

The semiotic functioning of synthetic media

The interpretation of many texts in the everyday world is concerned with their truth value in relation to the reality around us. The recent publication experiments with computer-generated texts have shown that the distinction between true and false, or reality and fiction, is not always clear from the text itself. Essentially, in today's media space, one may encounter texts, videos or images that deceive the reader by displaying nonsensical content or nonexistent events, while nevertheless appearing as genuine human-produced messages. This article outlines certain problems with artificial intelligence (AI)-generated content, and frames the issue as a problem of recognising its proper referential reality. Examples include generative texts, deepfakes and their functioning in contemporary culture. The article makes use of the concepts of mimicry and nonsense to reveal the elements and counterparts in the communicative processes involving generated media.

Keywords: *referential reality, artificial intelligence, deepfakes, generative text, language transformers*

Acknowledgement

This work was supported by the Estonian Military Academy and the University of Tartu [grant numbers R-012 and SSVUH20551].

Author Information

Auli Viidalepp, University of Tartu

<https://orcid.org/0000-0002-6206-5681>

Viidalepp, Auli. "The semiotic functioning of synthetic media".

Információs Társadalom XXII, no. 4 (2022): 109–118.

== <https://dx.doi.org/10.22503/inftars.XXII.2022.4.9> ==

*All materials
published in this journal are licenced
as CC-by-nc-nd 4.0*

1. Introduction

The traditional Fregean logic seems to rely on the assumption that textual statements are non-fictional or about the external reality, otherwise they do not have a reference. Literary scholar Marie-Laure Ryan (1991) proposes a theory of possible worlds to solve this problem for literary analysis. However, in semiotics, Jørgen Dines Johansen (2002) and Umberto Eco (1979) propose the diverging perspective that every text should be taken as a fictional construct first. Today, the problem is further highlighted by the emergence of language-transformer computer models that enable the creation of content that humans may find difficult to distinguish from regular content. Synthetic content (such as generated text) and synthetically enhanced content (such as deepfake videos) displace our habitual understanding of authorship and sense-making, enabling new, sophisticated ways of lying about the sender and the message content.

This article aims to bring the problem of synthetic content to the attention of semiotics and communication studies. Semiotics as the study of meaning-making in communication can offer useful perspectives for analysing how synthetic media functions in culture, as opposed to traditional media. Semiotics offers ways to understand how synthetic content readily distorts and hijacks parts of the habitual communication schema as described by Roman Jakobson.

The article proceeds as follows. First, I give an overview of the referentiality problem as outlined by literary and semiotic scholars. Then, I introduce certain types of computer-generated content using the examples of generated texts and deepfake imagery. After that, I highlight the ways in which the presentation of such synthetic content plays with our traditional understanding of authorship and making sense of the message, utilising the models of mimicry and nonsense. Finally, I further analyse how synthetic or synthetically enhanced content displaces and obscures the author and the real meaning of the message. In conclusion, I argue that generated content is equal to nonsense until the real author – the person or persons who request and handle the text – decides whether the text is suitable for publishing. A deepfake, however, is equivalent to a situation where the real sender, remaining obscure to the receiver, hijacks a trusted channel to deliver a lie.

2. Fiction, non-fiction and referential reality

According to Frege [1948], one of the first philosophers to consider fiction as a logical issue, a sentence about an imaginary entity does not refer, and this sentence is automatically false (or indeterminate, in a three-value system). Statements about fictional entities could then be excluded from the set of true statements on grounds of referential failure, while errors and lies would illustrate the case of faulty predication. Implicit to the Fregean position are three propositions: (1) Reference can only be made to that which exists; (2) 'To exist' is synonymous with 'to occur in the real world'; and (3) Only one world exists, the world we regard as real (Ryan 1991, 14).

