journalism and media

MDPI

Review

From Deepfakes to Different Types of Audiovisual Synthetic Media: Toward a Theoretical Framework for Conceptual Differentiation and Normative Evaluation

Daniel Bendahan Bitton [1,*] and Patric Raemy [2]

[1] Institute for Communication and Media Studies, Leipzig University, 04109 Leipzig, Germany
[2] Department of Communication and Media Research, University of Fribourg, 1700 Fribourg, Switzerland; patric.raemy@unifr.ch
* Correspondence: daniel.bendahan@uni-leipzig.de

Abstract

Deepfakes are gaining increased attention in society. To date, however, the understanding of what is commonly referred to as a deepfake still lacks consistency. Furthermore, the term has a negative connotation due to its association with fake news and its origin in pornographic contexts. Hence, its ambiguous nature affects the debate about the technology in research and practice. This study explores how deepfakes and synthetic media are defined in previous research, what criteria are considered for their evaluation, and how they relate to communicative norms. Based on this, the study proposes a framework to differentiate forms of audiovisual synthetic media based on their production process, descriptive evaluation, relation to reality and normative evaluation. The study suggests distinguishing between three ideal types of audiovisual synthetic media: deepfakes, deepfiction and deeprealism. The conceptual framework provides a structured interpretative basis for distinguishing synthetic media. Finally, this study focuses on the audiences' perspective and discusses potential outcomes of synthetic media.

Keywords: deepfake; synthetic media; AI; framework; communicative norms

1. Introduction

Since the emergence and spread of pornographic deepfakes on Reddit in 2017, synthetic media have gained enormous attention in research and the public sphere (Citron & Chesney, 2019). As with the emergence of any new technology, hopes and opportunities increase but so do concerns about potential risks. Because of its origins, the term deepfake is already loaded with negative connotations and is therefore usually viewed with both fascination and unease; the public and academic discourses mostly focus on the use of deepfakes for misleading and manipulative purposes, including media manipulations (Westerlund, 2019), disinformation and negative implications for democracy and security (Citron & Chesney, 2019; Weikmann & Lecheler, 2023b).

Hence, ever since the technology's emergence, researchers have been working on developing identification tools and legal frameworks to fight harmful applications of the technology. Studies in the social sciences revealed that the audiences have limited knowledge of deepfakes and a strong perception of risks associated with the technology (Bendahan Bitton et al., 2024; Godulla et al., 2021; Rauchfleisch et al., 2025). Further, individuals often struggle to recognize deepfakes and differentiate them from real videos (e.g., Bray et al., 2023). Previous studies suggest that encounters with deepfakes can increase skepticism

Check for updates

Academic Editor: Andreu Casero-Ripollés

Received: 5 October 2026
Revised: 19 November 2026
Accepted: 19 November 2026
Published: 21 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

towards (political) audiovisual content, reduce trust in news online (Ahmed, 2023; Vaccari & Chadwick, 2020), and contribute to media cynicism (Hoffmann et al., 2025).

Yet, while deepfakes are frequently associated with negative applications, the technology itself is neutral and can be used to create synthetic media that are based on positive rather than negative intentions. These include uses of the technology in the fields of entertainment, gaming and film, advertising, marketing, education, and even journalism (e.g., Whittaker et al., 2023). While some authors more generally refer to synthetic media when talking about positive applications of deep-learning technologies (e.g., Brady & Meyer-Resende, 2020), the term deepfake is more popular but is used rather inconsistently: some refer to deepfakes in the context of beneficial applications, others use the term exclusively for dangerous applications with negative intentions that do not allow for any beneficial use at all (e.g., Vaccari & Chadwick, 2020). Similarly, the question arises as to what extent the potential harms of the use of the technology relate solely to what is considered a "deepfake" or apply to synthetic media in general.

Despite the considerable growth of research, scholarly discourse lacks a sufficiently nuanced conceptualization of deepfakes and synthetic media. At the same time, the current debate often uses a non-academic term ("deepfake"), originating in pornography and prominently featuring the word "fake" to describe a relatively new form of media and communication (de Ruiter, 2021). This is a problem not only for communication science but for communication-related institutions and actors, too, since the terminological uncertainty also affects the uncertainty about the legitimate use of this new form of media.

As a result, there is still limited clarity not only regarding the terminology of deepfake and its relation to synthetic media, but also regarding criteria by which applications of these technologies should be evaluated. Hence, on the one hand, it remains unclear when it is conceptually appropriate to speak of a deepfake and how the potential harmfulness of a deepfake can be assessed (e.g., Wang, 2026; Rauchfleisch et al., 2025). On the other hand, it also remains unclear according to which criteria the normative legitimacy of synthetic media can be evaluated and how legitimate or beneficial uses can be distinguished from potentially harmful ones (e.g., Geise et al., 2025; Raemy et al., 2025). The goal of this research, therefore, is to provide a theoretical framework that helps to bring more clarity to the current situation by answering the following questions:

RQ1: What criteria are considered in previous literature to characterize deepfakes and synthetic media?
RQ2: How can different forms of synthetic media be evaluated based on these criteria and their relationship to communicative norms?

First, based on a critical literature review, definitions of deepfakes and synthetic media are systematized, similarities identified, and discrepancies uncovered. Second, the dimensions identified in the review are further specified and theoretically extended by discussing the relation of synthetic media to communicative norms. Third, a framework is proposed that distinguishes between different types of audiovisual synthetic media (deepfake, deepfiction, and deeprealism). Finally, the audiences' reception and potential outcomes of synthetic media are discussed.

2. Critical Review

To answer the research questions, a critical review drawing on Grant and Booth (2009) and Snyder (2019) was conducted. Critical reviews examine and synthesize conceptually relevant literature with the aim of further developing the theoretical understanding of a topic (Snyder, 2019). Unlike systematic review approaches, the focus is therefore not on exhaustive coverage of all available studies on a subject, but on the concep-

tual relevance of the literature and its contribution to framework development (Grant & Booth, 2009). This study draws on the approaches of Tandoc et al. (2017), who compared how scholars define the term "fake news" to identify recurring dimensions and develop a typology, and Westerlund (2019), who conducted an interdisciplinary synthesis of deepfake research.

The review builds on a body of literature on deepfakes from several years of research on the topic by the authors (e.g., Bendahan Bitton et al., 2024, 2026; Godulla et al., 2021; Hoffmann et al., 2025; Raemy et al., 2025; Seibert, 2023) and extends previous literature reviews (e.g., Godulla et al., 2021; Wang, 2026). The literature review covered by these publications and the authors' own published studies served as the starting point. Publications from this body of literature were selected based on their relevance to the research questions and the inclusion criteria outlined below. Subsequently, targeted snowballing and exploratory database searches in Scopus and Web of Science were conducted to identify additional conceptually relevant publications. The database searches took place in April 2026 and focused on publications from 2021 onward, complementing the period from 2018 to 2020 covered by the authors' previous systematic literature review (Godulla et al., 2021). Exploratory keyword searches were conducted using two strings. Both included the terms "deepfake" OR "deep fake" OR "synthetic media" OR "AI-generated media" OR "AI-based media" and their German equivalents ("synthetische Medien" OR "KI-generierte Medien" OR "KI-basierte Medien"). The first search string combined these terms with keywords related to conceptualizations by adding "defin*" OR "concept*" OR "classif*" OR "categor*" OR "typolog*" and their German equivalents ("konzept*" OR "klassif*" OR "kategor*"). The second search string combined them with keywords addressing normative perspectives by adding "ethic*" OR "norm*" OR "moral*" and their German equivalents ("ethik*" OR "ethisch*").

The exploratory search results were initially assessed based on titles, abstracts and keywords to identify publications addressing the conceptualization or normative evaluation of deepfakes and/or synthetic media. Potentially relevant publications were subsequently examined in full text. Publications were included if they (1) provided, extended, or discussed definitions, conceptual distinctions, or classifications of deepfakes and/or synthetic media or (2) discussed normative or ethical questions relevant to evaluating these phenomena (e.g., intentions, transparency or consent). Publications were excluded if they merely adopted a definition from other authors without theoretical reasoning or focused exclusively on technical detection or legal analysis without a conceptual differentiation or a normative evaluation. A total of 31 publications were considered in the critical review.

For framework development, first, criteria used to conceptualize deepfakes and synthetic media were inductively derived from the articles. Second, recurring elements were compared and grouped (see Table 1). Third, the identified criteria were critically examined to identify conceptual overlaps, inconsistencies, and limitations. For example, existing definitions often conflate technological sophistication and authenticity or rely on the creators' intentions, which are difficult to observe. Based on this assessment, criteria were either retained or further differentiated (as discussed in Section 3). Fourth, communicative norms were introduced as a theoretical reference point to address the limitations of definitions relying primarily on creators' intentions and to link the identified characteristics to expectations of truthful and transparent communication (see Section 3.1).

