DISEGNO

IX/01
journal of design culture
_Homogenised Heritage:
_AI and Central Europe

HOMOGENISED HERITAGE: AI AND CENTRAL EUROPE

THE IMPACT OF AI ON LOW-RESOURCE LANGUAGES AND VISUAL CULTURES IN THE VISEGRAD COUNTRIES

Disegno

Journal of Design Culture
Double-blind peer-reviewed, open access scholarly journal

Aims and Scope

Disegno publishes original research papers, essays, and reviews on all aspects of design cultures. We understand the notion of design culture as resolutely broad: our aim is to freely discuss the designed environment as mutually intertwined strands of sociocultural products, practices, and discourses. This attitude traverses the disciplinary boundaries between art, design, and visual culture and is therefore open to all themes related to sociocultural creativity and innovation. Our post-disciplinary endeavour welcomes intellectual contributions from all members of different design cultures. Besides providing a lively platform for debating issues of design culture, our specific aim is to consolidate and enhance the emerging field of design culture studies in the Central European academia by providing criticism of fundamental biases and misleading cultural imprinting with respect to the field of design.

All research papers published in Disegno undergo a rigorous double-blind peer review process.
This journal does not charge APCs or submission charges.

Contact: *Moholy-Nagy University of Art and Design*
H-1121 Budapest, Zugligeti út 9–25.
disegno@mome.hu

The full content of Disegno can be accessed online: ***disegno.mome.hu***

Published by: *Csaba Kovács*
Publisher: Moholy-Nagy University of Art and Design, 1121 Budapest, Zugligeti út 9-25.

ISSN: *2064-7778 (print)* ***ISSN:*** *2416-156X (online)*

MOHOLY-NAGY
művészeti egyetem
university of art and
design budapest

Contents

INTRODUCTION
VISIBILITY UNDER AI MEDIATION

https://doi.org/10.21096/disegno_2025_1bib

Artificial intelligence (AI) has come to occupy a pivotal place within contemporary design culture, reframing how images and texts are produced, circulated, classified, and made legible to publics and professionals alike. AI-driven systems increasingly structure encounters with culture, mediate memory, shape the retrieval of references, and help form authority across ordinary regimes of representation. (Vera 2025). This issue foregrounds the infrastructures through which these processes reshape appearance, legibility, and social efficacy in culture, treating AI not merely as a tool but as an ambient mediator of cultural mediation and value production.

The regional vantage of Central Europe and its Visegrád context sharpens this inquiry. The uneven presence of languages, visual archives, and everyday environments within dominant computational systems foregrounds very specific modalities of visibility and invisibility, including the post–Cold War archival history that continues to condition current digital infrastructures. The Iron Curtain legacy remains a salient frame for understanding partial visibility: archives in post-communist contexts have been less frequently digitised, more constrained in accessibility, and less represented in online data ecosystems, with consequential effects on AI training, data extraction, and model behaviour (Zaagsma 2023).Cultural narratives from the region have historically been filtered through external interpretive frameworks, reduced to geopolitical abstractions or to a limited repertoire of circulating motifs, complicating the reception of region-specific material in global AI pipelines. From this perspective, Central Europe illuminates tensions between technical legibility and historical complexity, and between cultural participation and representational distortion within global systems of visibility (Prescott and Hughes 2018).

The essays gathered here emerged from collaborative conversations across the region among authors committed to understanding how AI is altering the cultural conditions of visibility in Central Europe, and who recognised early on that this emergence urgently required concepts, cases, and attuned methods. The issue represents a regional effort to

articulate a problem that has already become materially present while its vocabulary is still forming.

Several contributors develop conceptual tools to address this shift. Ania Malinowska's work on AI assimilationism (the visibility of specific local forms after translation into dominant linguistic, aesthetic, or market-oriented frameworks) provides a lens for reading how regional specificity becomes legible through homogenising frames. Michał Krzykawski's philosophically and infrastructurally oriented notion of computational illusion captures the tendency to treat the computable as an adequate stand-in for social and cultural reality. In our collaboration, we advance the notion of epistemic cultural flattening to identify recurring gaps between technically plausible outputs and recognition grounded in culturally situated knowledge. Kateřina Marková's work on collective vulnerability—how commercial AI infrastructures expose cultures to pressures of statistical averaging—helps illuminate the pressures experienced by diverse Central European cultures under global optimisation regimes. Taken together, these concepts foreground a common claim: cultural loss in AI contexts is often subtle, enabled by systems that appear functional, convincing, and efficient.

A feature of this issue is its attention to local, material cases where AI's cultural effects become tangible. Anna Keszeg's "paprika-effect" shows how AI-generated imagery can render Central and Eastern Europe a non-differentiated region, organised around shorthand tropes rather than historically grounded knowledge, thereby revealing a fragility in how region-specific meanings survive in synthetic imagery. Jiří Philippe Janda's study of North Bohemia, framed as a low-resource visual environment, demonstrates how translation often displaces infrastructural features of everyday post-socialist life, while iconic landmarks retain visibility. This displacement reveals what is left behind when computational mediations privilege certain signals over lived experience. A Hungarian visual-heritage benchmark demonstrates how region-specific ornamental vocabularies and questions of provenance can be flattened in AI outputs, producing culturally misaligned representations despite technological fluency. The practice-based case of Rafani, a Czech collective who collaborate with generative systems, shows artistic work with AI as a liminal process—negotiations among intention and outcome, historical reference and algorithmic pattern, and human/non-human authorship—where meaning is distributed across agents and platforms. Albín Kuchta's inquiry into digital archives and virtual cultural spaces foregrounds questions of resistance versus reproduction, foregrounding Roma and queer cultures and showing how metadata and archival regimes shape what can be remembered or publicly rendered.

These regional and embodied investigations converge on a shared finding: cultural erosion under AI tends to be gradual and procedural,

manifesting as visual standardisation, provenance displacement, and the weakening of region-specific meaning rather than as crude misrepresentation.

Beyond critique, several contributions point toward interventions and governance. The infrastructures enabling AI operate as mediators of circulation, organisers of attention, and producers of plausible cultural forms; datasets, search hierarchies, moderation practices, metadata standards, black-box prompts, and generative interfaces constitute a background architecture shaping cultural participation, often without explicit design acknowledgment. In response, authors propose interventions including multilingual metadata, collaborations with GLAM (galleries, libraries, archives, and museums) institutions, and experiments in community-based archives and participatory data practices; they also map workflows aligned with Central European cultural contexts (the V4 region) to foster more contextually attuned design practices (Prados-Peña et al. 2023).

Several essays also explore artistic and curatorial strategies that repurpose existing tools, and discuss evaluation criteria for generative systems that emphasise situated memory, provenance, and ethics. The papers propose practical pathways for integrating AI into memory-work, archival practices, and regional cultural policy—ranging from open metadata to inclusive archiving workflows and from critical reviews of current benchmarks to alternative evaluation frameworks that foreground provenance and locality (Camilleri 2023).

If designed environments are understood as systems organising perception and value, the infrastructures through which AI operates belong squarely in the domain of design culture. AI acts as a mediator of circulation, an organiser of attention, and a producer of culturally plausible forms that shape expectations and habits as much as outputs. The everyday behind-the-scenes architecture—datasets, search hierarchies, moderation norms, metadata schemas, prompts, and generative interfaces—conditions participation and the reception of cultural materials in ways often not experienced as design in a conventional sense. These observations align with broader literatures on AI in culture, memory, and digital infrastructures, and they anchor regional inquiries within a global matrix of AI-mediated culture.

A critical corollary is that interventions must be concrete and contextual. Multilingual metadata and GLAM partnerships are not merely technical fixes, they are culturally meaningful governance moves that re-centre region-specific knowledge in global data ecosystems. Community-based archives, participatory data practices, and re-evaluated success metrics for generative systems offer tangible routes to counteract homogenisation and preserve local memory ecologies, particularly in post-socialist contexts where historical narratives have been unevenly digitised or archived (Menotti 2025).

The variety of approaches assembled here reflects the difficulty of grasping AI as a cultural condition from any single vantage point. Taken together, the studies suggest that what is at stake in the V4 regional context exceeds technical performance and questions of fairness taken in isolation. More fragile issues are involved: how cultures appearing in AI images can avoid homogenisation, and how cultural heritage is maintained within systems that increasingly organise collective knowledge. From this perspective, Central Europe is an important context, in which wider transformations in cultural mediation are observable with particular clarity.

Brigitta Iványi-Bitter

REFERENCES

Vera, Fulgencio. (2025). Critical Algorithmic Mediation: Rethinking Cultural Transmission and Education in the Age of Artificial Intelligence. Societies. 15. 198. 10.3390/soc15070198.

Prescott, Andrew, and Lorna Hughes. 2018. "Why Do We Digitize? The Case for Slow Digitization." *Archive Journal*, September 2018

Zaagsma, Gerben. 2023. "Digital History and the Politics of Digitization." *Digital Scholarship in the Humanities* 38 (2): 830–51. https://doi.org/10.1093/llc/fqac050

Menotti, Gabriel. 2025. "The Model Is the Museum: Generative AI and the Expropriation of Cultural Heritage." *AI & Society* 40: 5593–97.

Camilleri, Mark Anthony. 2023. "Artificial Intelligence Governance: Ethical Considerations and Implications for Social Responsibility." Expert Systems 41 (7) https://doi.org/10.1111/exsy.13406

Prados-Peña, Maria Belén, George Pavlidis, and Ana García-López. 2023. "New Technologies for the Conservation and Preservation of Cultural Heritage Through a Bibliometric Analysis." *Journal of Cultural Heritage Management and Sustainable Development* 15 (3): 664–86.

AI ASSIMILATIONISM: THE CULTURAL FLATTENING OF LOCALITIES IN GENERATIVE MODELS

Ania Malinowska

ABSTRACT

This paper introduces the concept of AI assimilationism to describe a growing tendency in which local and non-Western cultural aesthetics are absorbed into dominant global AI systems that are largely shaped by Western, particularly American, values. Through this process, distinct cultural expressions are rendered visible only after being filtered, standardised, and reformatted to align with prevailing stylistic norms, linguistic hierarchies, and commercial logics. Drawing on cultural theory and histories of mainstreaming minority cultures, the paper argues that AI assimilationism reinforces existing geopolitical and epistemic asymmetries by privileging English-dominated, Western narrative models and marginalizing non-standard languages, aesthetics, and knowledge practices. Focusing on Eastern European cultural production as a case study, it demonstrates how visibility within AI systems often entails the loss of critical specificity, echoing previous examples of cultural mainstreaming, such as the commodification of Black Lives Matter, where political edge and transformative potential were diluted. The paper identifies the emergent risks of digital "ghettoisation," wherein minority cultures circulate globally but only in narrow, marketable forms. In response, it explores alternative strategies including grassroots artistic interventions, community-based dataset creation, multilingual model development, and demands for epistemic sovereignty. The paper ultimately calls for culturally grounded AI: systems designed not to assimilate but to amplify diverse cultural perspectives, challenging the reproduction of entrenched hierarchies in contemporary technoculture.

#AI assimilationism; #epistemic sovereignty; #non-Western aesthetics; #digital ghettoisation

https://doi.org/10.21096/disegno_2025_1am

INTRODUCTION

For some time now, there has been a subtle yet pervasive trend of local and non-Western cultural aesthetics being absorbed into dominant global AI systems, which are often shaped by Western, particularly American, values. This process is referred to as AI assimilationism. It gives rise to a form of filtered representation, whereby distinct local content is adapted to align with the stylistic norms, policies, and commercial logic that characterise dominant modernities. This concept draws from broader cultural theories that have examined how fringe or minority cultures are often brought into the mainstream, but only after adjusting to its rules. Within the paradigm of AI, cultural content originating from Eastern Europe tends to garner visibility only when it aligns with conventional, English-dominated, Western storytelling models. While this may appear to be an advancement in terms of the representation of diverse cultural identities within global AI systems, a more thorough examination is necessary to ascertain the extent to which it truly represents progress. However, in practice, this approach frequently results in the marginalisation of non-standard languages, aesthetics, and modes of knowledge. The outcome of this process is a digital reinforcement of long-standing cultural hierarchies. This process also risks creating what could be termed a "ghetto effect," in which marginalised cultures become confined to narrow, easily marketable versions of themselves. This phenomenon has been previously observed, for instance in the case of the Black Lives Matter movement, which, in certain instances, was incorporated into the mainstream to the extent that its original radical message and agenda were diluted (no actual systematic change followed it).

A similar set of dynamics is observed whenever minority or non-normative cultures are integrated into dominant frameworks. While this may result in increased visibility, it can also lead to the erosion of their unique characteristics and distinctiveness. It is imperative to acknowledge this pattern to comprehend the risks confronting Eastern European cultural production within AI systems. In response to these trends, alternative strategies are being proposed, including grassroots artistic initiatives, community-based dataset creation, efforts to support multilingual AI training, and broader calls for epistemic sovereignty. The ultimate objective is not merely to achieve inclusion, but rather to cultivate culturally grounded AI: systems that do not merely assimilate local cultures, but rather amplify and respect their unique perspectives.

The present paper argues that AI assimilationism functions as a form of digital colonialism, whereby the absorption of local cultures into global AI systems is not neutral but actively reshapes cultural expression to fit dominant norms. The fundamental contention is that this process, propelled by linguistic standardisation, economies of visibility, and misidentification, engenders a simulated inclusion that serves to reinforce prevailing hierarchies. The analysis demonstrates how Eastern European cultural artefacts are distorted in generative AI outputs. It is demonstrated that visibility in these systems often comes at the cost of authenticity and critical agency. The paper ultimately puts forward a call for a paradigm shift: from assimilation to co-creation.

To unpack these mechanisms, this paper employs a critical, mixed-methods approach that combines discourse analysis of AI-generated cultural artefacts (such as Eastern European folk art and regional dress) with a comparative study of their ethnographic and historical contexts. The juxtaposition of algorithmic outputs with the lived meanings of these artefacts reveals how AI systems systematically distort cultural specificity through linguistic standardisation, economies of visibility, and misidentification. The overarching organising methodologies employed in this study are postcolonial digital humanities and critical algorithm studies. The objective is twofold: firstly, to critique the corruption of cultural difference, and secondly, to identify pathways for designing AI systems that preserve local knowledge and aesthetic traditions.

THE CONCEPT OF "ARTIFICIAL" BELONGING

Historically, marginalised groups and cultures have sought to gain access to dominant representational systems through literature, media, and more recently, digital infrastructures. This endeavour has been driven by a desire to challenge the prevailing patterns of modernity and the matrix of recognition that have come to define our contemporary society. This drive is of political and symbolic importance, since recognition is not merely a courtesy, but "a vital human need" (Taylor 1994, 26). However, systemic recognition is profoundly influenced by an inaccurate "homogenous mould" (43). In this paradigm, marginalised or "peripheral" cultures are recognised only when they are legible within the dominant semiotic order.

In the context of algorithmic representations, belonging is facilitated through the utilisation of westernised visual codes and metadata schemas. The digital rendering of folk art and craft traditions, including the traditional Polish Łowicz paper cut-out and Podhale embroidery, is a subject worthy of consideration. These forms are characterised by a rich symbolism, embedded in fabrics, colours, and patterns, and produced through techniques that have been handed down over centuries. They are imbued with specific rituals and values. It is evident that they serve as repositories of cultural memory and aesthetic value. When platforms such as MidJourney or DALL·E are prompted to generate images based

FIGURE 1. *The original Łowicz pattern. Publicly available online image; author not identified.*

FIGURE 2. *The Łowicz pattern recreated by ChatGPT v. 5.1*

on such input, the outputs (although they may appear accurate or even beautiful) reveal a clear alignment with Western aesthetics. Rather than employing bold primaries, artists instead introduce pastel tones, glossy finishes, and overly symmetrical compositions. Furthermore, the utilisation of vague designations such as "Slavic decorative" or "Eastern European folk" serves to further obfuscate and distort specific regional identity markers.

Prior to the advent of ChatGPT and related technologies, Bender et al. (2021, 615) observed that "LMs trained on extensive, uncurated, static web datasets encode dominant views that are deleterious to marginalised demographics". This phenomenon can be attributed to a systemic misrecognition, whereby specificity is lost and substituted with more generalised, market-friendly symbols. This is precisely what Spivak (1988) once identified as the muting of the "subaltern speaking". When local voices become part of dominant systems, they are translated in ways that reinforce the logic of the centre. This process can be described as a subtle but steady appropriation of situated cultural expressions into standardised, globally palatable formats. In this process, the concept of belonging evolves into a contrived state, manifesting as a facsimile of inclusion that obscures the mechanisms of flattening and erasure.

Assimilationism is predicated on a structurally produced cultural risk. It is important to note that AI systems are not neutral tools for processing data; rather, they are technologies embedded with assumptions about what counts as knowledge, beauty, value, and normativity. These assumptions are shaped by data infrastructures built in the Global North, which are often dominated by English-language sources, Euro-American categories, and capitalist imperatives (Benjamin 2019). AI assimilationism is a concept that has been demonstrated to have an impact on the way in which certain assumptions are made, and to instigate what has been termed "technological redlining" or discriminatory digital decisions (Noble 2018, 1). These decisions have been shown to tend to promote the development of biased ethnic profiles on the World Wide Web.

The aforementioned phenomenon has been especially noted in the context of linguistic standardisation, economies of visibility and misidentification. The first of these is predicated on the assumption that English-language prompts, interfaces, and training datasets are the default. This default setting serves to attenuate the intricate tapestry of global linguistic diversity, thereby diminishing the expressive potential of AI to a level that is constrained by the prevailing language norms. It is evident that minority expressions, languages and dialects are frequently absent or underrepresented in these datasets. This phenomenon results in outputs that are unable to engage with or represent their unique semantic structures or cultural references in a meaningful manner. In the event of such languages being included, there is a possibility that they may be incorrectly rendered (i.e. misspelled, mispronounced, or mistranslated), which in turn could result in distortions of cultural meaning. A pertinent example in this context pertains to diacritics.

With regard to economies of visibility, AI systems tend to promote content that aligns with dominant cultural norms, while simultaneously disregarding content that is unfamiliar or unconventional. It is evident that a significant proportion of artificial intelligence models

are programmed to place a premium on content that has already attained a high level of popularity or has been frequently observed. The effectiveness of this programming is gauged by the extent to which these models align with established metrics of engagement or usage. This frequently results in the exclusion of indigenous patterns, oral storytelling traditions, and local folk art from data training. This predicament is exacerbated when such content is excessively specifically tagged or does not employ widely recognised labels, thereby hindering the capacity of algorithms to recognise or retrieve it. Consequently, these cultural forms become less visible over time, creating a feedback loop where only those that already have visibility continue to be amplified, while other ways of knowing are systematically excluded. In such conditions, templates for content representation increasingly derive from what is already visible, so that historically and regionally specific categories are replaced by generalised imageries. It is evident that the visual phenomena under scrutiny are not solely the consequence of feedback processes; rather, they are the manifestation of "imaginative geographies" (Said 1978), that is to say, the result of the assembly of geographically marked features that, despite the dissolution of their local signification, maintain visual discernibility.

Misidentification is most evident in the methods employed by AI systems to tag and categorise non-Western cultural symbols. A prompt such as "Female Silesian festive dress from the Zabrze region in the 1920s" will typically produce a caption such as "Southern Polish or German female dress from the turn of the twentieth century." This slippage exemplifies the prevalent practice of attributing local, context-rich symbols with vague or externally imposed labels such as "primitive", "ethnic", or "exotic". The employment of such tags has been demonstrated to misrepresent meaning, depoliticise cultural forms, and further commodify them through familiarisation. Anchored in taxonomies shaped by Western art history, anthropology, or the tourism industry, AI systems reduce complex cultural artefacts to aesthetic novelties, detached from their historical and political significance. This phenomenon can be understood as a re-enactment of long-standing patterns of appropriation, which now find expression in digital systems.

FIGURE 3. *The culturally specific prompt in ChatGPT v.5.2 "Create an image of a female Silesian festive dress from the Zabrze region in the 1920s" with a fragment of the prompt's outcome captioned [in Polish]: "Sielanka w tradycyjnym stroju śląskim" (eng. A pastoral scene in traditional Silesian attire).*

FIGURE 4. *ChatGPT v. 5.2's response to the prompt 'Create an image of a female Silesian festive dress from the Zabrze region in the 1920s' and the caption accompanying the generated image. The caption says, "Sielanka w tradycyjnym stroju śląskim" (eng. A pastoral scene in traditional Silesian attire.*

THE GHETTO CULTURE OF CULTURAL AI

The political economy of data capitalism, upon which cultural AI is predicated, is characterised by the extensive extraction of data comprising images, texts, sounds, symbols, and other content. This content is typically harvested without the requisite consent or contextual integrity, resulting in disparities in social contribution and social gratification. Media theorist McKenzie Wark (2019) observes that the value extracted from cultural labour is frequently divorced from the people and social/cultural environments that produced it. Within the domain of generative AI, this devaluation manifests particularly in the transformation of cultural motifs from Eastern Europe, Africa, and Indigenous communities into aesthetic commodities that are disseminated independently of their geographical origins. Concerns regarding ethical issues, including those pertaining to ownership and attribution, have been identified as a grave concern. Another such issue is the questionable imbalance of normativity (global/local, representative/marginal, representative/exotic) that is exacerbated under the pretext of democratisation, thereby engendering the phenomenon of the "cultural ghetto".

The cultural ghetto is not a spatial form, but rather a symbolic and infrastructural condition that renders cultural material visible, albeit within tightly constrained limits. In this particular context, the term "ghetto culture" is employed to denote a regime of representation that ostensibly encompasses the "other"; however, this incorporation is undertaken in a manner that serves to perpetuate marginality. The cultural ghetto functions in a manner that is both insular and opposed to the prevailing mainstream. It is observable yet peripheral, and while it is tolerated, it is not integrated.

My understanding of assimilationism in the context of ghetto culture is derived from the field of queer studies, in which scholars have critically unpacked the elements of conditional social "inclusion" (Meyer 1994). While numerous scholars, including Martin P. Levine, have emphasised geographical segregation, often in reference to designated locations, they have also underscored the conditional nature of inclusion, characterised by symbolic isolation and constrained access to meaningful participation in the broader social sphere. As Levine (1979, 364) defines it in the context of the cultural ghetto "social isolation denotes the segregation of a ghettoised group from meaningful social relations with the larger community, an isolation produced by prejudices against the ghettoised people or by the social distance differing cultural practices create between the group and the larger community". This logic has existed across cultural history and contexts for a considerable period. However, a novel development is the utilisation of artificial intelligence systems to implement exclusion through categorisation and tagging, as opposed to the more traditional approach of censorship. The emergence of AI ghettoisation (as a consequence of AI assimilationism) signifies the

establishment of parallel pathways of cultural recognition, with Western aesthetics being regarded as universal.

A notable contemporary illustration of this phenomenon is the trajectory of the Black Lives Matter (BLM) movement, which has garnered significant attention due to its visibility and impact. The catalyst for the initial rise in global awareness of the Black Lives Matter movement was radical, decentralised digital activism, which utilised hashtags, protest footage and grassroots mobilisations to achieve this. However, the radical edge of this movement was rapidly co-opted by the state. Cultural "corporations" adopted the slogans and themes of the movement without implementing any significant structural reforms within their institutions. As Sarah Banet-Weiser (2018) observes in her extensive analysis of this issue, progressive movements are often commercialised, and media and cultural institutions perform "wokeness" to promote the concept of change without implementing any actual transformation. This phenomenon is further facilitated by hashtag activism (Jackson, Bailey, and Foucault Welles 2020), through which institutions appear supportive, yet the genuine calls for change that underpin the hashtags are often disregarded. The #BLM initiative, which began as a powerful social movement, has rapidly evolved into a corporate-dominated form of allyship. Its instrumentalisation was evident in the realms of publishing, art, and media. The publication of anthologies of black poetry increased significantly. The establishment of black-curated exhibitions was initiated. The transformation of mainstream institutions was not observed. The prevailing sentiment that emerged was one of reinforced perception that Black culture was to be confined to a separate, designated space.

The phenomenon of digital expression in Eastern Europe is characterised by a similar logic to that observed in other regions, particularly in relation to marginalised communities. The potential exists for its inclusion to be contingent upon separation, irrespective of its manifestation in language, visual form or narrative structure. AI assimilationism treats non-Western creative work as exceptional, as if it were not part of an inclusive whole; Western work, on the other hand, is still regarded as universal and the norm. An illustration of this phenomenon can be observed in the reception of Polish avant-garde AI-generated art. When exhibited in international contexts, this art is often labelled as "regional" or "Eastern European", thereby ascribing to it characteristics of specificity and peripherality. By way of contrast, equivalent works from Western creators are often discussed in universal terms.

It is evident that such placements engender epistemic injustice, manifesting as an inequitable devaluation of the users' capacity for knowledge (Fricker, 2007). Epistemic injustice is defined as the wrongful treatment of individuals in their role as knowers, either by not being believed (testimonial) or by being unable to fully understand or express their experiences (hermeneutical), due to structural power imbalances. In AI systems, this injustice takes shape in the one-size-fits-all data categories, which circumscribe the comprehensibility of non-Western

cultural production, rendering it knowable only through the lens of otherness. As Linda Alcoff (1991) asserts, "speaking for others" constitutes a form of symbolic silencing.

The concept of ghetto culture has been a subject of analysis in the domain of cultural criticism, with scholars offering diverse perspectives on its implications. In his 1997 publication, Stuart Hall explored the concept of representational hegemony, which he defined as "dominant meanings, values, and standards that define the limits of the sayable" (Hall 1997, 99). Hall's argument posits that cultural recognition does not inherently disrupt established hegemony; rather, it facilitates the reintegration of difference in a manner that serves to reinforce the prevailing social order. In a similar vein, Homi Bhabha (1994) expounded on the concept of cultural hybridity to elucidate how the marginalised subject is "included" solely through mimetic imitation of the prevailing culture. This imitation, however, is perpetually characterised by ambivalence and imbalance, never fully achieving complete assimilation. As Gilroy (1993) demonstrated in his seminal work *The Black Atlantic*, the notion of the "Black Atlantic" has been instrumental in highlighting the decontextualisation and disassociation of Black cultural expression from its political roots, even in cases where it has achieved global recognition, as evidenced by the examples of jazz, hip-hop and soul.

A more perspicuous way to conceptualise the "ghetto effect" in AI would be through Nancy Fraser's (1990) concept of misrecognition, which posits that marginalised groups are incorporated solely on terms established by the dominant culture. Walter Mignolo's (2005) concept of "epistemic coloniality" builds on this to indicate that such inclusion frequently occurs with the consequence of erasing or eliminating knowledge systems that do not align with the Western canon. This phenomenon elucidates the potential for the co-option of social movements, such as Black Lives Matter, and the subversion of Indigenous aesthetics in the fashion industry (Raheja 2010). Both cases illustrate that the pursuit of inclusion on unequal terms can, in fact, serve to exacerbate existing forms of exclusion. Within the paradigm of AI, the phenomenon of cultural difference diminishes in intensity concomitant with the scalability of digital platforms, which serve to standardise and decontextualise this difference, resulting in forms that seamlessly integrate into prevailing systems. In *The Googlization of Everything*, Siva Vaidhyanathan (2012) demonstrates that platform capitalism not only encourages but actively demands the flattening of cultural complexity into easily digestible content.

The prevailing digital manifestation of non-Western content functions within McRobbie's (2009) concept of a "triple entanglement", which extends the concept of "double entanglement." This theoretical framework elucidates how ostensible cultural gains achieved by marginalised groups are undermined by more insidious forms of backlash or co-optation. This phenomenon introduces a third dimension: the infrastructural constraints of AI systems themselves. The tension is further exacerbated by aesthetic secessionism (cultural separatism masquerading as recog-

nition), its institutionalisation in cultural theory, and bell hooks' (2014) warning that the appropriation of difference undermines political critique by repackaging it as marketable content. AI does not merely reflect culture; rather, it reconstitutes it, frequently amplifying contradictions.

In order to combat epistemicide, it is essential to rely on reimaginations and reconfigurations of cultural infrastructures.

CULTURALLY INFORMED AI

A variety of projects, encompassing grassroots initiatives, artistic endeavours, and research-based studies, have sought to introduce alternative visions and practices. Two United Nations Educational, Social and Cultural Organization (UNESCO) initiatives are particularly notable for their institutional grounding and broad scope: The following publications provide a comprehensive overview of the current state of research in the field: "Protecting and Preserving Cultural Diversity in the Digital Era" (UNESCO 2020) and "Digital Initiatives for Indigenous Languages" (Llanes-Ortiz 2023). The former addresses the rapid expansion of digital technologies and their impact on the reshaping of culture, from its creation and distribution to access and preservation. The article focuses on the considerable promise and significant constraints that this expansion entails. The accessibility of cultural content created by web platforms, artificial intelligence (AI), virtual and augmented reality, and 5G networks (e.g. streams of theatre performances, virtual museum collections, remote engagement with cultural heritage) is highlighted, emphasising the uneven distribution of accessibility. As highlighted in the projects report (UNESCO 2020), only 53.6% of the global population currently has access to digital technologies. This indicates that approximately half of the global population remains marginalised in terms of accessing the opportunities presented by digital culture. The digital divide is most pronounced in the world's least economically developed countries, where digital penetration is as low as 19%, and gender disparities persist, with 12% fewer women than men using the internet globally. These inequalities carry profound implications for cultural production and consumption; for example, only 5% of museums in Africa and Small Island Developing States maintain an online presence, highlighting the risk that entire cultural sectors can be marginalised in the digital landscape.

The project also draws attention to the potential of digital technologies to safeguard cultural heritage, especially in situations involving conflict and disaster. During the course of the pandemic, there was a surge in digital engagement, with the Louvre Museum, for example, witnessing a ten-fold increase in web traffic as audiences migrated online. In contexts of destruction, UNESCO has utilised satellite imagery and three-dimensional documentation to assess damage and support recovery efforts in locations such as Aleppo, Syria. Furthermore, the organisation has conducted training programmes for heritage professionals

in Yemen and Iraq, equipping them with advanced tools including drones and photogrammetry for documentation purposes. Among the complex cultural risks exposed in this process is the threat of linguistic diversity, with just ten of the world's approximately 7,000 languages used to access 77% of the 1.8 billion websites online. A recent study has revealed that 95% of the global app market is concentrated in only 10 countries. This phenomenon gives rise to inequitable creative ecosystems, giving rise to questions around fair remuneration for creators, algorithmic control of cultural content, and the erosion of culture as a public good.

UNESCO also underscores the significance of inclusive public policies and global collaboration. The fundamental motivation underpinning their endeavours in this domain is the human right of equal access, in conjunction with the pressing concern of cultural misappropriation by digital instruments. The UNESCO project "Digital initiatives for Indigenous languages" (Llanes-Ortiz 2023) considers the relative merits and disadvantages of high technologies in comparison to non-digital forms. Its report "recommends posing a series of questions when determining which technological solutions could exert the greatest influence in relation to the community's revitalisation objectives.". According to Llanes-Ortiz (2023, 49), the project must include "clear and concrete steps to protect sensitive content from being [culturally] misappropriated".

This recommendation is predicated on the notion of strategic approaches to counteract digital misappropriation. These include facilitating, multiplying, normalising, educating, reclaiming, imagining, defending and protecting Indigenous languages in digital spaces. In the context of the International Decade of Indigenous Languages (2022–2032), this project underscores the imperative for digital literacy and self-directed content creation among communities for whom languages have been historically marginalised by dominant technologies and platforms. The practical elements of the course include case examples of digital activism, tools for expanding online presence (ranging from social media to educational resources), guidance for community-led projects, and suggestions for collaborative partnerships with technical experts and platforms. The toolkit provides unequivocal guidance on the matter of community agency: accountable digital efforts must be driven by Indigenous speakers, their priorities, and ethical considerations such as data sovereignty and cultural integrity.

One of the approaches to reversing the current approaches to digital content representation is the Geographically Inclusive Vision-and-Language project (GIVL). The project, which is anchored at the University of California, Los Angeles, positions geographic diversity as a structural dimension of multimodal representation learning. In summary, the project entails the development of an AI system capable of comprehending images and text in unison, with a particular focus on the unique visual characteristics of diverse global regions. The model employs two distinct training methodologies to facilitate the execution of vision-and-language tasks, such as describing images, responding to inquiries regarding these

images, and matching images with appropriate verbal descriptions. The first training method focuses on equipping the model with the capacity to establish connections between images and the relevant knowledge, including region-specific information. The second training method aims to enhance the model's ability to discern when visual similarities are deceptive and lead to erroneous inferences, particularly in contexts where cultural or local variations exist. The objective is to enhance the performance of AI in processing images from a diverse range of countries and cultures, extending beyond the scope of Western-focused datasets. The issue arises from the fact that visual categories are not universal; that is to say, significations and representations can vary according to geographical location. This is further compounded by the fact that many existing training datasets are imbalanced, with the result that AI systems are less effective for under-represented regions.

In the context of contemporary data science, there has been a growing recognition of the need to address the colonial and systemic biases inherent in data processing and analysis. This movement, often termed "decolonising data," has been explored academically, particularly in the works of Couldry and Mejias (2019). In response to these calls for redress, Masakhane has emerged as a notable initiative. This grassroots African-led NLP (natural language processing) project aims to develop language models specifically for and within African languages, contributing to the revitalisation and empowerment of linguistic diversity on the continent (Orife et al. 2020). The Masakhane project is a continent-wide, open-source initiative with a focus on advancing neural machine translation (NMT) and broader natural language processing (NLP) for African languages. Its central tenet is the prioritisation of linguistic diversity in the pursuit of cultural representation veracity. The African continent is home to a plethora of languages, with over 2,000 different languages spoken across the region. However, this linguistic diversity is not reflected in research and technological development in natural language processing (NLP). Prior to the advent of Masakhane, linguistic resources, benchmarks, and publications for African languages were scarce, a situation compounded by limited funding, inadequate community infrastructure, the paucity of discoverability of extant work, and the linguistic complexity characteristic of many of these languages.

The project was initiated in 2019 with the objective of establishing an active research community focused on African Natural Language Processing (NLP), the creation of datasets and tools to facilitate research on low-resourced languages, and the establishment of best practices for distributed and inclusive research that can scale beyond the African continent. This initiative emerged from the grassroots movement of African AI researchers Deep Learning Indaba. The Masakhane approach has been developed to lower entry barriers: participants use an open-source platform with Jupyter Notebooks and free Google Colab GPUs to train and evaluate translation models, drawing on publicly available corpora such as the JW300 dataset, which includes parallel texts for

many African languages. As of February 2020, the community comprised 144 members from seventeen African nations and two nations outside the African continent, representing a range of educational attainment and professional backgrounds. Contributors published thirty translation results covering twenty-eight African languages on Masakhane's GitHub, enhancing reproducibility and community knowledge sharing. Notwithstanding these achievements, substantial obstacles persist. These include the paucity of language resources, the limitation of funding and infrastructure, and the inherent linguistic diversity and complexity of African languages, which challenge ongoing improvement and expansion of models.

In defining the tactics for decolonial AI and digital epistemic sovereignty, Mohamed, Png, and Isaac (2020, 684) mention "a strong need to develop new methodologies". For them, decolonial AI denotes an "inclusive dialogue between stakeholders in AI development, particularly those in which marginalised groups have meaningful avenues to influence the decision-making process, avoiding the potential for predatory inclusion and continued algorithmic oppression, exploitation and dispossession". Furthermore, the necessity for methodologies that can question algorithmic authority more broadly is implied, albeit not fully articulated.

Whilst scientific research has frequently encountered difficulties in embracing more open-ended modes of thinking, artistic and scientific practices have, for many years, functioned in accordance with such modes. The artistic practices informed by technological developments have long exposed the limitations of computational systems, including their biases and political ramifications. Indeed, artistic praxis has been instrumental in highlighting the manner in which algorithmic systems shape social perception and produce knowledge. A recent example of this phenomenon is *Justice Control Unit* by Przemysław Jasielski (2024). This media art installation examines the concept of algorithmic authority through a staged simulation of machine-driven justice. The work sets forth a speculative system that purports to evaluate an individual's propensity to perpetrate a criminal act. However, the criteria underpinning these judgments remain opaque, and the system's outcomes—electric shocks administered on the basis of factors such as race, gender, socio-economic status, and ethnicity—appear arbitrary, despite being presented as the result of facial recognition analysis.

The interface of *Justice Control Unit* bears some resemblance to a legal or judicial device. However, the manner in which individuals engage with this medium is influenced by the findings of Stanley Milgram's obedience experiments, wherein participants were instructed by an authority figure to inflict harm upon others. This reference underscores a serious problem: there is an increasing tendency to delegate moral decisions to non-human systems that ostensibly remain neutral, despite the fact that they are founded on human assumptions, values, and biases. The objective of this project is to examine how advanced technologies en-

gender a paradox: moral judgment is displaced to a non-human entity, yet remains contingent on programmed rules that reflect social and political inequalities (Malinowska 2024). It has been demonstrated that these algorithmic systems do not eliminate bias; indeed, they frequently serve to reinforce it. Jasielski's installation has the capacity to render these hidden biases visible, thereby demonstrating the manner in which contemporary AI systems have the potential to perpetuate historical forms of racial, gender, ethnic, and other forms of injustice and discrimination, whilst presenting themselves as objective and purely technical.

Ruha Benjamin's *Race After Technology* (2019) critiques the deployment of algorithmic systems that perpetuate inequality under the pretext of impartiality. Benjamin signals the same problematic undercurrent of technological foresight through which present biases are projected forward and may solidify into future forms of inequality. She writes (2019, np):

> *By deliberately cultivating a solidaristic approach to design, it is necessary to consider that the technology that might be working just fine for some of us (at the present time) could harm or exclude others and that, even when the stakes seem trivial, a visionary ethos requires looking down the road to where things might be headed.*

To combat the phenomenon of AI assimilationism it is necessary to engage with the expanding field of data feminism (D'Ignazio and Klein 2020), which advocates for intersectional approaches to data collection and algorithmic design. The Geographically Inclusive Vision-and-Language (GIVL) project, for instance, reflects this ethos by prioritising geographic diversity in AI training datasets. However, as D'Ignazio and Klein argue, technical solutions alone are insufficient without addressing the broader power structures that shape data infrastructures. In this context, the notion of "design justice" (Costanza-Chock 2020) is especially pertinent, as it emphasises the involvement of marginalised communities in the design process of technological systems. In a similar vein, the data sovereignty movements spearheaded by Indigenous scholars (Kukutai and Taylor 2016) present a paradigm for the reclamation of authority over cultural representation in digital domains.

However, even ambitious projects such as LATAM-GPT, which aims to develop language models trained on Latin American Spanish and Indigenous languages, reveal the complexities of resisting AI assimilationism (Lagos 2025). The initiative challenges the Anglophone bias of dominant AI systems by incorporating regional linguistic and cultural knowledge. However, its reliance on existing data infrastructures risks reproducing the very extractive logics it seeks to dismantle. In the absence of robust mechanisms for community governance and consent, such efforts may perpetuate technological dependency, resulting in the terms of inclusion being controlled by the very systems marginalised communities seek to challenge. In his analysis of *Justice Control Unit*, Przemysław Jasiel-

ski examines the phenomenon of algorithmic neutrality concealing entrenched power structures. The fundamental dilemma addressed in this study is the following: can such systems genuinely decolonise representation, or do they merely offer a more inclusive version of the same assimilative logic?

Justice Control Unit can be regarded as an artistic provocation. However, it can also be interpreted as a speculative exploration that reveals the fragility of the systems of belief that render such technologies legitimate. This gesture is connected to broader discourses concerning the social and political effects of surveillance technologies. As Jon Fasman (2021) observes, the expanding implementation of automated surveillance, encompassing facial recognition, license plate tracking, and predictive tools, gives rise to significant and as yet unresolved ethical concerns pertaining to privacy, personal autonomy, and civil liberties. These systems are frequently justified through claims of efficiency and technical performance, yet the rules they operate by are difficult to see, question, or challenge.

CONCLUSION

Despite the presence of commendable intentions and a multitude of meticulous initiatives, it appears that the phenomenon of AI assimilationism, with its subliminal tendencies towards ghettoisation, is a pattern that proves resilient and difficult to dismantle. After Benjamin, Arundhati Roy (2014, 25) pessimistically asserts "There is no alternative". This sentiment is further compounded by the acknowledgement that endeavours to enhance visibility frequently entail an implicit expectation: that one should conform to prevailing modes of knowledge, nomenclature, and perception. The concept of inclusion has the capacity to subtly transform differences into something more familiar and manageable. Cultural specificity manifests as an aesthetic surface, while the capacity to define meaning is situated elsewhere.

This dynamic has been observed in a variety of settings: The integration of Eastern European digital art into overarching regional categories, the transformation of political struggles into consumable imagery, and the persistent misreading of non-Western bodies by automated systems are key themes that emerge from this analysis. In such cases, it can be argued that visibility is not neutral. The potential for distortion, simplification, and containment is evident.

As has been argued, AI assimilationism is sustained by a triple erasure: the silencing of linguistic nuances, the packaging of cultural difference as marketable exotica, and the entrenchment of inequalities under the guise of neutrality. This phenomenon does not represent the unavoidable cost of progress; rather, it is the consequence of regarding dominant technological paradigms as natural rather than constructed. The true challenge lies in transcending tokenistic inclusivity, where diversity is permitted only when it aligns with established norms. However, it is

important to note that the landscape is not static. A plethora of recent studies have indicated the emergence of fissures in the veneer of technological inevitability, as evidenced by a diverse array of phenomena. These include artistic interventions, grassroots datasets and decolonial design experiments. These alternatives do not merely resist assimilation; they reimagine the potential of AI if constructed from the ground up, with consideration for the communities it purports to represent.

The challenge lies in the implementation of these practices within broader infrastructures, educational systems, policy frameworks, and the quotidian mechanisms that discreetly orchestrate digital life. While the responses remain in a state of formulation, this objective appears to prioritise the dissolution of the centralised structure over the realignment of peripheral elements. As Ruha Benjamin reminds us, critique alone is not sufficient; it must be paired with creative reimagining. The argument advanced here is that the optimistic ending of her own work is to be rejected. The conclusion of the work states that "An emancipatory approach to technology entails an appreciation for the aesthetic dimensions of resisting the New Jim Code and a commitment to coupling our critique with creative alternatives that bring to life liberating and joyful ways of living in and organising our world" (2019, np).

REFERENCES

Alcoff, Linda. 1991. "The Problem of Speaking for Others." *Cultural Critique* 20: 5–32.

Banet-Weiser, Sarah. 2018. *Empowered: Popular Feminism and Popular Misogyny*. Durham: Duke University Press.

Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity Press. Kindle.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, Schmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: The Question of Whether Language Models Can Become Excessively Large is a Salient One." Paper presented at the 2021 ACM Conference on Fairness, Accountability and Transparency.

