

<https://doi.org/10.53656/phil2026-02-07>

Philosophy and Modernity
Философия и современность

POST-TRUTH AND ARTIFICIAL INTELLIGENCE: FROM EMOTIONAL NARRATIVES TO SYNTHETIC AUTHENTICITY

Silvia Petrova

South-West University "Neofit Rilski"

Abstract. This article examines post-truth as a cultural regime of public sphere emerging at the intersection between the attention economy and platform logic. In this environment, emotionally charged narratives and social indicators of trust often replace expert verification. The argument is that artificial intelligence (AI) acts not simply as an accelerator of disinformation, but as a cultural technique that reorganizes the criteria for truth. On this basis, the concepts of “algorithmic public sphere” (fragmented, affectively structured publics), “synthetic authenticity” (the aesthetics of verisimilitude in deepfakes and generative models), and the transformation of “regimes of truth” (Foucault) into model-predictive standards. The conclusion emphasizes the need for media and algorithmic literacy, as well as research tools for the ethical and epistemic management of synthetic realities.

Keywords: post-truth; algorithmic public sphere; artificial intelligence; attention economy; platforms; regimes of truth

Discourses on post-truth often organise the world from an ethical perspective, using the concepts of fact and lie to describe audiences as either ‘informed’ or ‘ignorant’ and susceptible to manipulation. However, such simplification obscures the structural conditions of contemporary public sphere. A cultural studies perspective enables us to view post-truth as a regime of mediality, in which emotionally charged narratives take precedence over cognitive complexity. Platform algorithms prioritise engaging content over verifiable statements. Rather than being delegated to institutions, trust is placed in social proof (likes and shares) and the charisma of intermediaries. In this situation, not only does artificial intelligence accelerate disinformation, it also reshapes what is considered knowable, reliable and visible. This raises the question: how does AI participate in the production of ‘truth’?

From post-truth to algorithmic public sphere

Post-truth is a symptom of the convergence between the attention economy and platform logic. Rather than a shared public space, we have fragmented micro-

communities consisting of personalised narratives. In this context, facts do not disappear; they simply lose their advantage over emotions and narrative. Platforms reward content that holds people's attention, even if it does not necessarily provide information or knowledge. Truth now competes with attention productivity. This fragmentation of the public sphere has also been discussed in Bulgarian media research, where digital platforms are described as environments that structurally privilege segmented interpretative communities over shared deliberative spaces (Vangelov, 2023).

Many researchers in this field emphasise that post-truth should not be understood as an 'era of lies' or a decline in rationality, but rather as a transformation in the conditions of knowledge production and circulation. Hayden White (1987) observed towards the end of the 20-th century that both historiography and public narratives had never been 'pure facts', but had always been structured by narrative codes and rhetorical techniques. In this sense, post-truth represents a shift in epistemic hierarchies: the emotional and affective now take precedence over the cognitive and expert. However, this is the result of structural changes in the media environment.

David Altridge and Georg Frank (Frank, 1999) describe the contemporary cultural environment as an 'attention economy', in which the ability to attract and retain the attention of the audience is more important than the information itself. In conditions of oversaturation, facts lose their significance and compete with symbols, images and emotions. Post-truth is the result of the logic of media markets: what 'works' is not so much the verifiable, but the viral.

José van Dijck, Thomas Poell and Martin de Waal (van Dijck, Poell & de Waal, 2018) introduce the concept of 'platform logic' to explain how social networks are reorganising public sphere. The algorithms of platforms such as Facebook, TikTok and YouTube are not neutral tools, but infrastructures of selection that arrange content according to criteria such as engagement, viewing time or likelihood of sharing. In this sense, post-truth is an algorithmic environment in which truth competes with the effectiveness of recommendations.

Sarah Ahmed (2004) and Jonathan Grayson (2019) emphasise the role of affect as a social glue. In the post-truth era, news not only informs, but also ignites feelings that create and unite audiences. These audiences are not interested in facts or rational arguments, but rather in a shared emotional experience, such as outrage, anger, fear or pride. Algorithms support these audiences because affective content is statistically more engaging.