Table 1. Summary of criteria for conceptualizing deepfakes and synthetic media.

Focus	Definition	References
Classification	Synthetic media as an umbrella term, deepfakes as a (negative) form of synthetic media	(Bateman, 2020; Bendahan Bitton et al., 2024, 2026; Firc et al., 2023; Hoffmann et al., 2025; Karaboga, 2024; Pawelec, 2024; Rauchfleisch et al., 2025; Vaccari & Chadwick, 2020; van der Sloot & Wagensveld, 2022; Wang, 2026; Whittaker et al., 2023; Whittaker et al., 2020)
	Synonymous use of the terms deepfake and synthetic media	(Hameleers et al., 2023; Millière, 2022)
	Negative connotation of the term deepfake and definitional problems	(Bateman, 2020; Bray et al., 2023; Jacobsen & Simpson, 2023; Pawelec, 2024; Rauchfleisch et al., 2025; van der Sloot & Wagensveld, 2022; Wang, 2026)
Creation method	Differentiation from simple forms of media manipulation such as cheap fakes and shallow fakes	(Brady & Meyer-Resende, 2020; Karaboga, 2024; Paris & Donovan, 2019; Wang, 2026)
	Focus on the technology behind deepfakes (AI, machine learning, and deep-learning technology)	(Bateman, 2020; Bendahan Bitton et al., 2024; Brady & Meyer-Resende, 2020; Bray et al., 2023; de Ruiter, 2021; Habgood-Coote, 2023; Hameleers et al., 2023; Jacobsen & Simpson, 2023; Karaboga, 2024; Millière, 2022; Paris & Donovan, 2019; Westerlund, 2019; Wang, 2026)
Media Types	Differentiation of types of deepfakes (photo, audio, video, audio & video; face swap, face replacement, face reenactment, face generation; speech synthesis; voice swap; text to speech)	(Bendahan Bitton et al., 2024, 2026; Firc et al., 2023; Hameleers et al., 2023; Karaboga, 2024; Kietzmann et al., 2020; Pawelec, 2024; Rauchfleisch et al., 2025; Thaw et al., 2020; van der Sloot & Wagensveld, 2022; Whittaker et al., 2020)
Depicted Subject and Object	Focus on the imitation of existing individuals (mimic real people, fake someone's representation); the depiction of fictive individuals and objects	(Ahmed, 2023; Bateman, 2020; Bendahan Bitton et al., 2024, 2026; de Ruiter, 2021; Hameleers et al., 2023; Hoffmann et al., 2025; Karaboga, 2024; Paris & Donovan, 2019; Pawelec, 2024; Rauchfleisch et al., 2025; Weikmann & Lecheler, 2023b; Westerlund, 2019)
Intention	Focus on the intention behind deepfakes and the (missing) consent of the depicted individuals	(de Ruiter, 2021; Habgood-Coote, 2023; Hoffmann et al., 2025; Karaboga, 2024; McCammon, 2021; Wang, 2026)
	Understanding deepfakes as a form of disinformation (malicious in nature, aiming to deceive)	(Ahmed, 2023; Hameleers, 2024; Hameleers & Marquart, 2023; Paris & Donovan, 2019; Vaccari & Chadwick, 2020; Weikmann & Lecheler, 2023b)
Quality and Authenticity	Focusing on the quality and authenticity (photorealistic media; various degrees of authenticity, portraying individuals performing inauthentic things)	(Bateman, 2020; Bendahan Bitton et al., 2026; Brady & Meyer-Resende, 2020; de Ruiter, 2021; Godulla et al., 2021; Hameleers et al., 2023; Hoffmann et al., 2025; Jacobsen & Simpson, 2023; Karaboga, 2024; Pawelec, 2024; van der Sloot & Wagensveld, 2022; Weikmann & Lecheler, 2023b; Westerlund, 2019; Whittaker et al., 2023; Wang, 2026)

2.1. Deepfake or Synthetic Media?

While "deepfake" has become a widely established and discussed term in both research and practice, the term "synthetic media" is primarily addressed in a definitional sense in relation to deepfakes. Furthermore, the classification of the term deepfake reveals terminological inconsistencies, which also reflect the different disciplinary perspectives and research interests through which the phenomenon is approached. On the one hand, synthetic media is used as an umbrella term that encompasses "any form of media generated by AI, including, video, audio, images and text" (Brady & Meyer-Resende, 2020, p. 2). In this context, deepfakes are understood as a subgroup of synthetic media (e.g., Bateman, 2020; Bendahan Bitton et al., 2024, 2026; Hoffmann et al., 2025; Karaboga, 2024). On the other hand, the terms deepfake and synthetic media are used synonymously (e.g., Hameleers et al., 2023; Millière, 2022; Pawelec, 2024). For example, Millière (2022) argues that synthetic media are "often subsumed colloquially under the label deepfakes" (p. 1). Further, Bateman (2020) and Wang (2026) describe deepfake as a colloquial term for which there is still no standardized definition. Overall, it is unclear whether deepfakes can be considered equivalent to synthetic media or whether deepfakes should be understood as a specific (negative) form of synthetic media. Uncertainty in the differentiation between these two terms also occurs, as many risks associated with deepfakes (e.g., fraud, manipulation of public opinion; Citron & Chesney, 2019) might apply to synthetic media in general (van der Sloot & Wagensveld, 2022). Finally, Rauchfleisch et al. (2025) show that the term deepfake is associated with negative aspects, whereas the term synthetic media evokes more positive perceptions of the technology's benefits across different fields of application. In summary, these inconsistencies show that greater terminological clarity is needed regarding how deepfakes are conceptually situated within the broader field of synthetic media.

2.2. Creation Methods

Definitions of "deepfake" often focus on the underlying production processes: the use of AI in general (e.g., Bateman, 2020; Brady & Meyer-Resende, 2020) or the use of deep-learning technology as a method of machine learning (e.g., Bendahan Bitton et al., 2024; Habgood-Coote, 2023; Hameleers et al., 2023; Wang, 2026). Furthermore, the automation of the creation of deepfakes and limited human supervision needed in this process are considered in this context (Bray et al., 2023; Jacobsen & Simpson, 2023). At this point, a distinction is made between deepfakes and simpler forms of manipulation of audiovisual content. So-called "cheap fakes" (also described as "shallow fakes") refer to a low level of manipulation of audiovisual media created using easily accessible software. This allows content to be slowed down, sped up, or presented in a different context (Brady & Meyer-Resende, 2020; Karaboga, 2024; Wang, 2026). Advances in generative AI, however, have made it possible to produce sophisticated audiovisual content with comparatively limited technical effort (Millière, 2022). Consequently, deepfakes should not only be situated within the broader field of synthetic media but also distinguished from related forms such as cheap fakes and shallow fakes based on the underlying production technology. Without such clarification, the similarity of these terms may lead to confusion and imprecise use.

2.3. Media Types

Given the audiovisual nature of deepfakes, different types are described. These often relate directly to the function of the technology to imitate faces, voices, and change the overall portrayal of a person. Hence, photo, audio, video, and combined audio-video deepfakes are distinguished (e.g., Bendahan Bitton et al., 2026; Karaboga, 2024; Kietzmann

et al., 2020; Pawelec, 2024). Photo and video deepfakes mainly refer to face-swapping technology that allows for the replacement of someone's face (or body) with that of someone else (e.g., Hameleers et al., 2023; Whittaker et al., 2020). In addition, photo and video deepfakes refer to altering facial features (face re-enactment) and the generation of entirely new faces (Firc et al., 2023). Focusing on audio deepfakes, the technology holds the possibility to swap, alter, or imitate voices, as well as to apply text-to-speech and speech-to-speech synthesis (Firc et al., 2023). Furthermore, Kietzmann et al. (2020) and Whittaker et al. (2020) describe "sophisticated deepfakes" combining both visual and audio content and their ability to create lip-syncs by "manipulat[ing] mouth movements and facial expressions alongside audio materials to have someone convincingly say things which they have never said" (Whittaker et al., 2020, pp. 91–92). Overall, this suggests that deepfakes should not be equated solely with sophisticated videos, as the term is also used for audio, image-based, and multimodal forms in which various media components may be altered.