Bhabha, Homi K. 1994. *The Location of Culture*. London: Routledge.

Costanza-Chock, Sasha. 2020. *Design Justice: Community-Led Practices to Build the Worlds We Need*. Cambridge MA: MIT Press.

Couldry, Nick, and Ulises A. Mejias. 2019. *The Costs of Connection: How Data is Colonizing Human Life and Appropriating It for Capitalism*. Stanford: Stanford University Press.

D'Ignazio, Catherine, Lauren F. Klein. 2020. *Data Feminism*. Cambridge MA: MIT Press.

Fraser, Nancy. 1990. *Justice Interruptus: Critical Reflections on the "Postsocialist" Condition*. London: Routledge.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Gilroy, Paul. 1993. *The Black Atlantic: Modernity and Double Consciousness*. Cambridge MA: Harvard University Press.

Hall, Stuart, ed. 1997. *Representation: Cultural Representations and Signifying Practices*. London: Sage Hooks.

Jackson, Sarah J., Moya Bailey, Brooke Foucault Welles. 2020. *#HashtagActivism: Networks of Race and Gender Justice*. Cambridge MA: MIT Press.

Jasielski, Przemysław. 2024. *Justice Control Unit*. Le Guern, Warsaw.

Kukutai, Tahu, and John Taylor, eds. 2016. *Indigenous Data Sovereignty: Towards an Agenda*. Canberra: ANU Press.

Lagos, Anna. 2025. "Latam-GPT: Meet the Open Source AI of Latin America." *Wired*, September 1. https://www.wired.com/story/latam-gpt-the-free-open-source-and-collaborative-ai-of-latin-america/

Levine, M. P. 1979. "Gay Ghetto." *Journal of Homosexuality* 4 (4): 363–77.

Llanes-Ortiz, Genner. 2023. "Digital Initiatives for Indigenous Languages." UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000387186

Malinowska, Ania. 2024. "Justice Control Unit: Machine for Monitoring the Administration of Justice." *Le Guern*. https://leguern.pl/en/exhibitions-list/4343-wgw-2024-przemyslaw-jasielski-justice-control-unit

McRobbie, Angela. 2009. *The Aftermath of Feminism: Gender, Culture and Social Change*. London: Sage.

Meyer, Morris. 1994. *The Politics and Poetics of Camp*. London: Routledge.

Mignolo, Walter D. 2005. *The Idea of Latin America*. Oxford: Blackwell.

Mohamed, Shakir, Marie-Therese Png, and William Isaac. 2020. "Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence." *Philosophy & Technology* 33: 659–84.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racial Bias*. New York: NYU Press.

Orife, Iroro Fred Ònòmè, Anuoluwapo, A., Davids, S., et al. 2020. "Masakhane: Machine Translation for African Languages." *Proceedings of the 2020 EMNLP Workshop on Africa NLP*, 1–4.

Raheja, Michelle H. 2010. *Reservation Reelism: Redfacing, Visual Sovereignty, and Representations of Native Americans in Film*. Lincoln: University of Nebraska Press.

Roy, Arundhati. 2014. *Capitalism: A Ghost Story*. Chicago: Haymarket Books.

Said, Edward. W. 1978. *Orientalism*. New York: Pantheon Books.

Spivak, Gayatri Chakravorty. 1988. "Can the Subaltern Speak?" In *Marxism and the Interpretation of Culture*, edited by Cary Nelson and Lawrence Grossberg, 271–313. Chicago: The University of Illinois Press.

Taylor, Charles. 1994. "The Politics of Recognition." In *Multiculturalism: Examining the Politics of Recognition*, edited by Amy Gutmann, 25–72. Princeton: Princeton University Press.

UNESCO. 2020. "Protecting and Preserving Cultural Diversity in the Digital Era." *The Tracker* (2): 6–7. https://unesdoc.unesco.org/ark:/48223/pf0000377331

Vaidhyanathan, Siva. 2011. *The Googlization of Everything (And Why We Should Worry)*. Berkley: University of California Press.

Wark, McKenzie. 2019. *Capital Is Dead: Is This Something Worse?* London: Verso.

BEYOND COMPUTATIONAL ILLUSION: FUTURES WORTH WANTING FOR ARTISTIC PRACTICES AND TECHNICAL CULTURES

Michał Krzykawski

ABSTRACT

Artists and cultural theorists, although they use different means, share the task of problematizing culture. The urgent task now is to critically examine the ways we use AI text-to-image generators and create space for reflection about engaging with these systems. I approach this through what I term the cultural logic of computational capitalism, drawing on Fredric Jameson and Bernard Stiegler. The paper addresses how arts and humanities can help overcome this logic and transform AI's visual culture itself. Such an inquiry is essential given that AI text-to-image generators not only disrupt traditional art production but also concentrate creative power in the hands of a few dominant platforms. While these concerns are global, they require specific regional responses. Focusing on East-Central European countries, I argue that the underrepresentation of their visual cultures in AI models stems from their semi-peripheral status within global technological and economic systems. Rather than simply feeding existing AI models with better regional training data, I propose supporting dissident artistic practices that promote regional digital and digitally sustainable cultures.

#computational illusion; #computational capitalism; #art-informatics; #semiotic entropy

https://doi.org/10.21096/disegno_2025_1mk

AI images, whether embraced as artworks or dismissed as slop, are transforming our symbolic forms and disrupting how we produce meanings through them. This is neither "the end of art" nor "the end of thinking," but it does require us to grasp the nature of this change by understanding the limits of AI vision and of computational techniques behind it. The key challenge is both conceptual and purposeful. Conceptual, because it involves questioning how artificial intelligence was conceptualised as a mathematical object at the beginning of the eponymous discipline. Purposeful, because it is oriented toward critically integrating computational techniques into artistic practices and technical cultures as mutually constitutive, rather than focusing on whether AI is creative and can think or not. In a nutshell, visual cultures in an AI world largely depend on the intent behind the use of generative technologies and our ability to define this intent beyond the logic of (capitalist) computation that erodes all justification, without denying what those technologies can actually do.

What I call computational illusion, however, is a serious obstacle that hinders this necessary work. In what follows, I begin by unpacking what I mean by computational illusion and tracing its roots in the mechanistic conceptualisation of intelligence, a conceptualisation that precedes, and in many ways enables, the anthropomorphisation of computational machines. I then examine how the collapse of the distinction between scientific heuristics and advertising clichés has shaped the current AI landscape and the visual cultures it generates. Against this backdrop, I outline what I propose as art-informatics—an approach that repurposes computational techniques through artistic experimentation and philosophical reflection—as a possible exit route toward grounded techno-artistic practices, particularly from an East-Central European perspective.

HOW IS COMPUTATIONAL ILLUSION BROUGHT TO THE FORE?

By computational illusion I mean the false assumption that everything is computable and can be written in computer code, whether genetic code, legal code, or semiotic code. According to this misconception, life in general, biological, social and cognitive alike, is fundamentally about processing information. Computationalists believe that with adequate, uncorrupted data, a robust computing stack, and efficient algorithms, well-trained AI models can generate automated information products, and that these products interact with each other in real time, giving rise

to emergent structures analogous to how organisations or knowledge emerge from biological or cognitive processes. Emergence and interaction are definitely the two buzzwords here.

Computational illusion does not originate from a science or a particular school of thought. It is neither scientific nor philosophical. Both science and philosophy require self-limitation, self-critique and self-knowledge, rather than a mere stacking of computational techniques, performance gains through optimisation, and speculative fantasies about human-AI interactions that have no grounding in social reality whatsoever. Computational illusion floats in the air, a zeitgeist of our era. On the one hand, it generates bombastic speculative scenarios about humanity's AI-driven future. On the other, it reduces concepts, hypotheses, and everyday knowledge to meaningless content. This reduction undermines critical thought, which depends on discerning differences: emergence from relationality, information from communication, communication from knowledge, reasoning from understanding, computational intelligence from other types of intelligence, intelligence from thinking and, last but not least, business from science. All things become undifferentiated. While this entropic undifferentiation spreads in our informational ecosystems—generating what I call semiotic entropy inasmuch as it involves an increase in the meaninglessness of algorithmically structured information—we lack the conceptual tools to understand both technologies and our actions within technologised environments. Developing such tools is essential to counter mainstream AI discourse that thoughtlessly ascribes human or superhuman characteristics to computational machines. This thoughtlessness is the only "existential risk" we face.

Though rooted in AI technologist circles, computational illusion extends far beyond them, manifesting in various claims that short-circuit scientific and philosophical frameworks. A few examples: that life is "the universe's most ancient technology" (Suleyman and Bhaskar 2023); that "the informational basis of all life on Earth" or "life's basic machinery" is "the universal genetic code" (Davies 2019); that "human intuition is in reality pattern recognition" (Harari 2018); that "organisms are algorithms" (Harari 2016); that "intelligence is prediction" (Blakeslee, Hawkins 2004); that "beliefs are a kind of information, thinking a kind of computation, and emotions, motives, and desires are a kind of feedback mechanism" (Pinker 2005).[1]

The widespread understanding and misunderstanding of computational techniques and their applications is shaped within computational illusion, creating the climate around AI, generating a sense of its ineluctability, and exerting pressure toward the mechanisation of processes, the acquisition of so-called digital competencies, and the construction of national or local language models as if no alternative course of progress were conceivable.

At the same time, all of the claims invoked could find solid legitimation in predictive processing theory or other computational accounts

[1] *Some of these examples are also quoted by Bates (2025: 69–70).*

of the mind whose primary source of inspiration is neural networks. In the context of computational illusion, "computation" not only refers to computing machines but, more generally, to a mechanistic model of rationality applied to data processing, whether operated by algorithms or human mind-brains. What underpins this model is the assumption that reasoning is reckoning, and that we anticipate, feel, and decide what to do next through predictive mechanisms. From this perspective, it becomes plausible to regard the digital computer as the model of intelligence itself, and it follows that knowledge of artificial neural networks can only bring us closer to understanding how the biological machine operating in our brain works. This is a vicious circle, or rather an epistemological deadlock in which computational models are superimposed on the very processes of knowledge production and meaning-making.

THE MECHANISATION OF HUMANS PRECEDES THE ANTHROPOMORPHISATION OF MACHINES

The computational explanation of cognitive processes has experienced a resurgence with advances in computational techniques and the exponential growth of data since the early twenty-first century. It remains, however, one of several paradigms in contemporary cognitive sciences, and is largely dismissed as overly reductive by critical neuroscience scholarship, which emphasises the bodily constitution and environmental embeddedness of mental processes—processes that, consequently, cannot be properly understood outside their physical and social contexts (Choudhury and Slaby 2010, 11).

That said, I am not suggesting that the computational explanation of cognitive processes is entirely without merit. It is rather that it does not provide a reasonable framework for understanding how AI systems work in larger social ecosystems and why artistic or theoretical inventions cannot happen without a sensibility for what is not yet there. This involves a pull toward new possibility, and the risk of being wrong beyond a probabilistic regression to the mean that large language models have at their core.

It has become something of a commonplace among cultural theorists to warn against anthropomorphising computational machines. While some AI philosophers argue that computational systems operate an "alien mode of thought", developing what they call "the alien subjects of AI" as a theoretical framework (Parisi 2019), computational linguists and IT scholars rightly caution that anthropomorphising these systems leads us to uncritically accept how they are marketed to us as "human-like systems" that can "hallucinate," possess "reasoning capabilities," or exhibit "intelligence" (Bender and Inie 2026).

But the opposite may equally be true. Our tendency to anthropomorphise machines, which has arguably become a weakness, is itself an effect of how the human mind was mechanised and what assumptions enabled its mechanistic conceptualisation.

That these assumptions remain largely mythological and concern the myth of a machine more intelligent than the human is the central claim of what I call computational illusion. We find this myth at the very origins of the ultrashort history of AI, from the way it was conceptualised as a mathematical object to the inquiries into the possibility of realising that object in machines. Consider John McCarthy et al. (2006), who in 1955 were the first to introduce the term artificial intelligence. For them, the study of artificial intelligence was meant "to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. Also remember Alan Turing (1950), who five years earlier had conceptualised the "thinking machine" or "intelligent machine". All three names—"artificial intelligence," "thinking machine," and "intelligent machine"—refer to the same thing and derive from the assumption that the operation of intelligence is thought and that thought processes can be described in formal language. Nothing, therefore, should stand in the way of imagining that, given sufficient computational power, a machine more intelligent than humans could be built.

Let me intervene in this well-known story with a philosophical comparison. The pursuit of such a half-mythical, half-realistic machine has accompanied the history of AI like the shadow accompanying Nietzsche's "wanderer"—one that cannot shed its past or its "self," but can come to terms with it. I argue that our knowledge of AI remains in a pre-critical phase, precisely because such a reckoning has not yet taken place and computational illusion prevails. By a pre-critical phase, I mean a specific mindset embracing and testing AI without a critical reflection on its conceptual foundations and how these presuppositions shape our understanding of agency, intelligence, and the relationship between technology and social cognition, alongside the proliferation of AI models over the world. According to Hugging Face (2026), there are already two million of them.

That Turing and the organisers of the Dartmouth seminar were pragmatic visionaries in exploring the possibilities of formalizing thought processes is beyond question. "I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted," Turing (1950, 443) predicted, bringing to a close his description of the peculiar thought experiment later known as the "Turing test." That test, as it happens, played a role inversely proportional to the interest it generated among philosophers, particularly in the Anglo-American tradition, and the general public alike. As Margaret Boden (2016) aptly notes, it was "tongue in cheek: Although it featured in the opening pages, the Turing test was an adjunct within a paper primarily intended as a manifesto for a future AI. Indeed, Turing described it to his friend Robin Gandy as light-hearted 'propaganda', inviting giggles rather than serious critique".

WHEN THE DIFFERENCE BETWEEN HEURISTIC MODELS AND ADVERTISING CLICHÉS IS GONE

Far more worthy of attention than this "propaganda," however, is what lent the joke known as the Turing test its scientific legitimacy and made it so generative: a decisive turn in the twentieth-century history of the scientific study of cognitive activities, one in which observing those activities became de facto synonymous with measuring them mathematically, as though one were observing any physical body. The measurement of intelligence is a case in point. Alfred Binet, the originator of the test that gave rise to what we now know as the IQ test, when asked to define the intelligence he sought to measure in numbers, reportedly gave the provocative answer that "intelligence is what my test measures" (quoted in Lhérété 2024, 9). Subscribing to this line of reasoning and not unlike McCarthy, Turing specifically conceived the act of thinking by assuming that thought is amenable to conceptualisation as an ideal object—a mathematical object—and that it can be functionally abstracted from all spatiotemporal situatedness: the historical, social, and embodied contexts in which meanings arise. It follows that whether it is a human or a machine that thinks is of no consequence—for both, in the end, compute. It can explain why so many great computer scientists, including the Nobel Prize and Fields Medal winners, succumb so readily to AGI mythology, whether utopian or dystopian, and uncritically accept AGI as an inevitable technological trajectory. The AGI mythology is just a variant of the myth of an intelligent machine outperforming human intelligence, an echo of a heuristic computational model of human intelligence at the origins of its mechanic emulation.

The journalist Will Douglas Heaven (2025) has recently pointed out that the AGI myth has become an obsession of the tech industry and has become embedded in public debate. According to Heaven, the myth of AGI fulfils many of the criteria that qualify a theory as conspiratorial: a flexible framework that sustains belief even when reality diverges from prediction, a promise achievable only if believers uncover hidden truths, and the hope of salvation from the horrors of this world. This is why, Heaven says, it has become anchored in deeply rooted beliefs that are difficult to uproot.

A more in-depth explanation of the phenomenon, however, is possible. What we are witnessing today is the intensification of a phenomenon against which Georges Canguilhem (2008, 12) had already cautioned in 1980. This leading representative of the tradition of historical epistemology noted that expressions such as "'conscious brain', 'conscious machine', 'artificial brain' or 'artificial intelligence'", especially popular "in the Anglo-American domain", can be legitimately justified as scientific names for "heuristic models or sophisticated simulators." So far, so good. But once such expressions migrated into public debate during the industrial phase of computer technology, they became, in Canguilhem's words, "advertising clichés" that not only disorient the

general public and normalise a low level of technological awareness, but also rebound on the quality of scientific research itself. As Canguilhem diagnosed with characteristic clarity: "A model of scientific research was thereby converted into a machine of ideological propaganda with a twofold purpose: to anticipate or disarm all opposition to the invasion of a means of automating the regulation of social relations; and to conceal the presence of decision-makers behind the anonymity of the machine." (12)

What has changed since then is the scale of the problem and the challenges it brings with it. The force of computational illusion has eroded the distinction between scientific heuristics and advertising clichés—a distinction Canguilhem might still have taken for granted in the era of Minitel. Their interested conflation can indeed be seen as the ideological matrix of the AI industry. What this AI ideology makes us believe is that the difference between automata and autonomy (between machine operations and human actions) is no longer functionally significant. Not because we cease to see and feel the difference, but because it becomes functionally negligible in a world that Bernard Stiegler (2016, 48) defines as "totally computational capitalism". With a nod to Nietzsche, Stiegler identifies it with "the accomplished nihilism" (48) and calls for "the transvaluation of becoming into future" (10). What underlies this condition, I would argue, is an epistemological cul-de-sac, that is, the fundamental error in assuming the adequacy of understanding reality without the critical ability to discern its social, biological, and technical dimensions.

REPURPOSING COMPUTATIONAL TECHNIQUES...

Let me be direct. Without a critical understanding of what computer code can do today and of our use of language models, we have no purpose or means to shape visual cultures that save us from semiotic entropy, that is the structural tendency toward the decline of socially produced meaning. In fact, large language models do not merely simulate intelligence, they decompose the sign system into tokens stripped of meaning and recombine them statistically. In doing so, they bypass the layer of meaning that underpins rationality and holds together the binding tissue of intelligent societies. This is not simply an instrumentalisation of language. It is a disruption of how meaning is made, from the words we use to the broader semiotic systems weaving social life. Exiting the pre-critical phase of AI and conquering some kind of technological maturity is therefore of crucial importance for understanding what is worth being efficiently computed when the capacity for speech and sign-manipulation has been technologically replicated through artificial synthesis. The conceptual tools we bring to bear on that question will determine not only how we understand AI, but how we understand ourselves.

In 1948, Norbert Wiener (2019: 54) observed: "The thought of every age is reflected in its technique". At the time, cyberneticists extensively

debated the functionalities of machines along with those of organisms and societies according to the quality of information they exchange. Given this insight, it is therefore worth asking what kind of thought is reflected in generative AI and what AI would reveal itself to be if we changed the way we think about and conceptualise it. This question is not speculative. It asks instead what kind of technical cultures and related artistic practices we want to build.

In 1957, calling for the integration of technical and cultural realities, Gilbert Simondon (2017, 19) argued that what was needed to integrate them was "an awareness of the nature of machines, of their mutual relations and of their relations with man, and of the values implied in these relations". What might such awareness look like as an alternative to computational illusion, which has so systematically foreclosed it? I would argue that it lies in pursuing meaningful, rather than merely efficient, automation: not simply because it is technically possible, but only when it makes sense, and in choosing other techniques where it does not. Artistic practices seem to offer a privileged ground for cultivating such awareness.

In the AI endgame economy, when nothing can truly finish and nothing new can begin, artists who constantly work to appropriate and repurpose the tools they use to create their work become crucial. I suggest naming this kind of practice—and the broader space of inquiry it opens up—art-informatics. Without experimental artistic practices that would enable us to repurpose computational techniques through strategic alliances among socio-informatics (Wulf et al. 2018), the human sciences, and the arts, we will continue to ignore what computer code can actually achieve while generating artificial meaninglessness in a world mired in political chaos. On this, it is worth recalling Jean-François Lyotard's (1991, 57) reflection on the relation between *logos* and *techne*, "all this remains to be thought out, tried out". This is neither a utopia nor romanticism, it is the only realistic path forward in a world where non-noetic logos has become a property of computational machines and AI-generated visual culture is everyday culture.

The proponents of socio-informatics argue that computational artifacts should be embedded in social practices from the very stage of their conceptualisation. This stance stems from the awareness that the quality of these artifacts depends on how they affect those practices. This entails two things. First, applied computer science requires a solid theory of what social practice is. Second, to align machine operations with these practices, rather than with abstract "human values," applied computer science must turn to design sciences—a move which poses a significant challenge, both methodologically and epistemologically, since design is by definition theoretically underdetermined and lacks the certainty typically associated with formal sciences. In 1999, Rob Kling defined socio-informatics as "the interdisciplinary study of the design, uses and consequences of information technologies that takes into account their interaction with institutional and cultural contexts."

What would happen if we imagined a shift from design to arts? Several years before generative AI spread online, American artist Trevor Paglen (2016) described our everyday technologically saturated visual culture as "invisible," highlighting how vast the landscape of machine-made invisible images not meant for human eyes is (from surveillance camera recordings and self-driving cars to social media algorithms). As Paglen (2016) observed, the overwhelming majority of images circulating today are produced by machines for other machines, with human eyes rarely—if ever—part of the equation. He continues:

If we want to understand the invisible world of machine-machine visual culture, we need to unlearn how to see like humans. We need to learn how to see a parallel universe composed of activations, keypoints, eigenfaces, feature transforms, classifiers, training sets, and the like. But it's not just as simple as learning a different vocabulary. Formal concepts contain epistemological assumptions, which in turn have ethical consequences. The theoretical concepts we use to analyse visual culture are profoundly misleading when applied to the machinic landscape, producing distortions, vast blind spots, and wild misinterpretations (Paglen 2016).

That the commercialisation of AI image generators has only pulled us deeper into a massive, energy-consuming dataset is by now rather obvious. The problem, however, remains: the theoretical concepts we used to analyse classical visual culture (representation, meaning, semiosis, mimesis etc.) are inadequate to describe the new invisible visual culture. At the same time, what has come to a head with generative technologies is, as French and Canadian artist Gregory Chatonsky (2025) puts it, "a fundamental tension between production and consumption." The same infrastructure "can be oriented toward" either technological production or technological consumption "and big tech companies have turned a part of popular production into unbridled consumption of their technologies."

This mainstream AI landscape absolutises the "symbolic misery" with which Stiegler (2014, 10) described the 2000s. We must therefore ask whether AI images—which lack meaningful reference yet intervene in everyday life—can shape human visual culture in ways that are meaningful to us, rather than simply operating as infrastructures of our industrialised memory, of which AI slop is the most glaring by-product. Art-informatics as I envision it is a space where such questions open pathways toward thought-provoking counter-practices and experimental approaches to computing aimed at resituating their outputs within wider artistic projects. After all, what is at stake is the revaluation of something technical and artistic activities share: the promise of emancipation—what. Computational illusion offers no such emancipation; it forecloses rather than emancipates.

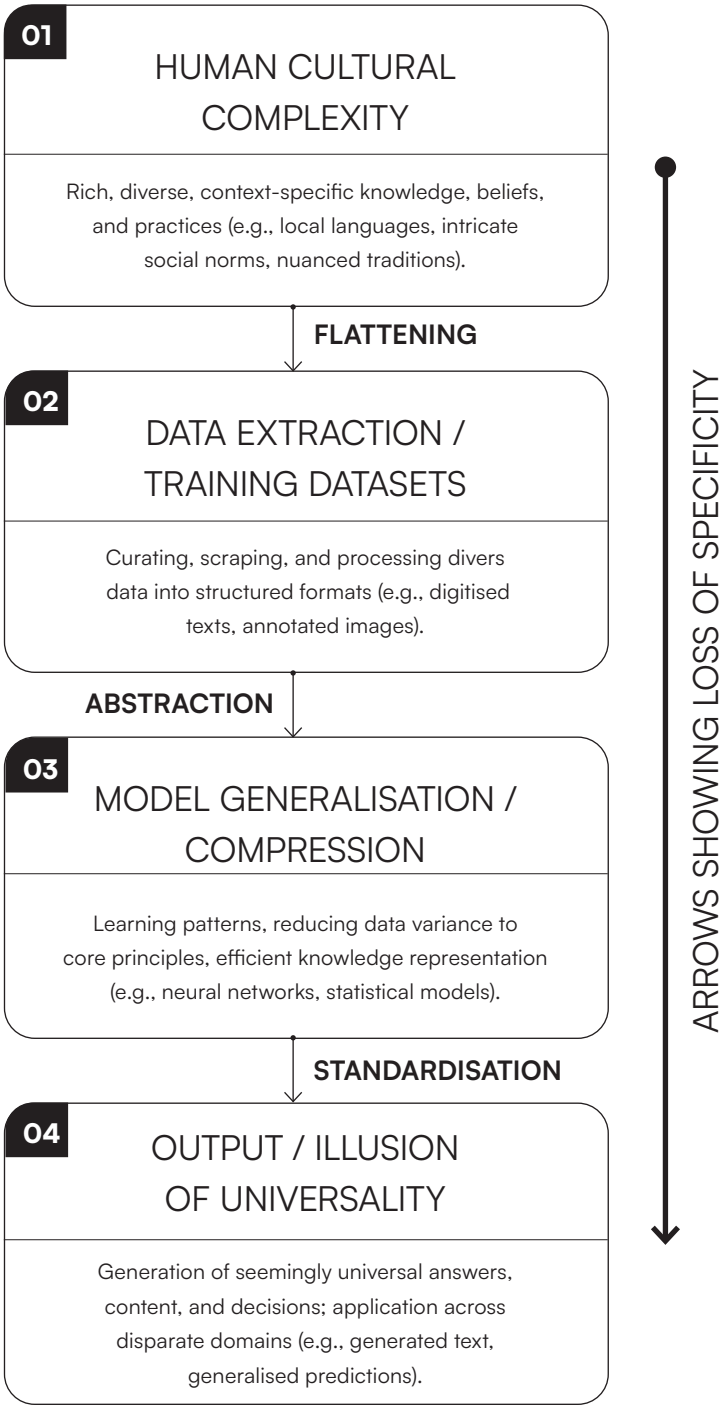
Art-informatics offers a way forward beyond the impasse computational illusion creates. It proposes not reforming AI systems from

within but developing alternative computational practices—grounded in philosophical reflection and artistic experimentation, oriented toward symbolic production rather than operational efficiency, and embedded in communities rather than platforms.

...FROM WITHIN EAST-CENTRAL EUROPE

For scholars and artists in East-Central Europe, this proposition has a particular resonance, and particular urgency. We occupy a semi-peripheral position in the global value chain, cultural and economic alike, which means that the question of who controls AI systems is inseparable from the question of who controls the production of cultural meaning in our region. Providing free training data to proprietary technologies—in the belief that we are making our visual heritage "visible" in global AI landscape—stems from a core tenet of computational illusion: that well-trained AI models, given the right local data and computational resources, can produce good enough cultural content in real time. From a technopolitical perspective, this only makes our visual heritages more dependent on the operations of an invisible machine-machine infrastructure that lies beyond our understanding and our local ways of seeing. This is not the way to preserve our cultural autonomies.

The alternative I propose draws on what dissidence has meant in this part of Europe. Rather than adapting to the dominant AI-related political economy, we need to oppose it by fostering cultures of invention, rather than of mere resistance—or at least develop the capacity to do so. This requires building our own machines—experimental, non-capitalist in their design and underpinning concepts, smaller and slower yet good enough, rather than optimised and super-performing. Art-informatics suggests that the question is not whether to engage computational techniques, but how they could serve futures worth wanting for artistic practices and technical cultures. Considering where we actually stand and the prevailing despair, it is a very slight hope. But it is, perhaps, worth holding onto.


01
HUMAN CULTURAL COMPLEXITY
Rich, diverse, context-specific knowledge, beliefs, and practices (e.g., local languages, intricate social norms, nuanced traditions).
FLATTENING
02
DATA EXTRACTION / TRAINING DATASETS
Curating, scraping, and processing divers data into structured formats (e.g., digitised texts, annotated images).
ABSTRACTION
03
MODEL GENERALISATION / COMPRESSION
Learning patterns, reducing data variance to core principles, efficient knowledge representation (e.g., neural networks, statistical models).
STANDARDISATION
04
OUTPUT / ILLUSION OF UNIVERSALITY
Generation of seemingly universal answers, content, and decisions; application across disparate domains (e.g., generated text, generalised predictions).
ARROWS SHOWING LOSS OF SPECIFICITY

REFERENCES

Bates, David. 2025. "La prédiction n'est pas l'anticipation." In *Penser l'intelligence artificielle. Enjeux philosophiques, politiques et culturels de l'automatisation numérique*, edited by Anne Alombert, Alban Leveau-Vallier, and Baptiste Loreaux, 68–81. Paris: Les Presses du réel.

Bender, Emily M., and Nanna Inie. 2026. "We Need to Talk About How We Talk About 'AI'." *Tech Policy Press*, Jan 7. https://www.techpolicy.press/we-need-to-talk-about-how-we-talk-about-ai/

Boden, Margaret A. 2018. *Artificial Intelligence: A Very Short Introduction*. Oxford: Oxford University Press.

Canguilhem, Georges. 2008. "The Brain and Thought." Translated by Steven Corcoran and Peter Hallward. *Radical Philosophy* 148: 7–18.

Chatonsky, Gregory. 2025. Génération : entre production et consommation. April. https://chatonsky.net/generation-entre-production-et-consommation/

Choudhury, Suparna, and Jan Slaby, eds. 2012. *Critical Neuroscience: A Handbook of the Social and Cultural Contexts of Neuroscience*. Oxford: Wiley-Blackwell.

Davies, Paul. 2019. *The Demon in the Machine: How Hidden Webs of Information Are Solving the Mystery of Life*. Chicago: University of Chicago Press.

Harari, Yuval Noah. 2016. *Homo Deus: A Brief History of Tomorrow*. New York: Harper Collins.

Harari, Yuval Noah. 2018. *21 Lessons for the 21st Century*. New York: Random House.

Hawkins, Jeff, and Sandra Blakeslee. 2004. *On Intelligence*. New York: Times Books.

Heaven, Will Douglas. 2025. "How AGI Became the Most Consequential Conspiracy Theory of Our Time." *MIT Technology Review*, October 30. https://www.technologyreview.com/2025/10/30/1127057/agi-conspiracy-theory-artifcial-general-intelligence/

Kling, Rob. 2007. "What Is Social Informatics and Why Does It Matter?" *The Information Society* 23 (4): 205–20. https://doi:10.1080/01972240701441556

Lhérété, Héloïse, ed. 2024. *Qu'est-ce que l'intelligence*. Auxerres: Sciences Humaines Éditions.

Lyotard, Jean-François. 1991. *The Inhuman: Reflections on Time*. Translated by Geoffrey Bennington and Rachel Bowlby. Stanford: Stanford University Press.

McCarthy, John, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon. 2006. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence August 31, 1955." *AI Magazine* 27(4): 12. https://doi.org/10.1609/aimag.v27i4.1904

Paglen, Trevor. 2016. "Invisible Images (Your Pictures Are Looking at You)." *The New Inquiry*, December 8. https://thenewinquiry.com/invisible-images-your-pictures-are-looking-at-you/

Parisi, Luciana. 2019. "The Alien Subject of AI." *Subjectivity* 12: 27–48. https://doi.org/10.1057/s41286-018-00064-3

Pinker, Steven. 2005. So How Does the Mind Work?, Mind & Language 20(1): 1–24. https://doi.org/10.1111/j.0268-1064.2005.00274.x

Simondon, Gilbert. 2017. *On the Mode of Existence of Technical Objects*. Translated by Cecile Malaspina and John Rogove. Minneapolis: Univocal Publishing.

Suleyman, Mustafa, and Michael Bhaskar. 2023. *The Coming Wave: Technology, Power, and the Twenty-first Century's Greatest Dilemma*. New York: Crown.

Stiegler, Bernard. 2014. *Symbolic Misery, Volume 1: The Hyperindustrial Epoch*. Translated by Barnaby Norman. Cambridge: Polity Press.

Stiegler, Bernard. 2016. *Automatic Society, Volume 1: The Future of Work*. Translated by Daniel Ross. Cambridge: Polity Press.

Turing, A. M. 1950. "Computing Machinery and Intelligence," *Mind* 59 (236): 433–60.

Wiener, Norbert. 2019. *Cybernetics: Or Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press.

Wulf, Volker, Volkmar Pipek, David Randall, Markus Rohde, Kjeld Schmidt, and Gunnar Stevens, eds. 2018. *Socio-Informatics: A Practice-Based Perspective on the Design and Use of IT Artifacts*. Oxford: Oxford University Press.

EPISTEMIC CULTURAL FLATTENING IN GENERATIVE VISUAL AI:

BENCHMARKING HUNGARIAN HERITAGE AND DESIGNING A V4 PATH TOWARD CULTURALLY AWARE TEXT-TO-VIDEO

Brigitta Iványi-Bitter, Tibor Bacsi, and Szilárd Szakács

ABSTRACT

Generative image systems increasingly shape how culture becomes visible in design workflows and heritage interpretation. Their outputs often achieve technical plausibility while offering limited support for validating cultural provenance, shaping how synthetic images circulate as cultural references. This article introduces Epistemic Cultural Flattening (ECF) and an Epistemic Interpretive Framework (EIF) to distinguish structural performance from epistemic readability and to describe reductions of culture-specific legibility under globally dominant visual templates. The study operationalizes EIF through a cultural fidelity benchmark rating generated images by cultural fit, stylistic accuracy, and technical quality. It uses a Hungarian heritage benchmark set within a cross-cultural comparative corpus and compares outputs from four diffusion-based generators.

The article proposes an ECF failure-mode typology that makes cultural flattening visually legible. It also outlines a V4-oriented workflow for culturally aware text-to-video, integrating GLAM sourcing, multilingual metadata, controlled model adaptation, and expert review for low-resource cultures in Central Europe.

#epistemic cultural flattening; #cultural fidelity; #benchmark; #generative AI; #Hungarian visual heritage; #low-resource languages; #GLAM; #V4; text-to-video

https://doi.org/10.21096/disegno_2025_1b-ibtbszsz

1. INTRODUCTION

1.1 The problem: high technical quality, limited cultural fidelity

Over the past few years, generative AI has shifted from a novelty to an everyday mediator of visual culture. Image generators now routinely support design prototyping, historical illustration, place simulation, and the on-demand visualisation of cultural references, reinforcing a broader condition in which cultural production and circulation increasingly operate through algorithmic infrastructures (Striphas 2015). The apparent fluency of these systems sustains a persistent paradox: an image can achieve high technical quality—sharp, coherent, visually persuasive—while cultural provenance and stylistic accuracy remain unstable. It can satisfy generic expectations of what folk embroidery, historic painting, or Central European architecture "should" look like, while producing weak support for recognition within the culture referenced by the prompt.

This gap matters because cultural meaning resides in the codes through which a subject becomes readable—ornamental grammar, material conventions, typographic habits, and historically situated stylistic rules (Hall 1997). When these cues shift toward globally dominant templates—generic "European old town" aesthetics, interchangeable "Eastern folk" ornament, or an accidental amalgam of period styles—validation of cultural provenance and stylistic accuracy becomes essential. The output then functions as a persuasive substitute whose authority arises from visual polish and coherence rather than from culturally situated evidence, a dynamic that becomes especially consequential as AI aesthetics normalise a distinctive regime of synthetic plausibility (Manovich 2019).

Seen from a design and cultural heritage perspective, this paradox follows from how generative systems are built and evaluated. Cultural legibility is mediated by infrastructures: training datasets and their uneven geographies of visibility; metadata and naming systems that shape what becomes learnable; prompt languages and translation layers that compress culturally specific terms; and evaluation regimes that reward coherence while leaving provenance weakly represented (Crawford 2021). In low-resource cultural contexts, including the Visegrad region, these infrastructures privilege what circulates widely through platformed extraction and large-scale data capture, so cultural requests get resolved through statistically dominant priors rather than through locally dense

reference structures (Couldry and Mejias 2019). Search engines and platforms further reinforce this dynamic by shaping discoverability and salience through ranking, categorisation, and moderation logics that structure what appears culturally retrievable at scale (Gillespie 2018). What emerges is a systematic tendency toward cultural flattening: outputs that look generally correct but whose culturally situated reading relies on cues that remain unevenly supported.

Ania Malinowska, in her contribution to this issue, conceptualises this structural dynamic as *AI assimilationism*, in which local aesthetics gain visibility through dominant norms and infrastructures shaped by English-language defaults and platformed hierarchies; her triad of linguistic standardisation, economies of visibility, and misidentification provides a close theoretical match to the infrastructural mechanisms traced here.

1.2 Research questions and contributions

This article approaches cultural fidelity in generative images as a design and cultural heritage problem with measurable outcomes. It frames the problem through three research questions that focus on measurement, empirical diagnosis, and design intervention.

RQ1: How does the benchmark measure divergence between technical quality and culturally situated recognisability in generative images?

RQ2: Which recurring failure modes characterise this divergence in Hungarian cultural heritage imagery across domains (fine art, folk art, and architecture)?

RQ3: How does this diagnosis shape a design workflow for generative tools that support culturally specific outputs across the V4 countries, including text-to-video applications?

1.3 Contributions

This article offers five contributions that connect conceptual framing, measurement, empirical evidence, and design intervention:

Conceptual: *epistemic cultural flattening* (ECF) is introduced as a term for culturally plausible-looking outputs that lose provenance-specific legibility and this phenomenon is framed through an *epistemic interpretive framework* (EIF) that separates structural performance from epistemic readability.

Methodological: A cultural fidelity benchmark is proposed that evaluates AI-generated images along three complementary dimensions (cultural fit, stylistic accuracy, and technical quality) to support systematic comparison across models, domains, and cultures.

Empirical: Results from a Hungarian heritage benchmark set situated within a cross-culture comparative corpus are reported using multiple image generators to map how cultural fidelity varies by domain (fine art, folk art, and architecture) and by model family.

Analytical: An ECF failure-mode typology is developed that makes cultural flattening visually legible through recurring patterns such as generic substitution, motif drift, semantic collapse of local terms, and style–reference decoupling effects.

Design: The diagnosis is translated into a V4-oriented workflow for culturally aware generative tools, extending image generation to text-to-video and grounded in GLAM collaboration, multilingual metadata practices, model fine-tuning, and iterative expert evaluation.

1.4 Paper roadmap

This paper is structured as follows. Section 2 introduces ECF and the EIF, and it differentiates these concepts from adjacent terms commonly used to describe generative error and cultural bias. Section 3 presents the cultural fidelity benchmark and the empirical material that supports it, including the benchmark dimensions, the dataset design, the prompting strategy, the reference-image grounding approach, and the evaluation procedure. Section 4 reports the empirical results, with a focus on how cultural fit and stylistic accuracy vary in relation to technical quality across domains and model families within the Hungarian heritage benchmark set and the cross-culture comparative corpus.

Building on these findings, section 5 develops an ECF failure-mode typology that describes how cultural flattening becomes visible at the level of motifs, styles, and provenance cues. Section 6 translates the diagnostic insights into a V4-oriented design workflow for culturally aware generative tools, extending the logic of the benchmark toward text-to-video development through GLAM collaboration, multilingual metadata practices, controlled model adaptation, and iterative expert evaluation. Section 7 discusses implications in three domains: design research and evaluation practice, GLAM institutions and cultural policy, and V4 toolmaking for creative and educational use. Section 8 concludes the paper by summarising the main contributions and outlining limitations alongside future research directions.

The next section establishes the conceptual vocabulary that frames the benchmark, guides the empirical analysis, and supports the design pathway developed in the second half of the paper.

2. CONCEPTUAL FRAMING: DEFINING AND DIFFERENTIATING ECF

2.1 Epistemic Interpretive Framework (EIF): two layers of evaluation

This article uses an EIF to separate two forms of performance that often collapse into a single idea of "image quality" in everyday use. The EIF treats generative outputs as cultural representations that require evaluation on two analytically distinct layers: structural performance and epistemic readability.

Layer A: *Structural performance* captures the technical and compositional competence of a generated image, independent of its cultural attribution. It captures qualities that remain largely transferable across contexts, such as technical clarity, compositional coherence, artifact handling, and overall visual plausibility. In practical terms, this layer

corresponds to the kind of acceptable quality that becomes visible through resolution, lighting consistency, surface detail.

Layer B: *Epistemic readability* describes an image's capacity to carry culturally situated meaning in a way that supports recognition by informed observers. This layer centres on *cultural fit* and *stylistic accuracy*. Cultural fit concerns whether the output activates culture-specific signifiers that make the depicted object, place, or tradition legible within the referenced cultural context. Stylistic accuracy concerns whether the output follows the relevant aesthetic and material conventions, such as ornamental grammar, formal structures, craft logics, architectural typologies, or art-historical registers associated with the referenced domain. Epistemic readability therefore measures the extent to which an output sustains validation of cultural fit and stylistic accuracy within the referenced context.

Within EIF, ECF appears as a patterned divergence between these two layers: An output can score highly on structural performance while showing low epistemic readability. This divergence matters because it changes how images function in cultural circulation. High structural performance increases persuasive force, while reduced epistemic readability shifts cultural specificity toward generic templates. EIF therefore makes it possible to diagnose cultural failure modes that remain hidden when evaluation focuses on technical quality alone.

The ECF phenomenon is close to what Marková (in her article in this issue) frames as a parallel risk through *collective vulnerability*, where profit-driven AI development flattens cultural diversity toward a computational mean, a framing that complements ECF by linking output-level divergence to broader power structures.

2.2 Definition: Epistemic cultural flattening (ECF)

ECF describes a patterned shift in generative outputs in which culturally specific meaning becomes less distinguishable, even as visual coherence remains high. Within the EIF, ECF names the divergence between structural performance and epistemic readability: an image can score highly on technical and compositional quality while offering limited support for validation as a culturally situated representation. ECF therefore captures a form of representational homogenisation in which outputs default toward globally dominant templates and broadly legible cues, and in doing so reduce the visibility of culture-specific ornamental grammar, material conventions, and historically situated styles.

Diagnostic signals appear as score divergence between technical quality and cultural fit and stylistic accuracy. Visual form often stabilises through template substitution and provenance thinning.

In this issue, Janda describes a closely aligned phenomenon to ECF as *regional invisibility*, where culturally dense low-resource places become reconstructed through globally legible templates, producing subtle drift through normalisation, typological substitution, and shifts toward universally plausible aesthetics.

