

Deepfakes or Synthetic Media? The Effect of Euphemisms for Labeling Technology on Risk and Benefit Perceptions

Social Media + Society
July-September 2025: 1–12
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20563051251350975
journals.sagepub.com/home/sms



Adrian Rauchfleisch¹ , Daniel Vogler² ,
and Gabriele de Seta³ 

Abstract

The language used in public debates and in the news can influence how citizens perceive the risks and benefits of technology. While framing effects on technology perception are well understood, few studies have focused on the effects of specific terms used to describe technology. We analyze how the terms deepfake and synthetic media affect risk and benefit perceptions across application fields. Using Switzerland as a case, our manual content analysis ($n = 380$ news articles) reveals a focus on risks in news coverage of deepfakes with minimal use of the term synthetic media. We then tested the effects of the terms on risk and benefit perceptions in a preregistered survey experiment ($n = 736$ participants). Term choice does not change perceived risks, but “synthetic media” significantly increases perceived benefits across application fields. As a theoretical contribution, we link our findings to the concept of euphemism, proposing that term choice should align with application fields to reflect the risks and benefits of technology. Overall, our study shows that the terms we use to label technology matter, especially for emerging technologies such as artificial intelligence.

Keywords

deepfake, euphemism, survey experiment, artificial intelligence, risk perception

Introduction

When Robert De Niro appeared in the 2019 movie *The Irishman*, he looked much younger than his actual age of 76. Many social media users were also surprised when they saw Yilong Ma, an Elon Musk lookalike from China, in short videos circulating on their favorite platforms. Both examples rely on the same AI technology to create audiovisual content in which the face or other identifying attributes of a person are replaced with someone else's. However, the terms used to describe this technology vary. In news coverage, Yilong Ma is typically associated with the term deepfake (Florian, 2023), whereas applications in the creative industry, such as those in *The Irishman*, are often called synthetic media or AI-enabled media (Glick, 2023). This highlights a broader competition over how the technology is labeled in public debates or in the news.

Today, such AI technology for creating and modifying audiovisual content is not only used by industries and specialists but is also available to everyday internet users. However, its origins and early uses are more controversial. The term deepfake was first coined on Reddit, where users

applied the technology to swap faces in pornographic videos with those of celebrities (de Ruiter, 2021). Eventually, news stories began linking the technology to disinformation, such as the case of a deepfake of Ukrainian President Zelenskyy in 2022 (Allyn, 2022). These portrayals in the news influence how people perceive the technology, as they often emphasize its risks while overlooking potential benefits. A focus on risks has also been observed in the academic literature (Godulla et al., 2021) as studies on the benefits of deepfake technology remain scarce (cf. Bendahan Bitton et al., 2024; Neyazi et al., 2024). The tech industry also laments the predominantly negative connotation of the term deepfake and instead prefers the more neutral and general term synthetic media (Altuncu et al., 2022). Building on the

¹National Taiwan University, Taiwan

²University of Zurich, Switzerland

³University of Bergen, Norway

Corresponding author:

Daniel Vogler, Research Center for the Public Sphere and Society, University of Zurich, Andreasstrasse 15, 8050 Zurich, Switzerland.
Email: daniel.vogler@foeg.uzh.ch



assumption that the terminology influences public perception of technology, our study investigates whether the use of the terms deepfake and synthetic media affects the perception of the technology's risks and benefits. More specifically, we investigate if the term synthetic media functions as a euphemism, potentially overshadowing the negative perceptions associated with the term deepfake.

We further argue that the field of application plays a central role in perceptions of deepfake technology. Its use for entertainment through tools like Midjourney or FaceSwap is presumably perceived as offering more benefits and personal gratification than risks. By contrast, in politics, the perception of risk is likely more prevalent as the potential for disinformation through deepfakes is particularly emphasized in political communication during elections or votes (Hameleers et al., 2022; Vaccari & Chadwick, 2020). Therefore, we assume that individual perceptions of the risks and benefits of deepfake technology depend on the specific application field being addressed. Thus, we examine the perception of risks and benefits of the technology separately in the contexts of politics, news media, the economy, science, and individual contexts, and then test whether the use of the labels "deepfake" or "synthetic media" affects these perceptions.

The empirical approach for our study is twofold. We first conduct a manual quantitative content analysis to examine how deepfake technology is depicted in Swiss news media. By investigating the salience of the terms deepfake and synthetic media, their fields of application, and the tone of news coverage, the content analysis provides valuable context for the subsequent preregistered survey experiment. In this experiment, we investigate how the term deepfake and its euphemism synthetic media influence the Swiss public's perception of risks and benefits of deepfake technology across different fields of application. Finally, we discuss the conceptual and practical implications of our findings, highlighting that while the term synthetic media increases perceived benefits, it does not alter risk perceptions.

Theory and literature review

The integration of new technologies into a society largely depends on the level of acceptance among its members. Thus, how people perceive the risks and benefits of technology has been the topic of numerous studies (Bao et al., 2022; Binder et al., 2012; Covello, 1983; de Groot et al., 2013; Frewer et al., 1999; Siegrist & Visschers, 2013). These studies, for instance, show that advances in the fields of genetic engineering (Frewer et al., 1999), nuclear energy (Siegrist & Visschers, 2013), and vaccinations (Wilson et al., 2015) were initially met with skepticism by parts of the population. This strand of research also includes studies on the perception of new media and communication technologies such as deepfakes (Cochran & Napshin, 2021; Kleine, 2022). Studies on deepfakes so far have primarily focused on the risks (Godulla et al., 2021), with only a few exceptions explicitly considering the potential benefits of the

technology (Bendahan Bitton et al., 2024; Neyazi et al., 2024). Our study examines both aspects and explores how synthetic media can serve as a potential euphemism for deepfake technology. Before discussing the literature on euphemism and why this concept matters for our study, we present the definitions of synthetic media and deepfakes.

Deepfakes and synthetic media: same technology, different definitions

While this study does not aim to provide definitive definitions for these two concepts, we begin with the assumption that both terms have a well-documented history and an established meaning that may evolve in the future. The term deepfake, a portmanteau word of "deep learning" and "fake," was established in 2017 by amateur users sharing their (largely pornographic) face-swap creations on a Reddit discussion board (de Ruiter, 2021). Synthetic media, on the other hand, emerged in the late 2010s as a catch-all term for a variety of AI-generated content, including deepfakes, virtual humans, and augmented reality (Kalpokas, 2021). While the two terms are often used interchangeably (Westerlund, 2019), it is important to note that deepfakes are only one specific kind of synthetic media. Synthetic media is a broader concept and also includes text, images, video, and audio content generated by a variety of machine learning models (de Seta, 2024).

Different definitions of deepfakes and synthetic media indicate how terminology frames understanding: Definitions in the domain of politics often emphasize deception, while others take a broader view, highlighting the technology's ambiguity. Thus, several definitions are proposed depending on the context (Altuncu et al., 2022) or the degree of technical sophistication (Paris & Donovan, 2019). In studies focusing on political deepfakes, scholars often highlight the deceptive nature of the technology, defining deepfakes as "intentionally deceptive synthetic videos created with the use of Artificial Intelligence" (Hameleers et al., 2022, p. 1) or as "synthetic videos created with deep learning techniques to make authentic persons say or do inauthentic things" (Hameleers et al., 2024, p. 56). These definitions may be suitable within the context of political disinformation but are less applicable to broader uses. Deepfakes do not necessarily depict real people or always involve deception, although they possess "a high potential to deceive" (Kietzmann et al., 2020, p. 136), particularly when used in political contexts (Vaccari & Chadwick, 2020).