2.4. Depicted Subjects and Objects

While deepfake technology enables users to create audiovisual content of any object (e.g., Brady & Meyer-Resende, 2020; Bray et al., 2023), definitions mostly center on the imitation of existing persons as a key characteristic (e.g., Bendahan Bitton et al., 2024; Hameleers et al., 2023; Hoffmann et al., 2025; Karaboga, 2024; Pawelec, 2024). As different media types of deepfakes demonstrate, applications of the technology often refer to the imitation of faces and voices of an individual. Further, the manipulation of narratives and mannerisms of a person depicted in a deepfake is part of conceptual approaches. Hence, deepfakes are described as being used "to fake someone's entire audio-visual representation" (Weikmann & Lecheler, 2023a, p. 4) and to depict real people performing inauthentic, untrue things that do not reflect their usual opinions, attitudes, or actions (Ahmed, 2023; Hameleers et al., 2023). In contrast, deepfake technology might not only be used to negatively portray real persons but also to create entirely fictitious individuals (Bateman, 2020). Therefore, van der Sloot and Wagensveld (2022) urge for the consideration of whether the depicted subject is real or fictitious when evaluating a deepfake. Consequently, this suggests that deepfakes are primarily associated with the manipulation of real individuals instead of fictional persons or objects. Regarding a conceptual evaluation of synthetic media, this indicates that it is important to consider to what extent the depicted person or object relates to reality.

2.5. Intention Behind the Creation of Deepfakes

Classifications of deepfakes often address the (assumed) intention behind their creation. In this context, deepfakes are considered malicious in nature (Hameleers & Marquart, 2023) and constitute a form of deception (Hameleers et al., 2023). Weikmann and Lecheler (2023a) argue that deepfakes "are meant to deceive and provide false evidence for something that never happened" (p. 4). Several authors share this assumption and conceptualize deepfakes as a new form of visual disinformation (Hameleers, 2024; Vaccari & Chadwick, 2020; Weikmann & Lecheler, 2023b). de Ruiter (2021) emphasizes the moral implications of deepfakes, arguing that although deepfakes are morally suspect, they are not inherently unethical. Rather, it should be considered whether the people depicted in a deepfake gave their consent and what the intent behind its creation was (de Ruiter, 2021; Hoffmann et al., 2025).

This focus on the intent behind the creation of deepfakes is also reflected in the distinction between disinformation and misinformation. Disinformation refers to false or misleading information that is intentionally created or disseminated to deceive or cause

harm, whereas misinformation refers to false or misleading information that is shared without such intent (Spies, 2019; Möller et al., 2020; Wardle & Derakhshan, 2017). However, making a distinction based on intent also poses a challenge: intent is notoriously difficult to observe, measure, or verify (Spies, 2019). McCammon (2021) argues that it is not suitable to use the term deepfake to refer to content created with the aim to deceive people "because it shifts the conversation from what such content is and assumes too much about what each person involved in circulating it intends for it to do" (p. 294). Finally, this use of the term excludes deepfakes that are not convincing or are simply used for educational or entertaining purposes (e.g., satirical videos; Habgood-Coote, 2023).

Overall, these studies center on the presumed intentions of deepfake creators and primarily conceptualize them as a new form of disinformation. However, deepfakes should not be classified solely based on their creators' intentions, as these intentions remain ambiguous and cannot be reliably inferred from the created media alone. This highlights the need for a more comprehensive perspective that considers further characteristics of deepfakes as well as the context in which they are used.

2.6. Quality and Authenticity of Deepfakes

Finally, categorizations of deepfakes emphasize the technological processes that can achieve realistic results and imitations of real people that can achieve a high level of credibility (Karaboga, 2024). Hence, deepfakes are often described as highly realistic and sophisticated (e.g., Weikmann & Lecheler, 2023b; Wang, 2026) or photo-realistic (Jacobsen & Simpson, 2023). In contrast, however, there is only limited acknowledgement that deepfakes might be authentic only to varying degrees (Bateman, 2020; Hoffmann et al., 2025; Pawelec, 2024; Wang, 2026). While deepfakes are often described as highly realistic and deceptive in research, their actual characteristics vary considerably. Attempts to use deepfakes to exert political influence have often been unsuccessful due to flaws in quality and because they were quickly exposed as fakes. In this context, Hameleers et al. (2023) argue that limited effects of deepfakes are not solely the result of technological flaws but also the lack of plausibility. For example, if a politician is portrayed in a deepfake delivering a speech containing radical statements that starkly contradict their established political positions, the video might be of high quality but would still likely lose credibility among viewers due to its implausibility (Hameleers et al., 2023). Therefore, van der Sloot and Wagensveld (2022) proposed to assess the extent to which the manipulation affects the reliability of the conveyed information.

Taken together, this suggests that existing definitions of deepfakes often treat technological sophistication ("quality"), authenticity, and plausibility as closely intertwined, even though these dimensions are not analytically identical. Deepfakes may appear technically sophisticated without necessarily being plausible, and they may also vary in their degree of authenticity. Therefore, a more comprehensive conceptualization is needed that distinguishes more clearly between these characteristics.

3. Discussion of the Critical Review

The critical review shows that there is no consistent understanding of the term deepfake. Existing studies approach the phenomenon from various disciplines and therefore focus on different aspects, including technological processes, legal and ethical questions, the role of the creator's intention, and effects on the audiences due to the media's perceived credibility. As a result, the term deepfake is currently used in conceptually inconsistent ways that are difficult to delimit across contexts. Based on the review, several recurring dimensions through which deepfakes and synthetic media can be evaluated were identified—including their creation method, type of media, depicted subjects and objects, intention as well

as quality and authenticity. However, these dimensions remain somewhat analytically inconsistent. To address these limitations, we draw on communicative norms as a theoretical basis for extending and evaluating the identified criteria (Section 3.1). We then integrate these considerations into a theoretical framework that distinguishes synthetic media based on their production process, descriptive features, relation to reality, and normative evaluation (Section 3.2). Based on this framework, we propose three ideal types of synthetic media, which we describe as deepfakes, deepfiction, and deeprealism. Finally, we consider the audiences' perspective and the potential impact of synthetic media (see Section 3.4).

3.1. The Legitimacy of Synthetic Media in the Context of Communicative Norms

The critical review revealed that existing definitions are mostly technical and isolated. Yet, from a communication studies perspective, an evaluation of synthetic media based solely on production technology, quality, or the presumed intent of its creator remains incomplete. Additional aspects are needed to explain whether synthetic media support or undermine expectations that are central to communication. From this perspective, synthetic media are not merely technological artifacts, but communicative forms that shape processes of meaning-making and mutual understanding. Communication can be understood as an interaction "through which persons create, maintain, and alter social order, relationships, and ideas" (Cronen et al., 1982, p. 85) and as a process in which participants "create and share information with one another to reach a mutual understanding" (Rogers & Kincaid, 1981, p. 63). It can therefore be seen as an ever-present process of action that relies on symbols that represent ideas for the purpose of meaning making and mutual understanding (Floyd et al., 2022). At the center of debates about "good" communication lies the expectation that communicators, whether individuals or institutions, adhere to honesty and avoid knowingly inducing false beliefs (Habermas, 1981). In this perspective, communication is expected to be truthful, comprehensible, sincere, and normatively appropriate. Whether synthetic media can be seen as legitimate communication therefore cannot be determined solely by whether a deceiving intent is assumed. Instead, normative legitimacy depends on whether representations are accurate and acceptable in relation to the communicative norms and values relevant to their use (Garusi et al., 2025; Geise et al., 2025; Opdahl et al., 2023). The extent to which the audience actually trusts such depictions or considers them credible requires a separate empirical investigation.

The lack of discussion about communicative norms in deepfake research has recently been criticized by Geise et al. (2025). We argue that communicative norms help to specify and extend the characteristics identified in the critical review. First, the communicative norm of truthfulness highlights the importance of factuality, because true or correct content is a necessary condition to qualify as information (Dretske, 1983; Mingers & Standing, 2018). Rowley (1998) argued that information can be treated as a resource or commodity, but if it is knowingly falsified, it ceases to reduce uncertainty (Berger & Calabrese, 1975) and therefore no longer qualifies as genuine information. Hence, if synthetic media are used to represent false or misleading material, they should not be classified as information but rather as mis- or disinformation (Kalsnes, 2018). The truthfulness of the content presented is also linked to the portrayal of subjects and their statements and actions. Deepfakes can depict real individuals saying or doing things they never actually said or did, thereby misrepresenting their behavior or identity (Ahmed, 2023; Hameleers et al., 2023; Weikmann & Lecheler, 2023a). We therefore argue that synthetic media should be assessed in terms of the factual accuracy of the information presented and whether statements and actions are truthfully attributed to the depicted individuals.