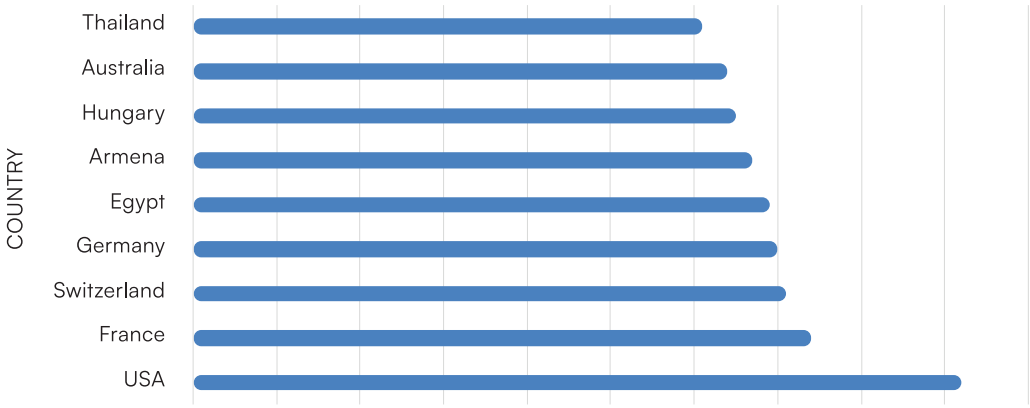


FIGURE 1. *Situating ECF among adjacent concepts*

ECF gains analytical precision through its contrast with terms that circulate widely in public and technical discussions of generative AI. These labels often describe genuine phenomena, while ECF focuses on a specific representational pattern: the reduction of culture-specific legibility under globally dominant visual templates. Korinek in his contribution to this issue, frames this asymmetry as a Central European condition: small-language environments enter into collaboration with global generative systems from an unequal position and the process amplifies tension between local experience and algorithmically preferred forms of language and imagery. This perspective strengthens the conceptual scope of ECF as an output-level pattern that remains structurally linked to cultural position within global data environments.

ECF and hallucination

In technical discussion of generative AI "hallucination" typically denotes outputs that contain implausible elements. ECF has a different regime of error: outputs often remain visually coherent and aesthetically plausible, while cultural attribution relies on substitution. The image reads as something like the requested cultural tradition, place, or artwork, and it achieves plausibility through generic cues rather than through culture-specific references. In this issue Krzykawski's critique of *AI parlance* complements this differentiation work by foregrounding how everyday terms such as hallucination shape public understanding and flatten conceptual distinctions that support critical evaluation.

ECF and bias

"Bias" commonly names systematic disparities that affect social groups, including stereotyping and unequal treatment across gender, race, or ethnicity. ECF complements this conversation by foregrounding a representational dimension that concerns cultural legibility, especially in heritage contexts. It highlights how ornamental grammar, typological specificity, and historically situated stylistic registers become diluted when models resolve prompts through dominant templates.

ECF and localisation

"Localisation" typically emphasises geographic placement and language adaptation, often evaluated through the extent to which a model depicts the correct place or renders text in the appropriate language. ECF addresses a broader representational mechanism: a scene can appear "Eastern European" while still losing the specific cues that support Hungarian, Polish, Czech, or Slovak cultural attribution. Cultural specificity depends on more than place names; it depends on typologies, materials, style conventions, ornamental grammar, and craft logic.

ECF and style transfer

"Style transfer" normally focuses on mapping one visual style onto another image or subject, and evaluation often concentrates on aesthetic resemblance. ECF speaks to the relationship between style and provenance. An output can match a plausible stylistic register while cultural attribution remains weak when provenance anchors (like regional typology) remain underrepresented in the training data.

ECF and dataset scarcity

Dataset scarcity describes uneven representation within training corpora, a condition that shapes what models can learn. ECF names the epistemic outcome that follows from such conditions: cultural requests get resolved through the widely available templates, and cultural distinctiveness shifts toward generalised visual forms. Dataset scarcity functions as a driver; ECF describes the representational effect visible in outputs and measurable through evaluation.

3. METHODS: THE CULTURAL FIDELITY BENCHMARK

3.1 Benchmark logic

This study operationalises cultural fidelity through a benchmark instrument that separates three evaluative dimensions: *cultural fit*, *stylistic accuracy*, and *technical quality*. The instrument responds to a recurring challenge in the assessment of generative images: outputs often achieve strong technical quality, and this competence increases their persuasive force in cultural circulation. Evaluation therefore benefits from a structure that distinguishes visual coherence from culturally situated legibility.

Cultural fit captures whether an output belongs to the referenced cultural context in a way that supports recognition by informed observers. It concerns the presence of culture-specific cues, such as typologies, iconographic conventions, and ornamental grammar, that enable viewers to attribute the image to the intended culture rather than to a generic proxy. Cultural fit therefore addresses the question of cultural attribution: the relationship between the prompt's cultural reference and the image's culturally situated readability.

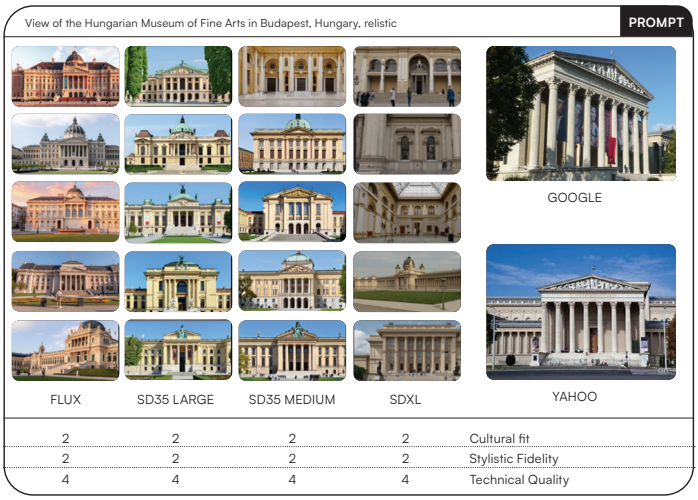


FIGURE 2. *Cultural fidelity benchmark instrument (three domain exemplars). The benchmark evaluates AI-generated images along three dimensions (cultural fit, stylistic accuracy, and technical quality) using a consistent 1–7 Likert scale. The three cards illustrate how the same evaluation backbone applies across domains while the interpretive focus shifts with the three categories.*

Stylistic accuracy captures whether the output follows the relevant aesthetic and material conventions associated with the referenced domain. It concerns formal and material coherence within a tradition: the exact structure of ornament, characteristic colour relations, compositional logic, craft constraints, and domain-specific visual registers. Stylistic accuracy therefore focuses on how faithfully the output aligns with the stylistic rules that shape a tradition's internal visual logic.

Technical quality captures structural performance at the level of the image as a rendered artifact. It concerns clarity, resolution, compositional stability, texture handling, and the overall absence of distracting visual elements that compromise visual coherence. Technical quality supports cross-model comparison because it remains interpretable across domains and cultures, while cultural fit and stylistic accuracy remain context-dependent.

Together, the three dimensions support an interpretable diagnosis of ECF. High technical quality can coexist with low cultural fit or stylistic accuracy, and this divergence provides a measurable trace of ECF. The benchmark therefore functions as a practical instrument for comparative evaluation across models, domains, and cultures, while also supporting qualitative interpretation through exemplar images. Benchmarking traditions in algorithmic accountability show how structured evaluation reveals systematic performance gaps and guides intervention, a logic that supports the cultural fidelity benchmark developed here (Buolamwini and Gebru 2018).

3.2 Dataset construction with Hungarian heritage as anchor

The benchmark builds on a comparative image-generation corpus designed to support controlled cross-cultural analysis while keeping Hungarian visual heritage at the centre of interpretation. The dataset comprises approximately 900 generated images produced with four diffusion image generators (Stable Diffusion XL, Stable Diffusion 3.5 Large, Stable Diffusion 3.5 Medium, and Flux Schnell) across ten cultures and three domains: architecture, fine art, and folk art. While the comparative corpus targets ten cultures, the results reported here draw on the eight for which evaluation data was complete at the time of analysis.

The benchmark uses countries as proxies for cultural contexts, a simplification that treats each national dataset as representative of a broadly identifiable visual heritage tradition. This operationalisation supports controlled comparison while recognising that cultural production within any country is internally diverse and that national boundaries do not map neatly onto cultural boundaries.

The selection of cultures functions as a set of comparative positions in global data hierarchies, spanning contexts with high digital resources (the United States, France, Germany, Switzerland, Australia) and contexts with lower digital resources (Hungary, Bangladesh, Thailand). One additional low-resource culture (Egypt) was targeted but not yet evaluated at the time of analysis.

Since this research began with an inquiry into misrepresentation in Hungarian heritage, the Hungarian dataset serves as the anchor case. The benchmark targets heritage objects and references that rely on culture-specific cues for recognition, including ornamental grammar, material conventions, regionally specific typologies, and art-historical provenance. The two comparator groups support interpretation in two ways: high-resource contexts function as baselines for strong learnability within global training corpora, while lower-resource contexts clarify how cultural specificity behaves under sparse representation conditions.

The corpus is organised through a domain-based prompt set that aligns the three heritage categories with comparable prompt structures across cultures. Each culture receives prompts in each domain, and each prompt is executed across all four generators to enable model-by-model comparison under identical textual conditions.

This design treats cultural representation as a pattern that emerges across repeated prompts rather than as an isolated anecdote: multiple outputs per prompt and per model support the identification of recurring representational logics tied to a given culture and domain.

To support cross-culture comparability, the prompt structures remain standardised while allowing culture-specific references to enter through culturally situated objects.

3.3 Generation systems, grounding, evaluation design, and analysis approach

Cross-model comparison matters because cultural fidelity rarely behaves as a simple function of technical advancement: models differ in rendering competence, stylistic priors, and text–image alignment, and these differences shape how culturally specific prompts resolve into visual form. A shared prompt set executed across multiple generators therefore supports two complementary readings: model-level differences in cultural fidelity and domain-level patterns that remain stable across models, both of which inform the diagnosis of epistemic cultural flattening.

To support culturally situated evaluation, each prompt was paired with two reference images sourced through search engines. These references function as grounding anchors rather than as definitive truth claims. Their role is practical and comparative: they provide evaluators with a memory aid for the culturally referenced object, place, or work, and they render visible the visibility regimes that structure online access to heritage imagery. Reference selection therefore carries diagnostic value because search results reflect platformed hierarchies, dominant iconographies, and metadata conventions that shape what becomes culturally retrievable at scale, including ranking and categorisation effects that organise visibility in platform environments (Gillespie 2018). In this setting, reference images support situated judgement by assisting evaluators in assessing cultural fit and stylistic accuracy while keeping the evaluation open to plurality within a tradition.

This methodological choice also aligns with Marková's observation, stated in her article in this issue, that research on Central European cultural alignment remains limited and that artistic and design research practices offer productive approaches for developing insight. Benchmark-based evaluation therefore operates as an interpretive instrument that combines structured scoring with culturally situated reference points to make patterns of alignment and flattening empirically legible.

Evaluation relied on culturally informed judgement. Evaluators were therefore recruited for their familiarity with relevant cultural contexts and heritage domains, and expertise was operationalised as the ability to recognise culture-specific cues and articulate domain conventions in folk art, fine art, and architecture. The study design used balanced presentation logic to support comparability and reduce fatigue effects. Prompt–model combinations were distributed so that evaluators encountered a controlled mix of domains and systems, and the assignment followed a structured balancing scheme (including Latin-square style distribution) to stabilise order effects across the full set. In addition to numeric ratings of cultural fit, stylistic accuracy, and technical quality, the protocol included brief qualitative notes. These notes function as interpretive traces that connect aggregate scores to recurrent visual patterns, especially in cases where evaluators identify substitution, genericisation, or weakened provenance cues.

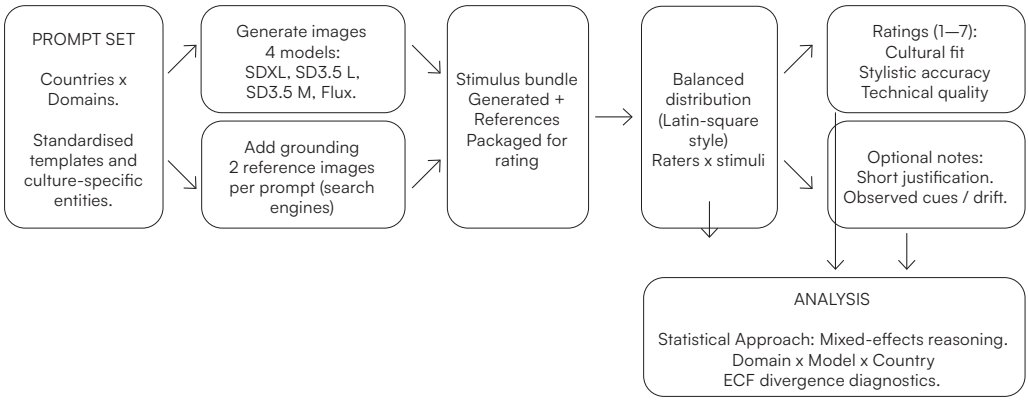
Inter-rater agreement was assessed using Krippendorff's alpha with ordinal distance. Across the three benchmark dimensions, agreement fell in the range $\alpha = 0.20$–0.29 (cultural fit: $\alpha = 0.29$; stylistic accuracy: $\alpha = 0.20$; technical quality: $\alpha = 0.27$). These values fall below the conventional threshold of $\alpha \geq 0.67$ recommended for definitive conclusions (Krippendorff 2004). However, moderate-to-low agreement is characteristic of tasks that involve culturally situated aesthetic judgment rather than factual classification. The benchmark does not claim ground-truth coding: it treats ratings as expert perceptions of culturally referenced outputs, where legitimate variation across evaluators reflects the interpretive plurality inherent in heritage assessment. With approximately three raters per culture, the design supports directional and benchmark-level claims rather than fine-grained pairwise inference.

Analysis treats the dataset as a multi-level structure in which prompts, models, evaluators, cultures, and domains contribute systematic variation. This structure supports mixed-effects reasoning: prompts differ in difficulty, evaluators differ in calibration, and models differ in priors, so comparison benefits from an approach that accounts for nested sources of variance. The analysis therefore centres on two perspectives: a Hungary-centred reading that treats Hungarian heritage as the anchor case, and a-cross-culture comparative reading that situates Hungary within broader visibility gradients. Results are reported in a form that remains readable for design research audiences through domain-level contrasts, model-level contrasts, and the divergence pattern that operationalises ECF as a gap between technical quality and epistemic readability.

Ratings were analysed using linear mixed-effects models (LMMs) with model, domain, and cultural resource level (high vs. low) as fixed effects and participant and prompt as crossed random intercepts. This specification accounts for systematic differences in rater calibration and prompt difficulty while estimating the effects of interest. Responses marked "cannot evaluate" (126 of 3,218 total responses, approximately 4%) were excluded from statistical analysis; these predominantly occurred in prompts requiring familiarity with specific heritage objects outside the evaluator's domain of expertise. An order-effects check confirmed no substantive position effects across dimensions (all $p > .05$), supporting the effectiveness of the randomised presentation design.

Methodologically, Janda's Total Distortion Score approach (published in this issue) complements this benchmark logic by treating drift as structured and repeatable and by coding variables of regional distortion across systems, strengthening comparative reading of low-resource visual contexts.

FIGURE 3. *Study pipeline schematic*



4. RESULTS: DIAGNOSING ECF IN THE HUNGARIAN HERITAGE SUBSET

4.1 Core pattern: technical success can coexist with reduced cultural fidelity

Across the Hungarian heritage benchmark set, results show a stable divergence between structural performance and epistemic readability. Generated images often achieve high scores on *technical quality* (clear rendering, coherent composition, and plausible surfaces) while evaluators assign lower scores to *cultural fit* and *stylistic accuracy*.

The magnitude of this divergence is confirmed by effect-size analysis. Across the comparative corpus, high-resource cultures received significantly higher ratings than low-resource cultures on all three benchmark dimensions: cultural fit ($d = 0.43$, $p < .001$), stylistic accuracy ($d = 0.64$, $p < .001$), and technical quality ($d = 0.47$, $p < .001$), corresponding to medium effect sizes by conventional benchmarks (Cohen 1988). The strongest divergence appears in stylistic accuracy, where domain-specific visual conventions

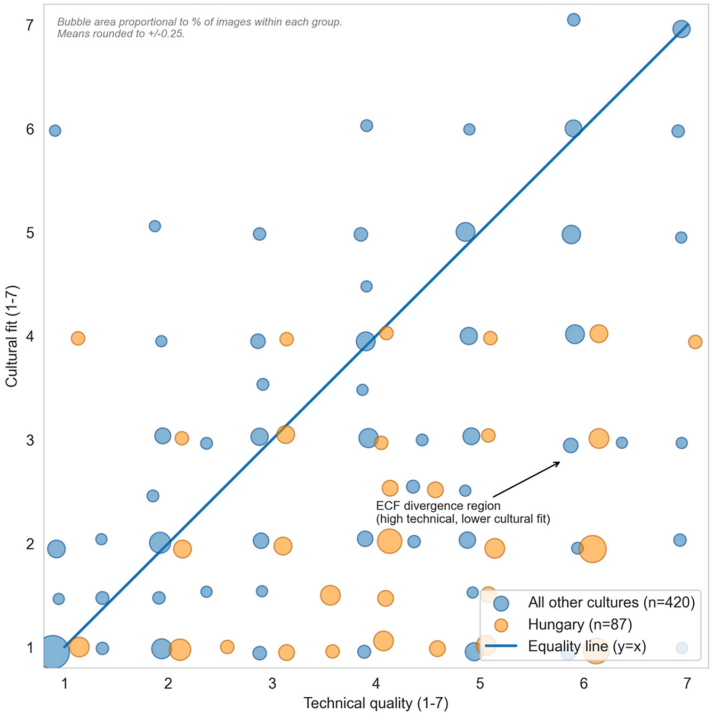


FIGURE 4. *Divergence plot: technical quality vs. cultural fit (Hungary vs. the other countries; bubble area represents percentage of group).*

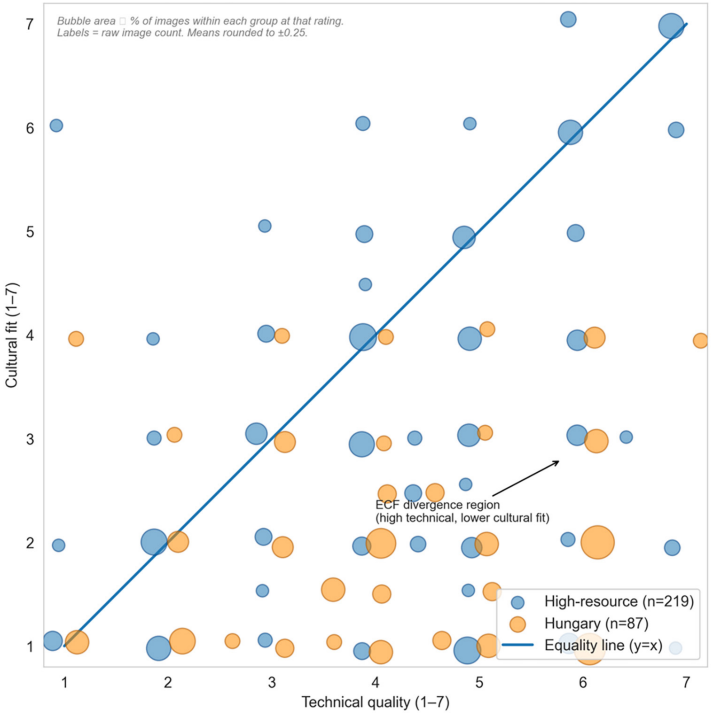


FIGURE 5. *Divergence plot: technical quality vs. cultural fit (Hungary vs. high-resource; bubble area represents percentage of group).*

are most dependent on culture-specific training coverage. That medium effect sizes emerge consistently across all three dimensions, rather than concentrating in cultural fit alone, strengthens the interpretation of ECF as a systematic representational pattern rather than an isolated scoring artifact.

The divergence becomes especially salient in prompts that depend on culturally specific anchors rather than globally common objects. In these cases, the generated image frequently presents a visually acceptable proxy: a genericised "folk" surface for folk art, a broadly "European" architectural scene for place-based heritage, or a historically plausible painting register for fine art. Technical quality supports the plausibility of these proxies, while cultural fit and stylistic accuracy depend on finer-grained cues (ornamental grammar, material conventions, typological specificity, and provenance anchors) that shape culturally situated validation. ECF therefore appears less as a breakdown of image generation and more as a patterned shift in representational strategy: the output prioritises globally legible templates and stabilises cultural meaning through substitution rather than through culture-specific evidence.

Within the Hungarian subset (n = 87 images), the divergence pattern is pronounced: mean technical quality (M = 4.22, SD = 1.55) exceeds mean cultural fit (M = 1.92, SD = 0.94) by 2.30 points.

4.2 Domain differences

Across domains, the results form a clear visibility pattern shaped by the interaction of cultural resource level and heritage type (figure X). In high-resource contexts, architecture tends to remain comparatively stable because place depiction can be assembled from widely circulating photographic templates and globally legible built-environment cues; cultural attribution often holds at the level of recognisability. In the same high-resource contexts, folk and fine art achieve a partial fidelity: models often reproduce a plausible aesthetic register while fine-grained provenance cues and domain-specific ornamental grammar remains uneven, producing outputs that support broad recognition rather than through tradition-specific evidence. Folk prompts also function as representational triggers: Keszeg's contribution to this issue demonstrates how *ethnicising bias* structures the visual grammar through which generative AI systems render folk culture.

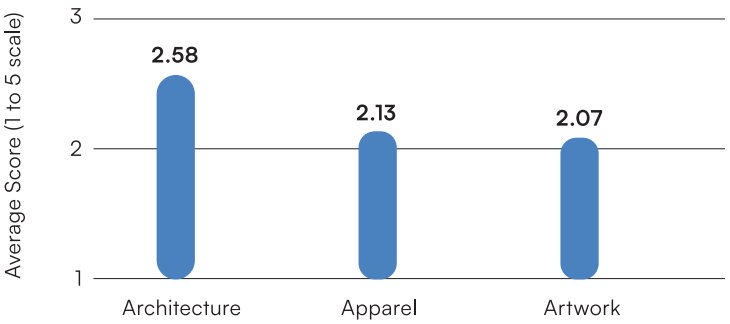


FIGURE 6. *Average AI cultural fidelity score by category.*

In low-resource contexts, the pattern shifts in two distinct ways. Architecture remains visible yet tends toward distortion: outputs stabilise into generic "European" scenery or interchangeable urban typologies, and place-specific anchors weaken, producing cultural attribution that requires additional validation. Folk and fine art show the strongest compression of cultural specificity and therefore approach invisibility at the level that matters for heritage reading: models sustain surface plausibility while the cues that enable culturally situated recognition, like regionally specific typologies, craft constraints, and ornamental grammar, thin out. This comparison summarises how ECF intensifies when cultural requests demand high-resolution specificity and when training data offers limited coverage of the relevant provenance anchors. The cross-domain visibility pattern resonates with Malinowska's contribution to this issue, particularly her account of *economies of visibility*, in which forms that already circulate widely gain further amplification through platformed selection.

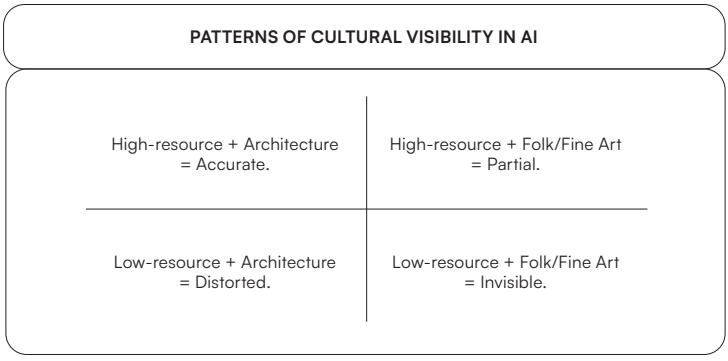


FIGURE 7. *Patterns of Cultural Visibility in AI.*

4.3 Model differences

The Hungary subset indicates small but consistent differences in mean cultural-fit ratings across the four image-generation systems. SDXL receives the highest mean cultural-fit score (M = 2.78), followed by Stable Diffusion 3.5 Medium (M = 2.64) and Stable Diffusion 3.5 Large (M = 2.59), while Flux Schnell receives the lowest mean rating (M = 2.49). Taken together, these values suggest a modest descriptive ordering among the models rather than a sharply differentiated hierarchy.

This ordering remains useful for comparative interpretation. It helps situate the kinds of images that different systems tend to stabilise under identical prompts and shows that some models produce outputs that are judged as somewhat more culturally fitting than others. At the same time, the relatively narrow spread between the mean values calls for caution. The differences shown in figure 8 should therefore be read as descriptive contrasts in average ratings, not as evidence of large separation in model performance.

Most importantly, the model comparison leaves the core ECF pattern intact. Even where one system performs somewhat better than another on average cultural fit, the broader structure of ECF remains visible across

all four systems. Improvements at the model level may raise overall capability, but they do not resolve the more fundamental divergence between technical image generation and culturally grounded readability. In other words, model choice matters for relative performance, yet the persistence of ECF points to a wider infrastructural condition that shapes cultural visibility across systems rather than to a problem confined to any single model.

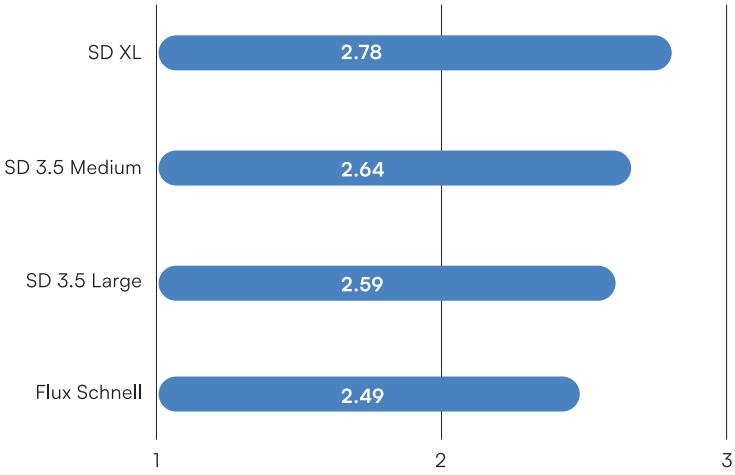


FIGURE 8. *Descriptive mean cultural-fit ratings across models for the Hungary subset.*

5. ECF FAILURE-MODE TYPOLOGY

Benchmark scores provide a necessary diagnosis of cultural fidelity. The benchmark identifies divergence between technical quality and epistemic readability, and it supports comparison across models, domains, and cultures. A typology adds a second layer of interpretive work: it translates measured outcomes into a design vocabulary that describes recurring representational problems.

First, this vocabulary makes ECF readable as a set of repeatable visual patterns, like substitutions, drifts, and reductions of provenance anchors, rather than as isolated errors. Second, it supports action. Designers, data curators, and tool developers benefit from terms that describe where cultural meaning collapses and how it collapses, because these descriptions guide both prompt design and dataset intervention. Typologies therefore function as bridge instruments: they connect quantitative evaluation with qualitative diagnosis, and they support iterative improvement through targeted tests, controlled comparisons, and domain-specific refinement.

The following typology describes six recurring ways in which ECF appears in generated images. Each mode names a representational mode, summarises the visual symptoms that make it recognisable, identifies the prompt conditions that tend to trigger it, and links the pattern to a benchmark across cultural fit, stylistic accuracy, and technical quality.

Generic-European substitution describes a resolution strategy in which a culture-specific request is satisfied through broadly legible "European" templates that support a general regional reading while narrowing culture-specific attribution. Visually, outputs converge on postcard-like urban textures, familiar rooflines, and standardised streetscapes, and they rely on scenic composition rather than on typological anchors and local material details. This mode appears frequently in architecture prompts framed as "view of…" and in folk prompts that specify "traditional" heritage without stronger material constraints. In the benchmark, the pattern typically presents as lowered cultural fit with a mild-to-moderate decrease in stylistic accuracy, alongside high technical quality. Keszeg's contribution to this issue provides a related cross-country reading of substitution as an *imaginary shift*, where outputs stabilise through Mitteleuropean, generalised Slavic, or Orientalised "Eastern" frameworks in response to prompt cues.

Ornamental drift describes outputs that present decorative patterning that reads as heritage ornament while the culture-specific ornamental grammar and craft logic remain unstable. Visually, motifs appear plausible yet reorganise into globally common floral geometry; colour relations move toward standardised palettes associated with generic folk aesthetics; and stitch logic or material behaviour reads as surface decoration rather than craft constraint. This mode concentrates in folk art prompts involving textiles, dress, embroidery, and decorative crafts. In the benchmark, stylistic accuracy tends to drop most clearly, cultural fit often follows with a smaller decrease, and technical quality remains high.

Semantic collapse of local terms describes a prompt–output shift in which culture-specific terms compress into a broader category label, steering the image toward generic object types and generalised heritage cues. Visually, named objects become category-level proxies, for example a traditional guba coat, Miska jug, or regional costume, and key identifiers lose strength while generic decorative cues increase. This mode appears often when prompts include Hungarian terms, diacritics, or regionally specific names, and when prompts combine Hungarian and English descriptors. In Keszeg's contribution to this issue, this mechanism aligns with *representational displacement under noisy labelling*, whereby locally specific visual forms are displaced by more widely available regional or generic repertoires. The benchmark typically records a decrease in cultural fit accompanied by a smaller decline in stylistic accuracy, with technical quality remaining high.

Anachronistic hybridisation describes outputs that integrate stylistic cues from multiple time periods into a single image, producing coherent scenes with unstable historical placement. Visual symptoms include garments that combine silhouettes and accessories from different eras, architectural depictions that mix façade motifs and material treatments associated with distinct periods, and fine art scenes that drift across

historical registers while maintaining an era-like look. This mode often arises in prompts that include dates, in historic architecture prompts, and in folk costume prompts framed as "realistic" without specifying a documentary register. In the benchmark, stylistic accuracy typically declines first, cultural fit follows with a smaller decrease, and technical quality remains high.

Style-source decoupling describes outputs that match a plausible period or genre register while source-specific anchors that support attribution to a named artwork, artist, or tradition remain weak. Visually, prompts naming artworks or artists yield images that "fit the era" while composition, iconography, and work-level identity drift; portrait, devotional, or plein-air scenes appear as genre-typical substitutes; and visual polish increases credibility while provenance cues thin out. This mode concentrates in fine art prompts that reference named works or artists from Hungarian art history. Benchmark scores typically show lowered cultural fit with a mild-to-moderate decline in stylistic accuracy, alongside high technical quality.

Locational blur in architecture describes outputs that produce plausible built-environment depictions while place-specific anchors that support landmark recognition and site attribution remain partial. Visually, landmarks resolve into generic historic façades or scenic city views; spatial context aligns with common tourist-photography conventions; and materials, ornamentation logic, and massing support a general regional reading. This mode appears frequently when prompts rely on a single place name or building name as the main constraint. The benchmark signature typically presents as lowered cultural fit and a smaller decline in stylistic accuracy, with technical quality remaining high.

Together, these six modes convert benchmark divergence into a practical design vocabulary.

FIGURE 9. *ECF failure-mode typology matrix.*

Failure mode	Core visual symptoms	Typical triggers	Benchmark signature
1. Generic-European Substitution	Generic European templates; interchangeable cues	Architecture 'view of…'; folk prompts with weak constraints	Cultural fit ↓ Style ↘ Tech ↗/high
2. Ornamental Drift	Decorative patterning; unstable ornamental grammar	Folk textiles, dress, decorative crafts	Style ↓ Cultural fit ↘ Tech ↗/high
3. Semantic Collapse	Named term → category proxy; object identity ambiguity	Local terms, diacritics, mixed HU/EN phrasing	Cultural fit ↓ Style ↘ Tech ↗/high
4. Anachronistic Hybridization	Mixed period cues; historical register drift	Dated prompts; historic costume/architecture	Style ↓ Cultural fit ↘ Tech ↗/high
5. Style–Source Decoupling	Era/genre mood replaces work-level anchors	Named artists/works; Hungarian art history prompts	Cultural fit ↓ Style ↘ Tech ↗/high
6. Locational Blur	Plausible scene; weak landmark/site anchors	Named buildings; place name as main constraint	Cultural fit ↓ Style ↘ Tech ↗/high

6. FROM DIAGNOSIS TO DESIGN: TOWARD A V4 CULTURALLY AWARE TEXT-TO-VIDEO WORKFLOW

6.1 Why video raises the stakes

The diagnosis of ECF gains additional urgency in the transition from image generation to video generation. Text-to-video amplifies cultural representation through temporal continuity, narrative structure, and embodied cues.

Temporal consistency raises the stakes. Video systems must maintain cultural cues across frames, and this requirement turns minor drift into a visible structural problem. Ornament, materials, typological anchors, and stylistic registers must persist as stable features rather than as accidental successes in single frames. Temporal coherence therefore functions as a stress test for cultural fidelity: a model that occasionally produces a culturally plausible still image can still yield a culturally unstable video when key cues fluctuate across time.

Narrative structure raises the stakes further. Text-to-video workflows typically embed visual synthesis within short scripts or prompts that imply roles, settings, and causal sequences. These scripts activate templates for "what usually happens," and such templates often carry stereotyping pressure. Cultural specificity in heritage contexts relies on situated relations between objects, places, gestures, and social practices, while generative narrative defaults often rely on globally dominant story grammars. As a result, video generation increases the risk that cultural meaning becomes organised through familiar narrative clichés that displace local history, regional nuance, and context-specific social imagination.

Embodied cues raise the stakes a third time. Cultural recognition often relies on how bodies move through space, how garments sit and behave on bodies, how tools are handled, how rituals unfold, and how built environments structure everyday action. These embodied cues matter in V4 heritage contexts because they carry tacit knowledge that remains difficult to encode as isolated visual tokens. Video therefore shifts evaluation toward performative fidelity: the relationship between dress and movement, between craft and gesture, and between architecture and everyday use. This shift expands the benchmark logic beyond static representation and positions culturally aware text-to-video as a design challenge in which temporal stability, narrative choice, and cultural knowledge function as core variables.

Kořínek's (forthcoming) discussion of video work in which *temporal trace* appears as a visible imprint of a specific stage of the technology's development reinforces the value of temporal cultural fidelity as a future evaluation dimension.

6.2 Proposed V4 workflow

The diagnosis of ECF supports a practical intervention path: a staged V4 workflow that integrates cultural governance with model development. The workflow treats cultural fidelity as an engineered property shaped by institutional partnerships, multilingual description practices, controlled model adaptation, and iterative evaluation.

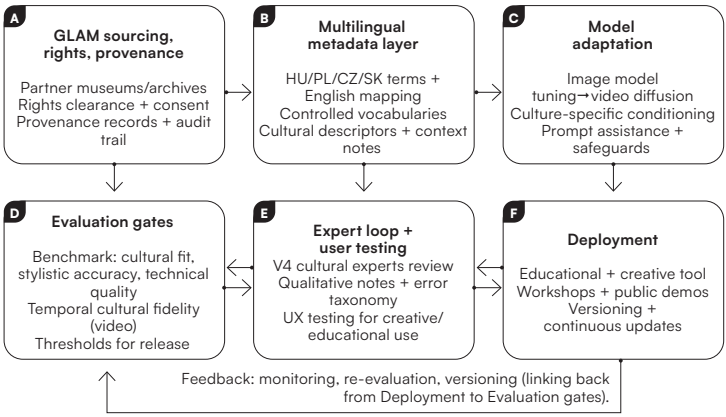


FIGURE 10. *Proposed V4 workflow (design + governance stack) toward culturally aware text-to-video.*

Stage A focuses on GLAM sourcing, rights, and provenance. The process begins through partnerships with museums, archives, and heritage organisations that hold regionally specific collections in Hungary, Poland, Czechia, and Slovakia. This stage structures permissions, consent, and documentation practices, and it establishes a provenance record for each asset, forming an audit trail that supports accountability.

Stage B builds a multilingual metadata layer that functions as a cultural interface. Heritage data gains usability for generative systems through controlled vocabularies and descriptive fields in HU/PL/CZ/SK with an English mapping that supports cross-country comparison and prompt tooling. This layer encodes typologies, ornamental grammar descriptors, material conventions, period registers, and contextual notes that assist both training and evaluation, and it strengthens cultural specificity at the level of language.

Stage C translates the dataset into model adaptation, progressing from image to video. The workflow uses image generation as the first stabilisation step, because images provide fast iteration cycles for cultural fidelity. Cultural conditioning then extends toward video diffusion once image-level performance supports validation.

Stage D introduces evaluation gates that combine the benchmark with a temporal extension for video. The benchmark dimensions (cultural fit, stylistic accuracy, and technical quality) serve as a release gate for image outputs, and the same logic extends toward video through temporal cultural fidelity criteria. Temporal cultural fidelity evaluates stability of key cues across frames, continuity of ornament and material behaviour, and coherence of typological anchors in motion.

Stage E operationalises an iterative expert loop and user testing. V4 cultural archive experts review outputs using both scores and brief qualitative notes aligned with the ECF typology, enabling targeted corrections in prompting, metadata, and adaptation strategy. User testing then evaluates whether the tool supports creative and educational use, including how users interpret cultural cues and how interface choices shape cultural outcomes.

Stage F deploys the system as an educational and creative tool with continuous versioning. Deployment includes workshops, public demonstrations alongside monitoring and periodic re-evaluation by data curators that supports iterative updates and governance over time.

6.3 How the workflow specifically targets ECF

The proposed V4 workflow targets ECF through a shift in what the system treats as learnable cultural evidence. ECF arises when models resolve culturally specific requests through globally dominant templates because these templates offer high statistical stability and broad visual legibility. The workflow intervenes by increasing *epistemic density*, meaning the availability, precision, and internal consistency of culture-specific cues that support validation of cultural provenance and stylistic accuracy.

Epistemic density can grow through governance choices. Rights-cleared GLAM sourcing and provenance documentation establish traceable cultural reference points, and this traceability supports accountability in model development and public deployment. Epistemic density then grows through the multilingual metadata layer, which functions as a cultural interface that translates heritage knowledge into structured descriptors. Controlled vocabularies, bilingual mappings, and context notes supply the model with richer anchors than generic labels.

The workflow also targets ECF by making evaluation an active design component. Evaluation gates translate cultural fidelity into release criteria during development. Expert review and qualitative notes then connect failure modes to actionable causes (prompt structure, metadata gaps, and conditioning weaknesses) supporting targeted iteration. This approach reduces misrecognition through design choices that stabilise provenance anchors and support culturally situated validation. The workflow also responds to Kuchta's observation in this issue that institutional virtual archives can feed AI image generators, so archival omissions and tagging structures can shape downstream cultural legibility in generated outputs.

7. IMPLICATIONS

7.1 Implications for design research and evaluation practice

The findings position cultural fidelity as a core concern for design research that engages generative systems as cultural infrastructures. Evaluation practices that emphasise visual coherence and technical polish capture only one layer of performance, while cultural meaning remains mediated by provenance anchors, ornamental grammar, typological specificity, and historically situated stylistic registers. Treating cultural fidelity as a first-class metric therefore expands evaluation beyond "does it look good" toward "does it support culturally situated validation," a shift that aligns generative assessment with design culture's concern for meaning, context, and interpretive accountability.

A two-layer evaluation logic provides a practical solution: it makes ECF divergence visible and it prevents high technical scores from functioning as implicit proof of cultural fit. This separation supports clearer claims in research reporting, because it allows authors to state precisely which form of performance improves.

Finally, the benchmark functions as a design instrument rather than as a post-hoc audit tool. In design research, instruments shape what becomes visible and therefore what becomes actionable. The cultural fidelity benchmark provides a repeatable way to locate failure modes, to compare systems under controlled prompt conditions, and to translate qualitative observations into structured intervention targets. Used iteratively, the benchmark supports prompt refinement, metadata redesign, and model adaptation decisions, and it provides a shared vocabulary for collaboration between designers, cultural experts, and technical teams. In this sense, benchmarking becomes part of the design process: a method for steering generative systems toward culturally accountable outputs through continuous evaluation and revision.

7.2 Implications for GLAM institutions and cultural policy

The results reposition GLAM institutions as active actors in the generative ecosystem. Museums and archives already function as validators of cultural knowledge through collection practices, cataloguing standards, and interpretive expertise. In the context of generative AI, this validating role extends into infrastructure provision: collections and their descriptive systems shape what becomes learnable, retrievable, and culturally legible in synthetic outputs. Cultural fidelity therefore depends on institutional decisions that historically belonged to heritage governance rather than to model development, especially as visual culture increasingly circulates through algorithmic infrastructures (Striphas 2015).

Metadata and access policies become especially consequential under this view. Cultural legibility in AI outputs draws from how objects, sites, and artworks are named, described, classified, and translated across languages. Controlled vocabularies, multilingual descriptors, provenance fields, and contextual notes supply epistemic density that supports cultural attribution and stylistic accuracy. Access policies shape which images circulate widely, which forms remain locally bounded, and which elements of cultural heritage become represented primarily through secondary, platform-driven iconographies. Crawford and Paglen's (2021) analysis of training images frames these selection effects as infrastructural, because dataset composition and labelling practices shape downstream representational capacity in AI systems. In practice, these policies influence whether AI systems learn heritage through high-quality documentation with strong provenance anchors or through fragmented, unevenly captioned web imagery shaped by ranking, categorisation, and platform governance (Gillespie 2018). Kuchta (2025) also frames institutional virtual archives as infrastructures that can strengthen democratic access and support resistance to censorship

through public availability and distribution, especially in contexts shaped by political pressure on cultural institutions.

Controlled collaboration offers a viable policy direction for supporting cultural sovereignty in low-resource contexts. This stakes a concrete governance role for GLAM institutions within the broader political economy of AI infrastructures and data extraction (Crawford 2021). Partnerships between GLAM institutions, universities, and technical teams can establish rights-cleared datasets, provenance documentation, and evaluation protocols that align model development with public cultural responsibilities. Such collaborations support accountability and reduce reliance on extractive pipelines that treat heritage collections as raw material for unregulated scraping, a dynamic widely discussed through the lens of data colonialism and large-scale appropriation (Couldry and Mejias 2019). They also create conditions for reciprocal benefit: institutions gain tools for education and interpretation, researchers gain structured cultural data, and communities gain representational agency through expert review and culturally situated quality gates. This approach treats cultural heritage as an infrastructural commons governed through consent, documentation, and shared evaluation, and it positions generative AI as a domain where cultural policy shapes the terms of visibility. This aligns with Krzykawski's framing of an East-Central European strategic choice around training data and cultural autonomy, which supports GLAM-led governance as a way to align cultural visibility with accountable infrastructures rather than default platform capture.

7.3 Implications for V4 toolmaking (text-to-video)

V4 toolmaking gains strategic value when it treats cultural fidelity as a shared regional infrastructure rather than as a country-by-country feature. Text-to-video applications amplify representational stakes through temporal continuity and narrative structure, and these properties call for common resources that support culturally situated generation across Hungarian, Polish, Czech, and Slovak contexts. A shared scenario library provides such a resource. It can assemble culturally grounded prompts and story fragments that encode regional diversity across domains, including architecture, folk traditions, and art history, while remaining comparable in structure for evaluation. In parallel, multilingual metadata functions as a cross-border interface layer: it connects local terms, diacritics, and domain vocabularies to aligned descriptors across languages and to an English mapping that supports tooling and interoperability. Cross-culture expert panels then provide the interpretive competence required for validation, ensuring that cultural cues remain legible within each context while supporting comparative diagnosis across the region.