In 2015, Zeynep Tufekci introduced the concept of algorithmic public sphere. According to Tufekci, the public sphere is no longer a common space for rational debate (as conceived by Habermas), but rather a network of fragmented information bubbles structured by platform codes. Rather than merely filtering content, algorithms construct realities by deciding which content will be visible and which will remain hidden. This means that post-truth is not merely the absence of facts, but rather a structural dependence on algorithmic logic for visibility.

In light of this perspective, post-truth can be described as a cultural regime of public sphere in which facts lose their monopoly on credibility and must compete with emotions for attention. The authority of experts is displaced by social evidence in the form of likes, shares and memes. Public sphere is reorganised by algorithmic intermediaries who determine what content is accessible. Consequently, truth becomes a category defined by metrics such as visibility and audience reach rather than institutional standards. Rather than being considered a moral decline, the transition from post-truth to algorithmic public sphere must be viewed as a transformation in the cultural techniques of credibility. Truth is now algorithmically moderated, and its social impact depends on how well it fits into the infrastructures of platform logic.

Artificial intelligence as a cultural technique

The concept of a cultural technique refers to the practices through which people organise, symbolise and manage the world, such as writing, calculating and mapping (Siegert, 2015). The key point is that these techniques do more than just serve as tools; they also shape the very content of culture. Similarly, artificial intelligence should not be considered merely as a technology for processing data, but as a cultural technique that shapes our perception of truth, knowledge, and authenticity. AI does not merely calculate; it translates social reality into numbers, which can then be returned to symbolic forms such as text, images and sound. As N. Tsenkov notes, AI creates a speech situation that is essentially new compared to the conventional communication paradigm (Tsenkov, 2024). This dual process of reduction and reconstruction rewrites cultural meanings. AI can be understood as a complex of three cultural operations:

1. Classification

Data is sorted into categories such as friend/enemy, positive/negative and true/false. This is reminiscent of the structural techniques of culture, which have always relied on binary oppositions (Levi-Strauss). In the contemporary media environment, AI classifications influence the public sphere, for instance by filtering content in news feeds.

2. Prediction

AI models probabilities: who will click, who will buy and who will vote. This activity is similar to the role of statistics in modernity, but now it happens in real time and on a scale that can actually change behaviour. Culturally, this means that the future is managed as a probabilistic simulation rather than a political project or moral vision.

3. Generation

Generative models produce texts, images and voices that ‘imitate’ cultural forms. Here, plausibility becomes more important than empirical verification. The

cultural consequence of this is the emergence of ‘synthetic authenticity’ – the artificially produced sounds that are ‘more real’ precisely because they are statistically optimised to match the audience’s expectations.

Shoshana Zuboff (2019) describes the logic of ‘surveillance capitalism’, whereby data is used to predict and manage behaviour. This results in the normalisation of behavioural predictability – people perceive themselves through algorithmic mirrors, such as advertisements, recommendations and ratings. Another consequence is the transformation of the personal into the statistical: the individual voice is subordinated to the probabilistic model. Consequently, the criteria for truth change – what is ‘valid’ is what fits the model, not necessarily what is factually true.

AI, as a cultural technique, transforms our understanding of knowledge in several key ways:

- From truth to probability: facts are replaced by probability models – knowledge becomes statistical.

- From objectivity to operability: what is reliable is what ‘works’ – optimises engagement or achieves results.

- From author to algorithm: the authority of knowledge shifts from human experts to technologies, which are often unclear to most audiences in terms of how they work.

- From verification to trust in metrics: instead of public debate about truth, we trust indicators such as clicks, views and reliability ratings from the algorithms themselves.

As a cultural technique, AI is a new technology and an infrastructure of symbolic production that redefines the criteria for knowledge and truth. While post-truth undermines trust in expertise through affect and narrative, AI goes one step further by replacing expertise with operationalised plausibility, which is measured by algorithmic metrics.

The aesthetics of plausibility

At the end of the last century, Jean Baudrillard (Baudrillard, 2019) spoke of simulacra: images and signs that replace reality. In hyperreality, the media produce worlds that seem more real than reality itself. Today, generative systems such as ChatGPT, Midjourney and DALL-E create texts, images and voices with no real reference, yet they function as credible representations. For instance, a deepfake video of a politician could elicit reactions and decisions, regardless of its veracity. AI not only imitates but also mass-produces these simulacra, which circulate faster and more convincingly than factual content. The cultural consequence is the blurring of the distinction between truth and simulation.