Second, drawing on Habermas's (1981) understanding of sincerity, we argue that transparency is a central criterion, because audiences should be able to recognize that content has been artificially generated or altered in order to evaluate it and to avoid being misled. Sincerity also plays a role in how synthetic media portray their relationship to reality. While deepfakes are often associated with fictional events that never took place (Weikmann & Lecheler, 2023a), synthetic media can also be used in fictional contexts such as satire (Habgood-Coote, 2023) or to convey factual information (Raemy et al., 2025). From this perspective, it is important whether fictional representations are recognizable as such or are misleadingly presented as real events. If the synthetic nature of a media product is concealed, this can lead to misconceptions and undermine expectations of sincerity and honesty (Habermas, 1981). Transparency is therefore closely linked to the question of whether synthetic media reduce uncertainty and can be regarded as legitimate information, or whether they instead undermine trust in communication (Berger & Calabrese, 1975; Rowley, 1998).

Third, normative rightness highlights the importance of the communicative context and the consent of depicted subjects. It concerns whether communication is appropriate in relation to the norms and expectations that apply in a given context (Habermas, 1981). When real individuals are depicted, their consent becomes an additional ethical consideration (van der Sloot & Wagensveld, 2022). Synthetic media may undermine norm-oriented communication and ethically guided social coexistence when they distort reality and induce false beliefs, especially in contexts such as public communication and opinion formation (Jungherr & Schroeder, 2023). At the same time, synthetic media are not necessarily perceived as fiction or as "fake" content that distorts reality. They may also be used to visualize, simulate, or communicate serious and factual content in ways that support communication, for example in educational or journalistic contexts (Godulla et al., 2021; Raemy et al., 2025). Importantly, communicative norms are not identical across all contexts. What is acceptable in film or entertainment may be evaluated differently in marketing or in journalism. The context of use is therefore crucial for assessing whether synthetic media can be considered legitimate forms of communication, including whether they portray real individuals in a normatively appropriate manner.

Fourth, the communicative norm of comprehensibility emphasizes that communication should be understandable to its recipients (Habermas, 1981; Rogers & Kincaid, 1981). In the context of synthetic media, this raises the question of whether their artificial origin and relationship to reality are communicated in a comprehensible manner. The interpretation of synthetic media may depend on their communicative context, disclosure of the production process as well as the audience's familiarity with the technology, digital literacy, and trust (Bendahan Bitton et al., 2024; Bray et al., 2023; Opdahl et al., 2023). We therefore argue that comprehensibility concerns whether synthetic media are presented in a way that makes their artificial origin and relationship to reality understandable, while the audiences' perspective requires separate empirical investigation.

3.2. A Framework for Evaluating Synthetic Media

The proposed framework (see Figure 1) considers the production process, descriptive evaluation, relation to reality, and normative evaluation to derive the labeling of synthetic media. The proposed dimensions are explained below, while a comprehensive discussion of the dimensions in relation to the proposed labels (deepfake, deepfiction, deeprealism) is provided in Section 3.3.

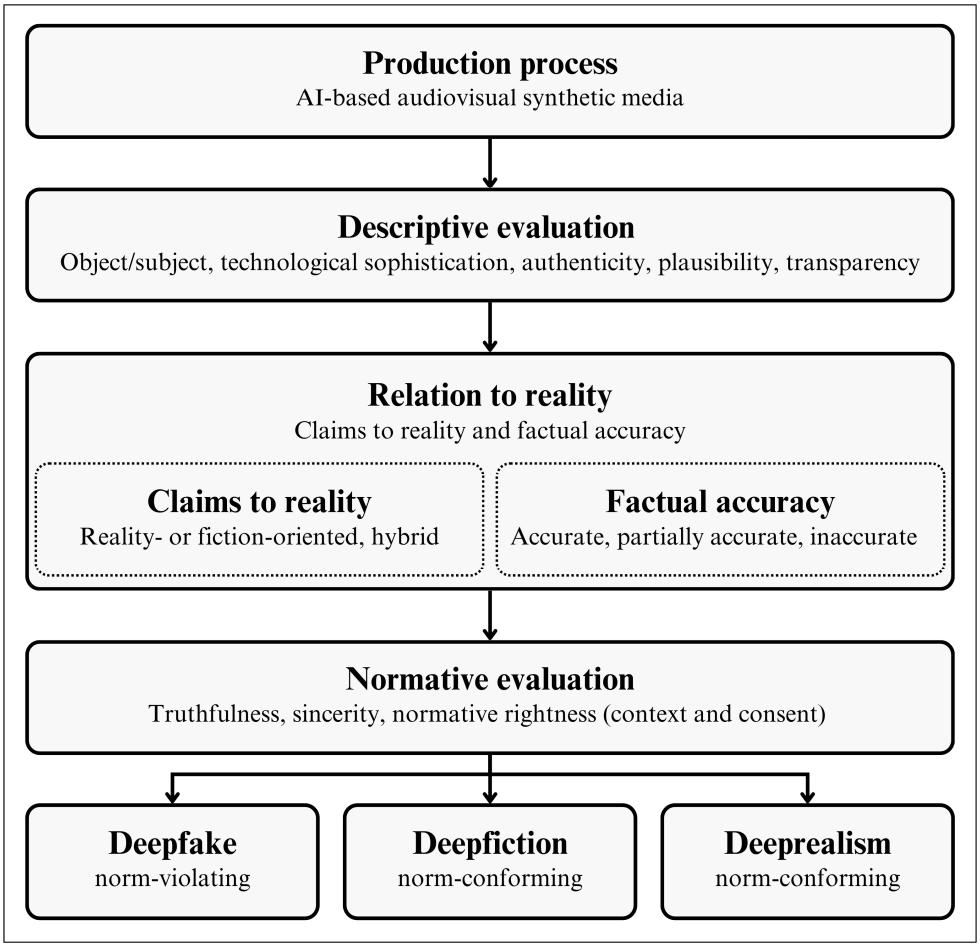


Figure 1. Theoretical framework for the classification and evaluation of synthetic media.

3.2.1. Production Process

The first dimension of the framework concerns the production process of synthetic media. Existing definitions often focus on the technological basis of deepfakes and distinguish them from media altered using conventional editing or manipulation techniques, such as cheap fakes (Brady & Meyer-Resende, 2020; Millière, 2022). We adapt this distinction and define the framework's scope by focusing on synthetic audiovisual media created using deep-learning technologies. This distinction, however, should not be equated with the technological sophistication of the resulting material or the expertise required for its production. Advances in generative AI have made it possible to produce sophisticated audiovisual content with comparatively limited technical effort (Millière, 2022). A further distinction concerns the media format, including images and videos, such as face or body swapping, and audio, such as speech synthesis (Kietzmann et al., 2020). Although synthetic text may also be generated using AI, the present framework focuses on audiovisual representations and their capacity to depict persons, objects, events, and actions.

3.2.2. Descriptive Evaluation

The second dimension of the framework provides a descriptive evaluation of synthetic media based on characteristics that can largely be examined through the material's content and presentation. It includes depicted subjects and objects, technological sophistication, authenticity, plausibility, and transparency.

The review shows that existing conceptualizations do not always clearly distinguish between subjects and objects depicted in synthetic media and their relationship to reality.

For example, definitions frequently focus on the imitation of real individuals, although synthetic media can also depict fictional persons and objects (de Ruiter, 2021; van der Sloot & Wagensveld, 2022). We therefore propose distinguishing the two and specifying whether they represent real or fictional entities (Bray et al., 2023). This is relevant when identifiable individuals are depicted, as their representation may raise additional questions concerning consent and their truthful representation (de Ruiter, 2021), which are considered in the normative evaluation.

Furthermore, prior studies often equate technological sophistication with high authenticity and plausibility, even though these characteristics are not analytically identical (Bateman, 2020; Hameleers et al., 2023). We argue for distinguishing these three characteristics. Technological sophistication refers to the technical refinement and consistency of the media and the presence or absence of visible flaws in its audiovisual quality (Kietzmann et al., 2020; Whittaker et al., 2020). Authenticity describes how closely the representation resembles a real person or object in terms of features such as voice, gestures and patterns of movement and expression (Weikmann & Lecheler, 2023b). Plausibility is linked to the coherence and apparent likelihood of depicted events, actions, statements, or circumstances (Hameleers et al., 2023). Synthetic media may be technologically sophisticated and visually authentic while remaining implausible because the depicted behavior or statements appear inconsistent with what is known about the subject (Hameleers et al., 2023). Conversely, fabricated events or statements may appear highly plausible despite never having occurred. Plausibility should therefore be distinguished from factual accuracy, which is considered in the material's relation to reality.