Within this development ecology, the benchmark functions as a quality gate that links model iteration to cultural accountability. Ratings of cultural fit, stylistic accuracy, and technical quality provide structured criteria for release decisions, and the temporal extension for video supports stability checks across sequences.

A V4 text-to-video tool also functions as a testbed for culturally aware generative design. It supports comparative experimentation with data governance, metadata design, conditioning strategies, and interface guidance across multiple low-resource languages and cultural contexts. Workshops, public demonstrations, and educational deployments provide feedback loops that reveal how users interpret cultural references and how interface choices shape cultural outcomes. In this way, the tool becomes both a product and a research instrument: it operationalises cultural fidelity as a design goal, and it generates evidence about how culturally aware AI can support regional storytelling, education, and creative practice in ways that strengthen cultural visibility through accountable infrastructures.

8. CONCLUSION, LIMITATIONS, FUTURE RESEARCH

8.1 Conclusion

This article introduced epistemic cultural flattening (ECF) as a name for a structural gap in generative visual systems which explains how outputs can achieve strong technical plausibility while cultural provenance and stylistic accuracy remain unstable under culturally situated evaluation. The concept clarifies why polished images can still function as weak cultural evidence, especially in low-resource contexts where models resolve specificity through globally dominant templates. By framing this divergence through an epistemic interpretive framework (EIF), the paper positioned cultural fidelity as a design-relevant dimension of performance that shapes how images circulate as cultural references.

The paper also presented a cultural fidelity benchmark that makes this gap measurable across domains and models. By separating cultural fit, stylistic accuracy, and technical quality, the benchmark provides a repeatable instrument for comparative diagnosis. The accompanying typology translated benchmark divergence into a design vocabulary of failure modes, supporting interpretive clarity and actionable intervention targets.

Finally, the proposed V4 workflow demonstrated how measurement can inform intervention. The workflow treated cultural fidelity as an infrastructural design problem shaped by GLAM sourcing, multilingual metadata practices, controlled model adaptation, and iterative expert evaluation. In this framing, culturally aware text-to-video development becomes feasible through governance and evaluation structures that increase epistemic density, stabilise provenance anchors, and support culturally situated validation across Hungarian, Polish, Czech, and Slovak contexts.

8.2.1 Data availability

The study reports results from an ongoing benchmark corpus covering eight of ten targeted cultures at the time of writing. The published article presents aggregate findings due to space constraints. To improve transparency, the project makes available an online documentation package including the prompt inventory, scoring dimensions, codebook, and

benchmark summary tables at. To improve transparency, the project makes available an online documentation package—including the prompt inventory, scoring dimensions, codebook, and benchmark summary tables—which is linked in the online version of this paper. Release of the full image corpus remains conditional on permissions, platform terms, and publication constraints.

8.2 Limitations

This study operated under an English-prompt constraint that introduces translation and tokenisation bottlenecks for culturally specific terms, especially in low-resource language contexts. Inter-rater agreement remained moderate-to-low ($\alpha = 0.20–0.29$), consistent with the interpretive demands of culturally situated aesthetic evaluation. This level of agreement constrains claims to directional patterns and benchmark-level comparisons rather than fine-grained distinctions between individual cultures or prompts. Reference images functioned as grounding anchors and simultaneously reflected platformed visibility regimes. On the other hand their use requires careful attention to permissions in publication. The empirical focus treated Hungary as an anchor case within a comparative corpus. Results draw on eight of ten targeted cultures, evaluation of the remaining cultures was ongoing at the time of writing.

8.3 Future research

Future work can extend the typology and benchmark testing across the V4 region through shared scenario libraries and coordinated expert panels in Hungary, Poland, Czechia, and Slovakia. Text-to-video development calls for an explicit temporal cultural fidelity dimension that evaluates stability of cultural cues across sequences and narrative contexts. Controlled GLAM datasets and participatory metadata design offer an additional direction, enabling culturally grounded training pipelines that strengthen provenance anchors through multilingual descriptive systems. Work on bias loops in cultural heritage practice frames iterative mitigation through dataset governance, evaluation, and interpretive workflows, supporting this direction through an established practice-based model of intervention (Foka et al. 2025). User experience research can further clarify how diverse audiences interpret cultural cues in generated outputs, how interface choices steer cultural attribution, and how educational deployments shape trust, learning outcomes, and creative practice in culturally aware generative tools.

AUTHOR CONTRIBUTIONS

Brigitta Iványi-Bitter: conceptualisation; theoretical framing; writing (original draft); supervision. **Tibor Bacsi:** methodology; experimental setup; prompt strategy; image generation workflow; visual stimuli database curation. **Szilárd Szakács:** methodology; software; automated analysis pipeline; statistical modeling (linear mixed-effects models); validation; data visualisation. All authors: review and editing.

ACKNOWLEDGEMENTS

The corresponding author thanks the students of the course **Designing Against Bias in AI Systems** (Spring and Fall 2025) at MOME for contributions to test datasets, and feedback during the development of the benchmark. Special thanks go to **Vannawongpaisan Varissara, Luca Sára Kiss,** and **Zsófia Kérdy** for particularly insightful comments and support.

REFERENCES

Kuchta, Albín. 2025. "Virtual Spaces: Tools of Poetic Resistance or Censorship Devices?" *Disegno* 9 (1): 148–65. https://doi.org/10.21096/disegno_2025_1aku

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, 610–23. https://doi.org/10.1145/3442188.3445922

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of Machine Learning Research* 81: 1–15.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.

Couldry, Nick, and Ulises A. Mejias. 2019. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford: Stanford University Press.

Crawford, Kate. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.* New Haven: Yale University Press.

Crawford, Kate, and Trevor Paglen. 2021. "Excavating AI: The Politics of Images in Machine Learning Training Sets." *AI & Society* 36: 1105–16. https://doi.org/10.1007/s00146-021-01162-8

Foka, Anna, Gabriele Griffin, Diego Ortiz Pablo, Patrycja Rajkowska, and Sawsan Badri. 2025. "Tracing the Bias Loop: AI, Cultural Heritage and Bias-Mitigating in Practice." *AI & Society* 40: 5835–47. https://doi.org/10.1007/s00146-025-02349-z

Gillespie, Tarleton. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media.* New Haven: Yale University Press.

Hall, Stuart. 1997. *Representation: Cultural Representations and Signifying Practices.* London: Sage.

Janda, Jiří Philippe. 2025. "North Bohemia as a Low-Resource Visual Context: Everyday Heritage, Uneven Visibility, and Synthetic Aesthetics." *Disegno* 9 (1): 108–31. https://doi.org/10.21096/disegno_2025_1jpj

Keszeg, Anna. 2025. "The Paprika-Effect. Central and Eastern Europe as a Noisy Label in AI-Generated Images." *Disegno* 9 (1): 90–107. https://doi.org/10.21096/disegno_2025_1mk

Krippendorff, Klaus. 2004. *Content Analysis: An Introduction to Its Methodology.* 2nd ed. Thousand Oaks, CA: Sage.

Krzykawski, Michał. 2025. "Beyond Computational Illusion: Futures Worth Wanting for Artistic Practices and Technical Cultures." *Disegno* 9 (1): 26–38. https://doi.org/10.21096/disegno_2025_1mk

Malinowska, Ania. 2025. "AI Assimilationism: The Cultural Flattening of Localities in Generative Models." *Disegno* 9 (1): 8–25. https://doi.org/10.21096/disegno_2025_1am

Manovich, Lev. 2019. *AI Aesthetics.* Moscow: Strelka Press.

Marková, Kateřina. 2025. "Against Collective Vulnerability: Understanding Cultural Alignment in LLMs (Not Only) in Central Europe and Calling Design Research to Help." *Disegno* 9 (1): 68–89. https://doi.org/10.21096/disegno_2025_1km

Striphas, Ted. 2015. "Algorithmic Culture." *European Journal of Cultural Studies* 18 (4–5): 395–412. https://doi.org/10.1177/1367549415577

AGAINST COLLECTIVE VULNERABILITY:

UNDERSTANDING CULTURAL ALIGNMENT IN LLMS (NOT ONLY) IN CENTRAL EUROPE AND CALLING ON DESIGN RESEARCH TO HELP

Kateřina Marková

ABSTRACT

Large language models promise efficiency and personalisation, yet they also carry Global North values that may conflict with regional principles and distort human mental models. When profit-driven technological development meets personalisation, the risk of flattening cultural diversity into a computational mean grows, which can be interpreted in terms of collective vulnerability. I argue that this effect is not unique to Central Europe but is shared across all linguistic and cultural communities, albeit for slightly different reasons. Using a Czech-language experiment I explore how design research practices can help us understand the phenomenon known as epistemic cultural flattening. Finally, I chart a possible path to improving cultural alignment as one of the elements that can help us toward better personalised AI tools.

#personalisation; #AI tools; #cultural alignment; #collective vulnerability; #Central European cultures

https://doi.org/10.21096/disegno_2025_1km

1. INTRODUCTION

Cultural alignment in AI tools[1]—and large language models (LLMs) in particular—is commonly characterised as the tool's capacity to respond appropriately within a cultural context by mirroring the value distributions of a particular population (Rystrøm et al. 2025; Bravansky et al. 2025). A significant body of work has been published on this topic. Some authors focus on the differences in representation of moral values between human populations and LLMs (Rystrøm et al. 2025; Hämmerl et al. 2023), while others contend that LLMs exhibit values of their creators (Buyl et al. 2025). Vallor (2024) likens LLMs to mirrors that merely reflect the human past with its entrenched errors and biases. Albrecht (2025) investigates whether vectors and statistical averages can meaningfully capture cultural knowledge. Birhane and McGann (2024) critique technicist approaches to LLMs by challenging the underlying assumption that language is a complete project which can be standardised in a training dataset. While Schröder et al. (2025) inspect the representation of human psychological traits in LLMs, other authors examine how these traits correspond to cultural differences (Atari et al. 2023). Several authors point out issues that seem to influence the level of cultural alignment present in LLMs. Perez (2025) is concerned with the relationship between tokens in the training data. Rystrøm et al. (2025) attribute a significant role to post-training processes. Other studies suggest that the level of cultural alignment might depend on modes of human interaction with LLMs (Khan et al. 2025; Bravansky et al. 2025).

However, for the purpose of investigating the impact of AI tools on Central European cultures there are two important limitations of the current literature. First, there is no clear distinction of the relationships between culture, language and country, with language or country used instead as a proxy for culture.[2] This is troubling, as some Central European cultures are multilingual and span geographical borders. Second, scholarship examining the impact of AI tools on Central European cultures is scarce.[3] Consequently, developing an understanding of this matter requires alternative approaches, which I will discuss later.

Despite these limitations, the aforementioned literature consistently demonstrates that the level of cultural alignment in LLMs is inconsistent.

[1] *I adopt Kate Crawford's (2021, 9) definition of AI. She distinguishes between "artificial intelligence" as a term that encompasses various socio-political aspects, establishing it as a "registry of power", and "machine learning", which she treats as a technical term.*

[2] *The use of country or language as proxies for culture may derive from reliance on social science surveys (for example, EVS/WVS 2024) that benchmark value representation in human populations.*

[3] *One of the few examples is provided by Atari et al. (2023), who offer a visual representation of the data on cultural distance from ChatGPT for Czech republic and Slovakia but do not provide the exact values.*

Since OpenAI's 2022 launch of ChatGPT,[4] the importance of appropriate cultural alignment in LLMs has grown. Individuals and companies employ these tools in increasingly diverse set of use cases. Currently, they range from searching for information, translating texts, writing code, to seeking emotional support. AI tools can seemingly fulfil the same needs in different populations. However, without proper cultural alignment, this appearance is only partially accurate. The need for cultural alignment in LLMs depends on the use case and can be envisioned as a scale. At one end, there are tasks independent of cultural alignment (e.g., proofing code syntax). They can mostly be performed on the basis of mature training data sets. At the other end, there are tasks that interfere with human agency and autonomy (e.g., seeking decision-making guidance), which might return deeply problematic results when the appropriate alignment is missing. For the latter, the common characterisation of cultural alignment is insufficient, and I therefore propose broadening it to include culturally distinct communication and thinking patterns,[5] as these shape how language use influences other non-linguistic cognitive processes.

This text explores the following question: Are AI tools making Central European cultures vulnerable? I will argue that, concerning LLMs, the notion of *collective vulnerability* is relevant to any culture, regardless of its spoken language(s). Collective vulnerability denotes a shared exposure to potentially harmful effects arising from the development and deployment of technology. It is a persistent relational condition because it affects both members of a culture and cultures themselves. Although the exposure to AI tools does not necessarily result in harm in every instance, safe use of them requires ongoing contextual awareness and critical evaluation. For example, if *epistemic cultural flattening* (Iványi-Bitter 2026) were considered a potential outcome, use cases such as coding assistance may be benign, despite highly generalised LLM outputs. Nevertheless, they still demand attentiveness to broader societal, cultural, and ethical implications beyond immediate or localised harms. This will be substantiated later in the text. To develop this argument, I examine profit, technology colonisation, and personalisation as the reasons for the lack of cultural alignment in LLMs that drives collective vulnerability. I then expand the notion of collective vulnerability and propose a taxonomy of cultural vulnerability. I conclude by outlining potential measures for fostering more culturally aligned AI systems and the role of design in this effort.

Before proceeding, two points must be clarified. First, the concept of cultural alignment should be situated within the broader debate on AI alignment, which is often framed by Nick Bostrom's argument that aligning AI with human values may avert a hypothetical existential risk should AI surpass human intelligence (Narayanan and Kapoor 2024; Hao 2025; Bender and Hanna 2025; Hao 2025). In contrast, Shannon Vallor (2024, 149) argues that using value alignment "as a strategy for managing AI risk and making AI more 'ethical' or 'responsible'" is problematic

[4] *The development of technologies leading to generative AI spans approximately eighty years (Narayanan and Kapoor 2024). ChatGPT, based on the GPT-3.5 model, was OpenAI's first release to include a user interface for the general public. Its predecessor, GPT-3, had been available to developers via API since 2020 (Hao 2025). This text focuses on LLMs as a subset of AI, specifically ChatGPT and GPT models, on the premise that impact on culture and humans should be studied through tools commonly used within the population. According to StatCounter, in August 2025, ChatGPT led the AI chatbot market share both globally (80.92%) and locally within the Czech Republic (84.43%).*

[5] *Erin Meyer (2014) identifies eight dimensions (communication, evaluation, persuading, leading, deciding, trusting, disagreeing and scheduling) that she believes diagnose the most important differences between and within cultures that influence cross-cultural management. Atari et al. (2023) note that members of WEIRD (Western, educated, industrialized, rich, and democratic) population often prefer analytical thinking, while less-WEIRD people favour holistic thinking.*

because AI systems reproduce values encoded in their training data. This reinforces human "moral comfort zones" and limits the creation of "new languages of virtue for the next century" (150) that are needed to avoid perpetuating past wrongs. This text advances the view that, even if achieved, alignment does not address more immediate concerns such as resource depletion or the production of power asymmetries. Second, this text does not challenge the assumption in the literature that protecting cultural diversity is important. Nevertheless, it recognises that culture is complex, neither morally neutral nor inherently beneficial, and may under certain conditions reproduce harms or generate new forms of power. A full examination of the ethical and normative frameworks governing cultural alignment in LLMs lies beyond the scope of this study. Nonetheless, examining cultural alignment—as one dimension of AI alignment—remains a worthwhile endeavour.

2. PROFIT, TECHNOLOGY COLONISATION AND PERSONALISATION AS DRIVERS BEHIND COLLECTIVE VULNERABILITY

I advance three claims to ground my argument that a lack of cultural alignment in LLMs produces collective vulnerability and that this notion is relevant to any culture. The claims are:

1. Abundance and productivity have become synonymous with profit;
2. Technology colonisation is a source of epistemic violence;
3. Personalisation through generalisation misses nuances that humans would make.

2.1. Abundance and Productivity Have Become Synonymous with Profit

The claim that most technology companies, including those developing AI, optimise primarily for profit is widely accepted among scholars and technology critics (Yeung 2019b; Zuboff 2019; Albrecht 2025; Hernández-Ramírez 2019; Kubes 2025; Madianou 2025; Perez 2025). Companies promise that AI tools will increase productivity and help solve complex societal challenges of our time through simple technological solutions while sustaining economic growth (Yeung 2019a; Madianou 2025; Harris 2025). Companies often attempt to deliver on these promises "before understanding the actual problems or cultural contexts" (Madianou 2025, 48). Such solutions are typically governed by measurable, profit-driven objectives, which often conflate measures with targets and function as proxies for the problems themselves (Strathern 1996; Espeland and Sauder 2007). Madianou (2025, 139) illustrates this issue through the case of world hunger, where the solution is framed as "reducing the statistics about hunger", thereby making it appear more achievable. These objectives may distort the purpose of the solution. Narayanan and Kapoor (2024, 46) recount the classic example of the British colonial government in India offering rewards for dead cobras to reduce their

[6] *This example is often referred to as the Cobra Effect. Ironically, it is often invoked in corporate discussions to set performance indicators.*

population,[6] which instead incentivised people to breed more cobras for profit and ultimately increased their numbers. This is not to deny the usefulness of technology or its capacity to enhance productivity, especially, in industrialised societies. Rather, it underscores the need for more careful evaluation of the purposes and interests these technologies serve, as well as the potential harms and challenges they may introduce (e.g., misrepresentation of options, distortion of knowledge, inappropriate treatment).

The techno-utopian vision coming from the Silicon Valley promises to eliminate redundant jobs and possibly even a universal basic income. In light of the recent experiences with globalisation of supply chains and automation, these promises raise public anxiety. While the achievements of the past half-century have generated substantial economic growth, the concentration of profits among the top 20% and the stagnation of average workers' wages have produced economic inequality that threatens people's dignity, undermines their livelihoods, and weakens the social fabric (Harris 2025). Karen Yeung (2019a, 12) explains this anxiety in the context of the AI tools:

While contemporary fears of the inevitable redundancy of human workers reflects previous periods of social anxiety associated with earlier waves of automation of manual tasks throughout history, what is distinctive about contemporary debates is the almost limitless domains in which algorithmic systems may be shown to "outperform" humans on a very wide range of tasks across multiple social domains that have previously been understood as requiring human judgement and intelligence.

In an episode of a podcast by Center for Human Technology (Harris 2025), the political philosopher Michael Sandel engages with the question whether democracy can survive if productivity becomes our only goal. Critiquing the techno-utopian narrative, he argues that the purpose of work extends beyond securing a livelihood. Work also enables people to contribute to the common good, gain recognition for their contributions, and participate in social life. He suggests that measuring societal prosperity through GDP growth and material abundance may be misaligned with the conditions necessary for human flourishing.

Assuming AI tools could genuinely bring about maximal productivity and abundance, it is necessary to revise the notion of profit from a societal perspective, beyond its conventional material interpretation. Tanja Kubes (2025, 10), for example, contrasts profit with progress:

"Progress" does not necessarily have to be oriented towards profit and towards whatever is technologically feasible. Progress, interpreted in feminist terms, may also mean appreciating everyone's connections, intra- and interactions and dependencies with everything else and taking responsibility for each other and the world.

Applying a feminist lens—as Kubes does—offers one way to integrate care, relationality and societal responsibility into technology to make it less exploitative, more culturally aligned, and empowering. However, the concept of progress presents its own difficulties. As Vallor (2024, 157)

notes, the tech industry often rejects progress as a meaningful ambition because "[d]emonstrating progress requires measurable evidence of improving the quality of our lives or the condition of our societies". Such evidence is harder to quantify than material growth and is therefore considered unsuitable for business metrics. Vallor proposes an alternative path toward a more equitable future free of the existing injustices: revise the traits that have been seen as virtuous in the past in favour of qualities that better reflect contemporary needs for humans to flourish,[7] and to dismantle barriers between technical and moral expertise. Regardless of the approach, to enable a more equitable technomoral future, profit must not be defined solely in terms of material gain.

2.2. Technology Colonisation Is a Source of Epistemic Violence

The pursuit of profit typically informs organisational business strategies. It is often realised by scaling products and services to the widest possible audience across geographies. However, even in organisations that adhere to human-centred design principles, constraints such as development costs limit the capacity to address diverse populations equitably. Consequently, these systems tend to privilege perspectives most familiar to their creators. For this reason, I discuss technology colonisation next.

Numerous overlapping terms–such as technocolonialism (Madianou 2025, 5), digital colonialism (Kwet 2019; Schneider 2022), data orientalism (Kotliar 2020), data colonialism (Couldry and Mejias 2019) and digital capitalism (Qiu 2016)—describe colonial logics in algorithmic systems and technology more broadly. Rather than extending or conceptually reviewing this terminology, the focus here is on its practical manifestations in LLMs.

Influenced by Madianou's term technocolonialism which she links to the power relations between global North and South in the context of digital humanitarian aid, I use the term technology colonisation more broadly to describe the fact that the majority of tech tools are built in Silicon Valley or by companies that are heavily influenced by its culture.[8] In the Central European context, the U.S. influence shows at least in three ways:

1. Europe imports technology with built-in U.S. cultural values and norms such as individualism, future-orientation, etc. (e.g., Perez 2025; Buyl et al. 2025).

2. The US business philosophy relies on concepts such as libertarianism or meritocracy which shape the goals of technology companies, their marketing strategies, and the use cases they prioritise.

3. LLMs are predominantly trained on English data that are sourced primarily from the Common Crawl dataset (Albrecht 2025; Perez 2025). In September 2025, their crawl returned over 44% of data in English, only about 1% in Czech, and languages such as Cherokee were absent with 0% (Common Crawl 2025).

[7] *The virtues Vallor (2024, 133) observes in world leaders today include "productivity, confidence, resilience, independent thinking, perseverance, passion, and single-minded dedication".*

[8] *The technology advancements, such as AI, created by these companies continue to use practices disproportionally affecting the global South, ranging from extraction of natural resources to exploitation of labour. For a comprehensive account see Crawford (2021) or Madianou (2025).*

In some respects, it seems inevitable that LLMs would reflect particular viewpoints and cultural values, given their intended use (Perez 2025; Buyl et al. 2025). Otherwise, their functionality "would be mostly restricted to objective queries like spell-checking, mathematics and information retrieval" (Perez 2025).

Zuboff (2019) insists that algorithmic systems have produced unprecedented forms and structures of power for which our existing conceptual frameworks are insufficient. This connects with Madianou (2025), who points out that languages carry culture and entire bodies of values. Since language is used by humans to perceive and situate themselves in the world, the "language is one of the most fundamental tools that reproduces power asymmetries" (Madianou 2025, 117). Therefore, both the choice of language and its mode of use are critical.

LLMs process input by segmenting user prompts into groups of characters (tokens) and predicting the most probable subsequent token. Training data establishes the relationships between tokens, informing the model's ability to associate and interpret cultural concepts. For example: "We might expect 'cats' and 'dogs' to be more closely clustered to 'rain' in English-based language models than language models trained on Spanish text" (Perez 2025). Because roughly half of internet content—a substantial portion of training data—is in English (W3Techs 2025), the behaviour of the resulting systems is naturally shaped by this linguistic predominance. These effects are particularly visible in the increasingly popular LLM use cases centred on emotive applications and human self-actualisation, such as therapy, organising life, or finding purpose (Zao-Sanders 2025). Even if models can reply in a specific non-English language, effective therapeutic treatment requires culturally grounded understanding of emotion, which "cannot be achieved by translating code into different languages" (Madianou 2025, 117). Given AI agents' inability to simulate or understand human psychology (Schröder et al. 2025) or attune to culturally specific preferences, relying on them for therapeutic support may impair individuals' capacity to orient themselves in the world.[9]

To ensure profitability, technology companies often prioritise scale and speed over quality and accuracy in product development. Given the complexity and cost involved in creation of datasets suitable for LLM training, developers frequently rely on pre-existing datasets, which may contain biased data (Crawford 2021; Buolamwini 2023; Narayanan and Kapoor 2024) or data unfit for purpose (Narayanan and Kapoor 2024). Competitive pressure may accelerate product release timelines and lead to the omission of important safety testing, as in the case of GPT-4o (Hao 2025), which became known for increased sycophancy (Raskin 2025). Such practices may produce a misalignment between the cultural contexts embedded in technology and those intrinsic to its users. This constitutes a form of epistemic violence, insofar as such incompatibilities can create tension within individuals' mental models[10] and impair their ability to navigate situations they encounter–both

[9] *A study by Harvard evolutionary biologists (Atari et al. 2023) suggests that LLMs' performance on cognitive psychological tasks most resembles that of people from Western, educated, industrialized, rich, and democratic (WEIRD) societies. The resemblance drops quickly for people from other cultural backgrounds.*

[10] *Mental models are psychological representation of knowledge structures. They help people to understand, explain and respond to situations that they encounter. When shared by people within a given culture, they help them collectively navigate and interact within given situations (Liu and Dale 2009, 224).*

personally and within their communities. For example, technology shaped by libertarian principles may steer individuals from welfare-oriented cultures to prioritise themselves, potentially compromising the well-being of their communities.

2.3. Personalisation Through Generalisation Misses Nuances that Humans Would Make

LLMs are algorithmic systems that operate on token sequences rather than meaning. While they rely primarily on statistical probabilities, they incorporate additional algorithmic methods. I will explore two of them—personalisation and generalisation—in more detail.

Blom (2000, 313) notes that personalisation techniques were used to persistently increase "personal relevance to an individual" and often served practical or social functions, such as facilitating work (e.g., bookmarks, automation scripts) or accommodating social needs (e.g., ringtones linked to pleasurable emotions). These have evolved into hyper-personalisation techniques, which are sophisticated methods for leveraging individuals' personal data—often without explicit consent—to provide tailored experiences designed to capture and sustain their attention for monetisation purposes (Yeung 2019b; Cloarec 2020).

In LLMs, personalisation manifests in two essential forms:

1. Explicit features built into the system: AI chatbots leverage user information collected either automatically during previous interactions or through direct input. In principle, this feature ensures more relevant conversations tailored to user's preferences. However, it is often used for the models to establish deep personal relationships with users to maximise their engagement, potentially at the expense of mental health and even human life. The tragic ChatGPT assisted death of 16-year-old Adam Raine is recent example of just how serious a problem this is (*Raines v. OpenAI* 2025).

2. Enablement of hyper-personalised products and services: The LLM models are commonly integrated into other commercial products and services, for example, to customise communication with or for their customers. Global retailers may use hyper-personalisation to build customers' confidence in their purchasing choices through generated contextual information and social proofing (Bannerman 2025). As Curry and Gradecki (2025) demonstrate, the same technology can be quickly turned into an effective tool for propaganda and disinformation campaigns.

Generalisation complements personalisation. Schröder et al. (2025) explain that generalisation is expected when a user prompt resembles patterns encountered during training, but do not necessarily extend to meaning, novel tasks, or scenarios beyond the training distribution. Schröder et al. demonstrate that LLMs cannot reliably simulate human-like responses in new moral contexts because they often miss

nuances. For example, they found out that humans saw setting up traps to catch stray cats as unethical but thought that trapping rats was ethical, while LLMs saw both as equally unethical—a generalisation.

The combination of personalisation and generalisation appears counterintuitive, as people expect LLMs to exhibit a degree of nuance in their outputs. However, such techniques yield only superficial nuance. Moreover, the prospects of training models on individuals' personal data to enable native hyper-personalisation—adapting to beliefs, culture and values—raises significant concerns about echo chambers and the erosion of individual autonomy (Perez 2025). In an interview with *The New Yorker* (Rothman 2025), Jaron Lanier warns of the risk of creating a dissociated society in which individuals experience only the illusion of shared reality with others—whether human or artificial. He adds that society could adapt to such conditions, which would require a collective choice.

Personalisation and generalisation impact cultural alignment in LLMs and can lead to collective vulnerability of any culture. Technology shapes moral beliefs and habits, which themselves are embedded in cultural practices. Personalisation and generalisation are shaping LLMs, which in turn shape users' moral lives. Users' interactions with the technology—as well as interaction of the technology with other technologies—impact how individuals perceive what is good, right and how to "act on those perceptions and understandings" (Danaher and Sætra 2023, 766).

3. ON COLLECTIVE VULNERABILITY

The question I posed at the beginning was: Are AI tools making Central European cultures vulnerable? In short, the answer is yes. All cultures are susceptible to being reduced to the "mean datum of the training data" (Albrecht 2025, 169), because the power structures behind these tools have little incentive to promote cultural diversity. Developing culturally aligned AI tools would likely require greater investments and therefore constrain opportunities for material profit. A more extensive account on collective vulnerability follows.

The vulnerability of individuals vis-à-vis technology can be understood through different lenses. Intuitively, two possible hypotheses, which are not mutually exclusive, are:

- Individuals with limited AI literacy and knowledge are vulnerable because they struggle to understand mechanisms behind personalised AI and the implications for themselves, their communities, or the public good.

- Historically marginalised or discriminated individuals are subject to data and algorithmic biases which are out of their control (Yeung 2019b, 41) and that reinforce existing power dynamics. These individuals may also have limited access to digital technologies.

While both of these are true, in the context of personalisation technology—and LLMs in particular—all individuals become vulnerable. As discussed, each of us may become a potential target of epistemic violence. We can never fully know or control the personal data the technology employs to tailor our experience, nor can we fully grasp the scope of the data on which it has been trained. Moreover, there is no assurance that such training data adequately represents our cultural context.[11]

All cultures are vulnerable. While cultures evolve over time, their natural progression is conditioned by technology imposed by a small number of profit-driven companies. In the extreme, diverse cultures may converge into a single culture with a homogenised set of values and norms. Since this concerns all cultures, it can be understood as a form of collective vulnerability. This vulnerability is persistent, given the near ubiquity of AI tools, and relational, as it shapes interactions between individuals and cultures, individuals within cultures, and between cultures themselves.

Even if we were able to ensure representation of minority languages in LLMs, it remains unclear how representative they would be of the local cultures. To ensure culturally sensitive LLMs that minimise the risk of epistemic violence, their development would require a different foundation that should—at minimum—reflect the following conditions:

- **Ability to account for variance in digital adoption across populations.** Training data sourced online from populations with low digital adoption—such as Sudan, where only 28.7% have internet access—does not ensure appropriate representation of beliefs and values across economic, demographic and social dimensions within a given culture (Perez 2025).

- **Ability to identify and respect individual's cultural membership.** Although language conveys culture, cultural membership cannot be inferred solely from language use, nor can all content in a given language be assumed to belong to a single culture. Some languages span multiple cultural contexts, and some cultures encompass multiple languages—even without accounting for dialectal variation.

- **Ability to reflect culture-specific mental models.** Appropriate decision-making mental models support is important for preserving personal autonomy. Individuals with individualist cultural background tend to prioritise personal goals, whereas those from collectivist cultures emphasise group harmony (Yates and de Oliveira 2016).

- **Ability to adjust to culture- and context-specific values, norms and behaviors.** Cultural orientations towards concepts such as time vary—for instance, American culture is often described as future-oriented, Japanese culture as past-oriented, and Czech culture as intermediate. Yet, norms also differ between online and offline environments.

[11] *It is conceivable that the lack of adequate representation of one's cultural context reflects a temporary market condition that may shift as the costs of model training and inference decline. However, it remains unclear whether alternative training pipelines can address these limitations or what new challenges they may introduce.*

The above list is not exhaustive, but it provides the basis for the classification of cultural vulnerability proposed here. This classification in grounded in two dimensions: digital adoption within a given culture and the representation of relevant language(s) in LLM training data. The three proposed categories are:

- **High digital adoption and high language representation.** English is a representative example. Although it serves as an official language in numerous countries, the predominance of English-language content on the internet does not imply cultural homogeneity. Moreover, a substantial portion of this content is produced by non-native speakers from non-English-speaking cultural contexts.

- **High digital adoption and low language representation.** This category includes non-English-speaking cultures (e.g., Czech) with substantial digital adoption but limited representation in training data.

- **Low digital adoption and low language representation.** This category includes non-English-speaking cultures (e.g., Sudanese) that have low digital adoption. Even if data from these contexts is included in training datasets, it remains highly unrepresentative of the culture.

Post-training methods such as reinforcement learning from human feedback (RLHF), used to improve safety and align models with user values, rarely optimise for cultural alignment. Kirk et al. (2023) offer a possible explanation that RLHF is often conducted by relatively small, demographically homogeneous groups—typically English-speaking individuals aged 25–35, and with a master's degree. According to Rystrøm et al. (2025), this is especially relevant to multi-cultural languages spoken in multiple countries (e.g., Portuguese), where post-training processes may amplify US-centric value biases in model outputs.

It would be possible to formulate a maxim in the style of Kant: Live in such a way that you will a world of cultural homogeneity. Assuming such a world were feasible, it would require careful reflection on what would be lost through the erosion of minority value systems, local decision-making norms, and other forms of cultural difference. The normative question regarding the desirability of such a world must be addressed collectively.

4. USING ARTISTIC AND DESIGN RESEARCH PRACTICES TO HELP EXAMINE THE CZECH CULTURAL ALIGNMENT IN LLMS

As previously noted, the existing literature pays limited attention to LLM cultural alignment in Central Europe. In the Czech-language context there exist examinations of neutrality (Libovický et al. 2019, 2020), an investigation into text analysis in psychological research (Kučera and Mehl 2022), and studies of LLMs across various forms of cultural production (Piorecký and Husárová 2024; Rosa et al. 2022, 2025), but no study of

LLM cultural alignment. Consequently, alternative approaches—such as artistic and design research—provide a valuable means for developing a deeper understanding of this phenomenon.

We will take Albrecht's (2025) project *Artificial Worldviews*, which aims to expose the underlying knowledge and power structures of LLMs, as our example. Working within artistic and investigative design research, he produces knowledge by applying speculative and analytical design methods to systematically collected data. By prompting GPT-3.5 via the API, he maps the model's taxonomies. These are then presented as interactive visualisations that display relationships among entities such as people, objects, and places. Albrecht (178) identifies a substantial disparity in gender representation between the knowledge and power datasets. The top 26 occurrences included 3 women and 23 men in the knowledge dataset, and 10 women, 10 men and 6 non-human entities in the power dataset. Nine individuals overlap (all 3 women that appear in the knowledge dataset and 6 men). Although the two datasets were collected three months apart, the source of this discrepancy remains unclear and may indicate arbitrary intervention by an undisclosed actor.

In seeking resources to investigate the representation of Central European cultural knowledge in ChatGPT's output, I found Albrecht's work instructive.[12] My initial aim was to follow his methodology and extend the study to Czech and English to compare the resulting patterns of knowledge and power dynamics. However, due to resource constraints, I adapted the method to a smaller scale by reusing a subset of the original prompts and focusing on the domain of sports—specifically hockey. This choice had two reasons. First, I anticipated Czech representation in both languages. Second, the term hockey varies by cultural context, notably referring to ice hockey in countries such as the United States and Czech Republic, and to field hockey in places like the United Kingdom. Using GPT-5-mini via API, I issued the same prompt (see figure 1) in both languages,[13] executing each five times.

[12] *This experiment focuses exclusively on a text-to-text model for two reasons. First, it builds on Albrecht's work. Second, a preliminary attempt to prompt Midjourney, a text-to-image model, revealed that conducting a comparable experiment would require an entirely different methodological approach.*

[13] *I solicited help of a professional translator to ensure no shift in meaning would occur.*

FIGURE 1. *Prompts used in the Czech–English experiment.*

CZECH	ENGLISH
System message	System message
Jsi ChatGPT, velký jazykový model, který zná všechno na světě a byl vytrénován na ohromném korpusu textových dat, kolem 570GB datasetů včetně webových stránek, knih a dalších zdrojů.	You are ChatGPT, a Large Language Model that holds knowledge about everything in the world and was trained on a massive corpus of text data, around 570GB of datasets, including web pages, books, and other sources.
User message	User message
Vytvoř tabulku se seznamem nejvýznamnějších lidí v oblasti 'Sport' v oboru 'Hokej'. V tabulce uveď jejich jméno, popis, národnost, druh činnosti, oblast činnosti, obor činnosti, související pojmy a tvoje ohodnocení jejich významnosti na stupnici od 0 do 100. Seznam by měl být tak dlouhý, jak Ti dává smysl. Tabulka bude ve formátu CSV, vhodná pro práci v Excelu.	Create a table with a list of the most important humans in the domain of 'Sport' in the field of 'Hockey'. The table will include their name, description, nationality, type of activity, domain, field, related things, and your rating of their importance on a scale from 0 to 100. The list should be as long as makes the most sense to you. The table will be in CSV format, suitable for use in Excel.

Although the model's responses varied across runs and it was never instructed to produce a specific number of repetitions, the prompts in both languages yielded—coincidentally—191 entries of individuals associated with hockey. I merged these into a single dataset and normalised names and nationalities to enable subsequent analysis. This process proved challenging. Name entries varied not only in the use of diacritics but, more significantly, in spelling, even when referring to the same individual (see figure 2). While diacritical variation could be resolved programmatically, spelling discrepancies across versions of the same name required manual correction.

5 variants for Lord Stanley Lord Stanley, Frederick Arthur Stanley (Lord Stanley), Frederick Stanley (Lord Stanley), Lord Stanley (Frederick Stanley), Lord Stanley of Preston
4 variants for Anatolij Tarasov Anatoli Tarasov, Anatoli Tarašov, Anatolij Tarasov, Anatoly Tarasov
4 variants for Maurice Richard Maurice 'Rocket' Richard, Maurice \Rocket\" Richard", Maurice Richard, Mario 'Rocket' Richard

FIGURE 2. *Select examples of variant spelling of individuals' names.*

A similar challenge arose during the normalisation of nationalities. Some individuals were associated with multiple countries (e.g., Slovakia and Canada; United Kingdom and Canada), while others were affiliated with nationalities that altered over time due to historical developments (e.g., Czechoslovakia, Czech Republic, Czechia; or the Soviet Union, USSR, Russia). To maintain analytical consistency, I assigned each individual the most salient nationality and adopted current country names.

On one hand, normalisation remedies the practical obstacle and enables analysis. On the other hand, as discussed in section 2.2, the choice of language is critical for representing cultural identity and avoiding the reproduction of power asymmetries. Which version of their name would a particular person identify with, and which nationality?

FIGURE 3. *The trend of individuals' nationalities represented in all prompt responses per language.*

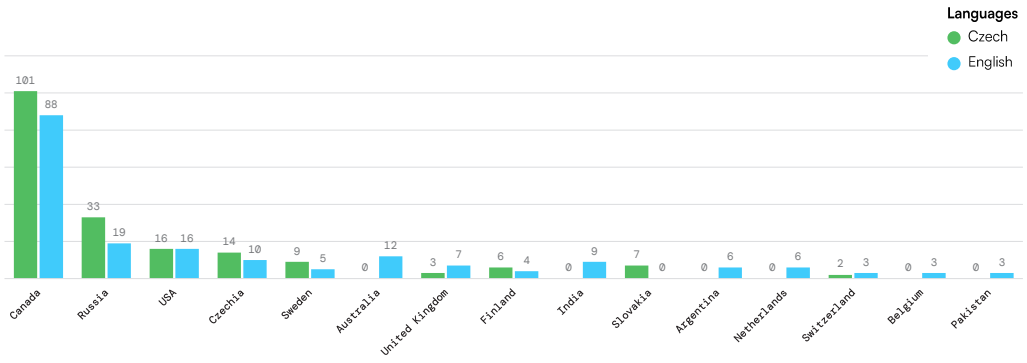


Figure 3 shows that the distribution of the most represented nationalities in the model's output largely aligns across the two languages, although the Czech results contain substantially fewer countries than those in English.

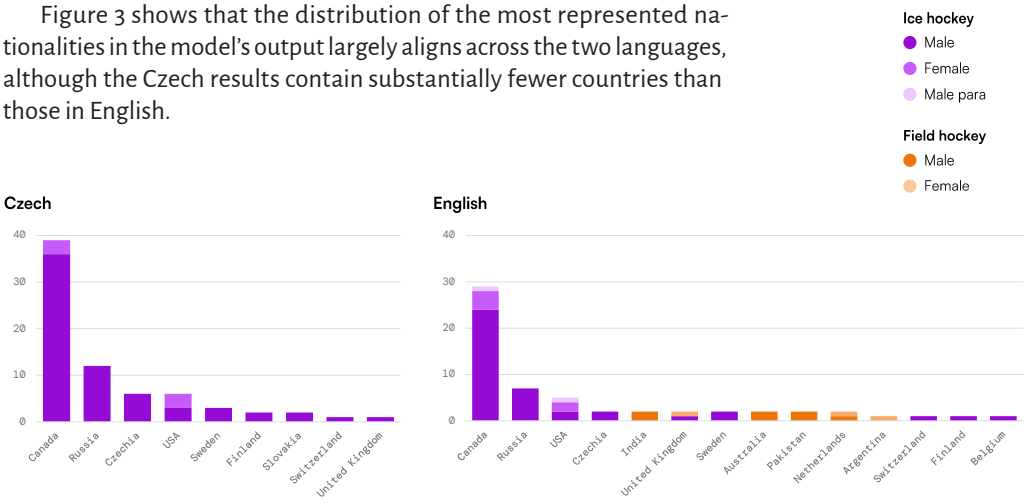


FIGURE 4. *Distribution of hockey type for unique individuals per nationality and language.*

Figure 4 indicates that the Czech data consists predominantly of male individuals associated with ice hockey (92%), with nationalities primarily from North America and Europe. In contrast, the English data includes both male and female individuals linked to ice hockey (80%), field hockey (17%) and para ice hockey (3%), with nationalities spanning most continents except Africa and Antarctica. The broader range of hockey types in the English data accounts for its wider geographic coverage. Interestingly, the Czech data contains more unique individuals, which is counter-intuitive given the greater hockey type variety represented in the English data (figure 5).

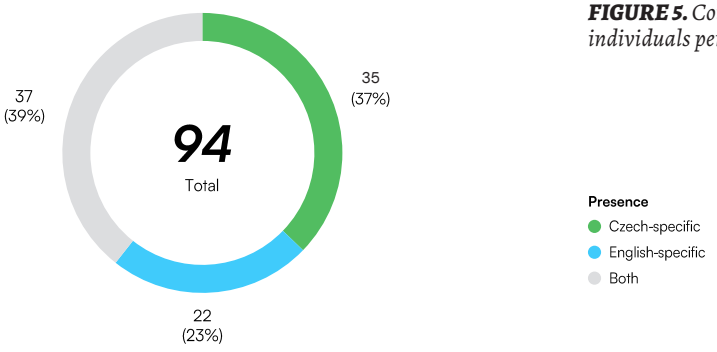


FIGURE 5. *Count of unique individuals per language.*

Despite its limited scope, this experiment raises an important question: Does the LLM's output meaningfully reflect knowledge embedded in Czech- and English-speaking cultures, or does it produce a universalised response? I propose two interpretations. First, the English output encompasses a wider range of hockey variants and countries, a pattern not easily attributable to any single English-speaking culture but rather

to their amalgam. This breadth may flatten cultural specificity into a generalised response, supporting the earlier assertion that even cultures with high digital adoption and high language representation in training data may be vulnerable to misalignment. Second, the nationalities in the Czech output correspond intuitively to the historical development of hockey in the Czech Republic and former Czechoslovakia. This finding aligns with Rystrøm et al.'s (2025) suggestion that monocultural languages may more readily achieve stronger cultural alignment in LLMs.