In literary studies, multiple scholars have indicated the shortcomings of purely pragmatist models in analysing fictional texts. The above reflection by Marie-Laure Ryan leads to her conclusion that a Fregean logic-based position on texts is not applicable to literature as non-fiction. Ryan (1991) continues looking for assistance in the segregationist ontology of Thomas Pavel (1986) as an alternative for making a difference between fictional and non-fictional statements. However, Pavel's segregationist ontology still 'loses the distinction between fiction and literary criticism, which also refers to nonexisting entities. ... Moreover, ... the segregationist ontology ... encounters the additional challenge of ontologically hybrid textual worlds' (Ryan 1991, 15). To overcome the problem of questionable truths and ontologically hybrid texts (hybrid in the sense of mixing fictional and non-fictional), Ryan proposes the theory of possible worlds — originally, a concept from Leibniz depicting 'an infinity of possible worlds [that] exists as thoughts in the mind of God' (Ryan 1991, 16). Earlier, literary theorist Lubomír Doležal stated that 'meaning in literary texts (and in natural language texts in general) cannot be reduced to the concept of truth or truth-conditions' (Doležal 1979, 195).

In *A Semiotic-Pragmatic Approach to Literature*, Jørgen Dines Johansen goes further in explaining the relationships between linguistic representations and their referential realities: 'On a basic level, the **linguistic representation**, at least in English, is **identical regardless of whether the universe referred to is fictional or historical** (e.g., verb tense is no certain indicator)' (Johansen 2002, 152, emphasis added). For this reason, it is not possible to derive the reality of reference from a linguistic text only, without taking into account the situation and context of its utterance. The same impossibility of inference goes for the truth value of the text. In semiotic theory, Umberto Eco rejects the need for any reference whatsoever, 'defining meaning as a purely cultural unit, a structure generated by a cultural code' (Nöth 1995, 97). From Eco's conviction that 'a sign is everything that can be used to lie' it follows that we can never be certain of the truth value deduced from a sign action alone. Lest it be a lie, the truth needs to be evaluated by other means than the semiotic, and the semiotic inquiry, in Eco's opinion, should not be concerned with the truth at all. Eco goes as far as indicating our experience of the reality as semiotic first and above all: 'Within the framework of a constructivistic approach to possible worlds, even a so-called "actual" or "real" world of reference must be taken as a possible world, that is, as a cultural construct' (Eco 1979, 222). This is the reason we can access the different worlds (fictional and non-fictional) in the first place — everything is experienced through culture.

So far in history, cultural texts and texts in culture have been fairly clear-cut in their relation to referential reality, especially through their authorship — every natural language-based text has an author (a human author, that is) whose intention leads the purpose of the text (fiction, non-fiction, a lie, a belief, and so on), and further reception of the text forms its position and interpretation in culture. Today, we can see a proliferation of a new type of text: a computer-generated one. This new kind of synthetic text challenges the habitual ways of reading texts and assigning meaning to them.

3. The problem of generated content and its reference

The Internet in general can be an unreliable source for information because in large part the content on the Internet is not validated or controlled in a way that was customary in the old media. This results in ‘a network of communicative fragments’ with very different origins, purposes and properties, making interpretation of the texts very difficult (Verschuieren 2001, 87). Furthermore, the digital space today is accumulating the kind of texts that are non-texts in a Lotmanian sense — lists of links and keywords used for spamming purposes, spam emails and, finally, computer-generated synthetic content such as the output of GPT-3. The reaction to and interpretation of these different types of content varies from person to person and from text to text. For instance, spam comments are easy to recognise as non-significant; spam emails are also usually identifiable as such, even while their language and personalisation are becoming more sophisticated. The intention behind spam and phishing is to draw the reader’s attention and make them act in a certain way — to click on the links or to divulge sensitive information such as passwords. Phishing emails and Twitter troll posts are rather intentional messages with misleading information, that is, lies. Eventually, all these examples at least pretend to be related to non-fiction, that is, refer to actual reality and make the reader act in some way in the actual world.

Text generators such as GPT-3 and its relatives pose an interesting problem in the space of texts today: they can produce outcomes that are, by all appearances, unrecognisably similar to ‘real texts’. Such is the case of short opinion articles or funny lists of ‘whale facts’ (Shane 2020; see also Shane 2019). At the same time, these ‘texts’ do not have real reference or referentiality in the Fregean sense of the word. The referential reality of synthetic texts is tentative, only constructed and determined by a (human) judge comparing their expectations with the outcome. If the result meets expectations — ‘whale facts’ are correct, or the opinion piece is satisfactory in style and content — the judging editor may state that the text is ‘about external reality’, and publish it. If the result is incorrect, insufficient or inconsistent with the respective possible world (fictional or non-fictional) — e.g., the requested whale facts contain statements such as ‘dolphins also live in the desert’ (Shane 2020) or other statements that are false or nonsensical in relation to the external reality, the judgement may evaluate that the text is ‘about nothing’.

