

WHEN CULTURE FAILS IN CODE: “CULTURAL HALLUCINATIONS” OF MULTIMODAL AI SYSTEMS

Introduction

This paper explores generative artificial intelligence (AI) as an emerging cultural agent, with a focus on the phenomenon of *cultural hallucinations*. This term builds on the better-known concept of *AI hallucinations* – cases where a model generates false or invented information. However, cultural hallucinations refer to outputs that may be factually accurate, but do not match the cultural patterns a reader or viewer expects.

As globalised generative AI systems – trained on large-scale, predominantly Western datasets and capable of producing text, images, audio, and video in many languages – become increasingly widespread, they play a growing role in shaping how cultural ideas are formed and discussed. In this context, the ability to interpret and critically assess AI-generated content through the lens of cultural patterns becomes essential. This is particularly relevant in education, media, and intercultural communication, where cultural accuracy, representation, and authenticity matter.

1. Multimodality and Cultural Identity

The effectiveness of any communication depends on the cultural backgrounds of the communicators. Cultural background, which can be described as “a set of semiotic systems, a set of systems of meaning, all of which interrelate” [2, p. 4], is inherently multimodal. It is shaped not only by the language we speak, but also by the music we grow up with, the architecture of our cities, and the food we eat [3]. These everyday experiences form *cultural patterns* – stable, culturally conditioned ways of expressing complex cultural concepts. Their multicomponent nature reflects the multimodal character of cultural identity and corresponds to *multimodality* as a natural feature of human communication that combines different semiotic resources – *modes* – to create meaning.

Kress defines a *mode* as “a socially shaped and culturally given semiotic resource for making meaning” [4, p. 79]. Because modes are socially constructed, they are also culturally relative: “What counts as a mode is what a community has developed into and uses as a mode” [1, p. 78].

This is why *multimodal literacy* – together with *AI literacy* and awareness of the risk of *cultural hallucinations* – has become a key element of contemporary communication.

2. Multimodal and AI Literacy as Essential Skills in Today’s World

Multimodal literacy refers to the ability to interpret, understand, and create meaning using multiple semiotic resources, such as visuals, sounds, gestures, and texts. However, as educational researcher Nik Peachy points out, traditional education often focuses excessively on textual analysis, potentially overlooking important multimodal dimensions of communication [6].

The emergence of multimodal generative AI models makes multimodal literacy even more relevant, as it applies to the interpretation of all types of texts – whether created by humans or by machines. This growing relevance brings multimodal literacy into close connection with AI literacy, which involves understanding how artificial intelligence functions, including an awareness of its strengths, limitations, and potential biases. In the context of generative models, AI literacy focuses on the ability to critically evaluate AI-generated content – not only by checking for factual inaccuracies in text, but also by

identifying culturally inappropriate or inconsistent multimodal elements that AI systems may produce.

3. Identifying Cultural Hallucinations: A Case Study

To illustrate the points discussed above about the importance of combining multimodal and AI literacy, let us consider a case study involving the Lithuanian cultural concept of *knygnešiai* – book smugglers who secretly transported books and newspapers across borders during the period of the Russian ban on Lithuanian publications (1864–1904) [5].

ChatGPT was asked to generate a video concept to introduce this cultural phenomenon to an international audience. The result was impressive: the script respected historical facts and followed the expected structure of the cultural pattern at all levels. The narrative accurately described the historical period, the political context, and the cultural significance of the *knygnešiai* for Lithuania. It also included detailed descriptions of the visual and audio components:

- **VISUALS:** Scenes of the Lithuanian countryside and daily life at the end of the 19th century. A *knygnešys* secretly preparing for a journey and carefully hiding forbidden books. A tense and dangerous journey through shadowy forests, across rivers, and past guarded borders. An encounter with Tsarist officials. Historical glimpses: authentic old books, documents, and photographs of *knygnešiai*.
- **SOUND EFFECTS:** The soft rustling of hidden bags, a distant dog barking in the night, the heavy footsteps of soldiers, and the quiet turning of book pages.

However, in the next step, when a video generation tool (InVideo AI) was used to visualise the script, something unexpected occurred. The visual content significantly diverged from cultural expectations: some frames included inaccurate clothing, modern-looking settings, or elements completely out of place in 19th-century Lithuania. These multimodal inconsistencies – particularly in the visual and spatial modes – are examples of what can be defined as *cultural hallucinations*. These are not simple factual mistakes; rather, they reveal deeper issues in how AI systems interpret and represent cultural patterns. Such hallucinations occur when multimodal content is generated using global or generic datasets that lack sufficient exposure to localised, culturally specific material.