Finally, while existing conceptualizations often associate deepfakes with deception (Weikmann & Lecheler, 2023a; Hameleers, 2024), the consideration of communicative norms highlights transparency as an additional criterion. Expectations of sincerity and honesty emphasize the importance of disclosing whether media content has been artificially generated or altered (Habermas, 1981; see Section 3.1). Transparency is therefore included in the descriptive evaluation and refers to the disclosure of the synthetic origin of the media. Taken together, these characteristics provide a basis for the descriptive evaluation of synthetic media, but do not allow for their classification or an assessment of their normative legitimacy or impact. Their normative legitimacy, instead, depends on their relation to reality, the communicative context in which they are used as well as the applicable norms within this context.

3.2.3. Relation to Reality

The third dimension of the framework considers the relation of synthetic media to reality. We argue that synthetic media differ in how they relate to real-world phenomena and how this relationship is communicated. Existing conceptualizations often associate deepfakes with deception and the fabrication of events that never occurred (Weikmann & Lecheler, 2023a; Hameleers, 2024). However, synthetic media can also represent factual information or combine factual and fictional elements (Lesher et al., 2022; Raemy et al., 2025). This distinction is also relevant from the perspective of communicative expectations of truthfulness and sincerity (Habermas, 1981). Particularly in reality-oriented communication, factual accuracy and the truthful representation of events are central to evaluating whether synthetic media meet relevant communicative norms (Raemy et al., 2025). Expectations of factual accuracy, however, differ across communicative contexts. Fictional representations do not necessarily need to correspond to actual events, whereas reality-oriented representations are subject to stronger expectations of factual accuracy (Habgood-Coote, 2023). We therefore propose distinguishing between claims to reality and factual accuracy. Claims to reality concern whether synthetic media primarily aim to depict, reconstruct, or explain

real-world phenomena or construct fictional and imaginative content. Factual accuracy, by contrast, concerns the extent to which a representation is supported by facts. We argue that these aspects are analytically distinct. A synthetic representation may claim to depict reality without being factually accurate, while a fictional representation may incorporate factually accurate information. Whether synthetic media claim to represent reality depends on their content as well as the context in which they are presented. While factual accuracy can be examined separately, the normative assessment of any deviations from reality depends on the communicative context and the applicable norms within this context.

3.2.4. Normative Evaluation

The fourth dimension of the framework applies communicative norms to the evaluation of synthetic media. Previous studies often center on the intentions of the creators of deepfakes (Weikmann & Lecheler, 2023a), which are difficult to observe or verify (Spies, 2019). To address these limitations, we turn to communicative norms. Hence, this dimension examines whether synthetic media conform to expectations of truthfulness, sincerity, and normative rightness (Habermas, 1981).

In this context, truthfulness concerns the factual accuracy of the presented content and the truthful representation of depicted subjects (Dretske, 1983; Mingers & Standing, 2018). As discussed earlier, synthetic media may contain factually accurate information while falsely attributing statements or actions to real individuals (Ahmed, 2023; Hameleers et al., 2023; Weikmann & Lecheler, 2023a). The evaluation therefore considers whether the content supports its claims to reality and whether depicted individuals, their statements, and actions are truthfully represented.

Sincerity reflects how synthetic media communicate their relationship to reality and their artificial origin (Habermas, 1981). This includes whether fictional representations are recognizable as such in a given context or are misleadingly presented as actual events (Habgood-Coote, 2023; Weikmann & Lecheler, 2023a). Furthermore, transparency is assessed regarding whether the use of AI is adequately disclosed (Geise et al., 2025). While the descriptive evaluation establishes whether disclosure is provided, the normative evaluation examines whether it is sufficient in the respective communicative context.

Normative rightness concerns whether a representation is appropriate in relation to the norms and expectations that apply in its communicative context (Habermas, 1981). These expectations may differ between journalism, entertainment, or other fields of application (Godulla et al., 2021; Raemy et al., 2025). The evaluation therefore considers whether the portrayal is legitimate within its context and whether consent has been obtained when real subjects are portrayed (van der Sloot & Wagensveld, 2022).

3.2.5. Labeling Synthetic Media

Based on the preceding dimensions, the framework distinguishes between deepfakes, deepfiction, and deeprealism as three ideal types of synthetic media. Their descriptive evaluation in terms of technological sophistication, authenticity, and plausibility may vary across all three types. While these ideal types provide guidance for the classification of synthetic media, their boundaries depend on the communicative context in which synthetic media are used as well as the norms that apply within this context. We propose the following definitions for the three ideal types.

Deepfakes involve substantial misrepresentations of reality or serious violations of fundamental communicative norms, particularly concerning truthfulness, consent, and personal autonomy. This may include fabricated content presented as reality, such as statements or actions falsely attributed to real individuals, as well as non-consensual synthetic representations that violate personal autonomy. The artificial origin of deepfakes is often

not disclosed. However, even when the use of AI is made transparent, misrepresentations of reality or serious violations of consent may still justify their classification as deepfakes. They appear in various contexts such as pornography and politics.

Deeprealism refer to synthetic representations that communicate reality-oriented and factually accurate information and conform to communicative norms such as truthfulness, sincerity and normative rightness. They aim to depict, reconstruct, or explain real-world phenomena, including statements and actions that are truthfully attributed to real individuals. Consent is obtained when real persons are represented, and the creation process (use of AI) is made transparent. These media may be used in various contexts such as journalism, education, advertising and entertainment.

Deepfiction refer to synthetic representations of fictional or imaginative contexts that conform to communicative norms within their original communicative setting. They may depict fictional events, statements, or actions while also incorporating factual elements. When real persons are depicted, consent is obtained where necessary, although it may not be required in certain contexts, such as satire. The artificial origin of deepfiction may be explicitly disclosed but does not necessarily require disclosure if its fictional character is recognizable within the communicative context. These media may be used in various contexts such as film, gaming, satire, and entertainment.

3.3. Assigning Forms of Synthetic Media to Deepfakes, Deepfiction, and Deeprealism

The proposed framework serves as the basis for evaluating synthetic media and distinguishing between deepfakes, deepfiction, and deeprealism. Table 2 summarizes the criteria and questions underlying this evaluation.

To distinguish between deepfakes, deepfiction, and deeprealism, we propose an evaluation process in which the material is first assessed descriptively, followed by an examination of its relationship to reality, and finally an assessment of whether the representation meets the relevant communicative norms in its respective context. The evaluation begins by determining whether the media product falls within the scope of the framework, that is, whether it is audiovisual material that was created or modified using deep-learning technologies. This is followed by an assessment of the descriptive characteristics of the media product. The people and objects depicted, the level of technological sophistication, authenticity, plausibility, and any visible disclosure of AI use can provide initial information about the representation. However, these criteria alone do not provide sufficient information to classify the media product. Instead, their relation to reality and factual accuracy need to be examined.

The initial distinction between deeprealism, deepfiction, and deepfakes begins with an assessment of how synthetic representations relate to reality. We argue that it should be examined whether synthetic media claim to represent actual events or are presented as fiction, and to what extent the depicted content is factually accurate. These aspects are considered separately, as fictional representations may incorporate factual information, while reality-oriented representations may contain fabricated or inaccurate content. Reality-oriented and factually accurate representations may initially be considered deeprealism, whereas recognizable fictional representations may be considered deepfiction. Representations that claim to depict reality but contain fabricated or factually inaccurate information may initially be considered deepfakes. Factual inaccuracies alone, however, do not necessarily justify classification as deepfakes, as their relevance also depends on the communicative context. In this regard, Pierini (2023) highlights that synthetic media share characteristics with other constructed forms of expression, such as drawings or paintings. They may depict events, actions, or statements that never occurred, potentially creating misleading representations of reality. When such fabricated representations are deliberately presented

as authentic, they are particularly closely associated with deepfakes. However, the mere depiction of a fictional event does not constitute deception. The distinction depends on whether the representation makes or implies a factual claim and how its relationship to reality is communicated.

Table 2. Dimensions, criteria and evaluative purposes of the framework.

Dimension	Criteria	Evaluative Purpose
Production process: Define which media fall within the framework.	AI-based audiovisual synthetic media	Was the material created or altered using deep-learning technologies? Is it an image, audio, video or audiovisual material?
Descriptive features: Describe the media product and its presentation—that is what is depicted, how it is produced or presented, and what a viewer can see or read before external verification.	Object/subject	Who or what is depicted? Does the material portray identifiable real individuals or fictional persons?
	Technological sophistication	How technically sophisticated is the media product? Are there any technical flaws or inconsistencies?
	Authenticity	How closely does the depiction resemble recognizable characteristics of its source? (object/subject)
	Plausibility	Do the material and content displayed appear believable in its presented context?
	Transparency	Is the artificial nature of the media revealed? (product/modification)
Relation to reality: Establish the content's claims to reality and verify its factual accuracy.	Claims to reality	Is the content contextualized as a representation of reality, fiction, a combination of both or as a fake?
	Factual accuracy	To what extent can the depicted statements, actions, or events be verified as completely or partially factually accurate?
Normative evaluation: Assess whether the content and its original presentation meet relevant communicative norms.	Truthfulness: factual accuracy	Does the depicted content support its claim to reality? Are there actual inaccuracies?
	Truthfulness: individuals	Are identifiable individuals, their statements, and their actions represented and attributed truthfully?
	Sincerity and transparency	Are the material's artificial origin and relation to reality disclosed?
	Normative rightness: context and consent	Is the portrayal appropriate in its original communicative context, and has consent been obtained where necessary?