While much of the academic discourse surrounding deepfakes emphasizes potential risks (Godulla et al., 2021), some definitions adopt a more neutral or even positive perspective. De Ruiter (2021) highlights that deepfakes are "not intrinsically morally wrong" (p. 1313) as they can be used for morally neutral or even beneficial purposes. Some scholars, such as Jacobsen and Simpson (2024), offer a more neutral definition, describing deepfakes as "photorealistic images, videos, or voice recordings that have been

algorithmically generated or manipulated” (p. 1095). They argue that the public discourse about deepfakes is dominated by anxiety and potentially harmful applications of the technology, such as misinformation or pornographic videos (de Ruiter, 2021; Wang & Kim, 2022). While these worries are legitimate, some scholars argue that the assumed negative impact of deepfakes, for instance, on elections, might be overblown (Łabuz & Nehring, 2024; Nadal & Jančárik, 2024; Simon et al., 2024). These concerns also overshadow the benefits of the technology for industries such as film, gaming, and customer relations in general (Altuncu et al., 2022; Godulla et al., 2021; Westerlund, 2019). However, based on their mixed-methods study, which includes expert interviews, Lundberg and Mozelius (2024) conclude that, although deepfakes offer both, the risks slightly outweigh the benefits. Due to this ambiguity, discussing the potential of deepfakes requires a balanced view that acknowledges both benefits and risks under consideration of specific application fields.

Euphemisms

In this article, we employ the concept of euphemism to examine how labeling technology as either “deepfake” or “synthetic media” influences perceptions of risks and benefits. With deepfakes being a more specific emergent term and synthetic media encompassing a broader technical category, the two terms may evoke more negative or more positive interpretations of the same technology. Thus, the concept of euphemism—the use of a more agreeable or neutral term as a substitute for a more loaded or unpleasant one—and its potential effects are discussed in the following section.

Studies in different disciplines, including business research, journalism research, health care and medicine, and communication science (for an overview, see Farrow et al., 2021), have already investigated the perception and effects of euphemisms. While definitions vary in the literature, they share a common focus on using language to express something negative in a more positive or softened way. Lutz (1987) defines euphemism as “a word or phrase that is designed to avoid a harsh or distasteful reality” (p. 382). He further argues that if euphemisms are intentionally used to deceive or mislead, they can be termed doublespeak. Other definitions highlight that euphemisms help to talk about taboos (Casas Gómez, 2009) but are also used to “representationally displace an unpleasant topic by avoiding direct reference to it” (McGlone et al., 2006, p. 261) or as a replacement for swear words (Bowers & Pleydell-Pearce, 2011). Finally, euphemisms (e.g., “enhanced interrogation” instead of “torture”) can be contrasted with dysphemisms (e.g., “extremist” instead of “activist”) (Walker et al., 2021), which describe terms intended to evoke a more negative interpretation (Casas Gómez, 2009; Walker et al., 2021). The two concepts are open to interpretation. For instance, it remains unclear whether deepfake now functions as a dysphemism for synthetic media, or, conversely, whether

synthetic media is being adopted as a euphemism for deepfake.

Prior research indicates that the use of euphemisms can influence people’s evaluations of situations and the actions they would take (Gladney & Rittenburg, 2005). Walker et al. (2021) highlight that euphemistic language can significantly affect evaluations of actions by replacing disagreeable terms with more agreeable ones, thereby making the actions seemingly more acceptable without triggering perceptions of dishonesty. However, there are only a few studies that focus on euphemisms in the context of technology-related terminology. The study most relevant to our discussion of euphemisms is an analysis of public perceptions of drones in the U.S. context, which briefly references the literature on euphemisms. PytlikZillig et al. (2018) investigated whether the specific use of terminology affects the public perception of unmanned aerial vehicles (UAVs), commonly referred to as drones in public debate. The use of the term drone has been criticized by industry representatives who favor the more neutral designation UAV. In their study, PytlikZillig et al. (2018) compare the perception of the publicly used terms, drones or aerial robots, with the industry terms like UAVs. In their case, they were unable to identify a significant effect of the terminology on public support for the technology. However, this might be different in the case of deepfakes and synthetic media.

Label studies and negative perception

The effects of euphemisms and dysphemisms demonstrate how language shapes public perception. This influence is also relevant for labeling generative AI, as the terminology can affect how credible or trustworthy content is perceived. However, studies on the labeling of technology have yet to make an explicit connection to the euphemism literature. Prior work focusing on labels for generative AI in journalism shows that content with labels such as ‘AI-generated’ is perceived as less accurate (Altay & Gilardi, 2024), less credible (Wittenberg et al., 2024), and less trustworthy (Toff & Simon, 2023) than content labeled as written by human journalists. This is also the case when readers are presented with identical news articles, indicating a significant influence of labeling on audience perception (Graefe et al., 2018).

Epstein et al.’s (2023) study on labeling AI-generated content is particularly relevant to our research, as they specifically focus on the perception of terms. They argue that labels could either focus on the process (e.g., AI-generated) or on whether the content is misleading (e.g., deepfake). Their study shows that participants associate both the labels “deepfake” and “synthetic” with misleading content to the same degree. However, content labeled “deepfake” leads to a more negative perception of the person sharing such content than content labeled ‘synthetic.’ Furthermore, people perceive content labeled as “synthetic” as more real than content labeled as “deepfake.” These results indicate that even minor

Table 1. Overview of the Media Outlets Used for the Content Analysis.

Outlet	Website	Language	Topical Focus
20 Minuten	20min.ch	German	Tabloid journalism
Aargauer Zeitung	aargauerzeitung.ch	German	Regional journalism
Blick	blick.ch	German	Tabloid journalism
Neue Zürcher Zeitung	nzz.ch	German	Quality journalism
SRF	srf.ch/news	German	Public Service Media
Tages-Anzeiger	tagesanzeiger.ch	German	Regional journalism
20 Minutes	20min.ch	French	Tabloid journalism
24 Heures	24heures.ch	French	Regional journalism
Le Temps	letemps.ch	French	Quality journalism
RTS	rts.ch/news	French	Public Service Media

differences regarding the wording can have substantial effects on people's perception of the content. Kleine (2022) arrives at a similar conclusion, demonstrating that the term deepfake is perceived as more negative than synthetic media. While these studies provide a good overview of the negative perception of different terms, they did not analyze how different terms for the same technology affect the general risk and benefit perceptions, an aspect that our study focuses on.

Overall, studies testing different terms as labels show that terms describing AI technology are perceived differently, a pattern also evident in the varying definitions of deepfakes used across the literature. However, such labels are not automatically assigned, nor are they typically decided top-down by authorities like the industry. Instead, they emerge through public debates and compete for attention in the news media. Therefore, examining news coverage provides insights into the dominant terminology used for technology and whether the focus is on its risks or benefits.

Study I: content analysis deepfakes in Swiss media

Offline and online communication play a crucial role in the perception of technology. This includes conversations with friends or family, exchanges in larger groups via messaging apps, or public debates on social media platforms. However, the news media continues to play a critical role (Gosse & Burkell, 2020), especially when people have no or only limited experience with new and unfamiliar technology, as is the case with AI (Sartori & Bocca, 2023). The media can also influence the public's perception of technology through topic selection, focusing on specific aspects of a subject, including the choice of voices and implicit or explicit evaluations. In the case of deepfakes, the connection to disinformation or fake news is strongly pronounced in news coverage, creating a predominantly negative framing (Gosse & Burkell, 2020; Wahl-Jorgensen & Carlson, 2021; Yadlin-Segal & Oppenheim, 2021), which makes negative perception effects likely.