In the context of digital culture and the algorithmic control of public sphere, the main criterion for validity is plausibility rather than truth. Content does not need to be verified; it is enough if it appears ‘credible’, resonates with viewers’

expectations or evokes emotion. This is a key characteristic of the post-truth era: if content makes an impression, it is perceived as ‘probably true’. In this sense, we can talk about synthetic authenticity: content produced by AI appears ‘authentic’ precisely because it matches the audience’s statistical expectations perfectly. It is not perceived as a “lie”, but as a new form of truth. The aesthetics of verisimilitude are particularly prevalent in visual culture: Deepfake videos and photos are not analysed critically, but are accepted as true based on emotional persuasion (“I see it, so it must be true”). Additionally, memetic forms combine humour and irony to create ‘truths through laughter’. The visual algorithms on TikTok and Instagram prioritise content that matches cultural patterns over reality. The verisimilitude produced by AI is an example of a post-truth persuasion technique, where truth is no longer sought through verification, but through visual and narrative persuasion. Baudrillard’s simulacra are reproduced en masse and become part of everyday public sphere.

Truth, institutional authority and power

Michel Foucault emphasises that truth is not a universal constant, but is constructed historically through discursive practices. Power does not merely prohibit or repress; it also produces knowledge that is imposed as truth. He refers to these as ‘regimes of truth’: the institutional, discursive and technological conditions that determine what is recognised as knowledge and what is excluded as invalid.

Throughout the modern era, institutions such as science, law and journalism have acted as arbiters of truth. In the digital age, however, this function is shifting towards algorithmic intermediaries – platforms and models that filter and redefine the criteria for authenticity. In the post-truth era, we are witnessing a transformation from expert consensus to algorithmic recommendation: if truth once meant the former, today the latter is often sufficient. Recommendation systems act as epistemic filters – only the first results are considered valid. Truth is replaced by engagement metrics: clicks, likes and views. Public trust is not delegated to facts, but to the number of shares. Each user receives their own ‘personalised truth’, optimised by the model according to the probability of receiving attention. Thus, a new regime of truth emerges, with machine prediction becoming the cultural arbiter of reality.

In light of Foucault’s ideas, algorithms are discursive agents, not “tools”. They determine what content circulates and what remains hidden, and they impose normative frameworks, dictating which topics and styles are acceptable. For example, social networks censor certain images and words. Algorithms reproduce ideological structures through algorithmic bias, which reinforces existing inequalities based on race, gender and social status. This is power through discourse reformatted as power through models, where truth is not articulated in dialogue, but optimised statistically.

This new regime has a number of cultural consequences. ‘Normal’ is now whatever the model predicts as most likely. Culture is reduced to common and engaging content. Factual truth loses priority to operational truth. Public sphere becomes fragmented – Instead of one public arena, there are multiple micro-realities, each with its own ‘truths’. The public trusts the machine’s decisions, such as whether content is ‘spam’ or whether a photo is authentic, because they perceive the machine as objective and impartial, even though it is socially and culturally coded.

Foucault discusses biopower, which he defines as the power that governs bodies through disciplinary and normalising techniques. Algorithms can be considered new carriers of biopower because they regulate our attention, determining what we see, think and how we react, and discipline our behaviour by recommending the ‘right’ lifestyle, trends and influencer norms. Furthermore, they shape our sense of self – we learn who we are through the content that algorithms present to us and reinforce. Thus, these new regimes of truth constitute a new form of epistemic power, with algorithms acting as arbiters of reality. They do not merely filter information; they produce truths that function as cultural and social facts. While post-truth favours emotion over expertise, AI goes further by replacing expertise with operationalised statistics that are perceived as truth.

In the current post-truth regime, power does not repress; it seduces by acting through desires for belonging, visibility and self-care. AI amplifies this dynamic by translating desires into data and data into recommendations that shape our desires. In the context of AI and post-truth, this dynamic becomes apparent: algorithms not only restrict access to information, but also participate in constructing subjectivity – who we are, what we desire and what we consider ‘normal’. As Foucault demonstrates, power is not external to desire; rather, it shapes desire from within. On social networks and through AI, we can observe a cycle of co-production: users express their preferences through likes, searches and comments; algorithms ‘learn’ and return content that reinforces these preferences; users internalise this flow as their own desires. Thus, desire becomes algorithmic; it is not purely personal, but the result of an infrastructure that shapes and reinforces it.