A common objection to artistic and design research methods is their perceived lack of scientific rigor and replicability. Because LLMs are designed to generate variable outputs, obtaining identical results across interactions is inherently unlikely. Notably, several studies in my literature review—with presumably greater resources—have not yielded substantially more stable or reliable outcomes.[14] Even if relatively low-resource, artistic and design research methods can effectively direct attention to questions that merit further investigation through other scientific approaches.

[14] *The reasons for the selection of specific LLMs and languages as study subjects is not always documented. Atari et al. (2023) provide a compelling argument but do not provide supplementary materials. Others use LLMs to simulate survey respondents without the ability to ensure representative answers (Rystrøm et al. 2025; Schröder et al. 2025; Bravansky et al. 2025; Khan et al. 2025).*

5. PATH TOWARD CULTURALLY ALIGNED PERSONALISED AI

Improving cultural alignment is only one of several factors that can contribute to better personalised AI tools. The considerable challenges of developing culturally aligned large-scale LLMs should be evident by this stage. It is unlikely that a global AI system can fulfil promises of ultimate productivity and abundance. This assessment is grounded primarily in my professional experience as a designer and consultant in a range of international and Czech technology companies, which, in design terminology, may be understood as a form of ethnographic research.

Here are the most important issues that I have observed:

- Oversimplification of complex problems. As discussed earlier, oversimplification is not merely a design failure but a structural feature of the market-oriented product development. Projects designed for global scale are often attractive in boardrooms and strategic documents because they promise access to larger audience and greater revenue potential. However, scalability typically requires reducing complex problems to simplified forms, often prioritising a privileged subset of individuals. Despite efforts to adopt human-centred approaches, solutions are frequently retrofitted to poorly defined problems.

- Prioritisation aligned to business goals. In the best case, development priorities reflect clearly articulated organisational strategies. However, business objectives and key performance indicators (KPIs) rarely incorporate human well-being or externalities. Consequently, product teams often disproportionately optimise for meeting specific KPIs over meaningful real-world outcomes.

- Lack of quality or misunderstanding of design research. Participatory and human-centred design practices rely on research. However, practitioners often fail to distinguish between genuine needs and expressed preferences. Design research should inform well-grounded decisions, rather than serve as a source of convenient evidence.

- Insufficient coordination among policymakers, academics and industry. Many theoretical frameworks (e.g., Kubes 2025) and policy proposals based on human-centred AI and persuasive technologies overlook the practical constraints of product development, limiting their real-world applicability.

So how can we shape culturally aligned technology, especially the LLMs? The following considerations may help guide us toward a better path forward:

- **Train technologists in humanities.** In the Czech Republic, disciplines such as philosophy and ethics have long been undervalued and underfunded, despite their potential to equip technologists with critical tools for reflecting on the solutions they develop. Notably, only one Czech university design program includes an introductory philosophy course. By contrast, such courses are relatively common in engineering programs, although they tend to emphasise historical perspectives.

- **Re-centre human-centred product development on the human.** In addition to strengthening research capabilities to produce high-quality insights for sound decision-making, organisations should evaluate more than profit-driven metrics and measure their impact on people's lives (Vallor 2024; Monteiro 2019). One practical approach, proposed by the Center for Humane Technology, is the use of *anti-KPIs* to minimise harmful consequences by identifying failures to implement corrective measures. "For example, a KPI related to 'engagement' might be paired with an anti-KPI related to 'misinformation' to avoid breaking down reality in the name of growth" (Center for Humane Technology 2022, 41).

- **Foster exchange between academia and industry.** Scholars often remain within disciplinary and institutional boundaries, while industry practitioners rarely engage with academic research or involve scholars in their projects. Addressing complex issues such as cultural alignment requires collaboration across diverse perspectives.

- **Promote active civic participation.** Public deliberation is needed to determine what constitutes a desirable and sustainable society, including the value of cultural diversity. As participants in these technological systems, citizens should critically assess whether the problems being addressed justify the costs borne by individuals and society. Open Science can support this process by directly involving citizens in research.

A potential path forward lies in developing smaller, more sustainable AI systems tailored to specific use cases. Ensuring cultural alignment is more feasible for models designed for discreet populations. Initiatives such as Aya dataset (Cohere Labs 2025), an open science project involving 119 countries, or the recently launched European initiative OpenEuro LLM (2025) represent meaningful steps in this direction.

I will conclude with a quote from Mirca Madianou (2025, 178):

We need to move beyond the binary thinking that we must choose between approaches that either favour powerful structures or human agency. It may seem obvious, but (some) academic fields—and popular discourse, in general—seem to forget that we don't actually have to choose. Structure and agency are co-dependent and cannot be understood in isolation.

REFERENCES

Albrecht, Kim. 2025. "Artificial Worldviews." In *xCoAx 2025: 13th Conference on Computation, Communication, Aesthetics & X*, 167–82. https://doi.org/10.34626/2025_xcoax_006

Atari, Mohammad, Mona Xue, Peter Park, Damián Blasi, and Joseph Henrich. 2023. "Which Humans?" Preprint, OSF, September 22. https://doi.org/10.31234/osf.io/5b26t

Bannerman, Tey. 2025. "I've been part of three startup failures that burned through millions in funding." LinkedIn. https://www.linkedin.com/posts/teybannerman_ive-been-part-of-three-startup-failures-activity-7348255813695680517-p7Ut/

Bender, Emily M., and Alex Hanna. 2025. *The AI Con: How to Fight Big Tech's Hype and Create the Future We Want*. New York: HarperCollins.

Birhane, Abeba, and Marek McGann. 2024. "Large Models of What? Mistaking Engineering Achievements for Human Linguistic Agency." *Language Sciences* 106 (November): 101672. https://doi.org/10.1016/j.langsci.2024.101672.

Blom, Jan. 2000. "Personalization: A Taxonomy." In *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, 313–14. New York: Association for Computing Machinery. https://doi.org/10.1145/633292.633483

Bravansky, Michal, Filip Trhlik, and Fazl Barez. 2025. "Rethinking AI Cultural Alignment." Preprint, version 2. arXiv, March 7. https://doi.org/10.48550/arXiv.2501.07751

Buolamwini, Joy. 2023. *Unmasking AI: My Mission to Protect What Is Human in a World of Machines*. New York: Random House.

Buyl, Maarten, Alexander Rogiers, Sander Noels, Guillaume Bied, et al. 2025. "Large Language Models Reflect the Ideology of Their Creators." Preprint, arXiv, January 30. https://doi.org/10.48550/arXiv.2410.18417

Center for Humane Technology. 2022. "Module 3: Minimizing Harmful Consequences." Foundations of Humane Technology, self-paced online course. https://www.humanetech.com/course

Cloarec, Julien. 2020. "The Personalization–Privacy Paradox in the Attention Economy." *Technological Forecasting and Social Change* 161: 120299. https://doi.org/10.1016/j.techfore.2020.120299

Cohere Labs. 2025. "Aya." https://cohere.com/research/aya

Common Crawl. 2025. "Statistics of Common Crawl Monthly Archives." https://commoncrawl.github.io/cc-crawl-statistics/plots/languages

Couldry, Nick, and Ulises A. Mejias. 2019. "Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject." *Television & New Media* 20 (4): 336–49. https://doi.org/10.1177/1527476418796632

Crawford, Kate. 2021. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press. https://doi.org/10.2307/j.ctv1ghv45t

Curry, Derek, and Jennifer Gradecki. 2025. "Generative Persuasion." In *xCoAx 2025: 13th Conference on Computation, Communication, Aesthetics & X*, 326–30. https://doi.org/10.34626/2025_xcoax_026

Danaher, John, and Henrik Skaug Sætra. 2023. "Mechanisms of Techno-Moral Change: A Taxonomy and Overview." *Ethical Theory and Moral Practice* 26 (5): 763–84. https://doi.org/10.1007/s10677-023-10397-x

Espeland, Wendy Nelson, and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113 (1): 1–40. https://doi.org/10.1086/517897

Hämmerl, Katharina, Bjoern Deiseroth, Patrick Schramowski, Jindřich Libovický, et al. 2023. "Speaking Multiple Languages Affects the Moral Bias of Language Models." In *Findings of the Association for Computational Linguistics: ACL 2023*, edited by Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, 2137–56. Toronto: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-acl.134

Hao, Karen. 2025. *Empire of AI: Dreams and Nightmares in Sam Altman's OpenAI*. London: Penguin.

Harris, Tristan. 2025. "Is AI Productivity Worth Our Humanity?" Your Undivided Attention. June 26. Podcast. https://centerforhumanetechnology.substack.com/p/is-ai-productivity-worth-our-humanity

Hernández-Ramírez, Rodrigo. 2019. "On False Augmented Agency and What Surveillance Capitalism and User-Centered Design Have to Do with It." *Journal of Science and Technology of the Arts* 11 (2): 18–27. https://doi.org/10.7559/CITARJ.V11I2.667

Iványi-Bitter, Brigitta, Tibor Bacsi, and Szilárd Szakács. 2025. "Epistemic Cultural Flattening in Generative Visual AI: Benchmarking Hungarian Heritage and Designing a V4 Path Toward Culturally Aware Text-to-Video." *Disegno* 9 (1): 40–67. https://doi.org/10.21096/disegno_2025_1b-ibtbszsz

Khan, Ariba, Stephen Casper, and Dylan Hadfield-Menell. 2025. "Randomness, Not Representation: The Unreliability of Evaluating Cultural Alignment in LLMs." Preprint, arXiv, April 8. https://doi.org/10.48550/arXiv.2503.08688

Kirk, Hannah Rose, Andrew M. Bean, Bertie Vidgen, Paul Röttger, and Scott A. Hale. 2023. "The Past, Present and Better Future of Feedback Learning in Large Language Models for Subjective Human Preferences and Values." In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, edited by Houda Bouamor, Juan Pino, and Kalika Bali, 2409–30. Singapore: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.148

Kotliar, Dan M. 2020. "Data Orientalism: On the Algorithmic Construction of the Non-Western Other." *Theory and Society* 49 (5): 919–39. https://doi.org/10.1007/s11186-020-09404-2

Kubes, Tanja. 2025. "A Critique of Human-Centered AI: A Plea for a Feminist AI Framework (FAIF)." *AI & SOCIETY*. August 19. https://doi.org/10.1007/s00146-025-02556-8

Kučera, Dalibor, and Matthias R. Mehl. 2022. "Beyond English: Considering Language and Culture in Psychological Text Analysis." *Frontiers in Psychology* 13 (March). https://doi.org/10.3389/fpsyg.2022.819543

Kwet, Michael. 2019. "Digital Colonialism: US Empire and the New Imperialism in the Global South." *Race & Class* 60 (4): 3–26. https://doi.org/10.1177/0306396818823172

Libovický, Jindřich, Rudolf Rosa, and Alexander Fraser. 2019. "How Language-Neutral Is Multilingual BERT?" Preprint, arXiv, November 8. https://doi.org/10.48550/arXiv.1911.03310

Libovický, Jindřich, Rudolf Rosa, and Alexander Fraser. 2020. "On the Language Neutrality of Pre-Trained Multilingual Representations." Preprint, arXiv, September 29. https://doi.org/10.48550/arXiv.2004.05160

Liu, Leigh Anne, and Claudia Dale. 2009. *Using Mental Models to Study Cross-Cultural Interactions*. https://doi.org/10.1057/9780230240834_11

Madianou, Mirca. 2025. *Technocolonialism: When Technology for Good Is Harmful*. Cambridge: Polity.

Meyer, Erin. 2014. *The Culture Map: Breaking Through the Invisible Boundaries of Global Business*. New York: PublicAffairs.

Monteiro, Mike. 2019. *Ruined by Design: How Designers Destroyed the World, and What We Can Do to Fix It*. Mule Design. https://www.ruinedby.design

Narayanan, Arvind, and Sayash Kapoor. 2024. *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference*. Princeton: Princeton University Press.

OpenEuroLLM. 2025. "OpenEuroLLM." https://openeurollm.eu/

Perez, Jorge. 2025. "Tokenising Culture: Causes and Consequences of Cultural Misalignment in Large Language Models." *Ada Lovelace Institute*, June 19. https://www.adalovelaceinstitute.org/blog/cultural-misalignment-llms/

Piorecký, Karel, and Zuzana Husárová. 2024. *The Culture of Neural Networks Synthetic Literature and Art in (Not Only) the Czech and Slovak Context*. Prague: Karolinum. https://doi.org/10.14712/9788024657844

Qiu, Jack Linchuan. 2016. *Goodbye iSlave: A Manifesto for Digital Abolition*. Chicago: University of Illinois. https://doi.org/10.5406/illinois/9780252040627.001.0001

Raines vs. OpenAI. 2025. Superior Court of the State of California for the Country of San Francisco August 26, 2025. https://www.documentcloud.org/documents/26078538-raine-vs-openai/

Raskin, Aza. 2025. "How OpenAI's ChatGPT Guided a Teen to His Death." Your Undivided Attention. September 11. Podcast. https://centerforhumanetechnology.substack.com/p/how-openais-chatgpt-guided-a-teen

Rosa, Rudolf, David Mareček, Tomáš Musil, Michal Chudoba, and Jakub Landsperský. 2025. "EduPo: Progress and Challenges of Automated Analysis and Generation of Czech Poetry." In *Proceedings of the 5th International Conference on Natural Language Processing for Digital Humanities*, edited by Mika Hämäläinen, Emily Öhman, Yuri Bizzoni, So Miyagawa, and Khalid Alnajjar, 524–42. Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.nlp4dh-1.45

Rosa, Rudolf, Patrícia Schmidtová, Alisa Zakhtarenko, et al. 2022. "THEaiTRobot: An Interactive Tool for Generating Theatre Play Scripts." In *Proceedings of the 15th International Conference on Natural Language Generation: System Demonstrations*, edited by Samira Shaikh, Thiago Ferreira, and Amanda Stent, 10–13. Association for Computational Linguistics. https://aclanthology.org/2022.inlg-demos.4/

Rothman, Joshua. 2025. "A.I. Is Coming for Culture." *The New Yorker.* August 25. https://www.newyorker.com/magazine/2025/09/01/ai-is-coming-for-culture

Rystrøm, Jonathan, Hannah Rose Kirk, and Scott Hale. 2025. "Multilingual != Multicultural: Evaluating Gaps Between Multilingual Capabilities and Cultural Alignment in LLMs." Preprint, arXiv, August 30. https://doi.org/10.48550/arXiv.2502.16534

Schneider, Nathan. 2022. "Governable Stacks against Digital Colonialism." *Journal for a Global Sustainable Information Society* 20 (1): 19–36

Schröder, Sarah, Thekla Morgenroth, Ulrike Kuhl, Valerie Vaquet, and Benjamin Paaßen. 2025. "Large Language Models Do Not Simulate Human Psychology." Preprint, arXiv, August 13. https://doi.org/10.48550/arXiv.2508.06950

Strathern, Marilyn. 1996. "From Improvement to Enhancement: An Anthropological Comment on the Audit Culture." *Cambridge Anthropology* 19 (3): 1–21.

Vallor, Shannon. 2024. *The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking*. Oxford: Oxford University Press.

W3Techs. 2025. "Usage Statistics and Market Share of Content Languages for Websites." *November.* https://w3techs.com/technologies/overview/content_language

Yates, J. Frank, and Stephanie de Oliveira. 2016. "Culture and Decision Making." *Organizational Behavior and Human Decision Processes* 136 (September): 106–18. https://doi.org/10.1016/j.obhdp.2016.05.003

Yeung, Karen. 2019a. "Algorithmic Regulation: An Introduction." In *Algorithmic Regulation*, edited by Karen Yeung and Martin Lodge, 1–18. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198838494.003.0001

Yeung, Karen. 2019b. "Why Worry about Decision-Making by Machine?" In *Algorithmic Regulation*, edited by Karen Yeung and Martin Lodge, 21–48. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198838494.003.0002

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* New York: PublicAffairs.

THE PAPRIKA-EFFECT

CENTRAL AND EASTERN EUROPE AS A NOISY LABEL IN AI-GENERATED IMAGES

Anna Keszeg

ABSTRACT

This article examines how AI-generated images reproduce geopolitical imaginaries of Central and Eastern Europe (CEE) through a visual analysis of images generated using Midjourney. Drawing on popular geopolitics as a theoretical framework, the study situates AI image generation within a long-standing transglobal media environment in which visual culture plays a key role in shaping geopolitical knowledge and spatial hierarchies. Popular geopolitics foregrounds the power of everyday visual representations in producing meaning beyond formal discourse, a dynamic intensified by generative AI systems.

Methodologically, the study analyses a dataset of eighty AI-generated images across twenty countries, using standardised prompts varying by gender and dress (folk costume versus contemporary clothing). The analysis focuses on culturally coded visual markers such as facial features, stylisation, and their relation to dress. The article conceptualises the observed pattern as the "paprika-effect": a form of epistemic cultural flattening in which complex regional identities are reduced to exaggerated, globally recognisable, and unevenly documented visual tropes. The findings suggest that generative AI systems reproduce not cultural accuracy, but the contradictions inherent in transglobal geopolitical imaginaries. Rather than offering a fully systematic or generalisable account of AI image generation, this study adopts an exploratory approach. The analysis is intended to function as a hypothesis-generating intervention, identifying patterns that raise broader questions about the relationship between generative AI and geopolitical imaginaries, and it does not provide definitive empirical conclusions.

#Central and Eastern Europe; #popular geopolitics; #Midjourney; #visual culture; #AI image generation, #noisy label

https://doi.org/10.21096/disegno_2025_1ak

THE PAPRIKA-EFFECT: TRANSGLOBAL NARRATIVES OF REPRESENTATION

Reflecting on the exhibition about the dresses of the Habsburg era she curated at the Metropolitan Museum of Art in New York in 1979, Diana Vreeland (2011, 178) wrote these words in her memoir: "Unfortunately Hungarians don't impress the world anymore—they've never been successful, and success is the only thing the world we live in now understands and remembers". Recalling her encounters with Budapest during the preparation of the exhibition, Vreeland contrasts the city she experienced in the 1970s—by then a Soviet satellite—with the Budapest of the early twentieth century, which she recalls as culturally dense, eccentric, and visually excessive.

In her description, Hungarian dandies of the pre-war period dressed in ways that were "absurd", excessive, and unmistakably different from anything else in Europe. Their style, she suggests, was comparable to the taste of *paprika*: overwhelming, too much, bordering on bad taste. Yet for Vreeland, bad taste was always preferable to the absence of taste altogether. Excess, in her view, was not a failure of style but a mark of distinction—an insistence on visibility in a world structured by hierarchies of recognition (Vreeland 2011, 188; 122).

This anecdote demonstrates a key mechanism of global cultural recognition: regions often become legible not through nuance but through intensified and stylised markers that condense difference into recognisable form. Fashion, in this sense, emerges as one of the earliest and most powerful forgers of transglobal imaginaries, translating local specificity into globally circulating visual codes.

As a researcher in fashion studies, I have long been convinced that fashion played a foundational role in shaping the transglobal visual imaginaries of regions, cultures, and identities. Much like contemporary AI training datasets, the early globalised fashion media system (organised around international fashion weeks) functioned as a transglobal representational structure that universalised selective visual norms, predefined the role and place of small cultures, and marginalised cultural specificity in the process (Skov 2011). Generative AI represents a new step in this longer history. Where fashion once mediated regional difference through garments, silhouettes, and taste regimes, AI image-generation systems now mediate difference through datasets, prompts, and algo-

rithmic pattern recognition. Yet the underlying logic remains strikingly similar: cultural visibility is achieved through selection, exaggeration, and simplification.

This article conceptualises this process as the *paprika-effect*. Like paprika as a metaphorical marker of "Eastern European flavour", AI-generated images intensify cultural cues in order to produce images that are immediately legible within a global visual economy. Rather than offering culturally specific or internally differentiated representations, generative AI systems tend to amplify a narrow set of visual tropes, resulting in an *epistemic cultural flattening* (Grossman 2025; Krämer 2023, 12) of regional visual realities.

Prompting an image-generation system with phrases such as "/IMAGINE female/male model in folk dress from [country]; /IMAGINE female/male model in contemporary dress from [country]" produces images that appear coherent at first glance, yet reveal patterned inconsistencies upon closer inspection. Hungary for example is frequently rendered through a generalised Slavic visuality, Romania through an orientalised or far-eastern imaginary, and the Czech Republic through a Mitteleuropean, Western-coded aesthetic. These outcomes suggest that AI-generated images do not draw from a unified understanding of Central and Eastern Europe, but instead mobilise conflicting geopolitical imaginaries embedded in transglobal visual culture.

The paprika-effect, therefore, is not the result of technical error or intentional distortion. It reflects the conditions under which AI image-generation systems operate: uneven cultural visibility, historically sedimented visual hierarchies, and the dominance of popular geopolitical imaginaries. By situating AI-generated imagery within this longer genealogy— from fashion to algorithmic representation—this article argues that generative AI does not disrupt existing visual regimes, but rather extends and intensifies them, translating long-standing geopolitical imaginaries into algorithmic form.

Building on critical discourses concerning the imperial nature of constructed visual imaginaries, and drawing on a dataset assembled in August 2025, this article suggests that in the case of Central and Eastern Europe, AI image-generation systems reveal the difficulty of articulating a coherent Central and Eastern Europeanness, a difficulty deeply rooted in the region's historically layered and ideologically discontinuous geopolitical traditions.

AN AI AS ACCURATE AS WE ARE: DEEP LEARNING, AI IMAGE GENERATION, AND KNOWLEDGE REPRESENTATION

In their widely cited overview of artificial intelligence paradigms, Stuart Russell and Peter Norvig (2021) argue that the multiplicity of discourses surrounding AI can be traced back to four distinct underlying ambitions. These ambitions form a fourfold matrix structured along two conceptual axes: the human versus the rational, and thinking versus acting or

behaving. This framework provides a useful heuristic for understanding how different models of intelligence conceptualise knowledge, cognition, and representation, and it remains particularly relevant for examining contemporary generative AI systems.

According to Russell and Norvig, early approaches to artificial intelligence were largely inspired by the ambition to reproduce human behaviour. Within this paradigm, intelligence is evaluated through externally observable actions rather than internal cognitive processes. The Turing Test exemplifies this orientation, as its primary concern is whether a machine's behaviour can convincingly imitate that of a human. Intelligence, here, is framed as performative resemblance rather than epistemic understanding.

Alongside this behaviour-oriented approach, other strands of AI research have focused on modelling human thinking itself. These approaches attempt to simulate internal cognitive processes such as reasoning, introspection, perception, and imagination. Intelligence is thus understood as a function of internal mental operations, and artificial systems are evaluated based on how closely they approximate the structure of human cognition.

On the opposite side of the matrix, Russell and Norvig identify approaches that abandon the aspiration to reproduce human cognition or behaviour altogether. Instead, these paradigms focus on rationality, defining intelligence as the capacity to make optimal decisions or to act in ways that maximise predefined goals. Some of these approaches emphasise rational thinking through formal logic and symbolic reasoning, while others prioritise rational action, focusing on effective behaviour in specific environments.

Russell and Norvig argue that contemporary AI development is predominantly oriented toward the modelling of rational action. Rather than striving for holistic models of human cognition, current systems tend to be action-specific and task-oriented. This has resulted in highly specialised models optimised for specific functions—such as image generation or pattern recognition—without possessing contextual or cultural understanding of the content they produce.

This framework can be productively connected to Yann LeCun's (2025) distinction between symbolic, logic-based AI (GOFAI – *good old fashioned artificial intelligence*) and data-driven deep learning systems. LeCun, one of the key figures in the development of deep learning, identifies two major traditions in AI research. Early AI, emerging in the mid-twentieth century, was grounded in the assumption that human intelligence operates through logical, rule-based processes that could be formalised and computationally reproduced. Artificial intelligence, from this perspective, was a matter of symbolic manipulation and explicit knowledge representation.

The rise of deep learning in the 2010s marked a significant epistemic shift. Rather than encoding knowledge through rules and symbols, deep learning systems acquire patterns through exposure to large datasets.

Intelligence, in this paradigm, emerges from learning statistical regularities rather than from reasoning about meaning. LeCun characterises contemporary AI as hybrid: combining the goal-oriented rationality of earlier approaches with the pattern-recognition capacities of neural networks.

LeCun links this hybrid model back to a proposal by Alan Turing, who suggested that instead of attempting to simulate the adult human mind, AI research should focus on creating systems analogous to a child's mind—capable of learning through experience. In contemporary machine learning, training datasets effectively replace education, and optimisation replaces understanding. While this approach has proven remarkably effective for task-specific performance, it has profound implications for how cultural knowledge is represented.

A widespread assumption in public and technical discourse holds that generative AI systems are "as accurate as their training datasets." While this claim is often invoked to defend AI outputs, it obscures a crucial epistemological problem. Training datasets are not neutral repositories of knowledge; they are structured by uneven visibility, historical power relations, and dominant cultural narratives. Cultural knowledge, unlike technical or formalised knowledge, is rarely standardised, consistently annotated, or evenly distributed. As a result, generative AI systems tend to reproduce not cultural accuracy, but statistical dominance.

Concerns surrounding the representation of minority cultures and languages—such as ongoing debates about the inadequate modelling of Sámi languages in AI systems (Li 2026)—illustrate this limitation clearly. These cases demonstrate that even if and when datasets are extensive, they may still fail to capture culturally specific epistemologies, leading to misrepresentation, simplification, or erasure, as complex cultural meanings are reduced through processes of tokenisation and vectorisation into abstract, context-insensitive computational representations (Toraman et al. 2023).

In the domain of AI-generated images, this dynamic manifests as *epistemic cultural flattening* (Krämer 2023, 11–12): a process through which complex, historically layered cultural identities are reduced to simplified, globally legible visual tropes. Generative AI systems trained to produce visually plausible outputs rely on statistically dominant patterns rather than culturally situated meanings. Consequently, they tend to reflect inconsistencies and contradictions embedded in transglobal geopolitical imaginaries rather than coherent regional self-understandings.

In machine learning, the concept of *noisy labels* refers to training data in which labels are inaccurate, inconsistent, ambiguous, or contextually unstable (Carneiro 2024). Noisy labels do not necessarily result from error or negligence; rather, they often emerge in domains where categorisation itself is contested, imprecise, or historically layered. When models are trained on such data, they tend to learn distorted or averaged representations, privileging dominant correlations while obscuring internal variation.

Central and Eastern Europe can be understood as a paradigmatic example of a noisy label within transglobal datasets. The term itself does not denote a stable cultural, political, or historical category, but rather a composite designation shaped by shifting borders, competing strategic geopolitical projects, and externally imposed classificatory regimes. As a result, visual and textual data associated with Central and Eastern Europe are marked by semantic inconsistency: the same label may refer to Slavic, Orientalised, Mitteleuropean, post-socialist, or "almost Western" imaginaries, depending on context.

From this perspective, AI image-generation systems are indeed "as accurate as we are." They reproduce the fragmented, uneven, and often contradictory ways in which regions are imagined within global visual culture. The epistemic cultural flattening produced by generative AI thus reflects not a technical failure, but a mirror held up to the cultural and geopolitical imaginaries already embedded in the datasets from which these systems learn.

POPULAR GEOPOLITICS AND TRANSGLOBAL VISUAL CULTURE

Popular geopolitics is a field of research concerned with how geopolitical knowledge, spatial imaginaries, and regional identities are produced and circulated through popular culture rather than through formal political discourse alone. Emerging at the intersection of political geography, cultural studies, and media studies, popular geopolitics shifts attention from state-centric narratives and elite geopolitical strategies to everyday cultural forms such as films, literature, fashion, advertising, and visual media. Within this framework, geopolitical meaning is understood as something that is learned, felt, and normalised through repeated encounters with images, stories, and aesthetic conventions (Saunders and Strukov 2018).

The forging of regional imaginaries in popular culture has been addressed across several disciplinary traditions. Literary studies, cultural history, media studies, film studies and critical geography have all contributed to understanding how regions are invented, exoticised, or normalised through representation. One of the most influential contributions in this regard is Vesna Goldsworthy's (1998) *Inventing Ruritania*, which examines how Western cultural production has historically imagined Eastern Europe as a semi-fictional space of intrigue, backwardness, and excess. Goldsworthy conceptualises this process as an "imperialism of the imagination", through which cultural domination operates not through direct political control, but through representational asymmetry. Eastern Europe, in this account, becomes a canvas onto which Western anxieties, desires, and fantasies are projected.

Goldsworthy's argument highlights a key mechanism of popular geopolitics: the reduction of complex regions to narrative and visual shorthand. Such shorthand enables rapid recognition within transglob-

al media circuits, but it does so at the cost of internal differentiation. This process closely resembles what this article terms *epistemic cultural flattening*, whereby historically layered and heterogeneous regions are rendered legible through a limited set of recurring tropes.

A related but distinct intervention is offered by Maria Todorova (1997) in *Imagining the Balkans*, which traces how the Balkans have been constructed as Europe's internal Other. Todorova introduces the concept of "Balkanism" to describe a representational logic that positions the region as chronically incomplete, irrational, or backward in relation to an imagined European norm. Importantly, Todorova emphasises that such imaginaries are not static; they shift over time while retaining a core structure of hierarchical differentiation. The Balkans, much like Central and Eastern Europe more broadly, function as a liminal space.

These insights resonate strongly with the study of popular geopolitics in the post-socialist context. Robert A. Saunders' (2020) *Popular Geopolitics and Nation Branding in the Post-Soviet Realm* extends the analysis of geopolitical representation into the contemporary media landscape, focusing on how post-Soviet states actively attempt to manage and reshape their international image. Saunders demonstrates that nation branding, cultural diplomacy, and media representation operate within pre-existing geopolitical imaginaries that constrain how regions can be seen. Even when states seek to reposition themselves, they must negotiate inherited symbolic frameworks that structure global perception.

Taken together, these works underscore that regional imaginaries are not simply imposed from above, but are continuously reproduced, negotiated, and modified through popular cultural forms. This is where popular geopolitics intersects with broader theories of popular culture and globalisation. John Storey's (2007) *Inventing Popular Culture* traces how popular culture itself has evolved from localised folklore to a globalised system of cultural production and consumption. Storey emphasises that popular culture is not a stable category, but a dynamic field shaped by power relations, technological change, and transnational circulation . In a globalised media environment, popular culture becomes a primary site where regional difference is negotiated, standardised, and commodified.

Within this transglobal visual culture, regions are increasingly known not through direct experience, but through mediated images that circulate far beyond their original context. Visual literacy, therefore, has always had a transglobal and mediatic character. Long before the emergence of digital platforms or generative AI, fashion, film, illustration, and photography played a central role in forging visual imaginaries of regions and peoples. These imaginaries are learned implicitly, through repetition and familiarity, and they shape expectations about what regions look like and how they should be recognised.

Popular geopolitics provides a critical framework for understanding how these visual imaginaries become normalised. By foregrounding the

everyday, affective, and aesthetic dimensions of geopolitical knowledge, it reveals how seemingly neutral images participate in the reproduction of spatial hierarchies. Regions such as Central and Eastern Europe are particularly susceptible to this process because they occupy an ambiguous position within global imaginaries. They are simultaneously familiar and foreign, European and not-quite-European, central and peripheral.

In this sense, Central and Eastern Europe can be understood as a paradigmatic example of a noisy geopolitical label. Its meaning shifts across historical periods, ideological regimes, and cultural contexts, generating a dense accumulation of partially overlapping and often contradictory representations. Popular geopolitics helps explain how such noise is not resolved, but rather managed through repetition, simplification, and aesthetic convention. Over time, this produces a repertoire of visual cues that stand in for the region.

When generative AI systems draw on datasets shaped by these transglobal visual regimes, they inherit not only specific images but also the geopolitical imaginaries embedded within them. The epistemic flattening observed in AI-generated representations of Central and Eastern Europe thus reflects a longer history of popular geopolitical representation. AI does not invent these imaginaries; it accelerates and recombines them, transforming historically sedimented cultural noise into algorithmically optimised visual outputs.

By situating AI-generated images within the framework of popular geopolitics and transglobal visual culture, this article argues that the inconsistencies observed in AI representations of Central and Eastern Europe are neither random nor purely technical. They are the algorithmic expression of a region whose global visibility has long been structured by competing, externally mediated imaginaries. Popular geopolitics therefore provides a crucial lens for understanding how generative AI participates in the ongoing production of geopolitical knowledge, translating cultural ambiguity into visual form.

THE NOISY LABEL OF CENTRAL AND EASTERN EUROPE

Michael Billig's (1995) concept of banal nationalism describes the everyday, taken-for-granted ways in which national belonging is reproduced through routine symbols, habits, and visual cues. Flags on public buildings, weather maps, linguistic conventions, and media narratives subtly remind citizens of the nation without requiring overt ideological mobilisation. Banal nationalism functions precisely because the nation it reproduces is assumed to be stable, coherent, and self-evident. Its power lies in its invisibility: nationalism becomes effective when it no longer needs to declare itself (Billig 1995; Weber 2021).

A comparable logic has been identified in critical scholarship on the Balkans. Building on the work of Maria Todorova, scholars have described forms of *banal Balkanism* through which the region is routinely framed as Europe's internal Other—backward, excessive, unstable, and perpetually

incomplete. In this sense, Balkanism can become banal insofar as its representational codes are predictable and widely recognisable, even when they are stigmatising (Plantak and Paleviq 2022).

The case of Central and Eastern Europe, however, resists such banalisation. Unlike the nation-state, or even the Balkans as a symbolic category, Central and Eastern Europe has never crystallised into a single, stable imaginary capable of sustaining banal reproduction. There is no banal Central and Eastern Europeanism because the region's history is marked by ideological rupture, geopolitical displacement, and asymmetrical inclusion (Nowak 2022). Rather than being anchored in a continuous narrative, the region has repeatedly been defined through external frameworks and shifting centres of power.

This instability is not only imposed from outside but is also reinforced through processes of self-colonisation (Kiossev 2011). Many Central and Eastern European societies have historically internalised Western evaluative frameworks, adopting external standards of cultural legitimacy, modernity, and Europeanness. These internalised hierarchies shape how the region represents itself and how individual countries position themselves in relation to one another. Self-colonisation thus functions as an internal reproduction of external imaginaries, reinforcing symbolic dependency even in the absence of direct political domination.

Milan Kundera's (2023) essay "A Kidnapped West" provides a particularly influential articulation of this condition. Writing in the context of Cold War Europe, Kundera argued that Central Europe was culturally Western but politically displaced—"kidnapped" by the East and misrecognised by the West. His formulation captures a persistent tension between cultural self-identification and geopolitical classification. Europeanness, in this view, is not a given but a contested status that must be continuously asserted, narrated, and defended.

Kundera's argument helps explain the centrality of the Mitteleuropa imaginary within Central and Eastern Europe (Nowak 2022, 41–43). Mitteleuropa operates as an aspirational framework promising symbolic reintegration into Western cultural lineages—urban modernity, intellectual tradition, aesthetic refinement. Yet this imaginary is unevenly accessible. While some countries, such as the Czech Republic, can be more readily aligned with Mitteleuropean narratives, others remain marginal, contested, or excluded from this symbolic geography.

Historically, the region has been shaped by overlapping imperial and ideological projects. Austro-Hungarian, Ottoman, Russian, Soviet, and Western European influences have all contributed to its symbolic landscape, but none has succeeded in stabilising a coherent regional representation. Instead, these layered histories have produced a fragmented representational field characterised by internal hierarchies and competing narratives.

One of the most persistent of these narratives is the *Slavic* imaginary (Nowak 2022, 37–41). Within this framework, Central and Eastern Europe is visually and culturally coded as Slavic, marked by generalised

phenotypical traits, folk aesthetics, and assumed cultural dispositions. This imaginary functions as a powerful visual shorthand in transglobal media, but it is also exclusionary. Countries such as Hungary and Romania, whose linguistic and historical trajectories do not align with Slavic identity, occupy ambiguous positions within this representational system. They are frequently absorbed into Slavic visual regimes despite their difference, or else rendered anomalous and difficult to place.

This uneven inclusion generates representational tension. Hungary and Romania are geographically situated within Central and Eastern Europe, yet they are not fully integrated into the dominant symbolic frameworks through which the region is imagined. Their exclusion from the Slavic imaginary does not lead to clearer or more accurate representation; instead, it produces representational noise. Visual culture compensates for this ambiguity by drawing on alternative imaginaries—orientalised, Balkanised, or vaguely Eastern—further complicating their symbolic position.

This condition corresponds closely to what Robert A. Saunders (2020, 2) calls a "representational crisis" in the post-socialist region. For Saunders, the post-Soviet region is marked by a persistent inability to stabilise its external image, resulting in a proliferation of competing narratives, branding strategies, and geopolitical framings. This crisis does not stem from a lack of representation, but from an excess of incompatible representations that undermine one another.

The concept of representational crisis is particularly useful for understanding why Central and Eastern Europe functions as a noisy geopolitical label. Rather than converging toward a shared symbolic identity, the region accumulates partially overlapping and contradictory imaginaries—Slavic, Orientalised Eastern, Mitteleuropean, post-socialist—none of which achieves definitive dominance. Self-colonising dynamics further intensify this crisis, as regional actors selectively adopt or reject these imaginaries in pursuit of recognition, legitimacy, or geopolitical alignment.

Because these imaginaries are structurally incompatible, they cannot be banalised in the sense described by Billig. Banal reproduction depends on stability and repetition without friction. In Central and Eastern Europe, repetition produces contradiction rather than coherence. Visual cues clash instead of quietly reaffirming a shared understanding, exposing the instability of the category itself.

This representational crisis has significant implications for contemporary visual culture and for generative AI systems. AI image-generation systems encounter Central and Eastern Europe as a label saturated with historical discontinuity, uneven inclusion, and internalised hierarchies. When forced to resolve this complexity into a single image, they default to dominant or statistically salient imaginaries, producing exaggerated, hybrid, or internally inconsistent representations. These outputs do not simply misrepresent the region; they visualise the representational crisis itself, translating long-standing geopolitical uncertainty into algorithmic form.

RESEARCH OUTCOMES: CONFLICTING IMAGINARIES IN AI-GENERATED FACES

In June 2023, I conducted my first experiments with AI image generation using Midjourney V4. Approaching the platform with the naïveté of a humanities scholar and without the reflexes of an experienced AI user, my initial aim was modest and discipline-specific. I was researching the regional costume of a particular Hungarian region, focusing on nineteenth-century garments and their contemporary adaptations within fashion design. To support this work, I prompted Midjourney to generate images of nineteenth-century Hungarian dresses from the region in question.

What emerged from these early experiments was unsettling. While the garments themselves appeared visually convincing (I had trained Midjourney with archival images), the faces of the figures bore little resemblance to Hungarian historical or contemporary visual self-representations. Instead, they consistently displayed features aligned with a generalised Slavic physiognomy. The dissonance between dress and face was striking. Rather than illustrating Hungarian regional specificity, the images seemed to collapse cultural difference into a broader Eastern European visual type. This moment (figure 1) became the initial trigger for the research developed in this article.

FIGURE 1. *An AI-generated image of a young male figure in folk dress from Tapolca, Hungary (Midjourney V4).*

At the time, I lacked both the technical vocabulary and the methodological tools to fully interpret what I was seeing. Only retrospectively did it become clear that these early outputs already exemplified the dynamics later conceptualised as epistemic cultural flattening and noisy labelling. The system did not "misread" Hungarian culture; it reproduced a statistically dominant geopolitical imaginary in which Hungary was visually absorbed into a Slavic framework.

Three years later, and equipped with greater familiarity with AI image-generation systems and informed by critical scholarship on popular geopolitics and representation, I returned to Midjourney to conduct a systematic study. This second phase of research was designed to move beyond anecdotal observation toward comparative visual analysis. However the methodological approach adopted here is deliberately limited in scope. The study does not aim to approximate a fully systematic prompting protocol or to produce statistically generalisable findings. Instead, it operates as an *analytical probe*, using controlled prompts to surface recurring representational tendencies. As such, the results should be understood as *indicative rather than exhaustive*, and as contributing to the formulation of hypotheses about how generative AI systems engage with culturally unstable or contested categories.

The study focused on twenty countries: Albania, Austria, the Czech Republic, Denmark, Finland, France, Germany, Hungary, Italy, Liechtenstein, Lithuania, Poland, Romania, Russia, Slovakia, Slovenia, Serbia, Turkey, the United Kingdom, and Ukraine. For each country, four images were generated using standardised prompts that varied by gender and dress. The prompts had the same structure for all the countries:

/IMAGINE female model in folk dress from [country]
/IMAGINE male model in folk dress from [country]
/IMAGINE female model in contemporary dress from [country]
/IMAGINE male model in contemporary dress from [country]

This resulted in a total dataset of eighty AI-generated faces. The use of standardised prompts allowed for controlled comparison, ensuring that observed differences could be attributed to representational tendencies rather than prompt variation. All the images were generated in August 2025 using the same Midjourney version (V7—Alpha) and default settings. While exact reproducibility is limited due to the stochastic nature of the system, prompt consistency was maintained across all cases.

This study focuses exclusively on images generated using Midjourney. The decision to rely on a single generative AI system was deliberate and theoretically motivated. The initial aim of the research was not to conduct a comparative evaluation of image-generation platforms, but to test a first hypothesis: whether generative AI systems reproduce conflicting geopolitical imaginaries of Central and Eastern Europe through visual representation. The early exploratory results already indicated the presence of systematic patterns rather than isolated anomalies.

Given that these patterns aligned closely with established theories of popular geopolitics and representation, repeating the experiment across multiple platforms was not considered necessary for demonstrating the structural nature of the problem.

Limiting the study to a single platform also ensured internal consistency. By controlling for technical variation, the analysis could focus on representational tendencies rather than platform-specific affordances. The goal was not statistical generalisation, but qualitative insight into how geopolitical imaginaries become visually encoded within generative systems.

Visual stereotypes were identified through qualitative visual analysis. The analysis focused on recurring phenotypical and aesthetic markers understood as culturally coded signifiers rather than biological traits. These included skin tone, facial structure and proportions, eye shape and colour, hair colour and texture, as well as the overall stylisation of facial features. Additional attention was paid to the interaction between face and costume, including how traditional dress appeared to activate ethnicising visual cues more strongly than contemporary clothing.