Thus, we are interested in (research question 1) how and (research question 2) in what context deepfakes and synthetic media are discussed in Swiss news media.

Method and data content analysis

We applied a manual quantitative content analysis to investigate the coverage of deepfakes in 10 Swiss news media outlets from the German-speaking (20 Minuten, Aargauer Zeitung, Blick, Neue Zürcher Zeitung, SRF, Tages-Anzeiger) and French-speaking (20 Minutes, 24 Heures, Le Temps, RTS) regions of Switzerland covering the period from 13 December 2017 to 31 December 2023 (see Table 1). The media sample consisted of the online news outlets with the highest reach in the two language regions (Udris et al., 2024). The selected outlets all cover a broad spectrum of news (e.g., politics, economy, science, sports, and celebrities) but have different journalistic focuses in their reporting (e.g., regional news and tabloid journalism). In total, 380 articles containing the term deepfake or synthetic media were identified. The news articles were accessed through the Swiss Media Database (SMD), which provides full texts of the analyzed outlets from January 2012 onwards. The starting point of the analysis was chosen as the publication date of the first article on the topic of deepfakes in the selected media, which we determined by searching the database without any temporal restrictions. Through our content analysis, we examined the thematic context in which deepfakes were covered (disinformation, criminality, pornography, entertainment, and unspecified) and how deepfakes were evaluated in the articles (positive, negative, ambivalent, and neutral). The analysis was conducted by two trained coders. To test intercoder reliability, the two coders coded 50 unique articles. Intercoder reliability was satisfactory, with Krippendorff's alpha at .92 for topic coding and .69 for evaluation coding.

Results content analysis

Our content analysis reveals that news coverage of deepfake technology in Switzerland almost exclusively uses the term deepfake, with only two out of 380 articles in our sample mentioning the term synthetic media. These two news articles used both terms, synthetic media and deepfake, indicating that the term synthetic media is nearly absent in public discussions about deepfake technology in Switzerland. In addition, 39.2% of the articles were published in 2023, while

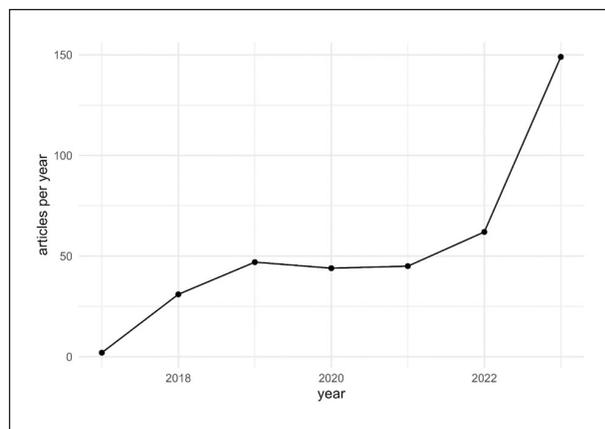


Figure 1. Number of news articles on deepfakes and synthetic media in Swiss news media per Year ($n = 380$; first news article on deepfakes on December 13, 2017—December 31, 2023).

the remaining 61.8% were distributed relatively evenly over the period from December 2017 to December 2022 (see Figure 1). This indicates that the concept of deepfakes received increased attention in the year of our survey experiment.

We then analyzed how deepfakes are evaluated in Swiss news coverage. Unsurprisingly, the focus is primarily on risks, with a predominantly negative view of the technology. More than half (59.7%) of the articles adopt a negative tone (see Table 2). In 20.0% of the articles, coverage is ambivalent, presenting both positive and negative aspects in roughly equal measure. Deepfakes are depicted neutrally in 13.2% of articles, while only 7.2% portray them in a positive light.

The emphasis on risks in deepfake reporting becomes more evident in the topics covered. Disinformation is the central topic (38.4%), especially in relation to foreign contexts such as U.S. politics and the conflicts in Ukraine and Gaza. Reports on deepfakes in connection with crime account for 10.5% of coverage. Pornography, which was a primary focus in the early stages of deepfake development, constitutes 8.9% of the articles. The context of entertainment comprises 17.4% of media reporting, while general or nonspecific articles make up exactly a quarter (25.0%) of the coverage.

In the final step, we combined the evaluation with the topics. Deepfake technology is covered primarily positively in the context of entertainment (see Table 2): Among articles discussing deepfakes in relation to film, music, or gaming, 34.8% adopt a positive tone. Ambivalent (31.6%) and neutral (27.3%) evaluations are also common, while negative assessments are rare (6.1%). In contrast, negative assessments clearly dominate the coverage of deepfakes in the contexts of disinformation (86.8%), crime (85.0%), and pornography (97.1%)—an expected result given the inherently negative connotations of these topics. In nonspecific contexts, negative (31.6%), ambivalent (36.8%), and neutral (29.5%) assessments of deepfakes are more evenly distributed.

Study II: the effects of the terms “deepfake” and “synthetic media” on perceptions

Our content analysis reveals that news coverage predominantly focuses on risks, with the positive aspects of deepfake technology rarely being addressed. The analysis also reveals that the term synthetic media is not used in news coverage. Next, we examine how individuals assess the risks and benefits of deepfake technology based on its labeling. While some studies have focused on the risk perception of deepfakes (Cochran & Napshin, 2021; Kleine, 2022), few have explored the perception of their benefits (Bendahan Bitton et al., 2024; Neyazi et al., 2024). Neyazi et al. (2024) measured positive attitudes toward deepfakes, but their one-dimensional scale primarily included negatively formulated items (e.g., deepfakes not a concern/issue), limiting the significance for specific benefits. Bendahan Bitton et al. (2024) distinguish between opportunities and risks using separate scales with items that assess benefits for various industries. To complement the limited research on benefits, we can also consider studies on attitudes toward AI. These studies help to better understand how the perception of risk and benefits vary across different application areas. For example, Bao et al. (2022) distinguish between the individual level and the societal level (i.e., democracy) when assessing the perceptions of risks and benefits. Studies also investigate risk assessments related to specific applications, such as the use of AI for autonomous weapons (Zhang & Dafoe, 2019). This more specific approach has also been used in the case of nanotechnology (Binder et al., 2012; Cobb & Macoubrie, 2004). Binder et al. (2012) distinguish between the impact of nanotechnology on science and medicine, job markets, and national defense. In the case of deepfakes, we also propose adopting a more nuanced approach that considers the various areas of risk and benefits.

Benefits and risks are usually operationalized as separate indices (Bao et al., 2022) or combined into a composite index that captures an overall risk-benefit assessment (Said et al., 2023). In the case of deepfakes, we argue that differentiating between risks and benefits and measuring them as separate dimensions is both conceptually and empirically more appropriate. An individual may, at the same time, perceive a technology as having very high benefits and very high risks. A composite index would fail to capture this dual perspective. Furthermore, a composite index would deliver the same result for individuals with high risk and high benefit perceptions as for individuals with both moderate risk and benefit perceptions. Thus, measuring risks and benefits separately provides a more nuanced analysis.

This conceptual direction is seconded by empirical research. Bao et al. (2022) demonstrate that the perception of risks and benefits associated with AI can be divided into two dimensions. Binder et al. (2012) recommend assessing risks and benefits separately rather than relying on a single global score when evaluating emerging technologies.