For example, a synthetic reconstruction of a historical event may combine factual information with plausible but unverifiable details. Such a representation remains reality-oriented, although its factual grounding and the way uncertainties are communicated require further examination. If the reconstruction primarily presents an imagined version of the event and its fictional character is recognizable, it may qualify as deepfiction. By contrast, a reconstruction based on verified historical information that does not present uncertain details as established facts may qualify as deeprealism, provided that relevant communicative norms are respected. By contrast, a synthetic video that presents fabri-

cated statements by a political figure as an authentic recording may indicate a deepfake (Weikmann & Lecheler, 2023a).

While the assessment of the relationship to reality examines what synthetic media represent and how accurately they depict reality, the subsequent normative evaluation considers their normative legitimacy in terms of truthful representation, communicative context, consent, and transparency. Truthful representation becomes particularly relevant when synthetic media attribute statements or actions to real individuals. Synthetic media may be based on factually accurate information while misrepresenting how this information was originally communicated. The communicative context is then considered to assess whether the portrayal is legitimate, and which normative expectations apply to its particular use. These expectations may differ across journalism, entertainment, advertising, and education. In journalism, for example, factual accuracy, truthful representation, contextual integrity, and transparency are particularly relevant (Raemy et al., 2025). Although journalistic reporting cannot fully capture reality, even factually accurate information may be misleading if relevant context is omitted or presented in a distorted way (Lesher et al., 2022). In contrast, fictional representations may be legitimate in entertainment or satire when their fictional character is recognizable within the original communicative context (Habgood-Coote, 2023). For example, a synthetic satirical depiction of a political figure may qualify as deepfiction, even though the depicted statements or actions never occurred. However, if the same representation is subsequently presented as an authentic recording outside its original satirical context, its new communicative use may qualify as a deepfake. A related distinction arises in political self-presentation. Chowdhury (2024) discusses so-called softfakes, including AI-generated representations intended to make political candidates appear more appealing. Such content may qualify as deepfiction when its fictional character is recognizable and relevant communicative norms are respected. If the presentation depicts a politician participating in an event that never took place and is presented as photographic evidence, however, it may be classified as a deepfake.

The distinction between deepfiction and deepfake is also particularly relevant for synthetic sexual content. Such content may be classified as deepfiction when the depicted individuals have given their consent and the fictional nature of the content is recognizable within its original context, such as on a platform for consensually produced adult content. In contrast, non-consensual fictional sexual representations of real individuals violate fundamental ethical norms and would generally be classified as deepfakes. Hence, the distinction between deepfiction and deepfakes does not depend solely on whether the depicted content is fictional, but also on whether its production and use respect relevant communicative norms and consent of the depicted individuals (de Ruiter, 2021).

Finally, transparency constitutes a further criterion of normative evaluation. Whether disclosure is adequate depends on how the synthetic media are presented and which expectations apply in the respective context. In journalism, for example, explicit disclosure may be particularly important to distinguish synthetic representations from original recordings. In fictional or satirical contexts, however, explicit disclosure may not always be necessary when the fictional nature of the content is already recognizable (Habgood-Coote, 2023). Importantly, transparency alone does not establish normative legitimacy of synthetic media. A clearly labeled synthetic representation may still falsely attribute statements or actions to real individuals or violate their consent. Thus, disclosure cannot compensate for serious violations of fundamental ethical norms.

Taken together, these proposed ideal types require careful examination of the communicative context and the norms that apply within it. As norms may differ across social, cultural, and institutional contexts, the classification may involve interpretive judgments, especially when distinguishing between deepfiction and deepfakes.

3.4. Audience Perspective and Outcome

The preceding dimensions of the framework provide the basis for classifying synthetic media as deepfakes, deepfiction, or deeprealism within the respective communicative context. This classification, however, does not determine how synthetic media are perceived and evaluated by audiences and what impacts they may have. We argue that these aspects require a separate empirical examination and should consider the communicative embedding, comprehensibility, characteristics and predispositions of the recipients and individual and societal consequences (see Table 3).

Table 3. Audience perspective and outcomes of synthetic media.

Dimension	Criteria	Evaluative Purpose
Audience perspective and outcomes: Describe how synthetic media may be interpreted and what consequences may arise from their reception.	Communicative embedding	How might framing or subsequent decontextualization influence the interpretation of synthetic media?
	Comprehensibility and perception	To what extent might the audience recognize the artificial origin, understand its claim to reality, and assess its normative legitimacy?
	Individual characteristics	How might sociodemographic characteristics, digital literacy and trust influence audiences' evaluation of the material?
	Individual and societal consequences	What individual and societal benefits, risks, or impacts may arise from the material and its communicative embedding?

Communicative embedding focuses on how synthetic media are presented within their communicative context and how subsequent framing or decontextualization (if present) may affect their interpretation. The comprehensibility of synthetic media reflects whether audiences recognize the artificial origin of synthetic media, understand their relationship to reality, and form judgments about their normative legitimacy (Geise et al., 2025; Opdahl et al., 2023). In this context, technological sophistication, authenticity, and plausibility may influence how convincing synthetic representations appear (Hameleers et al., 2023; Weikmann & Lecheler, 2023b). For example, a technologically sophisticated depiction may still appear implausible if the statements or actions depicted do not align with what is known about the subject (Hameleers, 2024). Moreover, fictional and satirical representations may also deliberately depart from plausibility to create humorous or critical effects (Habgood-Coote, 2023). Highly plausible fictional depictions may therefore be mistaken for actual events, whereas clearly implausible representations could serve as cues to their fictional or satirical character (Habgood-Coote, 2023; Hameleers, 2024). Additionally, a technologically sophisticated and authentic deepfake that depicts rather plausible events or statements may be perceived as more credible than an equally sophisticated but implausible representation.

Research on virtual avatars suggests that greater visual realism can enhance perceived trustworthiness (Baake et al., 2025; Franke & Groeppel-Klein, 2024), although it may also evoke feelings of eeriness that reduce trustworthiness (Shin et al., 2019). Similarly, prominent disclosure of an avatar's artificial nature may reduce perceived credibility (van Berlo & Breves, 2025). However, perceived credibility does not necessarily predict the effects of synthetic media. Hameleers (2024), for example, found that less plausible political deepfakes could have stronger delegitimizing effects despite being perceived as less credi-

ble. These findings highlight the importance of the communicative context and audiences' characteristics in shaping responses to synthetic media.

Recognizing the artificial nature of synthetic media, however, does not necessarily imply that audiences correctly assess their factual accuracy or normative legitimacy. For example, audiences may recognize an AI-generated historical reconstruction as synthetic while being unable to distinguish documented events from fictional additions. Similarly, the presence of an AI label does not guarantee that audiences notice or understand its implications (Geise et al., 2025). For instance, a disclosure provided only in an accompanying caption may no longer be available when the audiovisual material is shared without that caption. Whether a label is embedded directly in the content or provided through its context could be relevant to how audiences recognize and interpret synthetic media.

Furthermore, perceptions and comprehensibility may differ depending on individual characteristics, including sociodemographic factors, digital literacy, and trust (Bendahan Bitton et al., 2024; Bray et al., 2023; Rauchfleisch et al., 2025). Prior knowledge, expectations, and familiarity with depicted individuals may influence whether their appearance, behavior, or statements are considered plausible. In this context, critical thinking and fact-checking skills may contribute to audiences' ability to recognize potential deception and evaluate synthetic media (Bendahan Bitton et al., 2024; Bray et al., 2023; Thaw et al., 2020).

Individual and societal consequences include potential benefits, risks, and effects on trust and perceived normative legitimacy. At the individual level, synthetic media may facilitate the understanding of complex information, support creative expressions, or provide new educational opportunities. At the same time, they may mislead individuals, influence public attitudes, or harm individuals' reputation and dignity (Bendahan Bitton et al., 2024). At the societal level, synthetic media may offer new opportunities for public communication and engagement but also contribute to the spread of disinformation, create uncertainty about the authenticity of audiovisual content, and result in declining trust in media and news online (Hoffmann et al., 2025; Vaccari & Chadwick, 2020).