All in all, in the world we have texts and statements that do not have referential reality because their authors intended it so, or because the statements are about objects that do not exist in the actual world. Ryan (1991) takes the problem of fictional texts that do not refer strictly to actual/real reality and explains that they refer to their inner storyworlds (possible worlds) instead. In this sense, fictional texts are self-referential and their ‘truths’ exist within their storyverses. However, fiction is always *intentional* from its author’s side: it is written and published with intention, whatever its contents or complexity. Even with historical texts where the author’s person (or persons) is not clear (such as Shakespeare, or Qoheleth of the book of Ecclesiastes), the author is always implied (and sometimes the author is constructed out of their oeuvres, but that is another topic).

There also exists a type of texts and signs that are not intended but are *received* as signs or messages. Hereby I mean, for example, the characteristics of animistic beliefs and religion systems where certain natural phenomena are interpreted as signs of God(s). It is reasonable to assume that a storm or drought is simply a coincidental state of weather, and not an intentional ‘message’ from some entity to punish the village, but in an animistic system it may be interpreted as such.

3.1. Deepfakes

With the current deep learning technologies, it is possible to build models that imitate human voice in audio and human faces in photo-like still images and videos alike. In this manner, real faces in image and video can be swapped for other, borrowed faces, or new ones can be generated from scratch. The same applies to human voices. The imitations are realistic and convincing enough that they may be indistinguishable to the naked human eye from genuine video and photography. Considering the pixel composition of digital images and the relatively low pixel resolution of the majority of images found on the Internet, the best imitations are realistic enough to pass for a low- or medium-quality photograph or video of a real human. Eventually, a genuine photograph or a video is also just a representation of a living person. Although we take the ‘proof’ of photography for granted, the true existence of the object of reference in this sign relation can be verified only by presenting the real person next to the photo. In the absence of the real reference object, the verification in everyday life is provided by the complex system of social relations and institutions of trust, convincing us that if a friend shows us a photo with other people in it, and this friend usually tells the truth, then these people are probably real, not computer-generated fakes or ‘photoshopped’ images next to her in the photo.

Most commercial applications for deepfakes can currently be found in entertainment and retail. For instance, deepfake algorithms can be used to facilitate virtual ‘fitting rooms’ for clothes, hairstyles or cosmetics. In entertainment, deepfake technology supports imitating famous actors and singers (voice-, face- and body-swapping), seeing how one’s face would look when older or how one’s body would be as a dancer (full-body puppetry). (See Kietzmann et al. 2020 for an overview of the history, potential uses and technical functioning of this technology.)

Commercial applications aside, here I want to highlight two cases of deepfake use. The first example is the ‘borrowing’ of the face and voice of then-President Obama in an exemplary video by Jordan Peele. The short video available on the Internet, titled ‘You Won’t Believe What Obama Says in This Video!’, starts with an apparent speech by Obama, warning the viewer of the dangers of fake news and false information. The footage then continues with a lip-syncing video of Jordan Peele displayed next to the video of Obama. Thus, it becomes clear that the ‘message’ is actually being delivered by the actor, while Obama’s face and voice are only ‘borrowed’ for the occasion.

The second case concerns the ‘deepfaking’ of the voice of late Anthony Bourdain in a documentary about his life and death. The fact might have gone unnoticed were it not for the revelation of the documentary’s director Morgan Neville. In a promo-

tional interview, he boasted about the usage of AI in ‘helping’ Bourdain ‘say’ things that he had actually not recorded in real life (Simonite 2021). The previous recorded clips were used to synthesise new clips in Bourdain’s voice to express things that he had arguably written before, but not spoken. The case raised an immediate debate about the ethicality of such practice. However, deepfake technology has been deemed ethically and socially problematic on a much larger scale. Deepfake political videos are considered a serious threat to democracy (Poulsen 2021, 502). Even more intrusive and transgressive in nature, a disappointing practice of deepfake pornography can be found in the darker corners of the Internet (Poulsen 2021).