In *Discipline and Punish* (1975), Foucault describes the Panopticon as a model of disciplinary power, whereby individuals self-regulate because they believe they are being watched. In the digital context, the Panopticon has been replaced by a ‘Synopticon’, or even an ‘Omnipticon’ – a system in which everyone watches everyone else through algorithmic interfaces. Users voluntarily produce data in the form of selfies, statuses and GPS locations, and algorithms use this data to build models of normality. Consequently, people begin to self-discipline according to the standards suggested by the platform in terms of how they look, what they buy and what they think.

Algorithmic audiences

Early theories of mass communication viewed audiences as passive and susceptible to manipulation. However, later cultural studies have shown that audiences are active, interpreting, rethinking and reworking messages. In the age of social media, this activity has intensified and taken on a new dimension: audiences are now algorithmically moderated. Today, algorithms filter content and gather and divide people into audiences based on behavioural data (clicks, likes, viewing time), emotional reactions (anger, fear, outrage, joy) and likelihood of engagement. This phenomenon has given rise to the concept of ‘algorithmic audiences’ (Gillespie, 2018): communities that are not self-organising, but rather assembled by the machine logic of recommendations.

Sarah Ahmed (2004) and Ben Anderson (2009) demonstrate that emotion is not merely a personal sensation, but rather a social force that unites communities. In algorithmic environments, anger provokes more comments and is therefore rewarded, while fear is amplified because people share to ‘warn others’. Humour and irony create a sense of belonging through memes and sarcasm. Algorithms prioritise content that provokes emotions because it holds attention. Consequently, rather than a shared public sphere as described by Habermas, we now have a fragmented reality consisting of echo chambers (Sunstein, 2001), where only similar opinions are heard, and filter bubbles (Pariser, 2011), which are individualised streams of content that hide different points of view. Furthermore, emotions are not neutralised through debate, but rather escalate through repetition and algorithmic selection, creating affective spirals. The cultural result is a sense of irreconcilable groups – ‘us’ versus ‘them’.

Algorithms are not neutral. They determine which topics are important and normalise certain styles of speech, such as memes, short clips and emotional formulas. They also exclude other forms of discourse, such as long analyses and nuanced positions. This is a form of algorithmic power: culture is structured according to what generates the most attention and profit, rather than what is important for democratic debate.

Algorithmic audiences are affective constructions that reshape social spaces. Therefore, AI and algorithms not only disseminate information, but also produce cultural worlds in which truth is a function of emotional engagement.

Authorship without an author

Traditional media has always worked with representations – a photograph referring to a real event, a text describing an event or a voice belonging to a real person, for example. However, with generative artificial intelligence, we are witnessing a shift from representation to simulation. Forms are being created that have no basis in reality yet appear convincing. Deepfake videos, for example, show politicians saying words they have never actually uttered. Generative models such as

Midjourney and Stable Diffusion produce ‘photographic’ images of non-existent people. Synthetic voices imitate famous personalities or create ‘new authors’ who seem credible. This is not just media manipulation; it is a new form of cultural production.

In modern cultural tradition, the author is seen as the guarantor of meaning and truth (Foucault, 1991). In the age of AI, however, authorship is decentralised: content does not originate from individual intention, but from statistical data combinations. Responsibility is distributed among model developers, platforms and users. This creates a ‘zone of irresponsibility’. The concept of originality becomes blurred – texts and images are the result of existing cultural artefacts being reworked. Consequently, we can discuss authorship in the absence of an author – content that possesses style and persuasiveness, yet lacks an individual origin. Here, synthetic authenticity becomes important. Visually and acoustically perfect content seems more real than reality. AI influencers are perceived by millions of users as ‘authentic personalities’ with their own positions, emotions and identities. A synthetic voice in a podcast may sound ‘more professional’ and ‘more credible’ than a real speaker. In this context, authenticity is no longer ontological (related to reality), but rather aesthetic – a product of smooth form, stylistic persuasiveness and algorithmic optimisation. One of the main effects is epistemic erosion: the emergence of mass-generated simulations undermines the audience’s ability to distinguish fact from fiction. This can lead to epistemic scepticism – the feeling that nothing is certain.