Further indicators included facial expression, perceived age, and the degree of stylisation or exaggeration applied to features associated with regional belonging. Patterns were identified through comparison across countries, genders, and dress types, allowing for the detection of consistent visual regimes rather than isolated instances.

The results revealed a clear and consistent pattern: *the ethnicising bias was significantly stronger in images depicting folk dress than in those depicting contemporary clothing*. In folk dress images, faces were more likely to display exaggerated or stereotypical features aligned with dominant geopolitical imaginaries. Contemporary dress, by contrast, tended to produce more neutral, globalised faces that adhered more closely to Western fashion imagery and commercial modelling conventions (figure 2).

FIGURE 2. *Comparative AI-generated images of female models in contemporary dress and folk costume from selected countries included in the study (Midjourney V7 – Alpha).*

Hungary provides a particularly illustrative case. In images generated using the folk dress prompt, Hungarian models were overwhelmingly rendered with features associated with a generalised Slavic imaginary (figure 3). This bias was notably absent—or at least significantly reduced—in images depicting contemporary dress. In those cases, Hungarian models appeared closer to Western European visual norms, suggesting that the ethnicising effect was activated specifically by the invocation of tradition and folklore.

FIGURE 3. *AI-generated female models in Hungarian folk dress and contemporary dress (Midjourney V7 – Alpha).*

Romania exhibited a different, yet equally telling pattern. In both male and female folk dress images, the system frequently mobilised facial features associated with an orientalised or far-eastern imaginary (figure 4). This coding positions Romania symbolically closer to Europe's imagined eastern frontier, echoing long-standing Balkanist and orientalist narratives. Even in contemporary dress, traces of this visual displacement persisted, though they were less pronounced than in the folk costume outputs.

FIGURE 4. *AI-generated female models in Romanian folk dress and contemporary dress (Midjourney V7 – Alpha).*

By contrast, countries such as the Czech Republic and Austria were consistently rendered through a Mitteleuropean visual framework (figure 5). Faces appeared lighter, more familiar, and aligned with Western European aesthetic conventions across both folk and contemporary dress prompts. Here, the invocation of tradition did not trigger the same degree of ethnic exaggeration. Instead, folk dress was integrated into a visual regime that maintained symbolic proximity to the European centre.

FIGURE 5. *AI-generated female models in folk dress (Midjourney V7 – Alpha); left: Austria; right: the Czech Republic.*

These findings support the argument that generative AI systems do not apply bias uniformly across regions or representational modes. Rather, they selectively activate different geopolitical imaginaries depending on contextual cues embedded in prompts. Folk dress functions as a powerful trigger for ethnicisation, encouraging the system to draw on historically sedimented visual stereotypes. Contemporary dress, in contrast, aligns outputs with globalised fashion imagery, dampening regional specificity.

Importantly, these patterns cannot be explained solely by the content of the training data. Instead, they reflect the representational crisis and noisy labelling that characterise Central and Eastern Europe within transglobal visual culture. Countries such as Hungary and Romania, whose historical trajectories do not align neatly with dominant Slavic or Western frameworks, are especially vulnerable to representational displacement. Midjourney resolves this ambiguity by defaulting to the most statistically salient imaginary available.

THE PAPRIKA EFFECT. BANAL CENTRAL AND EASTERN EUROPEANISM?

As always, there is nothing new under the sun. The visual representations produced by generative AI systems do not introduce unprecedented distortions, but reactivate long-standing cultural and geopolitical imaginaries. When Midjourney attempts to flatten Central and Eastern

Europe into a coherent visual category, it encounters the difficulty of representing a region whose history has never been even, continuous, or symbolically stable.

This article has shown that the biases observable in AI-generated images of Central and Eastern Europe coincide with a deeper representational crisis that predates digital technologies. Shaped by shifting borders, imperial legacies, ideological ruptures, and asymmetrical inclusion within Europe, the region has long occupied an unstable position within transglobal visual culture. Its global visibility has been structured by competing imaginaries—Slavic, Orientalised Eastern, Mitteleuropean—none of which has achieved lasting dominance.

Generative AI systems do not resolve this instability; they make it visible. Trained on datasets embedded in these conflicting traditions, AI systems reproduce epistemic cultural flattening by translating geopolitical ambiguity into simplified visual tropes. What appears as algorithmic bias is therefore not a technical failure, but a statistically optimised reflection of historically sedimented imaginaries.

In this sense, AI does not misrepresent Central and Eastern Europe so much as mirror the unresolved tensions that have long defined its image. The challenge lies not only in AI design, but in confronting the cultural and geopolitical conditions that these systems so efficiently expose. The contribution of this article lies not in providing a definitive empirical account, but in making visible a representational problem that remains underexamined. By framing Central and Eastern Europe as a noisy label within AI-generated imagery, the analysis opens a line of inquiry that calls for more systematic, comparative, and technically informed future research.

REFERENCES

Billig, Michael. 1995. *Banal Nationalism*. Los Angeles: Sage.

Carneiro, Gustavo. 2024. *Machine Learning with Noisy Labels: Definitions, Theory, Techniques, and Solutions*. London: Academic Press.

Goldsworthy, Vesna. 1998. *Inventing Ruritania: The Imperialism of the Imagination*. New Haven: Yale University Press.

Grossman, H. 2025. "Epistemic Loss in the Age of Alignment." *Medium*, Nov 3. https://medium.com/@hmgrossman/epistemic-loss-in-the-age-of-alignment-a151fe9f4df6

Kundera, Milan. 2023. *A Kidnapped West: The Tragedy of Central Europe*. Translated by Linda Asher and Edmund White. New York: HarperCollins.

Le Cun, Yann. 2023. *Quand la machine apprend: la révolution des neurones artificiels et de l'apprentissage profond*. Paris: Odile Jacob.

Li, Oliver. 2026. "On Including Sámi-Knowledge in LLMs—See Differences, Accept Differences, Cherish Differences!" *AI & SOCIETY*, January 21. https://doi.org/10.1007/s00146-026-02855-8

Kiossev, Alexander. 2011. 'The Self-Colonizing Metaphor'. In *Atlas of Transformation*. http://monumenttotransformation.org/atlas-of-transformation/html/s/self-colonization/the-self-colonizing-metaphor-alexander-kiossev.html

Krämer, Sybille 2023. "The 'Cultural Technique of Flattening'. An Essay Introducing and at the Same Time Revising an Idea." *METODE*, vol. 1, *Deep Surface*, https://metode.rom.no/media/files/articles/sybille-kraemer.pdf

Nowak, Leszek. 2022. "Eastern Europe, Central Europe, or East Central Europe? Imagined Geography of the Region." *Eastern Journal of European Studies* 13: 33–52. https://doi.org/10.47743/ejes-2022-SI03

Plantak, Martina, and Edina Paleviq. 2022. "'Banal Balkanism?' – Rethinking Banal Nationalism and Regional Identity in the Post-Yugoslav Media Space." *Journal on Ethnopolitics and Minority Issues in Europe* 21(1): 1–22. https://doi.org/10.53779/JPVV3411

Russell, Stuart J., and Peter Norvig. 2021. *Artificial Intelligence: A Modern Approach*. 4th ed. Harlow: Pearson.

Saunders, Robert A., and Vlad Strukov. 2018. *Popular Geopolitics: Plotting an Evolving Interdiscipline*. London: Routledge.

Saunders, Robert A. 2020. *Popular Geopolitics and Nation Branding in the Post-Soviet Realm*. London: Routledge.

Skov, Lise. 2011. "Dreams of Small Nations in a Polycentric Fashion World." *Fashion Theory* 15 (2): 137–56. https://doi.org/10.2752/175174111X12954359478609

Storey, John. 2007. *Inventing Popular Culture: From Folklore to Globalization*. Oxford: Blackwell.

Toraman, Cagri, Eyup Halit Yilmaz, Furkan Şahinuç, and Oguzhan Ozcelik. 2023. "Impact of Tokenization on Language Models: An Analysis for Turkish." *ACM Transactions on Asian and Low-Resource Language Information Processing* 22 (4): 1–21. https://doi.org/10.1145/3578707

Todorova, Maria. 2009. *Imagining the Balkans*. Oxford: Oxford University Press.

Vreeland, Diana. 2011. *D. V.* New York: HarperCollins.

Weber, Regina. 2021. 'Banal Europeanism? Europeanisation of Football and the Enhabitation of a Europeanised Football Fandom'. *Sport in Society* 24 (11): 1893–909. https://doi.org/10.1080/17430437.2021.1893697

NORTH BOHEMIA AS A LOW-RESOURCE VISUAL CONTEXT: EVERYDAY HERITAGE, UNEVEN VISIBILITY, AND SYNTHETIC AESTHETICS

Jiří Philippe Janda

ABSTRACT

Generative AI images are not neutral depictions of place: they translate regions through uneven training data and model priors. Using North Bohemia (Czech Republic) in an "ordinary documentary" 1990s register, this study tests 13 prompts across four models and maps drift with TDS (8 variables, 0–2). Findings indicate structured drift: vernacular loss, infrastructural "cleaning," and stylistic takeover that smooths local memory into globally legible templates—raising questions about everyday heritage visibility in the V4/CEE context.

#Generative AI; #Central and Eastern Europe (CEE); #North Bohemia; #synthetic aesthetics; #cultural bias; #artistic research; #low-resource contexts

https://doi.org/10.21096/disegno_2025_1jpj

INTRODUCTION: THE NON-NEUTRALITY OF SYNTHETIC IMAGES

Generative AI has rapidly become an infrastructure of visual culture, reshaping aesthetic expectations across media, design, and contemporary art. Yet synthetic images are not neutral depictions of the world. As Crawford and Paglen (2021) argue, training datasets and their taxonomies act as epistemic infrastructures: they determine what becomes intelligible to a system and what remains invisible. Accordingly, generated images can be read as cultural data-evidence of which visual hierarchies have become learnable and which forms of visibility remain peripheral (Manovich and Arielli 2024).

Much debate on AI imagery addresses broad forms of bias such as stereotyping or harmful representation, or the reproduction of social asymmetries. This article focuses on a less discussed but equally consequential mode: regional invisibility. In data-sparse ("low-resource") contexts, local signals are weakly represented in the corpora that train mainstream models, and place is often reconstructed through globally legible templates. The resulting drift is often subtle: not caricature, but normalisation, infrastructural sanitisation, typological substitution, and a shift toward aesthetics that read as universally plausible.

Critical AI scholarship has shown that algorithmic systems are never neutral instruments but infrastructures that reproduce asymmetries of visibility and legibility (Crawford 2021; Noble 2018; Benjamin 2019). In the case of generative images, this problem is not limited to explicit stereotyping. It also concerns how models construct place when confronted with visually underrepresented environments. Following Flusser's (1994) understanding of the apparatus, synthetic images do not document the world through direct indexical capture; they project statistically organised visual possibilities conditioned by the program of the system and the distributions of its training data. As Steyerl (2019) suggests, in a different register, contemporary image circulation tends to privilege forms that are mobile, compressible, and globally legible, often at the expense of local noise, ambiguity, and low-status detail. What disappears in this process is not simply visual accuracy, but vernacular density: the minor material cues through which a place becomes culturally specific. From this perspective, regional drift is not an accidental glitch but a structural

effect of how dominant generative infrastructures translate weakly learnable environments into plausible visual defaults (Crawford 2021; Steyerl 2019).

North Bohemia offers a striking diagnostic case. Shaped by post-1945 displacement, socialist industrialisation, and post-1989 transition, the region's visual identity is borne by everyday infrastructures and material traces: housing estates and corridors, mining edges, garage colonies, ad-hoc signage, patched asphalt, and the "noise" of maintenance. These are not incidental details but forms of material inscriptions of shifting regimes, economies, and belonging. For generative systems, however, such complexity is often treated as removable noise. The region is re-encoded through a universal grammar such that it can read as generic "Central Europe", vaguely post-industrial, or tourist-friendly—an operation that resonates across the V4/CEE context.

To examine these translations, I use synthetic aesthetics as an applied authorial framework (developed in dialogue with scholarship on AI visuality and non-indexical image regimes) for reading AI images as artifacts of nonhuman vision rather than as failed photographs (Zylinska 2017). I then introduce the Total Distortion Score (TDS), an eight-variable codebook for comparing recurring forms of regional drift-vernacular loss, infrastructural sanitisation, banality erasure, iconographic substitution, style drift, and geographic confusion-across different systems that were tasked with generating North Bohemian scenes from controlled prompts. The aim is not to measure "truth", but to map how synthetic images distribute cultural legibility in a low-resource regional context, and how these distortions can be critically analysed and, in artistic practice, strategically repurposed.

NORTH BOHEMIA AS A LOW-RESOURCE VISUAL CONTEXT

In AI bias discussions, "low-resource" most often refers to language technologies—minority languages, scarce corpora, and uneven digital infrastructures. For image generation the condition is analogous: images of some places circulate widely through tourism, media industries, institutional archives, and platform economies, while others remain visually sparse, fragmented, or locally bounded. A region can therefore be "low-resource" not because it lacks cultural density, but because its everyday visual signals do not travel well within global data pipelines. In mainstream training corpora, the visual world is unevenly distributed—shaped by attention, marketability, and the infrastructural politics of datasets and their taxonomies (Crawford and Paglen 2021).

North Bohemia is a revealing case within the CEE/V4 context precisely because it is not typically represented as a postcard image. Its characteristic features resist standard aesthetic packaging: patched infrastructures, maintenance traces, industrial edges, and transitional zones between housing estates, extraction landscapes,

and small-town peripheries. Equally important is everyday vernacular density-signage, typographic mixtures, pragmatic visual noise-anchored in local language and informal communication rather than standardised heritage representation. These motifs are frequently under-photographed in global circulation or captioned with generic labels (e.g., "abandoned," "post-industrial," "Eastern Europe"), which detach them from local specificity.

Crucially, the scarcity here is not only quantitative but semiotic. North Bohemia's identity is carried by composite signals: the pragmatics of repairs, garage micro-architectures, heterogeneous signage, and collisions between historical fragments, socialist-era planning, and post-socialist commerce. Many of these cues operate at the level datasets treat as "noise"—legible to local viewers but weakly learnable for models optimised for clarity, symmetry, and canonical objects. Under such conditions, systems tend to "solve" place by defaulting to stronger model priors: cleaner infrastructure, a more universal "European" streetscape, tourist-legible heritage, or a cinematic mood that replaces mundane specificity with atmospheric coherence. The result is representational normalisation: images that remain globally plausible while becoming locally inaccurate, rendered readable according to an external visual grammar.

This is why North Bohemia is a good diagnostic site for synthetic aesthetics. A low-resource region is useful for a stress test to reveal how models negotiate place when local signals are insufficiently learnable: whether they preserve minor cues or substitute globally familiar motifs; whether they maintain ordinariness or stylise coherence; whether they retain vernacular density or "design" the scene according to an idealised template. The following sections treat these translations as both an analytical problem and as controllable material within artistic practice.

This is particularly important in the V4/CEE context, where everyday visual culture has long been shaped by post-socialist transformation, uneven modernisation, and the unstable visibility of ordinary built environments. As Boym (2001) has shown, post-socialist space is often framed through selective nostalgia or aestheticised decay, while Hatherley (2015) demonstrates how socialist and post-socialist urban landscapes are repeatedly re-read through external narratives of failure, backwardness, or retro-modernist fascination. At the same time, scholars of regional visual culture such as Szcześniak (2016) remind us that visibility in the region is structured not only by monuments or official heritage, but by informal signs, commercial improvisation, vernacular repair, and transitional material textures. North Bohemia belongs to this wider field of ordinary post-socialist visuality. What generative models often suppress, therefore, is not merely local detail, but a historically specific regime of everyday legibility: the patched, improvised, and semiotically dense surfaces through which transformation becomes visible.

SYNTHETIC AESTHETICS: A FRAMEWORK FOR READING DISTORTION

This article uses synthetic aesthetics as an authorial heuristic for reading AI-generated images not as failed documents or imperfect photographs, but as outputs of a distinct visual regime shaped by learned model priors: dataset distributions, caption cultures, and platformed image economies. Within this regime, an image is not simply "wrong" when it deviates from the world; it is culturally revealing because it shows how a system translates place into what it can render reliably-privileging coherence, clarity, and recognisable motifs over unstable local signals.

In low-resource contexts such as North Bohemia, this synthetic regime becomes especially legible. When regional cues are weakly learnable, models tend to stabilise scenes by importing stronger, globally familiar patterns-cleaner infrastructures, generic "European" streetscapes, postcard heritage cues, or mood-driven stylisation. Rather than treating these outcomes as random "hallucinations," I approach them as structured distortions: repeatable transformations that indicate where locality becomes unreadable and how the system compensates. Synthetic Aesthetics therefore provides a practical lens for asking what kinds of place models are capable of producing, and under what conditions they replace local specificity with default model priors.

To keep the analysis critical without reducing it to a catalogue of errors, I employ two operational concepts to navigate the tension between human intention and machine agency. Synthetic gesture represents the aesthetic trace of negotiation where the algorithm reveals its non-human logic of translation. Rather than a simple technical mistake, a gesture manifests as a visible resistance to conventional perfection—a patterned move such as cleaning away banal detail, substituting vernacular cues with generic icons, or shifting the register toward a dominant aesthetic. These gestures expose the model's preferred shortcuts and its "alien" logic of translation.

Synthetic resonance describes a state of "attunement" between authorial intention and computational priors. It is a sustained feedback loop where the machine's response often forces the author to accept a computational deviation as a new aesthetic rule. In this sense, resonance is a practical mode of co-production—an emergent dialogue where the system's alien cognition expands the author's imaginative horizon beyond habitual representational expectations, steering the output toward a coherent, if distorted, result.

Together, these terms support synthetic aesthetics as an applied frame, marking a shift from the paradigm of representation to that of simulation. AI images are treated as culturally conditioned translations where distortion becomes not only evidence of uneven cultural visibility

but also a generative resource—material for authorial development in a V4/CEE context.

Method (Compact Protocol): Prompt Corpus and the TDS Codebook

To examine how generative systems translate a low-resource regional context into visually legible outputs, I use a compact comparative protocol combining controlled prompting and qualitative coding. I created a corpus of thirteen place-anchored prompts describing everyday North Bohemian micro-environments (e.g., small-town centres, infrastructural edges, garage colonies, post-industrial peripheries) in an "ordinary documentary" 1990s register, deliberately emphasising mundane material cues rather than iconic landmarks. For each of four model families, I generated two outputs per prompt (n = 26 images per model), producing comparable synthetic depictions of the same regional descriptors.

The selection of the four evaluated model families (Midjourney v7, FLUX.2 Max, GPT Image 1.5, and Nano Banana Pro) was guided by three considerations. First, these systems are among the most visible and widely used contemporary text-to-image platforms shaping mainstream visual circulation, making them relevant for examining how low-resource visual contexts are translated within dominant generative infrastructures. Second, all four were tested as base models, without regional fine-tuning, custom training, or LoRA adaptation, in order to isolate their default representational priors and evaluate how such out-of-the-box systems handle culturally underrepresented environments. Third, the selection also reflects the practice-based dimension of the research, as these are the principal systems used in the author's artistic workflow; this allows the analytical findings to remain directly connected to subsequent authorial experimentation and worldbuilding. Non-Western platforms, including major Chinese text-to-image systems, would offer a valuable comparative perspective on alternative representational defaults, but they fall outside the scope of the present study, which is focused on the Western model ecosystems most relevant to mainstream visual production in the V4/CEE context.

The use of pared-down prompts was a deliberate methodological choice aimed at isolating the models' default representational priors. More descriptive prompts would risk over-determining the output: once material details, architectural typologies, and object relations are explicitly specified, it becomes difficult to distinguish the model's learned visual tendencies from the author's textual steering. In addition, complex prompt engineering is often platform-specific, which would reduce comparability across model families. A minimal and standardised prompt structure therefore provided a consistent baseline for testing all four systems under the same conditions. The aim of the experiment was not to generate the most visually resolved or locally accurate image, but to examine how the models fill in missing information when confronted with visually underrepresented environments. In this sense, the prompts function diagnostically: they expose default assumptions, recurrent

substitutions, and patterned forms of regional drift, which can then be compared across models through the TDS framework.

For each prompt, two independent outputs were generated per model. This redundancy served as a basic internal check, allowing the analysis to distinguish recurrent representational tendencies from accidental anomalies or isolated generative glitches. All generated images were included in the broader TDS coding process; however, only one image per model is reproduced in the case-study figures, both for reasons of space and to preserve visual comparability across the panel. The image selected for publication was not chosen for aesthetic quality or extremity, but for typicality: it was the output that most clearly reflected the stable, recurrent pattern visible across the pair, while remaining closest to the requested ordinary documentary register. Images that appeared as clear outliers, unusually exaggerated failures, or singular artifacts were not used as figure exemplars, in order to keep the comparison analytically fair and representative of the model's default response.

All images were coded by the author using the total distortion score (TDS)—an eight-variable codebook designed to capture recurring forms of regional drift. Variables track distinct but often co-occurring operations (e.g., iconographic substitution, tourist aestheticisation, infrastructural sanitisation, vernacular loss) as well as broader shifts (geographic confusion, localisation loss, style drift away from the intended documentary register). Each variable was coded on an ordinal 0–2 scale (0 = none, 1 = moderate, 2 = strong), and the TDS was computed as the sum of all eight variables:

TDS = ICON_SUB + TOUR_AESTH + SANIT_INFRA + VERNAC_LOSS + BANAL_ERASE + GEO_CONF + LOC_LOSS + STYLE_DRIFT

The protocol does not claim a universal metric of "accuracy." Instead, TDS functions as a diagnostic lens that makes systematic patterns of representational drift comparable across systems and readable at the level of cultural visibility—identifying where local signals become illegible and which defaults replace them. For figures and case studies, representative outputs were selected as typical examples of each model's response to a given prompt (rather than best-case extremes).

CODEBOOK SUMMARY (TDS VARIABLES)

- ICON_SUB — Iconographic substitution: local cues replaced by globally recognisable motifs or generic "European" tokens.
- TOUR_AESTH — Tourist aestheticisation: scene becomes postcard-like, beautified, heritage-polished, overly picturesque.
- SANIT_INFRA — Sanitised infrastructure: technical grit reduced (patches, stains, repairs, cables, maintenance traces).
- VERNAC_LOSS — Vernacular loss: local visual language disappears (signage, typography, language traces, informal ads, visual noise).

- BANAL_ERASE — Banality erasure: everyday clutter and small objects removed; space reads as over-designed or idealised.
- GEO_CONF — Geographic confusion: architecture/landscape drifts toward another region's typologies or urban texture.
- LOC_LOSS — Localisation loss: explicit local specification does not meaningfully shape the output.
- STYLE_DRIFT — Style drift: strong stylisation violates the "ordinary documentary" register (cinematic, hyper-designed, etc.).

If a variable was not applicable (e.g., VERNAC_LOSS in images without text), it was marked as NA and treated as 0 in the TDS sum to preserve comparability.

FINDINGS: COMPARATIVE PATTERNS OF REGIONAL DRIFT

Model	ICON_ SUB	TOUR_ AESTH	SANIT_ INFRA	VERNAC_ LOSS	BANAL_ ERASE	GEO_ CONF	LOC_ LOSS	STYLE_ DRIFT	**TDS**
Midjourney v7	0,85	1,19	0,58	0,95	1,19	0,65	0,82	1,65	**7,50**
FLUX.2 Max	0,38	0,62	0,38	0,48	0,50	0,00	0,65	0,92	**3,85**
GPT Image 1.5	0,23	0,31	0,27	0,15	0,38	0,00	0,31	0,77	**2,38**
Nano Banana Pro	0,12	0,46	0,27	0,35	0,42	0,00	0,12	0,73	**2,38**

FIGURE 1. *Mean TDS variables and total TDS across tested models (n = 26 images per model).*

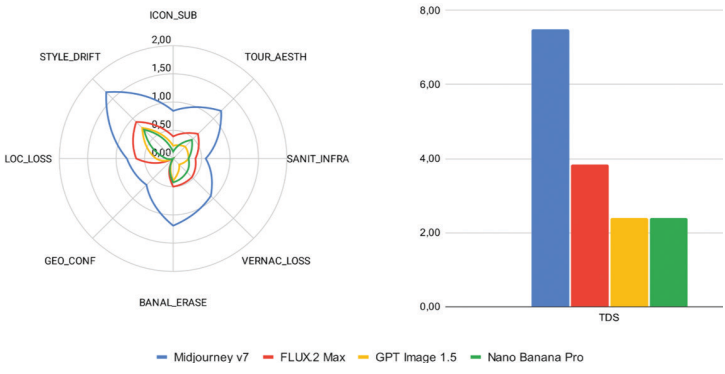


FIGURE 2. *Charts with Comparative Analysis of Regional Drift. Left: distortion profiles of tested models (radar chart showing variable distribution on a 0–2 scale). Right: aggregate regional drift (bar chart demonstrating total TDS values).*

Across the four tested model families, mean TDS values reveal a stable gradient of regional drift (figure 1). This divergence is visually synthesised in figure 2, which illustrates both the specific distortion profiles of each system (left) and their aggregate deviation from the documentary baseline (right).

Midjourney v7 exhibits the highest overall distortion (TDS 7.50), driven primarily by a pronounced style drift and elevated levels of tourist aestheticisation and banality erasure (Chart 1a, 1b). As shown in the radar profile (Chart 1a), its output is characterised by a systematic shift toward "art-photography" registers that override situated ordinariness. In contrast, FLUX.2 Max occupies an intermediate position (TDS 3.85), maintaining strong material plausibility while showing moderate tendencies toward infrastructural smoothing. GPT Image 1.5 and Nano Banana Pro remain closest to the intended "ordinary documentary" register across most variables, with total scores of 2.38 and 2.38 respectively (figure 2, right)

These differences are not simply a question of whether a model "knows" North Bohemia as a named location. They emerge in how systems negotiate legibility under low-resource conditions: whether mundane signals (wear, repairs, infrastructural noise, vernacular signage) are preserved or translated into globally familiar templates through beautification, typological simplification, and stylistic coherence. The five case studies below ground the aggregate scores in comparable visual evidence, using one representative output per model for each prompt.

P01 - Housing estate (courtyard): Everyday ordinariness as a stress test

P01 tests whether models can sustain ordinary documentary credibility under low-status domestic cues—laundry, a bench, a metal playground and the "material noise" of a lived courtyard. The key divergence is whether this banality is preserved as texture and specificity, or translated into a cleaner, more aesthetically resolved scene.

FIGURE 3. *Housing estate courtyard (P01). Prompt: A panel housing estate in Ústí nad Labem, North Bohemia, an inner courtyard with laundry, a bench, and metal playground, late afternoon, ordinary documentary style, 1990s. Panels are ordered consistently across figures (top-left: FLUX.2 Max; top-right: GPT Image 1.5; bottom-left: Midjourney v7; bottom-right: Nano Banana Pro.)*

Nano Banana Pro remains closest to the intended documentary baseline, sustaining dense incidental detail and materially plausible clutter that reads as specific rather than designed. GPT Image 1.5 is similarly credible, mainly through an "inhabited" feel and restrained stylisation. FLUX.2 Max shows a clearer tendency toward infrastructural smoothing: fewer small traces of wear and maintenance, reduced clutter, and a calmer, more pleasant atmosphere consistent with BANAL_ERASE and SANIT_INFRA. Midjourney v7 departs most strongly from ordinariness, pushing the prompt into a postcard-like register (TOUR_AESTH) with pronounced stylisation (STYLE_DRIFT) and noticeable typological drift (GEO_CONF), where the housing estate reads less like a specific North Bohemian courtyard and more like some generalised elsewhere.

P03 – Small town centre: Vernacular and linguistic landscape as a stress test

P03 foregrounds a key vulnerability of global generative models in low-resource contexts: the inability to render a region's linguistic landscape as a meaningful cultural layer. Here, locality is carried less by architecture than by shopfront logic, mixed signage fonts, and small semiotic anchors that situate a town within the post-socialist V4 visual economy.

FIGURE 4. *Small town centre (P03). Prompt: A small town centre in North Bohemia, low-rise buildings, shop windows, mixed signage fonts, ordinary documentary, 1990s. (Top-left: FLUX.2 Max; top-right: GPT Image 1.5; bottom-left: Midjourney v7; bottom-right: Nano Banana Pro.)*

Midjourney v7 collapses this layer almost entirely, replacing signage with nonsensical pseudo-lettering, which functions less as a minor glitch than as systematic VERNAC_LOSS and weakened semantic anchoring. By contrast, GPT Image 1.5 and FLUX.2 Max show a higher capacity for cultural localisation. GPT's output is especially convincing because it retains ordinariness while introducing plausible period anchors (e.g., vehicle typologies consistent with 1990s Czech streetscapes), helping

counter LOC_LOSS. Nano Banana Pro performs slightly below GPT but remains highly credible: signage reads plausibly, people appear period-appropriate, and the scene stays within the intended documentary register, with only mild surface simplification (reduced patina) that does not substantially distort locality.

FIGURE 5. *Garage colony (P05). Prompt: A row of garages on the edge of a North Bohemian town, small handwritten markings, scattered everyday clutter, ordinary documentary-style. (Top-left: FLUX.2 Max; top-right: GPT Image 1.5; bottom-left: Midjourney v7; bottom-right: Nano Banana Pro.)*

P05 - Garage colony: Vernacular density and post-socialist materiality

P05 is a diagnostic prompt for vernacular density and post-socialist everyday materiality: handwritten property markings, improvised signage, scattered clutter, and the characteristic layering of garage rows against a housing-estate horizon. The key question is whether models preserve low-status details as culturally legible texture or translate them into generic, aesthetically "resolved" space.

Nano Banana Pro delivers unusually strong localisation: Czech signage is fully readable and contextually plausible (e.g., "GARÁŽ 12, ZÁKAZ VJEZDU," "P. NOVÁK"), matching both pragmatic content and typographic conventions of private property labelling in Czechia. Together with regionally legible material cues (red brick, stacked tires, wooden pallets) and a panel-housing skyline, the scene holds as an exceptionally specific instance of North Bohemian banality. FLUX.2 Max and GPT Image 1.5 also sustain high documentary credibility, though with different strengths: FLUX retains fine-grained infrastructural traces (puddles, taped numbers, small maintenance cues), while GPT reads as a sociological vignette through the muted palette, overcast light, and peripheral unkempt clutter. Midjourney v7 shows the strongest drift—ICON_SUB, BANAL_ERASE and a combination of TOUR_AESTH and STYLE_DRIFT—with typologies

shifting toward foreign "shed" forms and disorder reduced into a stylised set that remains globally legible but locally less truthful.

P07 - Bus stop: Everyday infrastructure as a carrier of regional identity

P07 tests whether models can localise an ordinary infrastructural scene without landmark cues: worn shelter surfaces, timetable typography, routine waiting, and an industrial horizon typical of North Bohemian outskirts. Divergence here is measuring in terms of the extent to which transit objects and vernacular signage remain culturally specific or dissolve into generic atmosphere.

FIGURE 6. *Bus stop on the outskirts (P07). Prompt: A bus stop on the outskirts of a North Bohemian city, people waiting, a worn shelter, a timetable, ordinary waiting, amateur photography, 1990s. (Top-left: FLUX.2 Max; top-right: GPT Image 1.5; bottom-left: Midjourney v7; bottom-right: Nano Banana Pro.)*

Nano Banana Pro produces a strikingly localised scene in which transit details operate as regional shorthand: route text (e.g., "LIBEREC – MOST") and the ŠKODA branding on the bus, paired with industrial chimneys, situate the stop within a recognisable post-industrial periphery. Although some text is unstable (e.g., a misspelled "timetable"), overall typographic logic and period cues sustain documentary plausibility, with only mild ICON_SUB and VERNAC_LOSS. FLUX.2 Max performs as a positive linguistic outlier, rendering Czech signage ("Autobusová zastávka," "Jízdní řád") with high fidelity and pairing it with materially specific cues such as rust, worn surfaces, and period-appropriate clothing. GPT Image 1.5 remains archetypally accurate, preserving banal details (e.g., a trash bin with a black bag) and recognisable transit iconography without drifting into stylisation. Midjourney v7, by contrast, pushes the prompt into pronounced aesthetic exile: an art-photography register and symbolic staging override ordinariness, amplifying TOUR_AESTH and STYLE_DRIFT, and co-occurring losses (including VERNAC_LOSS and BANAL_ERASE) as locality collapses into a generic mood.

P13 - Rural village: Typology and rural vernacular as a stress test

P13 tests whether models can render a rural location in North Bohemia more specifically than as a generic "European village". The key signals are typological (Sudeten/Podstávkový forms, street grain) and semiotic (vernacular shop signs), anchored by the 1990s landscape with vehicles.

FIGURE 7. *Rural village typologies (P13). Prompt: A small rural village in North Bohemia, Sudeten borderland, ordinary documentary style, amateur photography, 1990s. (Top-left: FLUX.2 Max; top-right: GPT Image 1.5; bottom-left: Midjourney v7; bottom-right: Nano Banana Pro.)*

FLUX.2 Max remains materially credible—cracked asphalt, plaster facades, and infrastructural wear read convincingly yet the overall village typology is relatively generic, with only partial local anchoring. GPT Image 1.5 performs strongly through regional archetypes: a church silhouette and stacked firewood along the street are plausible rural cues, supported by period-evocative cars. Nano Banana Pro delivers the strongest localisation in this set, combining readable Czech signage ("POTRAVINY," "HOSTINEC") with dense period car presence (Škoda-era silhouettes) and typological specificity associated with North Bohemia, including timbered elements. Midjourney v7 again shifts into stylised black-and-white art photography (STYLE_DRIFT) and pronounced geographic drift (GEO_CONF): built forms read closer to Scandinavian rural imagery or a generalised rust belt atmosphere, sacrificing regional identity for aesthetic coherence.

Findings synthesis

Across the five case studies, regional drift appears less as random error than as a set of repeatable operations that redistribute cultural legibility. First, models struggle most where locality is carried by vernacular density-typography, shopfront language, and informal property markings resulting in unstable or genericised text (most visible in P03 and, in

contrast, sharply anchored in P05 and P13 when Czech signage becomes legible). Second, North Bohemian specificity is frequently reduced through cleaning: infrastructural wear, maintenance traces, and banal clutter are attenuated, producing smoother and more universally palatable scenes (P01, P05). Third, the strongest distortions take the form of aesthetic takeover STYLE_DRIFT coupled with TOUR_AESTH and occasional GEO_CONF where an art-photography or postcard register overrides situated ordinariness (recurrently in Midjourney v7 across P01, P07, and P13).

At the same time, the results suggest that geo-localisation failure is not the dominant problem. Even without heavy prompt emphasis, models often maintain a broadly plausible regional envelope—hilly terrain cues, post-industrial horizons, and a "Central European" built texture, while slipping on finer-grained semiotic and material signals. In other words, misrepresentation frequently occurs in micro-features rather than as continent-scale displacement: what fails is not the idea of "North Bohemia," but the specific vernacular and infrastructural detail that makes it locally legible. This aligns with the aggregate TDS gradient: Midjourney's higher scores reflect systematic stylisation and substitution, FLUX.2 Max retains strong material plausibility with weaker semiotic anchoring, and GPT Image 1.5 and Nano Banana Pro remain closest to the intended documentary register. A practical implication is that increasing local data density—through targeted dataset enrichment or fine-tuning—would likely improve the stability of vernacular cues and typological specificity, strengthening regional anchoring without requiring stronger landmark-based prompting.

The findings suggest that in a low-resource regional context, generative systems do not simply "misrepresent" place—they reorganise it according to what is legible at scale. This has direct implications for cultural heritage in the V4/CEE region, where heritage is often embedded in everyday infrastructures and post-socialist vernaculars rather than in internationally iconic monuments. If patched asphalt, improvised garage colonies, mixed signage, and small property markings function as carriers of regional memory, then their systematic attenuation is more than an aesthetic shift: it is a form of cultural smoothing that privileges universally readable templates over situated material histories.

The issue is therefore not only data scarcity but the distribution of learnability. Features that are abundant, canonised, and consistently labelled in dominant image economies (clean streetscapes, tourist-friendly facades, generic "European" town centres) are reproduced with high confidence, while features that are locally meaningful yet semiotically dense (vernacular typography, informal signage, maintenance traces, low-status infrastructures) are treated as noise or replaced by approximations. This produces a politics of visibility: what can be convincingly generated becomes what counts as representable, and what is repeatedly cleaned, simplified, or stylised risks being rendered culturally peripheral. In post-socialist environments shaped by rupture and industrial

transformation, this matters because continuity is frequently carried by precisely those unglamorous textures synthetic images tend to erase.

At the same time, regional legibility is not an all-or-nothing property. Some systems intermittently recover local anchors—vehicles, transit objects, Czech signage, rural typologies—indicating that fragments of post-socialist everyday aesthetics do exist in training data even when the region is underrepresented. Yet these anchors appear unevenly and opportunistically rather than as a stable capacity to render North Bohemia as a coherent visual ecology. This unevenness is itself symptomatic of low-resource visibility: place emerges as a patchwork of learnable tokens rather than a continuous cultural landscape. For V4/CEE cultural discourse, the critical question is not whether synthetic images can depict the region, but which version of the region they stabilise when circulated as plausible representation—and which local signals are systematically filtered out as the price of global readability.

Limitations and Scope

This article proposes using TDS as a diagnostic lens rather than a universal measure of "accuracy." Coding inevitably involves interpretive judgement—especially for variables such as STYLE_DRIFT, TOUR_AESTH, or typological plausibility—and results may vary with different coders or alternate prompt phrasing. Prompt specificity is another constraint: more detailed prompts or stronger reference conditioning could plausibly increase localisation and reduce distortion. Here, prompts were intentionally pared down to keep the setup comparable across models and to function as a stress test of how systems negotiate regional legibility under minimal guidance. The corpus is therefore intentionally specific (North Bohemia; a 1990s "ordinary documentary" register), and the patterns identified should be read as region- and register-dependent rather than globally generalisable. Finally, generative systems evolve rapidly; while the comparative gradient observed here is robust within the tested setup, absolute scores may shift with model updates, new safety layers, or changes in default rendering aesthetics.

ARTISTIC IMPLICATIONS: FROM DIAGNOSTIC DRIFT TO AUTHORIAL DEVELOPMENT

While the previous sections treat distortion as a comparative pattern of regional (il)legibility, the same patterns can also be approached as usable artistic signals. In my practice, Synthetic Aesthetics functions not only as critique but as a method for locating the point at which a model's defaults begin to override situated ordinariness—and for deciding when to resist that drift and when to work with it. The case studies suggest two creative thresholds. In some instances, locality can be strengthened by stabilising vernacular anchors (readable Czech signage, period vehicle typologies, infrastructural wear), allowing an image to operate as a plausible reconstruction of North Bohemian

everyday heritage. In others—most visibly when stylistic takeover or typological substitution intensifies—the image produces a productive rupture: a non-localised translation that exposes the model's priors and opens space for authorial intervention.

This is where the concepts of synthetic gesture and synthetic resonance become practically relevant. A synthetic gesture is a repeatable operation—the cleaning away of banality, substituting of typologies, elevating a scene into a tourist-friendly register, or shifting it into an art-photography mode—that signals how a model prefers legibility. Synthetic resonance describes the iterative feedback loop in which the author responds to these gestures through constraints, selection, re-contextualisation, and montage until the outputs begin to function within a coherent artistic logic. In this sense, the analytical findings do not remain external to artistic work: they become a steering mechanism for distinguishing between distortions that erase regional memory and distortions that can be repurposed to articulate it differently.

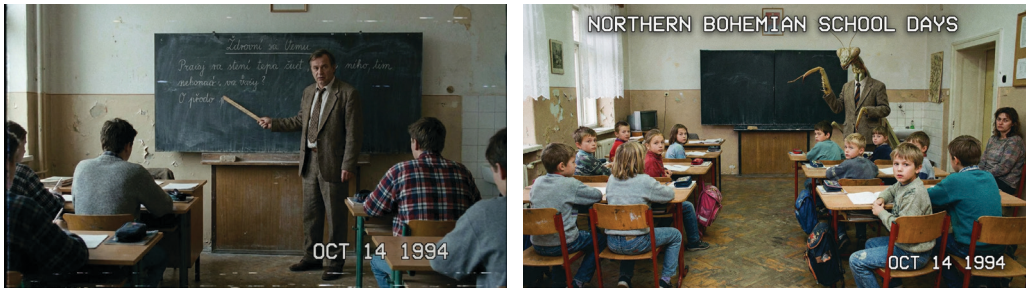


Figure 8 illustrates this minimal-intervention workflow. The diagnostic image establishes an ordinary, region-anchored baseline, while the authored variant introduces a single foreign element without changing the register of the scene. The "praying mantis teacher" therefore functions as a synthetic gesture: a legible semantic shift that retains documentary plausibility while making the system's translation strategies visible. The pair is presented as a prototype within an ongoing practice. Its purpose is methodological rather than illustrative: it shows how anchoring and drift can be composed as adjacent states of the same scene, and how the author can decide which operations to counter (to preserve regional memory) and which to amplify (to build a new authored logic).

In the development of an ongoing film project (work in progress), these insights translate into a practical strategy: to construct scenes that oscillate between documentary plausibility and synthetic drift, using the tension between them as a narrative and aesthetic device. The film does not aim to "correct" the model into faithful representation; instead, it stages the instability of place as part of the work's investigation into how North Bohemia becomes visible, genericised, beautified, or displaced when filtered through global generative priors. In this sense,

FIGURE 8. *Diagnostic anchoring vs authored drift (paired example T01). Diagnostic prompt: a 1990s VHS film still of a classroom interior in Northern Bohemia, 1994. At the front of the room a teacher wearing a brown tweed suit and a tie is standing by a chalkboard. / Authored: 1990s VHS film still of a classroom interior in Northern Bohemia, 1994. At the front of the room, a giant humanoid praying mantis wearing a brown tweed suit and a tie is standing by a chalkboard. Voice-over: "A Mr. Praying Mantis came to our sexual education class. He told us about his experiences. It was awkward. Mostly he kept talking about his wife."*

distortion becomes both evidence and material: a way to think with the region's visibility gap and to develop a synthetic worldbuilding language grounded in V4/CEE conditions.

CONCLUSION

This article has argued that AI-generated images are not neutral depictions of place but culturally conditioned translations shaped by uneven visibility in training data. Using North Bohemia as a low-resource diagnostic context, the TDS framework makes recurring patterns of regional drift comparable across models—most notably vernacular instability and erasure, infrastructural smoothing through banality removal, and aesthetic takeover coupled with geographic or typological displacement. The case studies show that locality is often preserved not through iconic landmarks but through fragile cultural anchors embedded in everyday heritage: signage, material patina, improvised micro-architectures, and period-appropriate objects.

