. J. Retailing 80 (2), 159–169.
Shumanov, M., Cooper, H., Ewing, M., 2022. Using AI predicted personality to enhance advertising effectiveness. Eur. J. Market. 56 (6), 1590–1609.
Song, K., Kim, K.S., Lee, S., 2017. Discovering new technology opportunities based on patents: text-Mining and F-Term analysis. Technovation 60–61, 1–14.
Stray, J., Halevy, A., Assar, P., Hadfield-Menell, D., Boutilier, C., Ashar, A., Bakalar, C., Beattie, L., Ekstrand, M., Leibowicz, C., et al., 2024. Building human values into recommender systems: an interdisciplinary synthesis. ACM Trans. Recomm. Sys. 2 (3), 1–57.
Suarez-Lledo, V., Alvarez-Galvez, J., 2021. Prevalence of health misinformation on social media: systematic review. J. Med. Internet Res. 23 (1), e17187.
Swire-Thompson, B., Lazer, D., et al., 2020. Public health and online misinformation: challenges and recommendations. Annu. Rev. Publ. Health 41 (1), 433–451.
Thorne, J., Vlachos, A., 2018. Automated fact checking: task formulations, methods and future directions. arXiv preprint arXiv:1806.07687.
Tomašev, N., Cornebise, J., Hutter, F., Mohamed, S., Picciariello, A., Connelly, B., Belgrave, D.C., Ezer, D., Haert, F.C.v. d., Mugisha, F., et al., 2020. AI for social good: unlocking the opportunity for positive impact. Nat. Commun. 11 (1), 2468.
Törnberg, P., 2018. Echo chambers and viral misinformation: modeling fake news as complex contagion. PLoS One 13 (9), e0203958.
Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., et al., 2023. Llama 2: open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288.
Valecha, R., Volety, T., Rao, H.R., Kwon, K.H., 2020. Misinformation sharing on Twitter during zika: an investigation of the effect of threat and distance. IEEE Internet Comput. 25 (1), 31–39.
Van Der Linden, S., 2022. Misinformation: susceptibility, spread, and interventions to immunize the public. Nat. Med. 28 (3), 460–467.
Van Dijk, J.A., 2005. The Deepening Divide: Inequality in the Information Society. Sage Publications.
Verma, G., Bhardwaj, A., Aledavood, T., De Choudhury, M., Kumar, S., 2022. Examining the impact of sharing Covid-19 misinformation online on mental health. Sci. Rep. 12 (1), 8045.
Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S.D., Tegmark, M., Fuso Nerini, F., 2020. The role of artificial intelligence in achieving the sustainable development goals. Nat. Commun. 11 (1), 1–10.
Vosoughi, S., Roy, D., Aral, S., 2018. The spread of true and false news online. Science 359 (6380), 1146–1151.
Walter, N., Cohen, J., Holbert, R.L., Morag, Y., 2020. Fact-checking: a Meta- analysis of what works and for whom. Polit. Commun. 37 (3), 350–375.
Wang, Y., Yao, Q., Kwok, J.T., Ni, L.M., 2020. Generalizing from a few examples: a survey on few-shot learning. ACM Comput. Surv. 53 (3). https://doi.org/10.1145/3386252. ISSN 0360-0300.
WHO, 2020. A coordinated global research roadmap: 2019 novel coronavirus. Tech. Rep. https://www.who.int/publications/m/item/a-coordinated-global-research-roadmap.
Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Brew, J., 2019. HuggingFace's transformers: state-of-the-art natural language processing. CoRR. abs/1910.03771. http://arxiv.org/abs/1910.03771.
Yuan, Z., Chen, A.Y., Wang, Y., Sun, T., 2025. How recommendation affects customer search: a field experiment. Inf. Syst. Res. 36 (1), 84–106.
Zhang, D., Pee, L., Cui, L., 2021. Artificial intelligence in E-commerce fulfillment: a case study of resource orchestration at Alibaba's smart warehouse. Int. J. Inf. Manag. 57, 102304.
Zhou, X., Zafarani, R., 2020. A survey of fake news: fundamental theories, detection methods, and opportunities. ACM Comput. Surv. 53 (5), 1–40.
Zhuravskaya, E., Petrova, M., Enikolopov, R., 2020. Political effects of the internet and social media. Annual Rev. Econom. 12, 415–443.