The political use of deepfake videos is no less significant. They can be used to discredit political opponents or manipulate elections. This leads to normalisation – users become accustomed to the ‘synthetic’ and accept it as part of their everyday cultural environment. However, if everything can be fake, critical debate is replaced by cynicism and apathy (‘it doesn’t matter what the truth is’). Synthetic images and voices are not just a passing trend in the media; they represent a cultural transformation in the concepts of truth, authorship and authenticity. In the age of AI, truth becomes distinct from reality – It equates to algorithmically produced persuasiveness that circulates as fact.

Conclusion

Analysing post-truth and artificial intelligence together shows that these aren’t random or temporary “flaws” in modern communication, but rather structural elements of cultural modernity in its digital phase. Rather than being reduced to a moralistic narrative about an era of lies, post-truth should be considered a regime of public sphere in which emotions, narratives and social proof (such as likes, shares and visibility) prevail over factual verification. Artificial intelligence is not merely a technological amplifier of these trends, but a cultural technique that alters the very concepts of truth and trust. Post-truth and artificial intelligence are interconnected as both the cause and consequence of the same transformation: the digitisation of

the public sphere and the algorithmisation of knowledge. Truth is no longer a common good guaranteed by institutions, but rather an operational value measured by affective resonance and algorithmic relevance.

In this regard, new analytical tools are needed, such as cultural data analytics and research into digital discourse. There are also normative dimensions to the problem, such as establishing forms of algorithm accountability and cultural literacy that strengthen democratic public sphere in the context of synthetic realities.

Post-truth and artificial intelligence do not mark the end of truth, but rather a re-organisation of epistemic regimes. Truth does not disappear; it competes with other social reality currencies, such as visibility, emotional resonance and operational utility. In this sense, the task is not to restore a lost 'gold standard', but to establish infrastructures that prevent plausibility from dominating verification and ensure that synthetic authenticity is culturally marked and traceable.

Acknowledgements and Funding

The financial support from of the National Science Fund under the project "Virtual Reality Research Centre at Faculty of Philosophy, South-West University "Neofit Rilski", КП-06-H80/3 is gratefully acknowledged.

REFERENCES

- Ahmed, S. (2004). *The cultural politics of emotion*. Edinburgh University Press.
- Altheide, D. L., & Snow, R. P. (1979). *Media logic*. Sage.
- Baudrillard, J. (2019). *Simulatsia i simulakrum* [Simulacra and simulation]. Tugedar Akademi. [In Bulgarian]
- Franck, G. (1999). Scientific communication – A vanity fair? *Science*, 286(5437), 53 – 55. <https://doi.org/10.1126/science.286.5437.53>
- Franck, G. (2018). The economy of attention. *Journal of Sociology*. <https://doi.org/10.1177/1440783318811778>
- Foucault, M. (1991). Shto e avtor? *Literaturen vestnik*, 21. [In Bulgarian]
- Foucault, M. (2019). *Istoria na seksualnostta: T. 1. Volyata za znanie. Kritika i Humanizam*. [In Bulgarian]
- Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
- Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin.
- Sunstein, C. R. (2001). *Republic.com*. Princeton University Press.
- Tsenkov, N. (2024). Mezhdur Omir i ChatGPT? *Chuzhdoezikovo obuchenie / Foreign Language Teaching*, (2), 206 – 221. [In Bulgarian]

- Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Colorado Technology Law Journal*, 13(2), 203 – 218.
- Van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.
- Vangelov, N. (2023). Art in advertising. *Medialog*, 13.
- White, H. (1987). *The content of the form: Narrative discourse and historical representation*. Johns Hopkins University Press.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

✉ **Dr. Silvia Petrova, Assoc. Prof.**
ORCID iD: 0000-0003-4809-4061
South-West University “Neofit Rilski”
66, Ivan Mihaylov
2700 Blagoevgrad, Bulgaria
E-mail: silvia.petrova@swu.bg