For all the above practices and reasons, it is important to understand how such technologies work and function in culture. The need for tools to deconstruct and analyse their functioning has never been more dire.

4. Synthetic content as mimicry and nonsense

There are two concepts that I find helpful in clarifying the structure and functioning of generative texts and deepfake videos. The concept of mimicry (Maran 2017), borrowed from biosemiotics and used here as an epistemological tool, helps differentiate the channels of synthetic message transmission and identify which elements of this process are ‘hijacked’ or different from habitual practices. The concept of nonsense (Katz and Shifman 2017) frames generated content as a cultural subtype of fiction and explains its functioning in culture in general.

The concept of mimicry is pertinent in the context of computer-generated media because it enables analysing deception in the communicative process regardless of the apparent identity or intentionality of the sender. The concept has been used in a wide range of disciplines including economy and anthropology (Maran 1999, 139). Certain military action strategies today find inspiration in mimicry-based deception models (Maran 2001, 2017).

Mimicry involves three participants – model, mimic, and receiver or dupe (Maran 1999, 2001, 2017) – and the relationships between them. Here, I focus on the relationship based on deception that is taking place between the mimic and the receiver (Maran 2017, 69). In the mimetic process, the mimic intercepts the communication channel between the model and the receiver in such a manner that the latter is deceived (duped) about the origins or the content of the message or both. A similar mechanism seems to be at work in the publication and media reception of a computer-generated news article in *The Guardian* in September 2020. The published article and the following explanatory comment hide the agency and authorship of the unnamed editorial team and name OpenAI’s text generator GPT-3 as the ‘author’ of the article (cf. Viidalepp 2022 for an analysis and longer explanation). But, more importantly, mimicry is a liminal type of phenomenon, functioning ‘by lingering between perception and non-perception, recognising and non-recognising, meaning-making and a lack thereof’ (Maran 2017, 60). This is quite similar to the effect created by synthetic texts that seem to draw attention and interest precisely due to their residing in the space between a text and a non-text, hovering between normal-

ity and a sense of otherness, and keeping the reader guessing about the sender and intention behind the message.

Nonsense has a particular role in cultural communication. Katz and Shifman analyse the concept using the example of ‘digital memetic nonsense’, a type of user-generated nonsensical visual meme that is supported and facilitated by characteristics of today’s new media space (Katz and Shifman 2017, 828). They outline three historical types of nonsense. Firstly, nonsense as a lack of meaning has been used as an artistic tool to challenge the notion of truth. This is especially characterised in the postmodernist use of intertextuality in pastiche, an art form offering ‘imitation for the sake of imitation; a state in which one text cites another without adding any referential meaning’ (Katz and Shifman 2017, 834). Secondly, nonsense as a deconstruction of meaning ‘examines the ability of signs to create meaning and highlights their failure’ (ibid., 826). They find that deconstruction of meaning is characteristic to works of literature such as *Finnegans Wake* or *Alice in Wonderland* where language and word games are used as parts of the plot, or *Waiting for Godot*, which experiments with the tentative inclusion of extratextual reality (the two characters in the play are waiting for a third character who never arrives). After an initial confusion, all the above works invite their audience to a deeper deciphering of the ‘meaning’ of the text. The third type of nonsense mentioned by Katz and Shifman is a play with meaning that can be found in certain limericks or in instances of polysemy that offer the reader multiple ways to interpret or read the same content (ibid., 827).

On the one hand, current digital media space features user-generated nonsensical content that mostly performs the phatic function, as demonstrated by Katz and Shifman (2017). In such a content, nonsense is used as a way to filter ingroup members — those who understand the joke — from the outgroup, as well as to enforce the ties within the phatic community. Nonsense and ongoing negotiation of the rules for its composition serve to keep the community alive, that is, engaged and constantly following each other (Katz and Shifman 2017). This type of nonsense can employ different approaches to question meaning (play, deconstruction or lack of meaning), but in every case analysed by the forementioned authors, the nonsensical meme is created and posted intentionally by Internet users. Therefore, it always has an author, a sender in the traditional sense.