4. The Reasons for Cultural Hallucinations

Let us now try to explore why *cultural hallucinations* occur. One key reason lies in how generative AI constructs its “understanding” of the world. Meaning-making in human communication is deeply embodied, socially shaped, and culturally situated. Human cognition is inherently multimodal – we perceive and make sense of our surroundings through a combination of senses: sight, hearing, touch, smell or taste. In contrast, generative AI systems do not perceive the world through sensory experience. Instead, they rely on statistical modelling. These systems process texts, images, sounds, or videos as sequences of data points – numerical vectors representing patterns in training datasets. They learn correlations between elements but lack embodied or culturally contextualised understanding. This means that AI-generated outputs are shaped not by lived experience, but by the frequency and co-occurrence of tokens in the training data.

Another central factor is *data bias*. Most global generative AI models (such as GPT, Claude, or Gemini) are trained primarily on large-scale, internet-based datasets dominated by English-language content and Western-centric cultural material. While these datasets are vast, they are not culturally balanced. Minority or localised knowledge systems are underrepresented or simplified. Consequently, when such models are prompted to generate

content related to culturally specific concepts – especially from smaller, or linguistically minor cultures – they often "hallucinate" representations that align more with dominant global discourses than with the authentic context.

These hallucinations can affect all modes of generation – text, image, sound, and video – and often manifest in subtle ways: stereotyped imagery, incorrect material culture, anachronisms etc.

Thus, *cultural hallucinations* reveal the structural asymmetries in global AI development. They emerge not simply from model failure, but from the absence of diverse, culturally specific data – and from the lack of interpretive frameworks that would allow AI to distinguish between global generalisations and local meanings.

5. Addressing Cultural Hallucinations: National AI Models

Given these insights, how can we address or minimise *cultural hallucinations*? One promising direction is the development and implementation of national or culturally specific AI models. Can the rise of national AI models help overcome the limitations of global models by incorporating culturally specific datasets? The main advantages of national LLMs include:

- **Supporting and preserving local languages**, especially those at risk of extinction or underrepresented in digital communication.
- **Reducing cultural hallucinations** by training AI on multimodal datasets that reflect local cultural, historical, and social contexts.
- **Enhancing cultural authenticity** in AI-generated educational, communicative, and creative content.
- **Improving national participation in global discourse** by enabling high-quality, culturally sensitive automatic translation and content creation.

By integrating culturally specific knowledge directly into their training pipelines, national models can significantly lower the risk of cultural hallucinations and increase the accuracy and cultural relevance of generated content.

However, national AI models also pose potential challenges. These include intensified international competition and geopolitical tensions surrounding technological development, as well as concerns related to data privacy, censorship, and ethical oversight. Without transparent frameworks for governance and accountability, national models may reinforce political agendas or introduce new forms of bias. Therefore, while national LLMs present a powerful tool for cultural representation, they must be developed within ethical and critically informed frameworks to truly address the limitations of global models.

Conclusion

To summarise, generative AI has already become a powerful cultural agent, influencing how collective narratives and identities are formed. However, this influence comes with the risk of *cultural hallucinations* – multimodal distortions of culturally expected patterns that result from biases embedded in globally dominant datasets and models.

Understanding cultural hallucinations draws attention to the fundamental differences between human and artificial cognition. While humans build a multimodal understanding of reality through lived sensory experience, AI relies on statistical pattern recognition. Therefore, in order to minimise cultural hallucinations and enhance cultural authenticity, we must develop *multimodal* and *AI literacies* that enable individuals to critically assess and creatively engage with AI-generated content.

The development of national and culturally specific AI models offers a promising approach to addressing cultural misrepresentation, preserving linguistic diversity, and supporting more inclusive digital representation. By incorporating locally grounded knowledge into training data, such models can help produce content that is both accurate and culturally meaningful.

Ultimately, the study of cultural hallucinations reveals not only the limitations of AI systems, but also important insights about ourselves – our cultural assumptions, our interpretive frameworks, and the semiotic resources through which we construct and communicate our identity.

REFERENCES

1. Bezemer, J., & Kress, G. (2014). *Touch: A resource for making meaning*. *Australian Journal of Language and Literacy*, 37(2), 77–85. <https://doi.org/10.1007/BF03651935>
2. Halliday, M. A. K., & Hasan, R. (1985). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Deakin University Press (Republished 1989 by Oxford University Press). <https://search.worldcat.org/title/813777493>
3. Hendriksen, B., & Kress, G. (2012, March 15). *Why adopt a multimodal approach?* [Video]. YouTube. <https://www.youtube.com/watch?v=rZ4rMVCWkQs>
4. Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge. <https://doi.org/10.4324/9780203970034>
5. Mukienė, D., et al. (2005). *The Lithuanian Word* [Multimedia project]. Retrieved May 16, 2025, from <http://www.spaudos.lt/knygnesiai/knygnesys.en.htm>
6. Peachey, N. (2023, November 2 approx.). *Multimodal literacy: Why and how* [Video]. YouTube. <https://www.youtube.com/watch?v=T592xTVpk6k>