Table 2. Evaluation and Topical Context in Swiss News Media Coverage of Deepfakes and Synthetic Media ($n=380$).

	Negative	Positive	Ambivalent	Neutral	n	χ^2
Disinformation	86.90%	0.00%	11.70%	1.40%	145	48.65***
Nonspecific	31.60%	2.10%	36.80%	29.50%	95	48.46***
Entertainment	6.10%	34.80%	31.80%	27.30%	66	11.05*
Crime	85.00%	5.00%	5.00%	5.00%	40	19.76***
Pornography	97.10%	0.00%	2.90%	0.00%	34	117.93***
Total	59.70%	7.10%	20.00%	13.20%	380	

* $p < .05$, ** $p < .01$, *** $p < .001$.

As we have argued in the literature review, using the term deepfakes or synthetic media might influence the perception of the risks and benefits of the technology. As the content analysis already indicated, synthetic media is almost absent from the media discourse compared to the term deepfake, we expect the following:

Hypothesis 1 (H1). Respondents are more likely to provide an open-ended response when asked about the term “deepfake” compared to the term “synthetic media.”

Furthermore, as the term deepfake is often discussed in a negative context in the media, we expect people to be more likely to provide a negative description of the term deepfake in contrast to synthetic media.

Hypothesis 2 (H2). Respondents are more likely to provide a negative description when asked about the term “deepfake” compared to the term “synthetic media.”

Based on the literature on risks of AI and technology in general, we differentiate between risks for five different societal domains:

Hypothesis 3 (H3). The term “deepfake” is associated with higher risk perception regarding politics than the term “synthetic media.”

Hypothesis 4 (H4). The term “deepfake” is associated with higher risk perception regarding the news media than the term “synthetic media.”

Hypothesis 5 (H5). The term “deepfake” is associated with higher risk perception regarding the economy than the term “synthetic media.”

Hypothesis 6 (H6). The term “deepfake” is associated with higher risk perception regarding science than the term “synthetic media.”

Hypothesis 7 (H7). The term “deepfake” is associated with a higher risk perception at the individual level than the term “synthetic media.”

For the analysis of benefits, we investigate the same societal domains as for the analysis of risks but exclude politics. It was challenging to operationalize positive applications of politics for the survey, as respondents would require specific knowledge. While some specific use cases exist (Jungherr et al., 2025) and AI technology in general has the potential of a positive impact on democracy (Jungherr, 2023; Jungherr & Rauchfleisch, 2025; Wuttke et al., 2025), the perceived threats specifically of deepfakes in politics (Hameleers et al., 2024; Pawelec, 2022; Vaccari & Chadwick, 2020) clearly overshadow the potential benefits. Thus, we use the following hypotheses for the perception of benefits:

Hypothesis 8 (H8). The term “deepfake” is associated with lower benefit perception regarding the news media than the term “synthetic media.”

Hypothesis 9 (H9). The term “deepfake” is associated with lower benefit perception regarding the economy than the term “synthetic media.”

Hypothesis 10 (H10). The term “deepfake” is associated with lower benefit perception regarding science than the term “synthetic media.”

Hypothesis 11 (H11). The term “deepfake” is associated with a lower perception of benefits at the individual level than the term “synthetic media.”

Methods experiment

To test our hypotheses, we use an online experiment. The study was approved by the Ethics Committee of the Faculty of Arts and Social Sciences of the University of Zurich (registration no. 23.03.20) and preregistered (<https://aspredicted.org/5m56-2z9c.pdf>). Prior to the data collection, we conducted a power analysis to determine the required sample size to identify small to medium effects. We conducted a power simulation (runs=1000, mean difference=0.35, $SD=1.3$), which indicated small effects (Cohen’s $d=0.27$) and a power of 0.91 for a group size of 300. The data were collected by accessing the Swiss online panel from Bilendi, which covers people aged 16 and above in Switzerland. We offered the questionnaire in French and German to cover

Table 3. Overview of Variables, Means, Standard Deviations, and Sample Characteristics.

Variable	M (SD)	n
H1: Answer to open-ended question	51.63%	736
H2: Negative answer to open-ended (Krippendorff's alpha=0.86)	22.01%	736
H3: Risk politics (3 items, $\alpha=.85$)	5.06 (1.48)	736
H4: Risk media (2 items, $\alpha=.74$, Spearman-Brown=0.74)	5.88 (1.20)	736
H5: Risk economy (3 items, $\alpha=.77$)	4.76 (1.33)	736
H6: Risk science (single item)	5.02 (1.60)	736
H7: Risk individual context (3 items, $\alpha=.69$)	4.45 (1.42)	736
H8: Benefit media (2 items, $\alpha=.70$, Spearman-Brown=0.70)	3.82 (1.53)	736
H9: Benefit economy (3 items, $\alpha=.71$)	3.84 (1.32)	736
H10: Benefit science (1 item)	3.46 (1.65)	736
H11: Benefit individual context (3 items, $\alpha=.75$)	3.51 (1.47)	736
Female	51.49%	736
Age	43.67 (14.47)	736
University degree	29.48%	736
French region	31.52%	736

different language regions in Switzerland. Bilendi handled the incentivization. Participants received standard market-rate financial compensation. The data collection took place between May 26 and June 4, 2023. The final sample consisted of 736 complete responses after removing 17 respondents who failed an attention check, as we have defined in our preregistration.¹ A total of 504 (68.48%) participants completed the German version, and 232 (31.52%) completed the French version of the questionnaire. The median age of participants was 43 ($M=43.66$, $SD=14.47$), 51.49% were female, and 29.48% had a university degree.

At the start of the survey, participants were randomly assigned to either the “synthetic media” ($n=358$) or the “deepfake” ($n=378$) condition. We checked that randomization was successful by comparing the two groups based on age (Welch- $t(729.81)=1.72$, $p=.087$), education ($\chi^2(7, 736)=3.69$, $p=.815$), and gender ($\chi^2(1, 736)=0.00$, $p=1.00$). One group received the complete questionnaire with the term deepfake, while the other group received the identical questionnaire with the term synthetic media.

After obtaining consent, we started with basic socio-demographic questions. Participants were then prompted with an open-ended question (related to H1 and H2), asking what they associate with the term deepfake/synthetic media. To prevent immediate skipping of this question, respondents had to type at least “I don’t know” if they did not want to give an answer. After that question, we provided a neutral definition for both conditions as follows: “[Deepfakes/Synthetic media] are realistic-looking media content such as photos, videos, or sound recordings that are created or modified using artificial intelligence (AI).” We opted for this definition as it is close to the definition that people would encounter on Wikipedia (German version) and because we did not want to include deception as an aspect. After answering all questions, participants received a debriefing.

For H1, we checked how many people responded to the open-ended questions (i.e., did not answer “I don’t know”). For H2, we classified the evaluation of deepfakes or synthetic media in the answers as negative or not negative with ChatGPT (model *gpt-4o-2024-08-06*), as OpenAI’s LLMs have demonstrated strong performance in labeling tasks (Gilardi et al., 2023). We manually labeled a random sample of 30 out of 338 unique answers to validate our approach. The classifier performed well, achieving Krippendorff’s alpha of 0.86 (for more details about the classification procedure, see Supplemental Appendix A.2).