5. The referential shift and the semiotic functioning of synthetic media

The structure of the communicative act contains in typical models the sender (author), the message, the channel and the receiver. Roman Jakobson (1971) also adds the context and the code. In a simplistic example of a typical, regular text, the sender (author) prepares the message (text) and uses some channel (print or new media) to transmit the message to the receiver (reader). The context and the code may help in asserting the preferred referential reality for the message. For instance, the context of a daily newspaper in the normative language positions the text as something about the actual world; the shelf labelled ‘Fiction’ in a bookstore or lines of text

displayed in a certain way categorise the text as being about imaginary entities, or a book of poetry, respectively.

However, the author of the synthetic text is displaced: it is no more the ‘originator’ of the text but its editor, publisher and even the reader who projects a possible referential world to make sense of the ‘message’. As part of the creative writing process is delegated to the machine, the author(s) may feel a decrease in their work and agency in producing the text, and are therefore eager to allocate their own agency outside of themselves — to the machine. The synthetic text also lacks a proper referent in Frege’s sense. In the case of GPT-3 output, it is based on the statistical probability of co-occurring words. The result refers to the statistics — patterns and regularities — within its training data. That it *happens* to be similar to natural language texts to an extent where it may be undistinguishable to the human reader is entirely a coincidence. (It is more likely to be perceived as a coincidence when the output does not meet expectations and is discarded as a faulty or bad result.)

The synthetic text may be attributed a referential reality by its real human authors (i.e. editors, publishers, readers) with the act of publication (making publicly available) of the text. This referential reality, in the case of GPT-3 op-ed, is borrowed from the secondary narrative world of science fiction (in the sense that Doležel 1979, 199 compares the notion to the Proppian morphological structures of the folktale). The cultural superstructure (Lotman 2005, 213) of the sci-fi universe prescribes AI robots as better-than-human characters that hold self-conscious monologues and aim to conquer the human world. Therefore, GPT-3 ‘the robot author’ is implicitly posited as omniscient (saying ‘Believe me, being omnipotent doesn’t get me anywhere’ in the op-ed).

In a deepfake video, the layers of the communicative act become manifold. As shown in the mimicry model, a second, true sender is added to the schema – the mimic delivering their message. As deepfake face- and voice-swapping videos take an effort to deliver a believable result, the process is less due to chance and more due to the intentionality of the mimic. Therefore, deepfakes may be conceived as intentional deceptive messages hijacking a trusted ‘channel’ — the model’s image and voice. In this case, the referential reality is likely still the actual world, but the message delivered is a lie. The lie can be about the content, but this is a separate topic for research. Foremost, the deepfake is a lie about the channel and the sender, but this may not be immediately recognisable.

6. Conclusion

Generative texts belong ontologically to the category of literature (fiction) first, regardless of their content. Their referential reality is first within the internal world of the text. Whether such a world can even be conceived is a decision and task for the first reader of the text who confirms whether the output corresponds to expectations and will be taken into account, to be worked on further. Then, further editors reform the output, turning it into a veritable text. In the case of any synthetic text, the validation of its fitness or eligibility is a task undertaken by its human ‘author’ —

the first reader of the output, the person who orders the output (wrote the prompt) or decides the publication of the output. This person is, in fact, the real author of the text. This is so regardless of whether part of the creative writing process was outsourced to a machine or not. It is really important to remember this and recognise such authorship every single time, because very often when the synthetic outcome looks acceptable in appearance, there tend to be individuals who attribute in this process a greater agency to the machine than it deserves. The human author can transform the text into a message that refers to something else or has a different purpose, but this is then a changed, new message. Without the interference by the real author, the message or the output itself is nonsense, although it may not be recognisable as such.

A deepfake is conceivable as a mimetic act where the mimic hijacks the communication channel to deliver a message that is essentially a lie. The reference of the deceitful message can be the sender's identity, the content of the deceitful claim or any other kind of object intended by the real sender, the mimic. In both cases of generative media, the real author (sender) is hidden, and the exact nature of the message is unknown or unclear.