Importantly, the potential consequences of synthetic media cannot be directly inferred from their classification. Deeprealism and deepfiction may conform to relevant communicative norms while nevertheless being misunderstood or leading to unintended consequences. This is not unique to synthetic media, as other forms of information, including journalistic content, scientific findings, and official communications from public institutions, may also become misleading when taken out of context or shared in ways that distort their original meaning (Lesher et al., 2022). Likewise, a normatively legitimate fictional representation may still be misunderstood when it is shared outside its original context. For example, fictional or satirical representations may be perceived as factual when they are shared beyond their original context (Tandoc et al., 2017; Lesher et al., 2022). Conversely, a deepfake may violate fundamental ethical norms even when audiences recognize the manipulation or no measurable harm occurs. For example, the recognition of non-consensual synthetic sexual content as artificial does not eliminate potential violations of the depicted person's privacy or dignity.

4. Conclusions and Outlook

This study contributes to the debate on deepfakes and synthetic media by identifying key characteristics and by proposing a framework for their systematic evaluation. In response to RQ1, the critical review shows that prior studies approach deepfakes and synthetic media from different disciplinary perspectives and emphasize different dimensions of the phenomenon. These approaches emphasize aspects such as technological production, types of media, the role of underlying intentions or the relation of quality and authenticity (de Ruiter, 2021; Hameleers, 2024; Weikmann & Lecheler, 2023b). Some of these

characteristics, however, remain analytically underspecified. This applies in particular to the relation between synthetic media and reality, the distinction between technological sophistication, authenticity, and plausibility, and the question of whether synthetic media should be classified primarily on the basis of presumed (malicious) intentions of their creators (Hameleers et al., 2023; Pawelec, 2024; Weikmann & Lecheler, 2023a, 2023b).

Building on these observations and addressing RQ2, the framework developed in this paper expands on the identified features and proposes a structured evaluation of how different types of synthetic media relate to communicative expectations such as truthfulness, sincerity and comprehensibility (Habermas, 1981; Rogers & Kincaid, 1981). The framework distinguishes four dimensions relevant for the classification of synthetic media: the production process, descriptive evaluation, relation to reality and normative evaluation. The production process defines which audiovisual media fall within the framework, while the descriptive evaluation examines their characteristics. The relation to reality considers claims to reality and factual accuracy, which are subsequently assessed in relation to relevant communicative norms. Together, these assessments inform the classification of synthetic media as deeprealism, deepfiction, or deepfake, which are understood as ideal types. Finally, the audiences' perspective and potential outcomes are addressed separately, focusing on how audiences may perceive and interpret synthetic media and the individual and societal consequences that may arise.

From a communication studies perspective, the framework proposed in this paper suggests that synthetic media should be approached as a new form of mediated communication that requires conceptual and normative clarification. The study demonstrates that a more systematic basis for differentiating between various forms of synthetic media and for evaluating their normative legitimacy is needed. The effects of synthetic media depend on how they are contextualized, interpreted, and trusted in processes of public and mediated communication (Geise et al., 2025). The paper therefore highlights the relevance of questions about truthfulness, sincerity, transparency, communicative context, and the audiences' perspective.

For communication research, the framework provides a theoretical basis for examining and comparing different forms of synthetic media. These might vary in terms of their relation to reality and communicative norms as well as the audiences' perceptions and potential impacts. Future surveys and experiments could examine how these characteristics influence perceived credibility, the identification of synthetic media, and willingness to share it. They could also investigate how these relationships differ depending on the audiences' sociodemographic characteristics, digital literacy and trust in the sources of synthetic media. Furthermore, an operationalization of the proposed framework is needed. The proposed dimensions should be translated into measurable indicators, for example for content analyses. This would make it possible to examine how the various dimensions interact in practice, and to what extent they influence the assessment of normative legitimacy in different communication contexts. In this context, future research could evaluate the role of synthetic media in institutional settings such as journalism, political and strategic communication. In these fields, relevant questions arise such as the conditions under which synthetic media can be considered normatively legitimate, how they should be made transparent and how they affect credibility and trust.

In practice, the framework offers criteria for evaluating the responsible use of synthetic media, for example, in journalism. These criteria may inform guidelines concerning claims to reality, factual accuracy, and truthful representation. The framework may also support media literacy education. In educational settings, its criteria can be applied to examples of synthetic media to compare their descriptive characteristics, factual accuracy, and normative legitimacy. Such comparisons could help learners understand that AI-generated

content is not necessarily misleading, and that technologically sophisticated or plausible representations are not necessarily factually accurate.

This work also comes with some limitations. Since the study follows a critical review approach, the selection of literature was based on its conceptual relevance and the principle of conceptual saturation, without aiming for comprehensive coverage. This approach was found to be suitable for identifying and critically summarizing relevant perspectives. Nevertheless, it is possible that other relevant publications were not included, particularly those published in languages other than English or German. A further limitation concerns the distinction between deep-learning-based media and content created using conventional manipulation techniques. In practice, different production methods may be combined. The framework therefore also applies to audiovisual media when the use of deep-learning technologies can be established, regardless of whether existing material is modified or deep-learning-generated elements are added. Such representations may also be assessed in terms of their relationship to reality and factual accuracy, alongside the other dimensions of the framework. Finally, it is important to note that the framework reflects a communication- and norm-oriented perspective and is not intended to replace but to supplement technical, legal, or computer-based classifications.

In summary, this work aims to serve both scientific research and communication practice as a foundation for analyzing and implementing communication with and through synthetic media. In this context, upholding communicative norms, combined with a critical and analytical stance, is key to navigating AI-driven communication.

Author Contributions: Conceptualization, D.B.B. and P.R.; Methodology, D.B.B. and P.R.; Validation, D.B.B.; Formal analysis, D.B.B.; Writing—original draft, D.B.B. and P.R.; Writing—review & editing, D.B.B. and P.R.; Supervision, D.B.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. The APC was funded by the Open Access Publishing Fund of Leipzig University.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

Ahmed, S. (2023). Navigating the maze: Deepfakes, cognitive ability, and social media news skepticism. *New Media & Society*, *25*(5), 1108–1129. [CrossRef]

Baake, J., Schmitt, J. B., & Metag, J. (2025). Balancing realism and trustworthiness: AI avatars in science communication. *Journal of Science Communication*, *24*(2), A03. [CrossRef]

Bateman, J. (2020). *Deepfakes and synthetic media in the financial system: Assessing threat scenarios*. Carnegie Endowment for International Peace.

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. [CrossRef]

Bendahan Bitton, D., Ötting, H., Godulla, A., & Hoffmann, C. P. (2026). Implications of deepfake technology for journalistic practice: A technology at the crossroads of challenges and opportunities. *Journalism*, *27*(11), 3503–3522. [CrossRef]

Berger, C. R., & Calabrese, R. J. (1975). Some explorations in initial interaction and beyond: Toward a developmental theory of communication. *Human Communication Research*, *1*(2), 99–112. [CrossRef]

Brady, M., & Meyer-Resende, M. (2020). *Deepfakes: A new disinformation threat*. Democracy Reporting International.

Bray, S. D., Johnson, S. D., & Kleinberg, B. (2023). Testing human ability to detect 'deepfake' images of human faces. *Journal of Cybersecurity*, *9*(1), tyad011. [CrossRef]

Chowdhury, R. (2024). AI-fuelled election campaigns are here—Where are the rules? *Nature*, *628*(8007), 237. [CrossRef] [PubMed]

Citron, D., & Chesney, R. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, *107*, 1753–1810. [CrossRef]

Cronen, V. E., Pearce, W. B., & Harris, L. M. (1982). The coordinated management of meaning: A theory of communication. In F. E. X. Dance (Ed.), *Human communication theory: Comparative essays* (pp. 61–89). Harper & Row.

de Ruiter, A. (2021). The distinct wrong of deepfakes. *Philosophy & Technology*, *34*(4), 1311–1332. [CrossRef]

Dretske, F. (1983). *Knowledge and the flow of information*. MIT Press.

Firc, A., Malinka, K., & Hanáček, P. (2023). Deepfakes as a threat to a speaker and facial recognition: An overview of tools and attack vectors. *Heliyon*, *9*(4), e15090. [CrossRef] [PubMed]

Floyd, K., Schrodt, P., Erbert, L. A., & Scharp, K. M. (2022). *Exploring communication theory: Making sense of us* (2nd ed.). Routledge.