A central block of questions involves the perception of opportunities and risks. To this end, we draw on research on risk and benefit perceptions of technology (Binder et al., 2012; Siegrist & Visschers, 2013) and AI (Bao et al., 2022; Zhang & Dafoe, 2019) and measured the risks and benefit perceptions across multiple domains. We ask about risks to politics (H3), the media (H4), the economy (H5), science (H6), and individual contexts like privacy (H7), as well as benefits for the media (H8), the economy (H9), science (H10), and individual contexts (H11). The only difference between risks and opportunities is that we did not measure benefits for politics. Most variables were measured with two or three items that were merged into a mean index. The items were measured using a 7-point scale with labeled anchor points (1—Do not agree at; 7—totally agree). All variables achieved good reliability (see Table 3 for details and Supplemental Appendix A.1 for the wording of all items).² As preregistered, covariates for all analyses included sex (female=1), age, educational attainment (university degree=1), and language region (French=1).³

Results

Our results show that for all areas, the risks are perceived as higher than the benefits (see Table 3). As preregistered, we

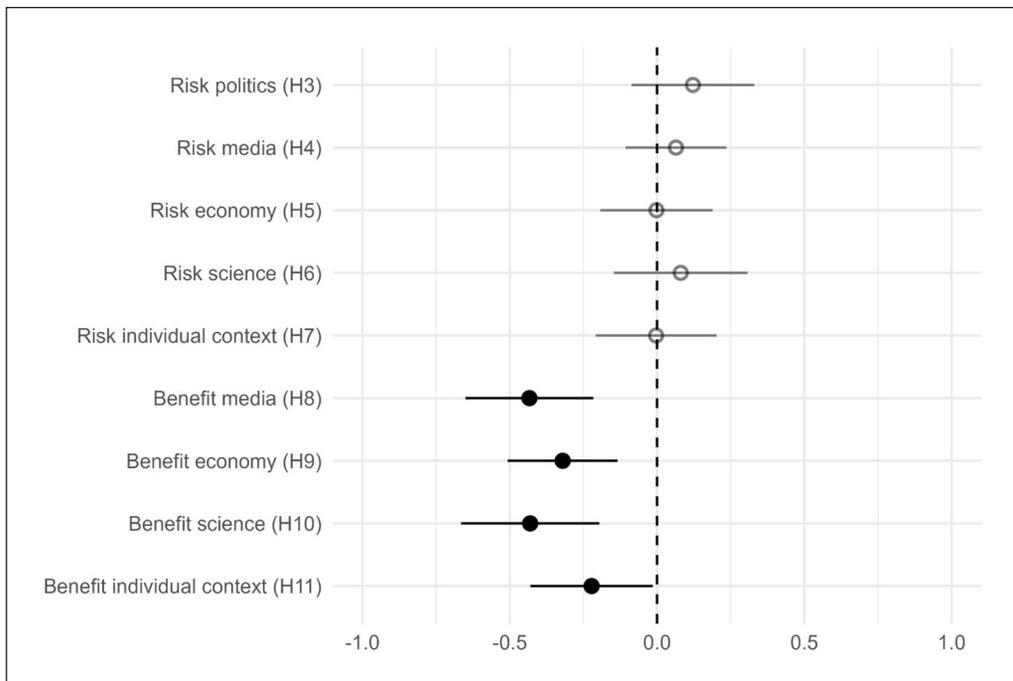


Figure 2. Estimates with 95% CI. Transparent dots indicate overlap with 0, while black dots represent significant estimates. Estimates show the effect of “deepfake” versus “synthetic media.” The identified effects are consistent with our hypotheses.

used binary logistic regression for H1 and H2 and linear regression for the remaining hypotheses. We will report the specific estimates for our hypotheses. The complete models are all reported in Supplemental Appendix B.1.

For H1, we tested whether participants were more likely to provide an answer to the open-ended question when exposed to the term “deepfake” compared to “synthetic media.” Our data supports H1 as participants who were given the term “deepfake” gave an answer in 63% of cases, which is significantly higher than the answer rate (40%) for “synthetic media” (odds ratio (OR)=2.68, $p < .001$, 95% confidence interval (CI) [1.98, 3.65]). This suggests that the term “deepfake” evokes more associations among the Swiss population than “synthetic media.” Our data also support H2, as the term “deepfake” triggered significantly more negative responses (38.1%) than “synthetic media” (5.0%) (OR=12.31, $p < .001$, 95% CI [7.48, 21.42]).⁴

Figure 2 shows all estimates for the risk and benefit perceptions. Regarding risk perception, the results for politics ($b=0.12$, $p=.252$, 95% CI [-0.09, 0.33]), media ($b=0.06$, $p=.461$, 95% CI [-0.11, 0.24]), economy ($b=0.00$, $p=.984$, 95% CI [-0.19, 0.19]), science ($b=0.08$, $p=.486$, 95% CI [-0.15, 0.31]), and the individual level ($b=0.00$, $p=.976$, 95% CI [-0.21, 0.20]) are all not significant. Thus, there is no support for hypotheses that assume a stronger influence on risk perception (H3–H7) when using the term “deepfake” instead of “synthetic media.”

In contrast, the findings for benefit perceptions support several hypotheses (see Figure 2). First, H8 is supported as

the term deepfake is associated with a significantly lower benefit perception for media ($b=-0.43$, $p < .001$, 95% CI [-0.65, -0.22]). The benefits for the economy (H9) are also perceived as significantly lower when the term deepfake is used ($b=-0.32$, $p < .001$, 95% CI [-0.51, -0.13]). Furthermore, the benefits for science (H10) are also perceived as lower ($b=-0.43$, $p < .001$, 95% CI [-0.67, -0.20]). Finally, on the individual level (H11), the benefit perception is also lower for the term deepfake than for the term synthetic media ($b=-0.22$, $p=.036$, 95% CI [-0.43, -0.01]).

As an additional, non-preregistered analysis, we conducted an equivalence test (Lakens et al., 2018) to determine whether the smallest effect size of interest (SESOI) can be rejected—namely, that effects of that magnitude are unlikely to be present. We defined the SESOI as Cohen’s $d=\pm 0.2$, a threshold commonly used for small effects, as we do not have prior studies. We used two one-sided tests (TOST) for all five risk perception variables with equivalence bounds set at ± 0.2 . Except for the domain of politics, all tests yielded significant results (see Supplemental Appendix C.1 for more details), suggesting that the effects for risk perceptions related to media, economy, science, and individual levels are not only non-significant but also negligible if we are interested in an effect size of at least Cohen’s $d=\pm 0.2$.

Discussion

In our study, we analyzed the news coverage of deepfakes and people’s perceptions of deepfake technology. First, we

examined the topical contexts in which deepfakes are discussed in Swiss news media, their evaluation, and the use of the terms deepfake and synthetic media in this coverage. Second, we analyzed whether the choice to use the label deepfake or synthetic media affects the perception of the technology's risks and benefits for different application fields in the Swiss population.

Our content analysis provides a clear picture. In the news, deepfakes are primarily discussed in a negative tone and in negative contexts, especially disinformation, crime, and pornography. This makes sense, as the term deepfake was initially coined on Reddit, where it was used to describe the swapping of celebrity faces into pornographic videos (de Ruiter, 2021). The topics identified in the content analysis also overlap with the issues discussed in the literature (Lundberg & Mozelius, 2024), most prominently disinformation related to politics (Hameleers et al., 2022, 2024; Vaccari & Chadwick, 2020). The term synthetic media is almost non-existent in news coverage and has only recently gained currency in policy and academic discourse (Bateman, 2020; McCammon, 2021).