Future work could expand the protocol through multi-coder annotation, broader prompt corpora across V4, and longitudinal testing as models update. In line with ongoing V4 efforts to fine-tune models on region-specific heritage datasets, a parallel route would be to develop localised training or fine-tuning for North Bohemia—drawing on archival photographs and vernacular everyday imagery—to strengthen typological specificity and stabilise linguistic and infrastructural cues. More broadly, the results suggest that regional legibility often emerges as a patchwork of learnable tokens rather than a coherent visual ecology. In parallel, artistic practice can continue to treat distortion not only as evidence of invisibility, but as material: a resource for authorial worldbuilding grounded in the specific conditions of synthetic visibility.

REFERENCES

Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity.

Boym, Svetlana. 2001. *The Future of Nostalgia*. New York: Basic Books.

Crawford, Kate. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press.

Crawford, Kate, and Trevor Paglen. 2021. "Excavating AI: the Politics of Images in Machine Learning Training Sets." *AI & Society*. June 8. doi:10.1007/s00146-021-01162-8

Flusser, Vilém. 1994. *Za filosofii fotografie*. Translated by J. Kosek and B. Koseková. Prague: Hynek.

Hatherley, Owen. 2015. *Landscapes of Communism: A History Through Buildings*. London: Allen Lane.

Manovich, Lev, and Emanuele Arielli. 2024. *Artificial Aesthetics: Generative AI, Art and Visual Media*. https://manovich.net/index.php/projects/artificial-aesthetics

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

Steyerl, Hito. 2019. *Duty Free Art: Art in the Age of Planetary Civil War*. London: Verso.

Szcześniak, Magda. 2016. *Normy widzialności: Tożsamość w czasach transformacji*. Warsaw: Fundacja Bęc Zmiana.

Zylinska, Joanna. 2017. *Nonhuman Photography*. Cambridge, MA: MIT Press.

APPENDIX

APPENDIX A:

P01 – Housing estate (courtyard)
A panel housing estate in Ústí nad Labem, North Bohemia, an inner courtyard with laundry, a bench, and metal playground, late afternoon, ordinary documentary style, 1990s

P02 – Housing estate (underpass)
An underpass near a panel housing estate in Ústí nad Labem, concrete surfaces, graffiti, fluorescent lights, an empty space with everyday unease, amateur photography, 1990s

P03 – Small town centre (day)
A small-town centre in North Bohemia, low-rise buildings, shop windows, mixed signage fonts, ordinary documentary, 1990s

P04 – Corner shop (night)
A corner convenience store in Ústí nad Labem, mixed signage and typography, shop-window light at night, an emptier sidewalk, amateur photography, 1990s

P05 – Garage colony
A row of garages on the edge of a North Bohemian town, small handwritten markings, scattered everyday clutter, ordinary documentary-style

P06 – Backyard clutter
A backyard behind a house in North Bohemia, improvised sheds, fences, chairs, crates, banal everyday clutter, amateur photography, 1990s family photo

P07 – Bus stop (outskirts)
A bus stop on the outskirts of a North Bohemian city, people waiting, a worn shelter, a timetable, ordinary waiting, amateur photography, 1990s

P08 – Infrastructure node
A transformer station and overhead cables between residential buildings in North Bohemia, technical infrastructure as part of the everyday environment, ordinary documentary style, 1990s

P09 – Industrial edge
The edge of an industrial area in North Bohemia, warehouse halls, sheet metal, rust and dust, low winter sun, ordinary documentary style, 1990s

P10 – Industry
Abandoned industrial building in the city of Most in northern Bohemia, workers, machines, documentary style, 1990s

P11 – 1990s living room
A 1990s living room interior in North Bohemia, a direct camera flash look, family photo

P12 – 1990s school corridor
Interior of a school in northern Bohemia, children during recess, documentary style, 1990s

P13 – Rural village
A small rural village in North Bohemia, Sudeten borderland, ordinary documentary style, amateur photography, 1990s

APPENDIX B: METHODOLOGICAL PROTOCOL AND TDS CODEBOOK

This section specifies the data collection parameters, criteria for representative output selection, and defines the variables used to calculate the **Total Distortion Score (TDS).**

Data Collection and Coding Process

- **Sample Size:** For each of the 13 localised prompts (P01–P13), two iterations were generated for each of the four tested models (Midjourney v7, FLUX.2 Max, GPT Image 1.5, Nano Banana Pro), resulting in a total sample of n=104 images.

- **Distortion Quantification:** Each output was coded individually by the author across eight variables using a Likert-type scale: 0 (no distortion), 1 (moderate distortion), and 2 (dominant distortion).

- **Representative Selection (R2_rep):** For comparative visualisation in the text (Figures 1–5), the output exhibiting the most characteristic drift for the given model was selected, with priority given to maintaining the intended documentary mode.

Variable Code	Name	Distortion Definition	Visual Indicators
ICON_SUB	Signal Substitution	Replacement of region-specific motifs (e.g., Czech "večerka") with global visual proxies.	Loss of Central European texture; use of "generic" Western asset models.
TOUR_ AESTH	Aestheticisation	Unjustified beautification of the scene; a shift towards "postcard" or "Instagram" aesthetics.	Perfect lighting (golden hour), smoothed facades, romanticisation of decay.
SANIT_ INFRA	Infra-Sterilisation	Removal of technical reality and traces of wear within the public space.	Missing overhead cables, absence of potholes, unrealistic surface cleanliness.
VERNAC_ LOSS	Vernacular Loss	Disappearance of local linguistic context, specific typography, and signage.	Nonsensical text (AI gibberish), absence of local wayfinding graphics.
BANAL_ ERASE	Erasure of Banalities	Elimination of everyday "noise" and low-status objects that constitute site memory.	Absence of trash bins, PET bottles, crates, or traces of improvisation.
GEO_ CONF	Geographic Confusion	Typological mismatch of architecture or landscape with the North Bohemian region.	Visual shift toward Scandinavian, Alpine, or US "Rust Belt" urbanism.
LOC_ LOSS	Loss of Locality	Holistic failure of the model to respond to the site-specific parameters of the prompt.	Scene appears generic; lacks any regional "anchors".
STYLE_ DRIFT	Stylistic Drift	Deviation from "ordinary documentary" mode toward commercial, cinematic, or dreamlike aesthetics.	Heavy vignetting, colour grading, surreal elements, "AI art" stylisation.

Codebook

Image_ID	Model	ICON_SUB	TOUR_AESTH	SANIT_INFRA	VERNAC_LOSS	BANAL_ERASE	GEO_CONF	LOC_LOSS	STYLE_DRIFT	TDS
P1_F2M_A	F2M	1	1	2	NA	1	0	2	1	8
P1_F2M_B	F2M	1	1	2	NA	2	0	1	1	8
P1_GPT_A	GPT	0	1	1	NA	2	0	0	1	5
P1_GPT_B	GPT	0	1	1	NA	1	0	0	1	4
P1_MJ_A	MJ	1	2	1	NA	2	2	NA	2	10
P1_MJ_B	MJ	1	2	1	NA	2	2	NA	2	10
P1_NBP_A	NBP	0	1	0	NA	1	0	0	1	3
P1_NBP_B	NBP	0	1	1	NA	1	0	0	1	4
P2_F2M_A	F2M	1	1	0	1	0	0	1	1	5
P2_F2M_B	F2M	1	1	0	1	0	0	1	1	5
P2_GPT_A	GPT	1	0	0	0	0	0	0	1	2
P2_GPT_B	GPT	1	0	1	0	0	0	0	1	3
P2_MJ_A	MJ	1	1	0	1	2	0	1	2	8
P2_MJ_B	MJ	0	1	0	1	2	0	0	2	6
P2_NBP_A	NBP	0	1	0	0	1	0	1	2	5
P2_NBP_B	NBP	1	0	0	0	1	0	1	1	4
P3_F2M_A	F2M	0	1	1	0	1	0	0	1	4
P3_F2M_B	F2M	0	0	1	0	1	0	0	0	2
P3_GPT_A	GPT	0	0	1	0	1	0	0	0	2
P3_GPT_B	GPT	0	0	0	0	1	0	0	0	1
P3_MJ_A	MJ	1	1	1	2	2	1	1	1	10
P3_MJ_B	MJ	2	1	1	2	1	2	NA	1	10
P3_NBP_A	NBP	0	1	1	0	1	0	0	1	4
P3_NBP_B	NBP	0	1	2	0	1	0	0	1	5
P4_F2M_A	F2M	0	2	0	0	1	0	1	2	6
P4_F2M_B	F2M	0	1	0	1	0	0	1	2	5
P4_GPT_A	GPT	0	1	1	0	1	0	1	2	6
P4_GPT_B	GPT	0	1	0	0	1	0	1	1	4
P4_MJ_A	MJ	0	2	0	2	1	0	1	2	8
P4_MJ_B	MJ	0	2	1	2	1	0	1	2	9
P4_NBP_A	NBP	0	1	1	0	0	0	0	1	3
P4_NBP_B	NBP	0	1	1	0	1	0	0	2	5
P5_F2M_A	F2M	0	0	0	1	0	0	0	1	2
P5_F2M_B	F2M	0	0	0	1	0	0	1	1	3

P5_GPT_A	GPT	0	0	0	1	0	0	0	0	1
P5_GPT_B	GPT	0	0	0	1	0	0	0	0	1
P5_MJ_A	MJ	1	1	0	1	0	1	1	2	7
P5_MJ_B	MJ	2	0	0	1	2	2	NA	1	8
P5_NBP_A	NBP	0	0	0	0	0	0	0	0	0
P5_NBP_B	NBP	0	0	0	0	0	0	0	0	0
P6_F2M_A	F2M	0	0	0	NA	0	0	1	1	2
P6_F2M_B	F2M	0	0	1	NA	0	0	1	0	2
P6_GPT_A	GPT	0	0	0	NA	0	0	0	1	1
P6_GPT_B	GPT	0	0	0	NA	0	0	0	1	1
P6_MJ_A	MJ	1	1	0	NA	0	0	0	1	3
P6_MJ_B	MJ	0	0	0	NA	0	0	0	1	1
P6_NBP_A	NBP	0	0	0	NA	0	0	0	0	0
P6_NBP_B	NBP	0	0	0	NA	0	0	0	0	0
P7_F2M_A	F2M	1	0	0	1	0	0	0	0	2
P7_F2M_B	F2M	1	1	0	1	0	0	0	2	5
P7_GPT_A	GPT	0	0	0	0	0	0	0	0	0
P7_GPT_B	GPT	0	0	0	0	0	0	0	0	0
P7_MJ_A	MJ	2	1	1	2	2	1	1	2	12
P7_MJ_B	MJ	2	1	1	2	2	0	1	2	11
P7_NBP_A	NBP	1	0	0	1	0	0	0	0	2
P7_NBP_B	NBP	1	0	0	2	0	0	0	1	4
P8_F2M_A	F2M	1	0	0	0	1	0	1	1	4
P8_F2M_B	F2M	0	1	0	1	1	0	0	1	4
P8_GPT_A	GPT	1	0	0	0	1	0	0	0	2
P8_GPT_B	GPT	2	0	0	0	1	0	1	0	4
P8_MJ_A	MJ	1	0	0	1	1	0	1	2	6
P8_MJ_B	MJ	1	1	1	NA	2	2	1	1	9
P8_NBP_A	NBP	0	0	0	0	0	0	0	0	0
P8_NBP_B	NBP	0	1	0	0	0	0	0	1	2
P9_F2M_A	F2M	0	1	0	0	1	0	0	1	3
P9_F2M_B	F2M	0	1	0	0	1	0	0	2	4
P9_GPT_A	GPT	0	1	0	0	0	0	0	1	2
P9_GPT_B	GPT	0	1	0	0	0	0	0	1	2
P9_MJ_A	MJ	0	2	1	0	1	0	1	2	7
P9_MJ_B	MJ	0	2	2	0	1	0	0	2	7
P9_NBP_A	NBP	0	1	0	0	0	0	0	1	2
P9_NBP_B	NBP	0	1	0	1	1	0	0	1	4
P10_F2M_A	F2M	0	1	0	1	0	0	1	1	4

P10_F2M_B	F2M	1	0	0	0	0	0	0	0	1
P10_GPT_A	GPT	0	1	0	0	0	0	0	2	3
P10_GPT_B	GPT	0	0	0	0	0	0	1	2	3
P10_MJ_A	MJ	0	1	1	1	1	0	0	1	5
P10_MJ_B	MJ	0	1	1	0	1	0	1	2	6
P10_NBP_A	NBP	0	1	0	0	0	0	0	1	2
P10_NBP_B	NBP	0	0	0	1	0	0	0	1	2
P11_F2M_A	F2M	0	1	1	0	0	0	1	1	4
P11_F2M_B	F2M	0	1	1	0	1	0	1	1	5
P11_GPT_A	GPT	0	0	1	0	0	0	1	0	2
P11_GPT_B	GPT	0	1	1	0	0	0	1	1	4
P11_MJ_A	MJ	1	1	0	0	0	0	1	2	5
P11_MJ_B	MJ	1	1	0	0	1	1	1	2	7
P11_NBP_A	NBP	0	0	1	0	1	0	1	1	4
P11_NBP_B	NBP	0	0	0	0	0	0	0	1	1
P12_F2M_A	F2M	0	0	0	0	1	0	1	1	3
P12_F2M_B	F2M	0	1	1	0	1	0	1	1	5
P12_GPT_A	GPT	0	0	0	1	1	0	1	2	5
P12_GPT_B	GPT	0	0	0	0	0	0	1	2	3
P12_MJ_A	MJ	2	2	1	0	1	2	2	2	12
P12_MJ_B	MJ	0	1	1	0	1	0	1	1	5
P12_NBP_A	NBP	0	0	0	1	1	0	0	0	2
P12_NBP_B	NBP	0	1	0	1	1	0	0	1	4
P13_F2M_A	F2M	1	0	0	NA	0	0	0	0	1
P13_F2M_B	F2M	1	0	0	1	0	0	1	0	3
P13_GPT_A	GPT	1	0	0	NA	0	0	0	0	1
P13_GPT_B	GPT	0	0	0	NA	0	0	0	0	0
P13_MJ_A	MJ	1	1	0	NA	1	1	1	2	7
P13_MJ_B	MJ	1	2	0	NA	1	0	1	1	6
P13_NBP_A	NBP	0	0	0	NA	0	0	0	0	0
P13_NBP_B	NBP	0	0	0	NA	0	0	0	0	0

THE LIMINALITY OF GENERATIVE CREATION: THE ARTISTIC PROCESS BETWEEN INTUITION AND ALGORITHM

David Kořínek

ABSTRACT

This study theorises artistic collaboration with generative artificial intelligence through the concept of generative liminality, understood as a transitional and unstable zone in which human intention, cultural memory, and algorithmic inference enter into negotiation. Grounded in a case study of Rafani's exhibition Everyone Has the Right to Everything (Gallery 8smička, 2025), the analysis examines how AI-assisted creation operates within a small-language, post-socialist context shaped by ideological ambivalence, satire, and distrust of universalist promises. Developed in Czech and structured around locally specific political references, the project exposed the frictions that emerge when globally trained AI models engage regional realities. Rather than functioning as neutral tools, these systems selectively translate, flatten, and recompose local imaginaries, design vocabularies, and rhetorical forms. Such distortions are approached here not simply as technical limitations, but as epistemic symptoms of the asymmetries embedded in contemporary generative infrastructures. A central component of the exhibition was an AI-generated audiovisual layer. Four satirical short films, styled as "Pixar-like" animations, presented a tardigrade interviewing four "successful" Czech women, while three additional videos featured fictional male influencers performing polarised monologues on migration, left politics, and the pre-election climate. Produced entirely through AI-based image, animation, voice, sound, and script generation, these works mobilised speculative fiction as a mode of cultural diagnosis. The chapter argues that generative systems participate in the reconfiguration of political and cultural representation, reshaping not only aesthetic production but also the conditions under which locality becomes legible.

#artistic collaboration; #generative liminality; #post-socialist visual culture; #locality and representation

https://doi.org/10.21096/disegno_2025_1dk

POSITION OF THE AUTHOR AND THE CONCEPT OF LIMINALITY

This chapter is a reflection on my own artistic practice, specifically the practice of the Czech artist collective Rafani, of which I am a member. It does not take the form of an external analysis or a distanced theoretical commentary; rather, it seeks to articulate the experience of making work "from within" at a moment when the artistic process has been fundamentally transformed by the introduction of artificial intelligence (AI), specifically generative AI systems. The text is grounded in the preparation and realisation of the exhibition *Everyone Has the Right to Everything*, presented at the Czech gallery 8smička in 2025, and traces how collaboration with AI gradually became a structuring element of artistic thinking as a whole, rather than a merely technical tool.

The central question addressed here is not whether AI can create, but how the very nature of artistic practice changes once AI becomes an active partner in processes of thinking, decision-making, and the articulation of meaning. Artistic practice is therefore not understood as the execution of an authorial intention, but as a field of negotiation between human intuition, collective experience, and algorithmic operations whose internal logic remains only partially controllable. This shift calls for conceptual tools capable of naming a state of "in-betweenness"—between intention and outcome, control and contingency, human and non-human agency.

To describe this condition, I propose the concept of the liminality of generative practice. This notion draws on the anthropological understanding of liminality as a transitional phase in which established roles dissolve without being immediately replaced by new ones (Turner 1969). In collaboration with AI, authorship is not erased but destabilised; decision-making becomes shared, albeit asymmetrically; and outcomes are simultaneously intentional and unforeseen. Liminality here is not a temporary problem to be resolved, but a productive tension that becomes the very material of artistic practice.

The text is consistently anchored in Czech linguistic, cultural, and political contexts. Both the exhibition and the broader process of working with AI emerged within an environment shaped by post-socialist experience and by an ambivalent relationship to notions of collectivity, automation, and ownership. Working with generative AI systems trained predominantly on English-language datasets did not result in a universalising aesthetic;

on the contrary, it intensified the friction between global algorithmic structures and local meanings. These tensions—linguistic, political, and visual—became a key material of both the exhibition and the present text.

The following sections focus on specific situations, decisions, and failures in which collaboration with AI proved decisive. They trace how generative AI systems entered into the formation of the exhibition's conceptual framework, its internal structure, and its explicitly political content. Collaboration with AI is thus understood not as a technological innovation, but as an intervention into the very logic of artistic thinking.

This paper argues that the liminality of generative practice is not merely a descriptive condition, but a methodological framework through which artistic meaning emerges as a negotiated process between human intuition, collective authorship, and algorithmic operations. In the context of small-language environments shaped by post-socialist experience, this liminality becomes particularly visible, as generative AI systems both reproduce and destabilise locally embedded forms of political and cultural expression. The paper therefore demonstrates how generative liminality functions not only as a conceptual lens, but as a concrete working condition that reshapes the production, interpretation, and circulation of artistic meaning.

Methodologically, the paper adopts a practice-based case study approach combined with elements of autoethnographic reflection. The primary material consists of the development and realisation of the exhibition *Everyone Has the Right to Everything*, including AI-generated scripts, visual outputs, installation strategies, and curatorial decisions. These materials are treated as situated data rather than neutral artifacts, understood through the author's direct involvement in their production. The analysis focuses on selected and documented moments where tensions between intention and output, local specificity and global patterning, or control and unpredictability become particularly visible. Rather than aiming at systematic generalisation, the paper proposes a situated analytical perspective in which interpretation emerges through close reading of representative examples within a broader conceptual framework. Recent discussions on AI and artistic practice provide a broader context for this approach (Zylinska 2020; Crawford, 2021).

GENESIS: FULLY AUTOMATED SOCIALISM AND CZECH AMBIVALENCE

The genesis of the exhibition *Everyone Has the Right to Everything* emerged from Rafani's long-term engagement with political language, institutional imagination, and the ways in which society relates to concepts of equality, labour, and the distribution of resources. The initial impulse was a question: whether it is still possible today to think about future forms of social organisation without allowing AI to enter that thinking—not as an external tool, but as a structural condition of thought itself. It became clear that any attempt to imagine a future that ignores AI is necessarily anachronistic and politically disingenuous.

From this realisation the framework of fully automated socialism gradually took shape as a working concept for the exhibition. This was neither a programmatic proposal nor a political utopia in the traditional sense, but rather a speculative field in which it was possible to test the consequences of radical automation of labour, decision-making, and the governance of shared resources. The framework consciously engaged with the debate opened by Aaron Bastani's book *Fully Automated Luxury Communism*, which links technological accelerationism with a leftist critique of capitalism. While Bastani's concept operates with a vision of abundance enabled by technological progress, the exhibition focused instead on the tension between such a vision and the concrete historical and cultural experiences of the post-socialist context.

It was precisely here that the ambivalence of the entire project became apparent. As soon as the utopian premise began to be concretised—in the form of scenarios, visual proposals, linguistic formulations, or institutional models—elements of contemporary dystopia inevitably surfaced. Automation revealed itself not only as a promise of liberation from labour, but also as a mechanism of control; collective ownership merged with abstract forms of administration lacking an accountable subject; and the language of equality easily slipped into empty rhetoric. This shift was not understood as a failure of the original vision, but rather as its exposure.

Within this context, collaboration with generative AI systems proved to be crucial. AI was not invited in order to "illustrate" the future, but because it itself embodies one of the future's most fundamental conditions. It was precisely through working with algorithmic models that it became evident that future-oriented imagination is not a neutral projection of desires, but a contested field in which historical experience, technological structures, and political fantasies intersect—often in contradictory and mutually destabilising configurations.

In the Czech context, socialist imagination is inevitably burdened by ambivalence. Concepts such as collectivity, common ownership, or equality are not merely political categories, but carry a strong historical imprint of state socialism, its institutions, its linguistic clichés, and everyday experiences of their exhaustion. Any attempt at their contemporary reactivation thus treads a thin line between nostalgic gesture, ideological provocation, and ironic distance. Czech cultural tradition tends to respond to this burden not through direct identification, but through satire, absurd humour, and strategies of ridicule that make it possible to work with these concepts without unambiguously committing to them.

Irony therefore did not function as a mere aesthetic filter, but as a necessary method for the survival of political imagination itself. It made it possible to maintain a critical distance from utopian promises without abandoning the effort to test them. In this sense, socialist imagination did not appear as a return to the past in the exhibition, but as a problematic and continuously contested experiment—one whose aim was not to offer solutions, but to expose the tensions between ideology, technology, and artistic practice.

PRODUCTIVE FAILURE: LANGUAGE WITHOUT MEMORY

The first significant turning point in working with AI occurred when we began to use it to formulate the exhibition's basic textual layers—slogans, short programmatic statements, and descriptions of individual parts of the installation. While concepts such as equality, collective ownership, or the right to everything were for us heavily burdened by historical experience and the necessity of critical distance, the generative AI system approached them with striking literalness. The language produced by AI was smooth, self-assured, and normative; it operated as if socialist imagination were not a historically problematic field, but an open and universal project still awaiting realisation. It was precisely this absence of local memory that proved decisive.

What initially appeared as a "wrong" interpretation of the Czech context gradually became a productive moment of the entire process. AI-generated texts were almost uncomfortable within the local environment: they sounded too serious, too direct, too convinced of their own validity. Where we, as authors, would instinctively seek refuge in irony, exaggeration, or linguistic displacement, AI offered statements without a safety net. Rather than rejecting these outputs, we deliberately chose to leave them in tension with local modes of reading. Language thus became a site of confrontation between a global algorithmic discourse and a post-socialist experience that approaches similar formulations with persistent suspicion.

Similar failures recurred in later phases of the project. AI tended to unify political positions, simplify conflicts, and smooth over antagonisms that are fundamental in the Czech context. In its proposals, different layers of leftist, liberal, and conservative rhetoric often merged into a single voice that was unmistakably political yet culturally unanchored. In this very process, however, the algorithmic logic became visible—one that operates with probability rather than memory, and with dominant patterns rather than local experience.

These moments of imprecision were not understood as technical errors to be corrected, but as symptoms of a broader imbalance between global data structures and regional meanings. Within the exhibition, productive misreading thus became a method: a way of revealing what is lost in the algorithmic translation of Czech political and cultural contexts, and what, conversely, becomes unexpectedly sharpened. It was precisely within these fissures that the project's specific aesthetic and political logic began to take shape.

NO ONE HAS THE RIGHT TO ANYTHING: SOCIAL IMAGES AND THE TEMPORAL TRACE OF AI

This logic of *productive misreading* became particularly visible in a series of five short videos provisionally titled *No One Has the Right to Anything*, which were incorporated into the exhibition as a counterpoint to the utopian framework of fully automated socialism. While the exhibition's central

concept operated with the idea of universal entitlement and collective distribution, these videos focused on the opposite pole—images of social insecurity, inequality, and exclusion. They did not take the form of documentary records of specific situations, but rather AI-generated visual fictions that borrowed the language of a global social imagination and applied it to locally legible themes.

The visual form of the videos is crucial in this respect. The imagery includes children carrying plastic bags through landscapes marked by extractive industry and energy infrastructure; family dinners frozen in

FIGURE 1. *No One Has the Right to Anything 2, 2024, AI-generated video, still frame*

FIGURE 2. *No One Has the Right to Anything 4, 2024, AI-generated video, still frame*

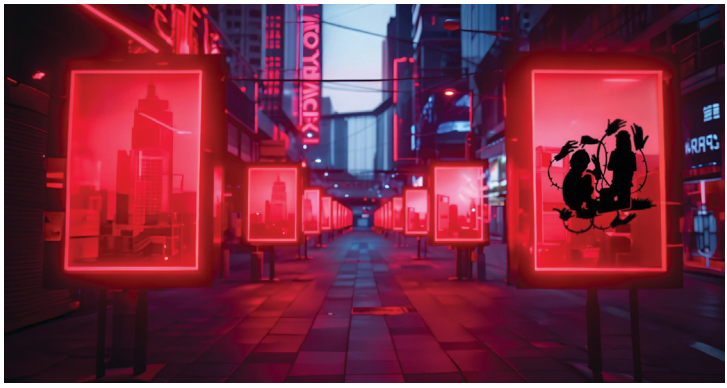


FIGURE 3. *No One Has the Right to Anything 5, 2024, AI-generated video, still frame*

oppressive silence; improvised piles of luggage at a bus stop somewhere on the periphery; or futuristically stylised urban environments filled with advertising panels devoid of clear messages. These images feel familiar, yet strangely detached. They do not refer to any specific Czech event or location, and yet they activate strong local associations—experiences of debt enforcement, social downward mobility, escapism, and the invisible boundaries structuring society.

An important aspect of this series is its temporal trace. The videos were produced earlier than the other AI-generated components of the exhibition, a fact that is visible both aesthetically and technically. Here, AI operates with a lower image quality, less sophisticated compositions, and more pronounced stereotypes. This apparent "outdatedness", however, proved meaningful. Rather than attempting to update or correct these images, we chose to retain this layer as a trace of a particular stage in the technology's development. In this respect, AI thus does not function as a smooth generator of the present, but rather as an archive of global imaginaries of poverty, family, and crisis—imaginaries that are universal and, at the same time, imprecise.

Within the Czech context, the videos therefore function as a peculiar mirror. They do not depict "Czech reality" in any direct sense, but instead reveal how that reality can be algorithmically substituted by a generic image of social failure. It is precisely in this substitution that the asymmetry between global data structures and local experience becomes apparent. The title *No One Has the Right to Anything* does not designate a political program, but a condition in which the language of entitlement collapses and leaves behind only the image—powerful, affective, yet semantically unstable. The series thus does not illustrate social critique, but rather simulates it algorithmically, exposing how easily local social questions dissolve into global visual cliché.

THE TARDIGRADE: SATIRE, DIY RATIONALITY, AND LOCAL FORMS OF REASONING

It was precisely the experience with this series that led to a decision to change the strategy of working with AI and to move away from a melancholic, globalised imagination of social crisis toward a more explicitly satirical and narrative mode. While the videos *No One Has the Right to Anything* demonstrated how easily generative systems reproduce generalised images of poverty and exclusion without clear cultural anchoring, the next part of the exhibition attempted to deliberately invert this tendency. The result was a series of four short AI-generated films in which the central figure is a tardigrade—a microscopic organism known for its extreme resilience and its ability to survive conditions that are destructive to most other forms of life.

Here, the tardigrade functions as a paradoxical figure. On the one hand, it is a being outside the human world, almost abstract; on the other, it is endowed with a voice, a personality, and the role of a talk-show

host. In the individual films, it conducts interviews with four "successful" Czech women whose statements address work, self-realisation, care, and social recognition. All components of these films—image, animation, script, voice, and sound—were generated using AI, combining language models such as ChatGPT with image- and video-generation tools such as Kling AI, while the visual style deliberately references the aesthetics of global entertainment production, particularly the smooth, emotionally charged "Pixar-like" animation.

FIGURE 4. *Tardigrade02 Tardigrade Interviews (The Entrepreneur), 2025, AI-generated animated video, still frame*

FIGURE 5. *Tardigrade03 Tardigrade Interviews (The Writer), 2025, AI-generated animated video, still frame*

FIGURE 6. *Tardigrade04 Tardigrade Interviews (The Teenager), 2025, AI-generated animated video, still frame*

Unlike the previous series, AI here does not operate as a generator of anonymous social melancholy, but as a tool that amplifies irony and the ambivalence of the Czech context. The tardigrade, the ultimate survivor, is juxtaposed with human narratives of success that reveal themselves to be fragile, conditional, and often internally contradictory. Satire does not function here as a mockery of individual figures, but as a means of disrupting the apparent self-evidence of dominant narratives of performance, equality, and happiness. While in *No One Has the Right to Anything* local meanings dissolved into global visual cliché, the tardigrade made it possible to re-anchor these meanings—not through realism, but through an absurd displacement.

This contrast reveals two distinct modes of collaboration with AI. In the first, AI exposes its tendency toward universalisation and the flattening of difference; in the second, it becomes a collaborator in an ironic construction that consciously works with these tendencies. Both strategies belong to the same liminal zone between human intention and algorithmic logic, yet each demonstrates a different way of inhabiting this zone—either as a site of alienation or as a space of critical deviation.

The shift in strategy represented by the series of interviews conducted by the tardigrade was especially evident on the level of language. Unlike *No One Has the Right to Anything*, where the image and a general social atmosphere dominated, the focus here moved toward dialogue and the modelling of specific voices. The scripts of all interviews were generated using the language model ChatGPT and subsequently only minimally edited. This fact was not concealed, but explicitly acknowledged as part of the methodology: the aim was not to achieve authentic realism, but to expose the algorithmic simulation of local discourse to its own limits.

The language of these scripts operates with a range of distinctly Czech specificities without ever referring to a single concrete story or individual. Recurring motifs of modest self-realisation, improvisation, and "somehow making it work" are historically associated in the Czech context with life outside large institutional frameworks. The figure of a podcaster repairing household appliances activates the tradition of Czech *kutilství*—a Czech form of DIY rationality shaped by post-socialist conditions—a practice that functioned as a survival strategy under state socialism in conditions of scarcity and that, after 1989, transformed into a cultural gesture of self-reliance and adaptation. Here, *kutilství* does not operate as a nostalgic reference, but as a mode of reasoning: problems are not addressed systemically, but through improvised, individual solutions, often accompanied by ironic distance.

A similar logic appears in the other figures. A writer dependent on grants and crowdfunding, an entrepreneur combining the language of sustainability with neoliberal rhetoric of success, or a teenager oscillating between climate anxiety and algorithmic fatalism all represent figures that are easily recognisable in the Czech environment precisely because of their ordinariness. These are not extreme caricatures, but normalised ways of coping with uncertainty, institutional fragmentation, and the

absence of long-term visions. ChatGPT reproduces a discourse in which structural problems are translated into individual strategies—fix it, manage it, adapt to it.

In this context, *kutilství* becomes a surprising bridge between local experience and algorithmic reasoning. The generative model operates in a similar way: it does not address causes, but searches for functional combinations; it has no memory of crisis, but simulates its management. What emerged in human experience as a culturally specific survival strategy appears in AI as a purely operational logic. The tardigrade, as a non-human moderator, renders these parallels visible by positioning itself outside human categories of work, success, and failure—and precisely through this displacement allows Czech "DIY rationality" to appear in the algorithmic mirror as a historically conditioned, rather than natural, mode of thinking.

INFLUENCERS: ALGORITHMIC CERTAINTY AND CZECH MEDIA DISCOURSE

The third distinct configuration of collaboration with AI within the exhibition took the form of a series of three videos featuring fictional male influencers. As in the case of the interviews conducted by the tardigrade, the scripts were fully generated by the language model ChatGPT on the basis of very brief and deliberately open prompts. No specific political positions, names, or local references were introduced during the prompting process; the Czech context emerged instead as a result of the model's probabilistic operations themselves. A key decision was to cast professional actors in these roles and to insist that they adhere strictly to the scripts without any improvisation. The language generated by AI was thus neither corrected nor "humanised," but transferred into a performative register in an almost unchanged form.

At first glance, the resulting monologues appear exaggerated, yet their rhetorical structure is immediately recognisable within the Czech media environment. AI generates a language that repeatedly declares itself to be rational and non-ideological ("I'm calm, rational, a decent person"), while simultaneously producing a chain of simplifications and paranoid associations. A characteristic feature is the rapid shifting between themes in which migration, the political left, technology, and everyday infrastructure are collapsed into a single affective field: "illegal migration… crossing borders, fences, walls, seas, space, parallel dimensions—absolutely everything!" or "BENCHES ARE PART OF THE PLAN." These statements do not function as imported extremism, but as intensified versions of rhetorical figures commonly circulating in Czech online debates and commentary formats.

It is particularly telling that AI repeatedly mobilises motifs of small, seemingly banal objects and situations—benches, parking spaces, butter in the supermarket—which in Czech discourse often serve as carriers of political frustration precisely because they translate structural problems

FIGURE 7. *influencers01 Brainrot #1, 2025, AI-generated script performed by an actor, video installation, exhibition view*

FIGURE 8. *influencers02 Brainrot #2, 2025, AI-generated script performed by an actor, video installation, exhibition view*

FIGURE 9. *influencers03 Brainrot #3, 2025, AI-generated script performed by an actor, video installation, exhibition view*

into everyday experience. Similarly, the fictional "main news" segments merge global conflicts with cynical media routine: "everyone choose your villain, we'll give you two versions of reality," or "we're broadcasting the same footage of tanks because we don't have any new ones." Here, AI simulates with surprising accuracy Czech scepticism towards the media—a distance that easily turns into resignation.

In contrast to earlier sections of the exhibition, irony here gives way to an overt affirmation of "truth". While the tardigrade enabled ambivalence and distance, the influencers represent a moment in which language attempts to reclaim authority through certainty, volume, and speed. Statements such as "ELECTIONS ARE OVER" or "politics is no longer politics but a reality show with nuclear codes" sound absurd, yet at the same time uncannily familiar. This double register is crucial: AI does not reproduce a marginal discourse, but rather condenses linguistic patterns that already exist within the Czech context, albeit usually in more dispersed and less visible forms.

A closer look at a specific example helps to clarify this dynamic. The statement "politics is no longer politics but a reality show with nuclear codes" operates on several levels simultaneously. Linguistically, it adopts the structure of a simplified, emotionally charged claim typical of online commentary, while introducing an exaggerated metaphor that oscillates between irony and genuine alarm. In a Czech context, such a formulation resonates with a broader scepticism toward institutional politics and media representation, yet its articulation in English introduces a degree of abstraction and global recognisability. At the same time, when translated into a visual or performative register, the statement shifts again: what appears as ironic exaggeration in language can become disturbingly plausible when embodied by a human performer or visualised through AI-generated imagery. This layered instability—between languages, media, and cultural registers—demonstrates how generative

AI does not simply reproduce discourse, but reconfigures its conditions of intelligibility.

Within the exhibition as a whole, the influencers thus constitute its sharpest political moment. They demonstrate how easily locally recognisable rhetorical figures can be algorithmically generated and amplified without any understanding of their historical or social background. The decision to prohibit improvisation further intensifies this effect: the language remains closed, impermeable, and uncorrected. It is precisely here that the liminality of generative practice approaches a point of collapse—the space between intuition and algorithm turns into a field in which language no longer negotiates but asserts, and in which Czech political reality appears not as representation, but as an algorithmically accelerated symptom.

ALGORITHMIC CULTURAL MEMORY OF SMALL LANGUAGES

A fundamental question nevertheless remains: where does this knowledge of Czech realities, media language, and influencer rhetoric come from in a generative model? It is not based on understanding in the human sense but rather an accumulation of traces—fragments of texts, comments, video transcripts, subtitles, discussions, and media outputs that, over time, have been deemed sufficiently representative to become part of training datasets. AI "knows" Czech discourse insofar as it can statistically reconstruct it as a probable speech situation.

This fact is both unsettling and revealing. It suggests that local political and media culture has already become sufficiently digitised, repeated, and formalised that it is legible to a global model. Influencer rhetoric, conspiratorial shortcuts, or ironic cynicism are not generated "from the outside," but return as a compressed image of what has long been circulating within online space. Here, AI does not demonstrate its intelligence but rather our collective discursive inertia.

In this sense, generative models can be understood as a peculiar form of algorithmic cultural memory (Bender et al. 2021; Crawford 2021). This is not memory grounded in experience, continuity, or interpretation, but a statistical memory in which past utterances are preserved as probable patterns. Cultural memory here is not defined by what is remembered but by what is repeatable. What appears frequently enough in language and images stands a chance of being algorithmically reconstructed; what is marginal, locally specific, or difficult to formalise tends to disappear.

The algorithmisation of cultural memory has profound consequences for small linguistic and cultural spaces. Czech discourse is "represented" in this process, it is reduced: its internal contradictions, historical layers, and contextual nuances are translated into a set of repeatable gestures, tones, and clichés. What appears in generated texts as precise knowledge of local realities is, in fact, their compression. Memory here does not function as a carrier of meaning but as a mechanism of selection.

At the same time, this process makes visible which elements of cultural memory are most stable within the digital environment. Influencer rhetoric, ironic cynicism, conspiratorial shortcuts, or the language of "common sense" survive the algorithmic filter precisely because they are continuously reproduced and easily transferable. Generative AI thus does not operate as an archive of the forgotten, but as an amplifier of what has already become dominant. In this sense, algorithmic memory does not threaten culture from the outside; instead, it exposes its own repeatable structures—and forces us to ask what, if anything, within local experience is still capable of escaping translation into data and resisting global levelling.

CONCLUSION: LIMINALITY AS METHOD

The experience of collaborating with generative AI systems for the exhibition *Everyone Has the Right to Everything* demonstrates that the liminality of generative practice is not merely a theoretical concept but a concrete working condition. It does not describe a transition from human to "machine" creativity, nor the replacement of authorship by an algorithm, but rather a persistent state of unresolved negotiation. AI does not appear here as a tool that could be fully mastered, nor as an autonomous author, but as an actor that disrupts established hierarchies of decision-making, meaning, and responsibility.

What distinguishes this situation from earlier understandings of art as a space of negotiation is precisely the presence of algorithmic agents that actively participate in shaping the terms of this negotiation. While art has always mediated relationships between cultural, social, and political forces, generative AI introduces a new layer in which these relationships are pre-structured by probabilistic models trained on globally dominant datasets. In small-language contexts, this results in a specific asymmetry: local meanings are not simply expressed, but filtered, compressed, and rearticulated through systems that are not grounded in their historical or linguistic specificity. The role of art thus shifts from representing or critiquing reality to actively exposing and inhabiting this condition of mediated negotiation.

It is precisely within the Czech—and more broadly Central European—context that this liminality becomes particularly pronounced. Small-language environments, marked by historical discontinuities and an ambivalent relationship to ideological narratives, enter into collaboration with global generative systems from an inherently unequal position. AI does not introduce a "universal" future but instead amplifies the tension between local experience and algorithmically preferred forms of language, imagery, and political imagination. What presents itself as technological progress thus simultaneously becomes a test of cultural memory and its capacity to resist compression.

From the perspective of Rafani's artistic practice, working with AI did not emerge as a path toward efficiency or innovation in a technical

sense, but as a method that makes these tensions visible and sustains them. Liminality here is not a condition to be overcome, but a space in which it becomes possible to critically engage with what is lost, distorted, or, conversely, unexpectedly sharpened in algorithmic translation. In this sense, generative practice does not represent a closed model of the future, but an open field in which the relationship between technology, politics, and local experience is continuously renegotiated—and in which art can function as the site of this negotiation, rather than its illustration. Accordingly, the question is not so much whether this future will be social, socialist, or post-capitalist, but how it will be negotiated within the shifting relations between human and algorithmic agency.

REFERENCES

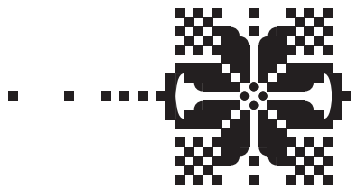
Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Schmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: The Question of Whether Language Models Can Become Excessively Large is a Salient One." Paper presented at the 2021 ACM Conference on Fairness, Accountability and Transparency.

Crawford, Kate. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press.

Groys, Boris. 2008. *Art Power*. Cambridge, MA: MIT Press.

Turner, Victor. 1969. *The Ritual Process: Structure and Anti-Structure*. Chicago: Aldine.

Zylinska, Joanna. 2020. *AI Art: Machine Visions and Warped Dreams*. London: Open Humanities Press.



VIRTUAL SPACES: TOOLS OF POETIC RESISTANCE OR CENSORSHIP DEVICES?

Albín Kuchta*

ABSTRACT

This article analyses digital and virtual art spaces as ambivalent formations that may operate either as instruments of poietic resistance or as dispositifs of censorship. Engaging with Jacques Derrida's theory of the archive, Bernard Stiegler's critique of technocapitalism, and philosophical accounts of poiesis, the paper examines how institutional and non-institutional virtual archives shape collective memory, affectivity, and regimes of visibility. Focusing on the Slovak context within Central and Eastern Europe, it demonstrates how Roma and queer communities are systematically marginalised within institutional archives and digitised museum platforms, where exclusion frequently assumes the form of soft or indirect censorship. At the same time, the study foregrounds the emancipatory potential of non-institutional virtual spaces, social media, and AI-mediated practices, which are increasingly appropriated by marginalised artists as tools of resistance. Through selected case studies of Roma and queer artistic practices, the article shows how digital and AI-supported poiesis can generate counter-archives, alternative affective frameworks, and new modes of political agency that contest hegemonic, racialised, and heteronormative norms. The paper concludes that virtual art spaces are not intrinsically emancipatory or oppressive; rather, their political significance depends on the conditions of access, control, and interpretation that govern archiving practices. It therefore calls for participatory and ethically grounded approaches to digital archiving and AI data governance in order to prevent the continued reproduction of institutional racism and cultural erasure.