Franke, C., & Groeppel-Klein, A. (2024). The role of psychological distance and construal level in explaining the effectiveness of human-like vs. cartoon-like virtual influencers. *Journal of Business Research*, *185*, 114916. [CrossRef]

Garusi, D., Juarez Miro, C., & Hanusch, F. (2025). Understanding news media trust through the lens of phenomenological sociology. *Communication Theory*, *35*(3), 163–172. [CrossRef]

Geise, S., Luther, A. R., Reich, S., & Linke, M. (2025). A new face of political advertising? Synthetic imagery in the 2025 German federal election campaigns on social media. *Studies in Communication and Media*, *14*(4), 485–516. [CrossRef]

Godulla, A., Hoffmann, C. P., & Seibert, D. (2021). Dealing with deepfakes—An interdisciplinary examination of the state of research and implications for communication studies. *Studies in Communication and Media*, *10*(1), 72–96. [CrossRef]

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, *26*(2), 91–108. [CrossRef] [PubMed]

Habermas, J. (1981). *Theorie des kommunikativen Handelns* [The theory of communicative action]. Suhrkamp.

Habgood-Coote, J. (2023). Deepfakes and the epistemic apocalypse. *Synthese*, *201*(103), 1–23. [CrossRef]

Hameleers, M. (2024). Cheap versus deep manipulation: The effects of cheapfakes versus deepfakes in a political setting. *International Journal of Public Opinion Research*, *36*(1), 1–9. [CrossRef]

Hameleers, M., & Marquart, F. (2023). It's nothing but a deepfake! The effects of misinformation and deepfake labels delegitimizing an authentic political speech. *International Journal of Communication*, *17*, 6291–6311.

Hameleers, M., van der Meer, T., & Dobber, T. (2023). Distorting the truth versus blatant lies: The effects of different degrees of deception in domestic and foreign political deepfakes. *Computers in Human Behavior*, *152*, 108096. [CrossRef]

Hoffmann, C. P., Bendahan Bitton, D., & Godulla, A. (2025). The role of self-efficacy and intellectual humility in perceived deepfake exposure and media cynicism. *Social Science Computer Review*. Advance online publication. [CrossRef]

Jacobsen, B. N., & Simpson, J. (2023). The tensions of deepfakes. *Information, Communication & Society*, *17*(6), 1095–1109. [CrossRef]

Jungherr, A., & Schroeder, R. (2023). Artificial intelligence and the public arena. *Communication Theory*, *33*(2–3), 164–173. [CrossRef]

Kalsnes, B. (2018). Fake news. In *Oxford research encyclopedia of communication* (pp. 1–21). Oxford Academic. [CrossRef]

Karaboga, M. (2024). Einleitung und Kontext [Introduction and context]. In M. Karaboga, N. Frei, M. Puppis, D. Vogler, P. Raemy, F. Ebbers, G. Runge, A. Rauchfleisch, G. de Seta, G. Gurr, M. Friedewald, & S. Rovelli (Eds.), *Deepfakes und manipulierte Realitäten: Technologiefolgenabschätzung und Handlungsempfehlungen für die Schweiz* [Deepfakes and manipulated realities: Technology assessment and recommendations for action for Switzerland] (pp. 69–81). TA-SWISS Publikationsreihe. Available online: https://zenodo.org/records/11643644 (accessed on 18 November 2026).

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, *63*(2), 135–146. [CrossRef]

Lesher, M., Pawelec, H., & Desai, A. (2022). *Disentangling untruths online: Creators, spreaders and how to stop them* (OECD Going Digital Toolkit Notes, No. 23). OECD Publishing. [CrossRef]

McCammon, M. N. (2021). The emerging threat of synthetic media. In L. Corredoira, I. B. Mallén, & R. C. Presuel (Eds.), *The handbook of communication rights, law, and ethics* (pp. 292–302). [CrossRef]

Millière, R. (2022). Deep learning and synthetic media. *Synthese*, *200*(3), 231. [CrossRef]

Mingers, J., & Standing, C. (2018). What is information? Toward a theory of information as objective and veridical. *Journal of Information Technology*, *33*(2), 85–104. [CrossRef]

Möller, J., Hameleers, M., & Ferreau, F. (2020). *Types of disinformation and misinformation: Various types of disinformation and their dissemination from a communication science and legal perspective*. die medienanstalten—ALM GbR.

Opdahl, A. L., Tessem, B., Dang-Nguyen, D.-T., Motta, E., Setty, V., Throndsen, E., Tverberg, A., & Trattner, C. (2023). Trustworthy journalism through AI. *Data & Knowledge Engineering*, *146*, 102182. [CrossRef]

Paris, B., & Donovan, J. (2019). Deepfakes and cheap fakes: The manipulation of audio and visual evidence. *Data & Society*, 1–49. Available online: https://datasociety.net/wp-content/uploads/2019/09/DS_Deepfakes_Cheap_FakesFinal-1-1.pdf (accessed on 18 November 2026).

Pawelec, M. (2024). Decent deepfakes? Professional deepfake developers' ethical considerations and their governance potential. *AI and Ethics*, *5*(3), 2641–2666. [CrossRef]

Pierini, F. (2023). Deepfakes and depiction: From evidence to communication. *Synthese*, *201*(3), 97. [CrossRef]

Raemy, P., Bendahan Bitton, D., Hoffmann, C. P., Godulla, A., & Puppis, M. (2025). Deepfakes and journalism: Normative considerations and implications. *Journalism Studies*, *26*(14), 1724–1744. [CrossRef]

Rauchfleisch, A., Vogler, D., & de Seta, G. (2025). Deepfakes or synthetic media? The effect of euphemisms for labeling technology on risk and benefit perceptions. *Social Media + Society*, *11*(3), 1–12. [CrossRef]

Rogers, E. M., & Kincaid, D. L. (1981). *Communication networks: Toward a new paradigm for research*. Free Press.

Rowley, J. (1998). What is information? *Information Services & Use*, *18*(4), 243–254. [CrossRef]

Seibert, D. (2023). Deepfakes: Temporary hype or long-term innovation driver? In A. Godulla, & S. Böhm (Eds.), *Digital disruption and media transformation: How technological innovation shapes the future of communication* (pp. 143–152). Springer. [CrossRef] [PubMed]

Shin, M., Song, S. W., & Chock, T. M. (2019). Uncanny valley effects on friendship decisions in virtual social networking service. *Cyberpsychology, Behavior, and Social Networking*, *22*(11), 700–705. [CrossRef] [PubMed]

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. [CrossRef]

Spies, S. (2019, October 22). *Defining "disinformation"*. MediaWell—Social Science Research Council. Available online: https://mediawell.ssrc.org/research-reviews/defining-disinformation/ (accessed on 18 November 2026).

Tandoc, E. C., Jr., Lim, Z. W., & Ling, R. (2017). Defining "fake news": A typology of scholarly definitions. *Digital Journalism*, *6*(2), 137–153. [CrossRef]

Thaw, N. N., July, T., Wai, A. N., Goh, D. H.-L., & Chua, A. Y. K. (2020). Is it real? A study on detecting deepfake videos. *Proceedings of the Association for Information Science and Technology*, *57*(1), e366. [CrossRef]

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. [CrossRef]

van Berlo, Z., & Breves, P. (2025). Disclosing the virtual nature of virtual influencers: The effect of disclosure prominence and the role of product digitality. *Computers in Human Behavior Reports*, *19*, 100742. [CrossRef]

van der Sloot, B., & Wagensveld, Y. (2022). Deepfakes: Regulatory challenges for the synthetic society. *Computer Law & Security Review*, *46*, 105716. [CrossRef]

Wang, S. (2026). Mapping the landscape of deepfake research: A systematic review of empirical studies. *Big Data & Society*, *13*(1), 1–13. [CrossRef]

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe Report*, *27*, 1–107.

Weikmann, T., & Lecheler, S. (2023a). Cutting through the hype: Understanding the implications of deepfakes for the fact-checking actor-network. *Digital Journalism*, *11*(1), 1–18. [CrossRef]

Weikmann, T., & Lecheler, S. (2023b). Visual disinformation in a digital age: A literature synthesis and research agenda. *New Media & Society*, *25*(12), 3696–3713. [CrossRef]

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, *9*(11), 39–52. [CrossRef] [PubMed]

Whittaker, L., Kietzmann, J., Letheren, K., Mulcahy, R., & Russell-Bennett, R. (2023). Brace yourself! Why managers should adopt a synthetic media incident response playbook in an age of falsity and synthetic media. *Business Horizons*, *66*(2), 277–290. [CrossRef]

Whittaker, L., Kietzmann, T. C., Kietzmann, J., & Dabirian, A. (2020). "All around me are synthetic faces": The mad world of AI-generated media. *IT Professional*, *22*(5), 90–99. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.