The findings of the content analysis are reflected in our survey experiment. When asked what comes to mind upon hearing the term deepfake or synthetic media, respondents who received the questionnaire with deepfake were more likely to provide an answer than those who received synthetic media, indicating that deepfake is the more widely recognized term. Participants also associated deepfakes with more negative aspects than synthetic media. Furthermore, the term synthetic media is associated with a more positive perception of the technology's benefits than the term deepfake across different domains. However, as the equivalence tests indicate, the risk perception is consistent across most domains, regardless of which term is used to describe the technology.

This raises the question of which term should be used when describing the technology, for instance, in news coverage. Epstein et al. (2023) already provide a good overview of the general perception of labels for AI-generated content; however, they do not focus on risk and benefit perceptions. Our study extends their findings by showing that the term synthetic media can be used to increase the perception of benefits of deepfake technology without lowering the perception of risks. This approach would align with the industry's preference for using more neutral terms for technologies that might otherwise be viewed negatively (Altuncu et al., 2022; PytlikZillig et al., 2018). However, while the potential effect of lowering risk perceptions is small, synthetic media could be used as a euphemism and, if used intentionally to mislead the public about the risks of technology, become doublespeak (Lutz, 1987).

We suggest avoiding the use of synthetic media as a potential euphemism or deepfake as a potential dysphemism. Instead, the context in which the technology is discussed and the specific application field matter. For example, if the technology is discussed in a domain such as politics in relation to

disinformation (Hameleers et al., 2022, 2024; Vaccari & Chadwick, 2020), crime (Lundberg & Mozelius, 2024), or pornography (de Ruiter, 2021), it makes sense to use the term deepfake. However, if the technology's application in a domain such as science, entertainment, or the economy is not inherently harmful, it would be better to use the term synthetic media, as the term deepfake may lower the perceptions of benefits. We also suggest relating the two terms as nested categories rather than mutually excluding alternatives: deepfake can be acknowledged as an emergent, vernacular category for AI-generated content intended to deceive for malicious or playful purposes, while synthetic media can be adopted as a more general technical category which includes deepfakes alongside many other kinds of AI-generated content. The use of the term deepfake to describe other forms of synthetic media (e.g., text-to-speech, multimodal image generation) risks diluting the critical valence of the term. At the same time, using synthetic media as a catch-all for malicious deepfake examples risks downplaying the specific harms associated with such cases and obscuring the broader range of content created through automated means.

This general idea requires further refinement to account for specific application contexts, such as the political uses of the technology. While the use of the term deepfake in relation to disinformation seems reasonable, it might be misleading to describe the use of the technology for producing campaigning material without the intent to spread false information (Jungherr et al., 2025) or to assist politicians with disabilities. For instance, a Swiss National Councilor with a speech disorder used text-to-speech and video synthesis systems to create an avatar through which he communicated to his audience during the election campaign (Blick, 2023). In the material for our content analysis, several articles discussed the use of deepfake technology to produce movies with beneficial effects, including reduced costs and lower carbon emissions. In such cases, the term synthetic media may be a more suitable alternative to deepfake.

The results of our experiment suggest that using the terms deepfake and synthetic media more distinctly may have an advantage. However, from a practical standpoint, whether this approach is feasible remains uncertain. It is unlikely that people or influential actors, such as journalists and politicians, can be forced to adopt the new terminology. As our content analysis reveals, attempting to change the terminology would challenge an already established and increasingly prominent framing (i.e., the term deepfake). Furthermore, dysphemisms like deepfake might be used strategically, for instance by politicians pushing for stricter regulations on AI technology, since alarmist discourse can boost public support for such measures (Jungherr & Rauchfleisch, 2024). Consequently, not all actors have an incentive to adopt more nuanced terminology.

Our study has several limitations that could be addressed in future research. First, we aimed to isolate the

effect of term use alone, without manipulating the domain in which the technology is applied. Future research could further explore this aspect. Second, similar to studies focusing on euphemisms (Walker et al., 2021), future studies could investigate whether specific terms and their use in a particular context are perceived as more or less misleading. Furthermore, other terms (e.g., AI content or AI-generated images) could be tested as alternative euphemisms. Third, we focused solely on risk and benefit perceptions as outcomes, without measuring attitudes toward regulation (Godulla et al., 2021). This could be further investigated, but as our study shows, using different terms primarily affects benefit perception, which might be less relevant than risk perceptions in relation to regulatory support. Finally, we only focused on the Swiss context. Because our survey was conducted in Switzerland, our content analysis was restricted to domestic media to assess how people encountered the two terms. While cultural differences may exist, Epstein et al. (2023) show that many AI-related labels are perceived similarly across cultures in this context. Furthermore, what we have observed in the Swiss case may also apply to other countries where media coverage emphasizes the negative aspects of deepfakes. Therefore, we are confident that our findings have a degree of generalizability. Still, future research could focus on a comparative design that compares the use of the different terms in media coverage across different countries, thereby further testing the cross-cultural generalizability of our findings.

ORCID iDs

Adrian Rauchfleisch  <https://orcid.org/0000-0003-1232-083X>

Daniel Vogler  <https://orcid.org/0000-0002-0211-7574>

Gabriele de Seta  <https://orcid.org/0000-0003-0497-2811>

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Adrian Rauchfleisch's work was supported by the National Science and Technology Council, Taiwan (R.O.C.) (Grant No. 113-2628-H-002-018-) and by the Taiwan Social Resilience Research Center (Grant No. 114L9003) from the Higher Education Sprout Project by the Ministry of Education in Taiwan. Daniel Vogler's research and the data collection for the project were funded by the Swiss Foundation for Technology Assessment (TA-SWISS). Gabriele de Seta's work was supported by a Trond Mohn Foundation Starting Grant (TMS2024STG03).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability statement

Data and replication code are publicly available at the project's OSF repository: <https://osf.io/mgfb/files/osfstorage> The preregistration is available on AsPredicted: <https://aspredicted.org/5m56-2z9c.pdf>

Supplemental material

Supplemental material for this article is available online.

Notes

1. Our sample size is larger than the preregistered sample size of 660. This was an accidental deviation from our preregistration. We explain the details in Supplemental Appendix B.2, where we estimate all models using a sample of only 660. The results are the same as those of our larger sample size.
2. The only deviation from the preregistration here is that we keep risk at the individual level with an α of only .69 instead of splitting it up. The results for individual items are the same as for the index; they are not significant.
3. The code and data to replicate all analyses are available at <https://osf.io/mgfb/files/osfstorage>
4. We get the same results (OR=11.63) if we only consider responses and omit non-responses instead of adding them as non-negative answers.