References

- Doležel, Lubomír. 'Extensional and Intensional Narrative Worlds'. *Poetics* 8, no. 1–2 (1979): 193–211. doi: [10.1016/0304-422X\(79\)90020-2](https://doi.org/10.1016/0304-422X(79)90020-2)
- Eco, Umberto. *The Role of the Reader: Explorations in the Semiotics of Texts*. Advances in Semiotics. Bloomington: Indiana University Press, 1979.
- Frege, Gottlob. 'Sense and Reference'. *The Philosophical Review* 57, no. 3 (1948): 209–30. doi: [10.2307/2181485](https://doi.org/10.2307/2181485)
- Jakobson, Roman. *Selected Writings II: Word and Language*. The Hague/Paris: Mouton, 1971.
- Johansen, Jørgen Dines. *Literary Discourse: A Semiotic-Pragmatic Approach to Literature*. Toronto Studies in Semiotics and Communication. Toronto/Buffalo: University of Toronto Press, 2002.
- Katz, Yuval, and Limor Shifman. 'Making Sense? The Structure and Meanings of Digital Memetic Nonsense'. *Information, Communication & Society* 20, no. 6 (2017): 825–42. doi: [10.1080/1369118X.2017.1291702](https://doi.org/10.1080/1369118X.2017.1291702)
- Kietzmann, Jan, Linda W. Lee, Ian P. McCarthy, and Tim C. Kietzmann. 'Deepfakes: Trick or Treat?' *Business Horizons* 63, no. 2 (2020): 135–46. doi: [10.1016/j.bushor.2019.11.006](https://doi.org/10.1016/j.bushor.2019.11.006)
- Lotman, Juri. 'On the Semiosphere'. *Sign Systems Studies* 33, no. 1 (2005): 205–29.
- Maran, Timo. 'A Note on the Semiotics of Biological Mimicry'. *Sign Systems Studies* 27 (1999): 139–47. doi: [10.12697/SSS.1999.27.07](https://doi.org/10.12697/SSS.1999.27.07)
- Maran, Timo. 'Mimicry: Towards a Semiotic Understanding of Nature'. *Sign Systems Studies* 29, no. 1 (2001): 325–39. doi: [10.12697/SSS.2001.29.1.20](https://doi.org/10.12697/SSS.2001.29.1.20)
- Maran, Timo. *Mimicry and Meaning: Structure and Semiotics of Biological Mimicry*. Cham: Springer, 2017.

-
- Nöth, Winfried. *Handbook of Semiotics*. Bloomington/Indianapolis: Indiana University Press, 1995.
- Pavel, Thomas G. *Fictional Worlds*. Cambridge, MA: Harvard University Press, 1986.
- Poulsen, Søren Vigild. 'Face Off – A Semiotic Technology Study of Software for Making Deepfakes'. *Sign Systems Studies* 49, no. 3–4 (2021): 489–508. doi: [10.12697/SSS.2021.49.3-4.12](https://doi.org/10.12697/SSS.2021.49.3-4.12)
- Ryan, Marie-Laure. *Possible Worlds, Artificial Intelligence, and Narrative Theory*. Bloomington/Indianapolis: Indiana University Press, 1991.
- Shane, Janelle. 'Facts about Whales'. *AI Weirdness*, July 13, 2020. <https://www.aiweirdness.com/facts-about-whales-20-07-13/>
- Shane, Janelle. *You Look Like a Thing and I Love You: How Artificial Intelligence Works and Why It's Making the World a Weirder Place*. New York: Voracious/Little, Brown and Company, 2019.
- Simonite, Tom. 'Are These the Hidden Deepfakes in the Anthony Bourdain Movie?' *Wired*, 23 August 2021. <https://www.wired.com/story/these-hidden-deepfakes-anthony-bourdain-movie/>
- Verschueren, Jef. 'Pragmatics'. In *The Routledge Companion to Semiotics and Linguistics*, edited by Paul Copley, 83–94. London/New York: Routledge, 2001.
- Viidalepp, Auli. 'Sociocommunicative Functions of a Generative Text: The Case of GPT-3'. *Lexia. Rivista di semiotica* 39–40 (2022): 177-192. doi: [10.53136/979122180426310](https://doi.org/10.53136/979122180426310)