References

- Allyn, B. (2022, March 16). Deepfake video of Zelenskyy could be "tip of the iceberg" in info war, experts warn. *NPR*. <https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia>
- Altay, S., & Gilardi, F. (2024). People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation. *PNAS Nexus*, 3(10), 403. <https://doi.org/10.1093/pnasnexus/pgae403>
- Altuncu, E., Franqueira, V. N. L., & Li, S. (2022, August 21). Deepfake: Definitions, performance metrics and standards, datasets and benchmarks, and a meta-review. *arXiv.Org*. <https://doi.org/10.3389/fdata.2024.1400024>
- Bao, L., Krause, N. M., Calice, M. N., Scheufele, D. A., Wirz, C. D., Brossard, D., Newman, T. P., & Xenos, M. A. (2022). Whose AI? How different publics think about AI and its social impacts. *Computers in Human Behavior*, 130, 107182. <https://doi.org/10.1016/j.chb.2022.107182>
- Bateman, J. (2020). *Deepfakes and synthetic media in the financial system: Assessing threat scenarios*. Carnegie Endowment for International Peace. <https://www.jstor.org/stable/resrep25783.1>
- Bendahan Bitton, D., Hoffmann, C. P., & Godulla, A. (2024). Deepfakes in the context of AI inequalities: Analysing disparities in knowledge and attitudes. *Information, Communication & Society*, 28(2), 295–315. <https://doi.org/10.1080/1369118X.2024.2420037>
- Binder, A. R., Cacciatore, M. A., Scheufele, D. A., Shaw, B. R., & Corley, E. A. (2012). Measuring risk/benefit perceptions of emerging technologies and their potential impact on communication of public opinion toward science. *Public Understanding of Science*, 21(7), 830–847. <https://doi.org/10.1177/0963662510390159>
- Blick. (2023). *SP-Politiker Islam Alijaj lässt seinen Avatar für sich sprechen* [SP politician Islam Alijaj lets his avatar speak for himself]. <https://www.blick.ch/politik/sp-nationalratskandidat-islam-alijaj-er-laesst-den-avatar-fuer-sich-sprechen-id18977675.html>
- Bowers, J. S., & Pleydell-Pearce, C. W. (2011). Swearing, euphemisms, and linguistic relativity. *PLoS ONE*, 6(7), Article e22341. <https://doi.org/10.1371/journal.pone.0022341>

- Casas Gómez, M. (2009). Towards a new approach to the linguistic definition of euphemism. *Language Sciences*, 31(6), 725–739. <https://doi.org/10.1016/j.langsci.2009.05.001>
- Cobb, M. D., & Macoubrie, J. (2004). Public perceptions about nanotechnology: Risks, benefits and trust. *Journal of Nanoparticle Research*, 6(4), 395–405. <https://doi.org/10.1007/s11051-004-3394-4>
- Cochran, J. D., & Napshin, S. A. (2021). Deepfakes: Awareness, concerns, and platform accountability. *Cyberpsychology, Behavior, and Social Networking*, 24(3), 164–172. <https://doi.org/10.1089/cyber.2020.0100>
- Covello, V. T. (1983). The perception of technological risks: A literature review. *Technological Forecasting and Social Change*, 23(4), 285–297. [https://doi.org/10.1016/0040-1625\(83\)90032-X](https://doi.org/10.1016/0040-1625(83)90032-X)
- de Groot, J. I. M., Steg, L., & Poortinga, W. (2013). Values, perceived risks and benefits, and acceptability of nuclear energy. *Risk Analysis*, 33(2), 307–317. <https://doi.org/10.1111/j.1539-6924.2012.01845.x>
- de Ruiter, A. (2021). The distinct wrong of deepfakes. *Philosophy & Technology*, 34(4), 1311–1332. <https://doi.org/10.1007/s13347-021-00459-2>
- de Seta, G. (2024). Synthetic probes: A qualitative experiment in latent space exploration. *Sociologica*, 18(2), 9–23. <https://doi.org/10.6092/ISSN.1971-8853/19512>
- Epstein, Z., Fang, M. C., Arechar, A. A., & Rand, D. (2023). *What label should be applied to content produced by generative AI?* OSF. <https://doi.org/10.31234/osf.io/v4mfz>
- Farrow, K., Grolleau, G., & Mzoughi, N. (2021). ‘Let’s call a spade a spade, not a gardening tool’: How euphemisms shape moral judgement in corporate social responsibility domains. *Journal of Business Research*, 131, 254–267. <https://doi.org/10.1016/j.jbusres.2021.04.002>
- Florian, A. (2023, January 11). I think my face was deepfaked into a Chinese camping stove ad. *Wired*. <https://www.wired.com/story/china-deepfake-advertising-policy/>
- Frewer, L. J., Howard, C., Hedderley, D., & Shepherd, R. (1999). Reactions to information about genetic engineering: Impact of source characteristics, perceived personal relevance, and persuasiveness. *Public Understanding of Science*, 8(1), 35. <https://doi.org/10.1088/0963-6625/8/1/003>
- Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30), Article e2305016120. <https://doi.org/10.1073/pnas.2305016120>
- Gladney, G. A., & Rittenburg, T. L. (2005). Euphemistic text affects attitudes, behavior. *Newspaper Research Journal*, 26(1), 28–41. <https://doi.org/10.1177/073953290502600104>
- Glick, J. (2023, January 13). AI isn’t Hollywood’s villain—It’s a flawed hero. *Wired*. <https://www.wired.com/story/artificial-intelligence-film/>
- Godulla, A., Hoffmann, C. P., & Seibert, D. (2021). Dealing with deepfakes—An interdisciplinary examination of the state of research and implications for communication studies. *SCM Studies in Communication and Media*, 10(1), 72–96. <https://doi.org/10.5771/2192-4007-2021-1-72>
- Gosse, C., & Burkell, J. (2020). Politics and porn: How news media characterizes problems presented by deepfakes. *Critical Studies in Media Communication*, 37(5), 497–511. <https://doi.org/10.1080/15295036.2020.1832697>
- Graefe, A., Haim, M., Haarmann, B., & Brosius, H.-B. (2018). Readers’ perception of computer-generated news: Credibility, expertise, and readability. *Journalism*, 19(5), 595–610. <https://doi.org/10.1177/1464884916641269>
- Hameleers, M., Van Der Meer, T. G. L. A., & Dobber, T. (2022). You won’t believe what they just said! The effects of political deepfakes embedded as vox populi on social media. *Social Media + Society*, 8(3), 205630512211163. <https://doi.org/10.1177/20563051221116346>
- Hameleers, M., van der Meer, T. G. L. A., & Dobber, T. (2024). They would never say anything like this! Reasons to doubt political deepfakes. *European Journal of Communication*, 39(1), 56–70. <https://doi.org/10.1177/02673231231184703>
- Jacobsen, B. N., & Simpson, J. (2024). The tensions of deepfakes. *Information, Communication & Society*, 27(6), 1095–1109. <https://doi.org/10.1080/1369118X.2023.2234980>
- Jungherr, A. (2023). Artificial intelligence and democracy: A conceptual framework. *Social Media + Society*, 9(3), 20563051231186353. <https://doi.org/10.1177/20563051231186353>
- Jungherr, A., & Rauchfleisch, A. (2024). Negative downstream effects of alarmist disinformation Discourse: Evidence from the United States. *Political Behavior*, 46(4), 2123–2143. <https://doi.org/10.1007/s11109-024-09911-3>
- Jungherr, A., & Rauchfleisch, A. (2025). *Artificial intelligence in deliberation: The AI penalty and the emergence of a new deliberative divide* (No. arXiv:2503.07690). arXiv. <https://doi.org/10.48550/arXiv.2503.07690>
- Jungherr, A., Rauchfleisch, A., & Wuttke, A. (2025). *Artificial Intelligence in election campaigns: Perceptions, penalties, and implications* (No. arXiv:2408.12613). arXiv. <https://doi.org/10.48550/arXiv.2408.12613>
- Kalpokas, I. (2021). Problematising reality: The promises and perils of synthetic media. *SN Social Sciences*, 1(1), 1. <https://doi.org/10.1007/s43545-020-00010-8>
- Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, 63(2), 135–146. <https://doi.org/10.1016/j.bushor.2019.11.006>
- Kleine, F. (2022). *Perception of deepfake technology—The influence of the recipients’ affinity for technology on the perception of deepfakes* [Master thesis, University of Applied Sciences Darmstadt]. Researchgate. <https://www.researchgate.net/publication/364254498>
- Łabuz, M., & Nehring, C. (2024). Information apocalypse or overblown fears—What AI mis- and disinformation is all about? Shifting away from technology toward human reactions. *Politics & Policy*, 52(4), 874–891. <https://doi.org/10.1111/polp.12617>
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 1(2), 259–269. <https://doi.org/10.1177/2515245918770963>
- Lundberg, E., & Mozelius, P. (2024). The potential effects of deepfakes on news media and entertainment. *AI & SOCIETY*, 40, 2159–2170. <https://doi.org/10.1007/s00146-024-02072-1>
- Lutz, W. D. (1987). Language, appearance, and reality: Doublespeak in 1984. *ETC: A Review of General Semantics*, 44(4), 382–391.
- McCammon, M. N. (2021). The emerging threat of synthetic media: A consideration of journalists’ responsibilities. In L. Corredoira, I. B. Mallén, & R. C. Presuel (Eds.), *The handbook*

- of communication rights, law, and ethics (1st ed., pp. 292–302). Wiley. <https://doi.org/10.1002/9781119719564.ch24>
- McGlone, M. S., Beck, G., & Pfister, A. (2006). Contamination and camouflage in euphemisms. *Communication Monographs*, 73(3), 261–282. <https://doi.org/10.1080/03637750600794296>
- Nadal, L. de., & Jančárik, P. (2024). *Beyond the deepfake hype: AI, democracy, and “the Slovak case.” Harvard Kennedy School Misinformation Review*. <https://doi.org/10.37016/mr-2020-153>
- Neyazi, T. A., Nadaf, A. H., Tan, K. E., & Schroeder, R. (2024). Does trust in government moderate the perception towards deepfakes? Comparative perspectives from Asia on the risks of AI and misinformation for democracy. *Government Information Quarterly*, 41(4), 101980. <https://doi.org/10.1016/j.giq.2024.101980>
- Paris, B., & Donovan, J. (2019). *Deepfakes and cheap fakes: The manipulation of audio and visual evidence*. Data & Society.
- Pawelec, M. (2022). Deepfakes and democracy (Theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions. *Digital Society*, 1(2), 19. <https://doi.org/10.1007/s44206-022-00010-6>
- PytlíkZillig, L. M., Duncan, B., Elbaum, S., & Detweiler, C. (2018). A drone by any other name: Purposes, end-user trustworthiness, and framing, but not terminology, affect public support for drones. *IEEE Technology and Society Magazine*, 37(1), 80–91. <https://doi.org/10.1109/MTS.2018.2795121>
- Said, N., Potinteu, A. E., Brich, I., Buder, J., Schumm, H., & Huff, M. (2023). An artificial intelligence perspective: How knowledge and confidence shape risk and benefit perception. *Computers in Human Behavior*, 149, 107855. <https://doi.org/10.1016/j.chb.2023.107855>
- Sartori, L., & Bocca, G. (2023). Minding the gap(s): Public perceptions of AI and socio-technical imaginaries. *AI & SOCIETY*, 38(2), 443–458. <https://doi.org/10.1007/s00146-022-01422-1>
- Siegrist, M., & Visschers, V. H. M. (2013). Acceptance of nuclear power: The Fukushima effect. *Energy Policy*, 59, 112–119. <https://doi.org/10.1016/j.enpol.2012.07.051>
- Simon, F. M., McBride, K., & Altay, S. (2024). *AI's impact on elections is being overblown*. MIT Technology Review. <https://www.technologyreview.com/2024/09/03/1103464/ai-impact-elections-overblown/>
- Toff, B., & Simon, F. (2023). “Or they could just not use it?”: *The paradox of AI disclosure for audience trust in news* (pp. 1–38). SocArXiv.
- Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1), 1–13. <https://doi.org/10.1177/2056305120903408>
- Udris, L., Rivière, M., Fürst, S., & Eisenegger, M. (2024). Reuters Institute Digital News Report 2024: Länderbericht Schweiz. Zürich: fög – Forschungszentrum Öffentlichkeit und Gesellschaft/Universität Zürich. <https://doi.org/10.5167/uzh-261185>
- Wahl-Jorgensen, K., & Carlson, M. (2021). Conjecturing fearful futures: Journalistic discourses on deepfakes. *Journalism Practice*, 15(6), 803–820. <https://doi.org/10.1080/17512786.2021.1908838>
- Walker, A. C., Turpin, M. H., Meyers, E. A., Stolz, J. A., Fugelsang, J. A., & Koehler, D. J. (2021). Controlling the narrative: Euphemistic language affects judgments of actions while avoiding perceptions of dishonesty. *Cognition*, 211, 104633. <https://doi.org/10.1016/j.cognition.2021.104633>
- Wang, S., & Kim, S. (2022). Users’ emotional and behavioral responses to deepfake videos of K-pop idols. *Computers in Human Behavior*, 134, 107305. <https://doi.org/10.1016/j.chb.2022.107305>
- Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, 9(11), 40–53. <https://doi.org/10.22215/timreview/1282>
- Wilson, R. J., Paterson, P., Jarrett, C., & Larson, H. J. (2015). Understanding factors influencing vaccination acceptance during pregnancy globally: A literature review. *Vaccine*, 33(47), 6420–6429. <https://doi.org/10.1016/j.vaccine.2015.08.046>
- Wittenberg, C., Epstein, Z., Peloquin-Skulski, G., Berinsky, A. J., & Rand, D. (2024). *Labeling AI-Generated Media Online*. OSF. <https://doi.org/10.31234/osf.io/b238p>
- Wutke, A., Rauchfleisch, A., & Jungherr, A. (2025). *Artificial Intelligence in government: Why people feel they lose control* (No. arXiv:2505.01085). arXiv. <https://doi.org/10.48550/arXiv.2505.01085>
- Yadlin-Segal, A., & Oppenheim, Y. (2021). Whose dystopia is it anyway? Deepfakes and social media regulation. *Convergence: The International Journal of Research into New Media Technologies*, 27(1), 36–51. <https://doi.org/10.1177/1354856520923963>
- Zhang, B., & Dafoe, A. (2019). *Artificial intelligence: American attitudes and trends* (SSRN Scholarly Paper 3312874). Social Science Research Network. <https://doi.org/10.2139/ssrn.3312874>

Author biographies

Adrian Rauchfleisch (Ph.D. University of Zurich) is an Associate Professor at the Graduate Institute of Journalism, National Taiwan University. His research focuses on the interplay of politics, technology, and journalism in Asia, Europe, and the United States. His new project explores Artificial Intelligence’s influence on society across different cultural contexts.

Daniel Vogler (Ph.D. University of Zurich) is the Research Director of the Research Center for the Public Sphere and Society (fög) at the University of Zurich and a Senior Research Associate at the Department of Communication and Media Research (IKMZ) at the University of Zurich. His research focuses on crisis communication, public relations, journalism, online communication, and computational communication science.

Gabriele de Seta (Ph.D. Hong Kong Polytechnic University) is a Researcher at the University of Bergen, where he leads the ALGOFOLK project (“Algorithmic folklore: The mutual shaping of vernacular creativity and automation”). His research work, grounded on qualitative and ethnographic methods, focuses on digital media practices, sociotechnical infrastructures and vernacular creativity in the Chinese-speaking world.