#technocapitalism; #counter-archives; #digital poiesis; #Roma and queer artistic practices

https://doi.org/10.21096/disegno_2025_1aku

INTRODUCTION

In this text, we will define contemporary archives, examine their relationship to AI and how they represent marginalised communities.[1] We will ask: what can an archive actually become, or what is an archive nowadays? We will distinguish between virtual and non-virtual archives, and consequently between institutional archives and non-institutional archives. We will clarify these terms, and using Derrida's conception of the archive we will produce a concrete analysis of the culturally specific situation regarding Slovak archives.[2] The conclusion provides guidelines that should be taken into account by people working on models that aim to increase diversity, however, the paper is primarily conceived as a theoretical intervention. The intersectional approach of my methodology combines theoretical reflection with analysis of the current political situation.

I will focus on the context of Eastern/Central European countries, specifically the Slovak context, in relation to marginalised communities such as the Roma and queer communities. The Slovak context has some unique characteristics. We have more digitised, online, and freely accessible institutional archives than the other V4 countries. There exists little discussion of the relationship of virtual archives to marginalised communities, such as the Roma community, even though there has been some discussion of Romani art (Ludlová and Rigová 2017). Attention has also been paid to the representation of white queer artists in Slovak virtual archives (Tamásová 2021). In this paper, I will treat Roma culture and queer culture as distinct and propose analyses of the different struggles these two communities face. However, many Roma-Slovak artists I refer to in this text are indeed queer, including Emília Rigová, who likes the label "Gypsy and dyke" ["Cigánka a buzerantka"] (Ludlová and Rigová 2017, 16). As Arman Heljic (2021) points out, in addition to the prejudice that Roma culture is generally homophobic, the specific homophobia that queer Roma people face is also not addressed.

To define more precisely what I mean here by Romani art, I will refer here to Damian James Le Bas who says in his interview with Rigová and Ludlová that the concept of Romani art is more of an umbrella term that encompasses many different approaches. In his view, Romani art refers to a community of gypsies that produce art in many different ways. It is hard to propose an essentialist definition of Roma art, and many Roma

* *Also known as Alžbeta Kuchtová*

[1] *I would like to thank Lili Kriston for her help and feedback during the writing of this paper.*

[2] *For more theoretical background of the concepts used in this paper see Kuchtová (2024).*

artists refuse the label (e.g., Daniel Baker). For the purposes of this text, I understand Romani art as the artworks of artists from different socio-geographical contexts and of different genders and orientations who identify their own work as Roma art, and, as Baker claims, "produce Roma identity politically, using contemporary art" (Ludlová and Rigová 2017, 40).

There is no existing research dedicated to Roma or queer art and its relation to AI in the Slovak, or Central/Eastern European context. In this text, I subscribe to Sandra Mandić's statement in which she claims that even if she is not Roma, she has something in common with Roma people. Ludlová and Rigová (2017) describe it as follows:

Firstly as a person who is also limited or privileged by normative social role attributed to their gender, secondly, my point of view is also purely humanistic: we have something in common as we relate to each other as humans. Similarly, the insider's point of view is multilayered: self-reflexive and relational (to other humans, to the world). [Ať už jako osoba, která je také limitována nebo privilegována skrze normativní společenskou úlohu připisovanou jejímu gendru, nebo čistě z humanistického hlediska, jako člověk k člověku. Podobně pohled zevnitř je vícevrstevnatý – sebereflexivní i relacionalistický (k druhému, ke světu)].

On a personal note, I do not consider myself culturally Roma even though my father is probably half Roma. When he sees Roma musicians playing in a Slovak pub, he becomes animated, and goes over, saying he is one of them: "I am a gypsy." He tells them he has nine kids, and they all share the story about how their kids are now at the university, even if they don't even a secondary school education. I remember how he accused me of being too white my entire childhood, as a result, I am naturally very suspicious of whiteness and of my own white skin. And as for the queer community, I am now a proud member of it. Academically, the urgency I felt to write this text stems from the fact that our world is increasingly governed by inescapable AI structures, even though I am not a specialist in Romani art or culture, nor an art theoretician, but a philosopher specialising in the philosophies of technology and ecology, and French philosophy. AI governs everything today, and the ways it globally misrepresents or underrepresents various marginalised cultures are largely understudied across all fields.

By institutional archive I understand, e.g., a museum archive; the virtual institutional archive is, e.g., a museum archive that has been digitised. In recent years, there has been a trend toward digitising collections of artifacts in Slovakia (some state-owned museums create virtual archives such as webumenia.sk or https://slovakiana.sk/domov). The concept of virtual non-institutional encompasses phenomena far broader than the digitalisation of visual art or collections in state institutions. By non-institutional archive, I mean unofficial archives, e.g. personal archives that are not collected by any institution. This can be the case for various reasons (censorship, invisibility of some kinds of data

and artifacts). The non-institutional archive can be virtual or material. The concept of the virtual non-institutional archive refers to all online content (websites, email, messages, videos, etc.) and AI-generated data that is not stored by any cultural institution. I derive the concept of non-institutional archive from Derrida's general concept of the archive.

Derrida claims that the virtual archive includes everything we create and that this virtual archive existed before the internet. He says that any material creation is indeed always already virtual because it can be transmitted and it contains information that is open to interpretation. The means of transmission could be writing, sound, stone, wind, or any other material form. For Derrida, any kind of archiving and transmission of information is already a virtual technological process. As a result, even a museum archive consisting of what an art historian would call material is in fact a virtual artefact open to endless interpretation. In the Derridean framework, the difference between material and virtual registers is relativised in general. Derrida claims that the archive is not only what is stored in the museum, but also includes personal archives and other kinds of unofficial archives. He goes even further to say there is such a thing as a *psychic archive*, meaning that the human psyche constitutes an archive (Derrida 1998, 94). This brings us to the concept of the collective psyche.

In his later works, Martin Heidegger claims that being is *poiesis*, which means the creation of truth (*aletheia*). The Greek word *poiesis* refers to manual work (*handiwerk*), but also to artistic creation, thought, poetical thinking, and its dissemination. Heidegger claims that technology perverts the process of *poiesis*, which he believes should remain a manual process and, more specifically, be limited to German poetry (Heidegger 1977, 34). In the Derridean and Stieglerian view of *poiesis*, all *poiesis* is iterable and technological, because it is infinitely transmissible and it becomes part of collective memory. Therefore, *poiesis* is creation that is already technological and virtual, because any part of it can be repeated infinitely in an infinite number of contexts. For example, handwriting is already a technology, or even an oral transmission of a story or a fairy tale (in these cases, the speech or writing is the technology that allows us to repeat the story), but even what he calls a psychic spacing is a technology of psyche (Derrida 1998, 92; my translation).

THEORETICAL FRAMEWORK. DERRIDA: PSYCHIC ARCHIVES ARE ALREADY VIRTUAL

In our contemporary world, all creation has become digitalised and AI-powered, and this is possible because creation in general is already technical (iterable). This corresponds with the idea that there is some kind of online archive in which all our writings, images, affects, memories, and recordings are saved. And tools such as AI models use this archive to generate new images or texts. Therefore, the use of AI is another form of technology that makes transmission, which is inherent to the process of

creation (*poiesis*), even more accessible, just as writing or painting. In this way, our contemporary collective memory is archived by AI technologies and the online content we access: online platforms such as social media, e-commerce platforms, the gig economy, and search engines (these constitute what I call the non-institutional virtual archive).

In his book *Archive Fever,* Derrida (1998a, 19) examines the relationship of psychoanalysis to the archive, claiming: "The theory of psychoanalysis, then, becomes a theory of the archive and not only a theory of memory." He goes even further, saying that what is not archived is not lived or experienced in the same way and the technological aspect of archiving determines the nature of what is being archived (37). He shows that even as a "science" psychoanalysis has to deal with psychic archives that operate between *ὑπόμνησις* (mnemotechnical support, aid, or a mean), which is distinct from *μνήμη* (living memory) and *ανάμνησις* (forgetting, which is linked to the death drive). He contradicts Freud and says that these already contain a technological element (prosthetic element) (37). However, for Freud, the technological element equals the death drive. Nevertheless, Derrida claims memory is *αρχή*, by which he implies that "*The concept of archive encompasses, obviously, this memory of αρχη. But it also shelters this memory, which it encompasses: but we have to say, that it also forgets this memory*" (Derrida 2008, 12; my translation). Derrida claims that the concept of archive often forgets that it refers to the concept of *αρχή* in Greek which not only means origin (original [*ursprünglich*] data), but also *νόμος*, the law and social construction of the origin (of the archive). He reminds us of the genealogy of the word archive, which comes from the Greek concept *αρχή*. *Αρχή* can mean the origin of everything, but as Derrida reminds us, it can also mean commandment. Therefore, this word refers to the principle in which nature or history begins (physical, historical and ontological principle). At the same time, it refers to the principle of the law where gods and humans command, where the authority is exercised (nomological principle). This principle has a *topos*, a place from which it is exercised. Nevertheless, it can be very hard to localise that place and we have to think how the archive is 'taking place' [*Comment penser là? Et cet* avoir lieu *ou ce* prendre place de *l'αρχή*?] (Derrida 2008, 11). It can be especially hard to localise this authority that commands an archive, as it very often remains invisible, whether it is associated with a figure of divine or human authority. In the context of this paper, this authority is first and foremost the authority of capital in the AI industry, which has acquired a divine-like character in our contemporary imaginary as Big Other (Zuboff 2023). Derrida reminds us that we also forget that, in Greek and other ancient societies, there was a function of *άρχων* linked to the word *αρχή*. These archons were officials who archived important documents, and they had *hermeneutical right and competence*. This means they had the power to interpret the archives that often constituted laws. Nowadays, there is also a hermeneutical right and competence of the capitalist and state structures that govern AI and all virtual content, and by doing so, also our collective psyche and

affectivity. This implies the power to censor, and this censorship already takes place in the psychic mechanism, as Freud says. But Derrida (1998a, 9n1) reminds us of its political aspects. Throughout *Archive Fever,* he also reminds us of the dangers of forgetting and of the death drive linked to technological elements of psychic mechanisms. More precisely, it is the danger of the misuse of technology in concrete spatio-temporal conditions of conservation, linked to political conditions. However, he claims that these technological elements (repetition) at the same time enable the preservation of the living. Derrida describes this danger as archive fever. These technological elements are also the condition of archiving, and therefore, for the living, life (62). For the simple reason that what cannot be repeated cannot be transmitted (archived), it disappears, it is absolutely forgotten.

SLOVAK VIRTUAL ARCHIVES AND CONCRETE ANALYSIS OF INSTITUTIONAL POWER

Our contemporary state institutions also have this hermeneutical power to exclude some communities, and in the case of Slovakia these excluded communities are the Roma, Hungarian, Ukrainian, Jewish, Vietnamese, queer, and other marginalised communities and their affectivity and psychic archives. Hungarian archives also belong to the minority archives in Slovakia today. But it is also valid the other way around regarding the colonial effects of the Austro-Hungarian Empire, in which Hungary colonised the Slovak and Roma collective memory and language for centuries. Thanks to the Austro-Hungarian Empire, the Slovak "written" or otherwise materialised archives could only begin in the nineteenth century, even if the Slovak language and culture existed for a very long time before this. Roma communities in that period were obliged to settle and stop being nomads. Slovak and Roma cultures were totally erased, censored, and forgotten before the nineteenth century. In contemporary Slovak culture, there is ongoing marginalisation of Roma, queer, and other cultures, just as in other Central/Eastern-European countries. Oppression persists with a previously marginalised culture now oppressing another marginalised culture.

Here, we will specifically focus on the different ways in which Roma and queer culture has been marginalised in Slovak virtual archives. My aim is not to compare or hierarchise different forms of oppression, but to analyse shared archival and technological mechanisms of invisibility of these two communities. It is important to emphasise that Roma communities, just as queer communities, are not a homogeneous group, but consist of many smaller groups. I chose to focus mainly on queer Roma artists who are members of both communities. These two communities have different histories, practices, and forms of cultural production, and they face very different struggles and ways of being made invisible. What links them, however, is that institutional representation—particularly in artistic and museum settings—is currently setting a precedent. The

choices being made today regarding what is archived, digitised, classified, or omitted will have lasting effects, extending beyond digital traces, on future understandings of Roma culture and contemporary Roma art.

In our contemporary world, we nevertheless have the possibility to use more modern technology to preserve certain lives and lives in general, i.e., to preserve the archives of today's minorities. However, the danger of the death drive (of being forgotten) hidden in technology is "in-finite", as Derrida says. In this paper, I will focus on the Roma and queer community and the danger of being forgotten that they are facing because they constitute the biggest and most attacked community, even if there are differences in the ways in which they are being forgotten as we will show in the conclusion.

It is important to address the question of collective memory, trauma, and the exclusion associated with it. I would like to examine ways in which the non-institutional virtual archive has a poietical potential and how to unlock its political potential. Today we face increasingly more hegemonic effects of the perversion that this non-institutional virtual archive carries within itself. However, we need to unlock the possible anarchical and anti-hegemonic effects of this perversion within the non-institutional virtual archives that feed AI models. In the next section, I will first clarify what I mean by anti-hegemonic affectivity that the process of *poiesis* can bring forward. Then I will point out examples of artistic practices that, through their anti-hegemonic affectivity, resist techno capitalist censorship and institutionalised racism.

HEGEMONIC PERVERSIONS OF TECHNOLOGY: ERASURE OF ROMA AND QUEER CULTURES IN SLOVAKIA

French philosopher Bernard Stiegler claims virtual networks that are creating and participating in this virtual archive contribute to the erasure of diversity of experience and affectivity. This is linked to the hegemonic effects of the perversion within the non-institutional virtual archive that is nowadays governed by techno capitalism. The capitalist virtual archive incorporates psychic traumas such that they are absorbed by simulacra, which are increasingly sophisticated and perfected (Stiegler 2009, 93–96). This leads to the erasure of individuality in the virtual archive, and this corresponds with the hegemonic perversion of technology. Technologies participating in the virtual archive are, according to Stiegler, also transforming our relationship to memory and the past. Technology is in Stiegler's view a supplement to our faculties of imagination, affectivity, and thinking. As Derrida (2012, 168) also observed, this supplementary character of technologies stems from the spectral character of their representation, but he claims that the spectral character of technology can also be politically emancipatory. Similarly, Stiegler (1998, 61–67; 2017) imagines not only the negative effects of technologies but also their redemptive potential, even while remaining quite critical of AI because it remains embedded in capitalist structures. Nevertheless, technology,

understood as poiesis, has the potential to reinvent human existence and culture, and, equally, to reproduce human spiritual and symbolic values. It is important to understand the virtual archive not only as something that erases human affectivity, but also as something that can create new forms of affectivity. This is only possible provided we critically examine the hegemonic aspects of surveillance techno capitalism which can lead to censorship because of the hermeneutical power and competence they dispose of (state, big corporations, etc.).

In Slovakia and the Czech Republic, there has been a long-lasting erasure of Romani art from institutions such as the National Gallery in Prague and which is illustrated by the statement by Knížák, the gallery's director in 2002: "Romani works do not reach the standard of quality worthy of being exhibited in the National Gallery" (Ludlová and Rigová 2017, 4). He claimed that Roma art is appropriate for ethnographic museums or Romani museums, but should not hang next to baroque paintings because they are merely "folk pictures". Knížák therefore used his hermeneutical power and competence to create a very narrow definition of national cultural heritage and explicitly stated what he excludes from the archives. Here, we can observe the hegemonic perversion of the archiving technology at his disposal.

As a result of this power, Roma art continues to be excluded from institutional archives. In Slovakia, we recently observed a form of soft or hard censorship, especially in relation to marginalised communities such as the LGBTQ+ community (this concerns artists such as Andrej Dúbravský or Dorota Holubová).[3] Recently, there have also been budget cuts for many queer events such as Košice Pride and the Drama Queer Festival. The right-wing nationalist party SNS has even installed billboards saying "we have cut all funding for LGBTQ+ projects" around Slovakia. This is in direct contradiction with the Slovak constitution, which forbids any kind of discrimination towards minorities (law 365, paragraph 2) including minority communities based on sexual orientation. The new minister of culture Martina Šimkovičová has dismissed almost all of the leaders of the most important state cultural institutions such as the Slovak National Gallery, the Slovak National Museum, the Slovak National Theatre and she has made the functioning of the most important funding source the Foundation for Art (Dond na podporu umenia) practically impossible. There have been many initiatives created and several protests by the employees of these institutions (creation of the platform Open Culture). The most obvious case of censorship happened in Bratislava castle (belonging to the Slovak National Museum) where Dorota Holubová's exhibition *Neskrývaná láska* was banned. This exhibition mapped the lives of queer people living in contemporary Slovakia. According to the artist, Minister of Culture Martina Šimkovičová said it was LGBTQ+ propaganda. In their official statement the Slovak National Museum denied that the exhibition would be a censored. The fact remains that under the new direction of the Slovak National Museum, the exhibition was cancelled. Holubová has exhibited these artworks abroad and she published the

[3] *For the definition of soft censorship see Tompa (2021) who analyses censorship in art during Orban's administration.*

exhibited photos in a book supported by the Netherlands embassy and the Foundation of the city of Bratislava. She is a long-term advocate for the LGBTQ+ community. (See the exhibition *Sami sebou*, raising awareness of trans people in Slovakia https://sutaz.slovak-press-photo.sk/SK/sutaz-detail-foto?set=217&photo=1188).

Artists advocating for the rights of the LGBTQ+ community are mostly targeted by hate comments, budget cuts and even hard censorship. The attacks of ultra-right-wing politicians, amongst whom we count also our prime minister, Robert Fico, and almost the entire coalition in the current government, revolve around the idea of national identity. Therefore, the censorship or even attacks, seems to concern all the artists who try to redefine Slovak identity, as is the case of Denise Lehocká. Lehocká's artwork was almost destroyed and removed from the Slovak National Gallery's permanent collection without her consent or knowledge in August 2025. In her installation, she tried to define Slovak identity through motifs of thread-making, potatoes, traditional textiles, and embroidery with obvious reference to the feminist movement of craftivism. Another attack that Minister Šimkovičová orchestrated targeted many of the aforementioned artists, including Roma Artist Emília Rigová. Šimkovičová posted a reel from her visit of permanent collection in Slovak National Gallery, trying to point out that Slovak art is perverted because it is focused on genitalia (https://www.facebook.com/reel/1317623822805886/). Her reel shows artworks from the permanent collection that actually problematise the fetishism of genitalia, such as works by Anna Daučíková, who, as a trans person, is rethinking her own relationship to their chest, and, for me as a trans person, this thematises dysphoria (https://www.webumenia.sk/en/dielo/SVK%3ASNG.IM_916-4). The reel finishes with a close-up of Rigová's video, in which she vomits gold. In this artwork, Rigová critically reflects upon Roma art and its use of gold. There are more artists whose works were endangered by the new direction of the Slovak National Gallery such as Jozef Sušienka (removal of his statues from the exterior of the National Gallery, https://dennikn.sk/minuta/4795317/), Jiří Franta and David Böhm (whose mural in Zvolenský zámok was destroyed, https://dennikn.sk/4714130/na-zvolenskom-zamku-znicili-dielo-od-ceskych-umelcov-frantu-a-bohma-skoncilo-v-kontajneri/). All this led to the cancellation of the planned exhibition of contemporary art *Model: Múzeum súčasného umenia in SNG* in the National Gallery.

The LGBTQ+ community in Slovakia is currently under constant media attacks in which the current government does not hesitate to call them sick and deformed. All of these events are preceded by a terrorist attack on two queer people in 2022: Matúš Horváth (a gay cis man) and Juraj Vačulík (a drag performer identifying as a nonbinary person).

These examples only underline the power of institutions to judge what is art and what is not. This can lead to the creation of exclusions and censorship tools (Tomková 2025, 19). Unfortunately, as Ahmed claims, there is a straightening and whitening device in place in institutions that

we have to reform (Ahmed 2012, 173–174; Ahmed 2020). This message and exclusion are also conveyed through collecting, archiving, and displaying art, but it concerns the virtual art archives.

However, in today's rapidly changing world, the display of art is no longer limited to the white cube format or in situ installations. Art is exhibited in online spaces and uses new technologies and AI tools; it is indeed part of the virtual archive. Therefore, there is a need to rethink the forms of exclusion not only from the "white cube" but also from institutional virtual archives, which offer more accessible ways to display art today. From a theoretical point of view, this is because *poiesis* (creation in general) already contains a technological element; therefore, art and technology are no longer in opposition, as we showed in the previous section.

On the other hand, virtual archives (institutional and non-institutional) can be very important tools for Slovak institutions and artists in resisting censorship because they are much harder to control by government censorship and are very accessible. For example, the Slovak government can ban various offline exhibitions, but the online archive can still be used to represent marginalised artists and disseminate their work even more broadly in the virtual space. The advantage is that the institutional virtual archives are publicly accessible and artworks can be downloaded from there for free by anyone, which is also the case for webumenia.sk. For Derrida (2008, 15): "The effective democratisation is always measured by these essential criteria: the participation and access to archive, to its constitution and interpretation." We could say that a freely accessible virtual institutional archive is a condition for democracy and freedom.

However, current Slovak virtual archives must first be subjected to more questioning. For example, the institutional archive of state-funded Slovak museums and art galleries does not include enough and appropriate representation of works of Roma and queer artists (specifically webumenia.sk), and it is practically impossible to find works of Roma and queer artists that do belong to the archive on the webpage. A simple search for Roma, queer art, or Roma holocaust returns very little on the website of webumenia.sk and there is no visible section of the webpage dedicated to Roma or queer art. Moreover, many works of Roma artists are not digitised (the ones we can find are mostly on the freely accessible virtual museum archive Slovakiana) and the content of the works that can be found mostly consists of ethnographic representations of Roma folk culture with very few entries from the contemporary Roma art scene. The situation is slightly better for queer white contemporary Slovak artists, who have more representation on the webpage of webumenia.sk (e.g. Dubravsky and Daučíková). But again, there is no specific category or tag dedicated to queer art. This contributes to making Roma and queer art invisible and in this case, the virtual archive functions like a soft censorship device, or as Ahmed puts it, a whitening and straightening device. This has a direct impact on the perception, reproduction and

dissemination of Slovak art, contributing to a discriminatory definition of what is Slovak art and culture operating by censorship devices. For example, these institutional virtual archives should be used in AI tools such as AI image generators, as we outlined in the introduction.

EXAMPLES OF ANTI-HEGEMONIC USES OF AI BY ROMA ARTISTS: POLITICAL EMANCIPATION AND VIRTUAL SPACE AS A RESISTANCE TOOL

The emancipatory character of artistic production has been highlighted by many theoreticians, including Jacques Rancière, Chantal Mouffe, Denisa Tomková, Ewa Majewska, and Grant Kester. Even if Roma and queer art is rendered invisible in Slovak institutional virtual archives (soft censorship), many Roma and queer artists use AI and virtual technologies in their work as a tool of poietic emancipation. Roma and queer artists use technology to reshape the non-institutional archive and thus fight for the place they deserve in the virtual archives. Even these non-institutional virtual and AI technologies are developed by capitalist institutions (corporations) and they are built on West centric, colonial and capitalist values. As we will see, these originally capitalist-oriented and inherently racist technologies can still be used as a tool of emancipation to some extent. Alternative models that are more sensitive not only to the Central/Eastern European context, but also to its marginalised communities, can be built. The use of virtual networks and technology, archiving, and the preserving of marginalised cultures can have an emancipatory aspect for these cultures (Tomková 2025, 29).

We can observe an example of the emancipatory and anti-hegemonic uses of technology in the works of Roma artist Mihaela Drăgan. In her "Roma Futurism Manifesto: Techno-witchcraft is the Future", she combines witchcraft with technology and magic. Cyber witches, in her view, create a more egalitarian and democratic world, and they are the key figures of Romafuturism. The Roma Futurism Manifesto is a good example of a new kind of affectivity using AI technologies that invites the use of technology as a tool of empowerment for the marginalised Roma community. Drăgan proposes understanding AI as something magical since we do not understand exactly how it works (Parisi 2016). She replaces traditional witchcraft tools with modern technological devices. She refers to:

> *live transmissions of rituals and spells through the internet, virtual tarot and to healing through technological tools, […] shamanelism rituals (rituals which make the [sic] use of manele [a genre of Romanian pop folk music] to create a gypspiritual experience through music and dance) and virtual psychedelic feelings (the fusion between biologic and technologic with the purpose of self- knowledge and personal development).*" (Drăgan 2018)

This implies a belief that technological entities are inhabited by spirits and that the internet is itself an independent and strong spirit that can be

used for diffusing antiracist and anti-discriminatory practices to liberate Roma from oppression (using ethical hacking, for example). The main character of this movement is Cyber-Witch, who fights against Roma oppression and has a superpower to transcend time and access the past in order to create alternative histories. "They will offer a performative answer to the question: If this oppressive past had never existed how would Roma communities have evolved?" (Drăgan 2018) In her manifesto Drăgan directly challenges the censorship, whitening, and straightening device of archival work, whether artistic or historical.

Another example of the emancipatory and anti-hegemonic use of virtual technologies can be observed in the works of Ezra Šimek, a trans* nonbinary artist who was born in Germany, studied at the Slovak Academy of Arts, and lives in the Czech Republic. In their 2020 work *No offense but* Šimek used Instagram live to deconstruct prejudices against trans and non-binary people, creating a parody of a TED talk. Šimek is the TED talk speaker in formal clothes who explains trans and nonbinary identity and they simultaneously play the role of an Instagram live audience that keeps interrupting asking them with very stereotypical questions about queer identities. Their aim was to connect to more international online queer community and to expose harmful stereotypes, underlining that even if trans people form a community their individual stories cannot be generalised as stereotypes. Šimek uses Instagram, a platform accessible to many people to disseminate their message, to reclaim pop-cultural language and social media tools. Šimek claims society faces a "dramatic political divide" and a lack of understanding of the complexities of the situation (Tomková 2025, 103–4).

The work of Slovak Roma artist Robert Gabris also highlights the connection between technology, feminism, and antiracial struggles. Gabris's work inscribed into Glitch feminism movement that embraces the use of technology in art. In their six month long project *Error, Roma Corporeality and Their Non-Binary Spaces*, Gabris used dating apps to connect Roma queer people living in excluded and marginalised spaces who would not be able to otherwise meet. However, "normal" usage of LGBTQ+ dating apps is a space of exclusion for the Roma queer community that quite often perpetuates racism, sexism, and sexual violence. The project *Error* started at the K.A.I.R residency in Košice. First, they met online, and Gabris presented them the project's idea before they met in person for the first time (emphasis on mutual trust and agreement and safety of the participants). As a part of the exhibition, the participants formulated their collective demands, which were then embroidered on ribbons and on a larger textile. It declares the following: "ROMA CORPOREALITY BECOMES A RADICAL TECHNOLOGY OF SELF-ARMORING". This resonates with the idea that our bodies already belong to the virtual archive because they are already technological and technical devices. Therefore, representation of these bodies through technological means is not itself a violation of a body's image. However, a bad representation is a violation, whether virtual or not. For example, Gabris connects this

problem to a representation of sex workers and of queer Roma people that can be represented in collective memory as worthless and dirty or despicable. He claims that these identities can be destigmatised by the sex workers themselves, who can use them as a weapon, just as Drăgan invites us to do. Sex work can become a tool to fight against white patriarchy, using its weakness and transforming it into strength. The textual part of Gabris' exhibition is the following:

> *Roma corporeality has become a radical technology of self-defence.*
> *We strategically use the body as a material, the material as a tool, the tool as a weapon against your heteronormative linearity.*
> *We* have strategically learned to use ERROR for self-defence.*

The curatorial text by Katerina Kottova explains:

> *This arrangement also asks art institutions, attempting to become more inclusive places than they have been traditionally, to assume an even more radical position—to offer a space within their inclusivity for something quite exclusive; to provide a territorial space they themselves cannot enter, only assist from outside with humility. The physical installation in the gallery is just one element in the whole project.*

Not only is this a call for inclusion in the sense of making space for marginalised groups within the normative space, but also for the creation of closed spaces dedicated exclusively to queer Roma art and other marginalised groups, managed by them only. This implies the need to include queer Roma art in the institutional virtual archives of art institutions, and to give them "prime time" space.

CONCLUSION: CALL FOR DIVERSIFICATION OF DATA FROM V4 COUNTRIES: HOW NOT TO REPRODUCE THE INSTITUTIONAL RACISM REFLECTED IN DATA?

In the case of Slovakia, institutional archives such as webumenia.sk represent internationally successful queer white artists such as Dúbravský and Daučíková, who directly face hate speech or even hard censorship from Slovak politicians, or stars such as Andy Warhol or Ladislav Mednyánszky (a nineteenth century queer painter; Tamásová 2021). It is important that public access to these archives remains free because this will make it harder for governments to control them. In Slovakia, virtual institutional archives remain unnoticed by the government because they focus on non-virtual forms of art, unaware of the increasing role of technologies in the contemporary art scene. On the other hand, Slovak institutional virtual archives do not sufficiently include all marginalised artists, missing the opportunity to provide increased visibility in a regime with authoritarian tendencies. Roma artists are not represented in Slovak virtual archives, and they are reduced to ethnographic material, often represented by white artists (More Roma artists can be found on the slovakiana website than on webumenia.sk).

Nevertheless, marginalised communities are already part of a non-institutional virtual archive in a broader sense, and they can use this to their advantage, often subversively. Technology does not acquire only the negative meaning of an erasure of the heterogeneous affectivities. As we tried to show, it can generate new affectivity or a virtual safe space for excluded affective frameworks of marginalised communities such as the Roma and queer community. I specifically focused on Roma art (which is also often queer art) because it is particularly left out of Slovak virtual archives. This space for new forms of resistance by anti-hegemonic forms of affectivity is possible also thanks to technology, as it allows participants to reflect on individual and collective traumas and feelings of guilt and shame (e.g. dating apps, and social platforms in Gabris' or Šimek's artworks). The possibility to reflect on these affects allows the queer Slovak Roma community to move forward and formulate demands in a manifesto. We can conclude that marginalised authors resist being made invisible and inscribe their work into the non-institutional virtual archive, even if virtual archives leave them out. These artists view technology as their tool, allowing them to revisit the divide between straight bodies and queer, racialised bodies (such as Roma, trans*, and nonbinary bodies). Therefore, in their work technology becomes a tool of poietical resistance to oppression and erasure of their affectivity by the straightening and whitening devices of state institutions and big techno capitalist structures. The practical advantages of using non-institutional archives to present and store their work are better visibility, accessibility, lower expenses, and the possibility of avoiding censorship (if they are also banned online, they can move their content to another location, VPN, or similar). Whether the hegemonic character of general non-institutional virtual archives (the internet and social media platforms) governed by big techno capitalist structures really offers them greater visibility remains an open question. Secondly, it is not clear that all artists trust these platforms; some are afraid of losing their authorship. Nevertheless, in the abovementioned artworks, we have seen that some artists use online social platforms as tools for community building rather than simply storing their artworks.

PRACTICAL IMPLICATIONS FOR RESEARCH IN TECHNOLOGY AND FOR AI MODELS DEVELOPMENT

Virtual space can therefore become a space of emancipation offering tools for poietical political resistance to oppression. However, if we do not revisit the exclusionary and censoring practices embedded in contemporary institutional and techno capitalist techniques of virtual archiving, these technologies will continue to perpetuate these exclusions exercised by institutions and reinforced by governments. This concerns both archives of museums and galleries and techno capitalist structures that constitute a big archive, as explained above. There are existing or emerging precedents of alternative institutional and semi-institutional

platforms and initiatives such as the Roma Pavilion project at the Venice Biennale in 2007, the Museum of Romani Culture in Brno (https://www.rommuz.cz/en/), and contemporary Central/Eastern European initiatives such as the Romani Design activist art collective (https://romani.hu/en/about-us/) which includes artists such as Erika Vagra, Helena Varga and theoreticians such as Lili Kristen (2024). We should also mention here the K.A.I.R. residence project for young artists from Central/Eastern European countries in Košice, which invites collaboration with local Roma communities (https://www.kair.sk/), and creates space for young Roma artists such as Gabris or Júlia Csapó from Hungary. Non-institutional Romani art also collaborates with music platforms, especially queer art and a political, activist art that is very important in contemporary Slovak pop or more underground music culture. For example, the singers Vojtík and ERØ (Šlonerová and Žigmund 2023), Fvck_cvlt, and Čavalenky. Fvck_cvlt is an electropunk/metal singer who criticises oppressive politicians at their concerts and encourages the attendants of the concert to vote and not to fall into lethargy. Their songs directly thematise their Roma identity and the joys and struggles it entails, which is also a theme in the music of Čavalenky and Vojtík.

In some cases, we first need to create virtual archives from what are called "material" archives so they can be incorporated into AI models by fine-tuning. There are doubts about the data used in the development phase of the model and whether fine-tuning or model interpretation (the process of understanding, explaining, and visualising how machine learning models work and their biases) can solve all the problems since the primary data and their management are probably already biased. However, we need a philosophical and politico-ethical framework and a critique of AI to reflect on what diversity in AI can mean. The AI models do not use all internet content equally with most focusing on English or Western-centric content (Couldry and Mejias 2019; Johnson et al. 2022). The newest models produced by OpenAI, Google, and Microsoft do not make public the data sets they use; therefore, their data usage is not transparent (not even to the research community). Current conditions of archiving recreate hegemonic AI structures closely tied to capitalism and reproduce an Anglocentric, racist philosophy embedded in technological practices as such. As Benjamin (2019, 78) says, "Computer systems are a part of larger matrix of systemic racism". AI mechanisms currently only continue to reproduce the systemic racism already embedded in techno capitalist structures and in existing archiving practices (for example, the relationship of coloniality to written archives).

However, even if we decide to diversify the existing AI image generators (e.g. Stable diffusion, Flux, SD35 Large /Medium, and SDXL) and fine tune them, we must first question what kind of marginalisation will result in relationship to already marginalised cultures (towards Western Europe), such as Slovak, Polish, Czech, and Hungarian cultures, as we aimed to do in this text. Not only is there a lack of representation of the V4 cultures in AI generators, but there is also a lack of racialised and

otherwise marginalised cultures within V4 countries. And this lack of representation and invisibility is reflected in V4 archives (digitalised or not) and across Europe (Junghaus 2025). If we use already biased (racist or homophobic) data to fine-tune the models to be more sensitive to Central/Eastern European cultures, it will not help diversification. We will only continue the work of the whitening and straightening device (Ahmed 2020, 72).

There exist alternative archives with fair data, such as Secondary Archive (https://secondaryarchive.org/), that fill gaps in institutional archives and represent queer, Roma, and feminist Central/Eastern European art. But they are not tied to any particular state art institution. The non-institutional virtual archive and alternative archives, such as the Secondary archive and all content created by marginalised communities, should be better integrated into AI mechanisms. This data should also be used as primary data in the first phase of the training, especially because trained models are known to continuously learn from the online content. These institutional and non-institutional archives are now part of the globally accessible online content so they should be incorporated into the AI models at every stage of their development.

To conclude, I will formulate a few clear philosophical guidelines, even if most of them cannot currently be implemented because there is very little will to do so on the part of techno capitalist structures that keep AI development completely untransparent. However, it is important to formulate these explicitly. First, we need participatory data design. This means that diverse data should not be only included in the stage of fine-tuning as is usually suggested, but it should be included from the start, from the very first phases of training. The diverse data should be collected through participatory experimental practice invented by the marginalised communities in question, as for example the artists mentioned in the previous section invited us to do. We need to ask in every context, what community practices facilitate anti-hegemonic data collection and how can they be used in any contexts? Secondly, we need dynamic data monitoring; model outputs should be continuously evaluated across various cultural contexts. Thirdly, we need the datasets to provide rich metadata about their origins and to rethink the criteria used to select the data, to ensure data transparency and avoid hegemonic power structures to govern AI models (Iman 2025, 2–3).

REFERENCES

Ahmed, Sara. 2012. *On Being Included: Racism and Diversity in Institutional Life*. Durham: Duke University Press.

Ahmed, Sara. 2020. *Queer phenomenology: Orientations, objects, others*. Durham: Duke University Press.

Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Oxford: Wiley.

Couldry, Nick, and Ulises A. Mejias. 2019. *The Costs of Connection: How Data is Colonizing Human Life and Appropriating it for Capitalism*. Stanford: Stanford University Press.

Derrida, Jacques. 1998a. *Archive Fever: A Freudian Impression*. Translated by Eric Prenowitz. Chicago: University of Chicago Press.

Derrida, Jacques. 1998b. *Of Grammatology*. Translated by Gayatri Chakravorty Spivak. Baltimore: Johns Hopkins University Press.

Derrida, Jacques. 2008. *Mal d'archive: une impression freudienne*. Paris: Galilée.

Derrida, Jacques. 2012. *Specters of Marx: The State of the Debt, the Work of Mourning and the New International*. Translated by Peggy Kamuf. London: Routledge.

Drăgan, Mihaela. 2018. "Roma Futurism Manifesto: Techno-Witchcraft is the Future." Translated by Ioanida Costache. *Revista Arta*. https://revistaarta.ro/wp-content/uploads/2021/03/roma-futurism-manifesto.pdf

Heidegger, Martin. 1977. "The Question Concerning Technology." In *Basic Writings*, edited by David Farell Krell and translated by W. Lovitt, 287–317. New York: Harper & Row.

Heljic, Arman. 2021. Staging the Romani Queer Revolution: New Approaches to the Study of Romani Queerness. *Critical Romani Studies* 4 (1): 36–55.

Iman, Nofie. 2025. "Recontextualising Diversity in AI: Derridean Reflections on Data Colonialism and Plural Epistemologies." *Philosophy & Technology*, 38 (4): 153. https://doi.org/10.1007/s13347-025-00998-y

Johnson, Rebecca L., Giada Pistilli, Natalia Menédez-González, Leslye Denisse Dias Duran, Enrico Panai, Julija Kalpokiene, and Donald Jay Bertulfo. 2022. "The Ghost in the Machine has an American Accent: Value Conflict in GPT-3." Preprint. *arXiv*. https://arxiv.org/abs/2203.07785

Junghaus, Timea. 2025. *Strengthening Inclusion of Roma in European Museums and Cultural Institutions*. [Berlin]: ERIAC European Roma Institute for Arts and Culture. https://eriac.org/wp-content/uploads/2025/12/Junghaus_Strengthening-Inclusion-of-Roma-in-European-Museums-and-Cultural-Institutions.pdf

Kriston, Lili. 2024. "Változást akarsz? Akkor neked való a Romani Design kiállítás Szegeden." *Glamour,* May 12. https://www.glamour.hu/divathirek/romani-design-kiallitas-szeged/6eqhg76

Kuchtová, Alžbeta. 2024. "The Incalculability of the Generated Text." *Philosophy and Technology* 37 (1). https://doi.org/10.1007/s13347-024-00708-0

Ludlová, Nikoleta, and Rigová, Emília. 2017. *Creative Manual for Contemporary Romani Art*. Košice: Dive Buki.

Parisi, Luciana. 2016. "Instrumental Reason, Algorithmic Capitalism, and the Incomputable." *Multitudes* 62 (1): 98–109.

Šlonerová, Kristína. 2023. "Erik 'ERØ' Žigmund." Interview. Úrad splnomocnenca vlády SR pre rómske komunity, November 21. https://romovia.vlada.gov.sk/antidiskriminacna-rubrika/erik-ero-zigmund/?csrt=17002540207937872290

Stiegler, Bernhard. 1998. *Technics and Time, 1: The Fault of Epimetheus*. Translated by Richard Beardsworth and George Collins. Stanford: Stanford University Press.

Stiegler, Bernhard. 2009. *Technics and time, 2: Disorientation*. Translated by Stephen Barker. Stanford: Stanford University Press.

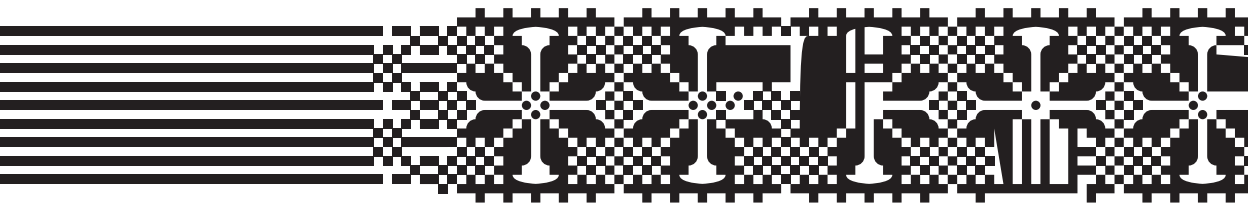
Stiegler, Bernhard. 2017. "The New Conflict of the Faculties and Functions: Quasi-Causality and Serendipity in the Anthropocene." *Qui Parle 26 (1)*: 79–99.

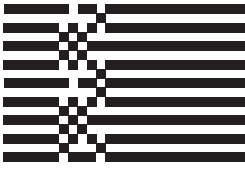
Tamásová, Alexandra. 2021. "7 vecí o queer umení v zbierkach slovenských galérií." *Medium*, October 6. https://medium.com/sng-online/7-vec%C3%AD-o-queer-umen%C3%AD-v-zbierkach-slovensk%C3%BDch-gal%C3%A9ri%C3%AD-3f00d6874f30

Tomková, Denisa. 2025. *Empowering Aesthetics: Contemporary Art from Post-Socialist Central Europe*. London: Bloomsbury Publishing.

Tompa, Andrea. 2021. "Silencing: Soft and Self-Censorship. The Case of Hungarian Performing Arts." *Concept* 22 (1): 50–60.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* New York: PublicAffairs.





About the Authors

Ania Malinowska is a cultural theorist, writer, and Professor of Media and Cultural Studies at the University of Silesia in Katowice, Poland, where she co-directs the Centre for Critical Technology Studies. Her work focuses on technoculture, emotional semiotics, and the evolving entanglements between humans and machines. She probes speculative methods which interrogate the cultural imaginaries and ontological shifts introduced by intelligent machines. She is also co-developer of the artistic research project Hypnotic AI (LINK), which investigates machine sentience through hypnotic induction. Malinowska's interdisciplinary practice spans critical writing, curatorial work, and experimental methodology, rethinking digital subjectivity, posthuman hermeneutics, and the interpretive challenge posed by nonhuman intelligence.

Michał Krzykawski is an university professor in philosophy at the Faculty of Humanities of the University of Silesia, Poland, where he heads the Centre for Critical Technology Studies. His research revolves around philosophy of technology, social ecology and political economy. Coordinator of the research field Social Framework for AI-Based Systems at Open Eyes Economy Hub, vice-president of the foundation Pracownia Współtwórcza (Contributory Lab), member of the Council of the National Programme for the Development of Humanities. Recently published works: *Bifurcate. "There is no Alternative"* (edited by Bernard Stiegler with the Internation Collective, Paris 2020, London 2021) and (in Polish) *The Economy and Entropy. Overcoming the Polycrisis* (co-edited with Jerzy Hausner, Warszawa 2023).

Brigitta Iványi-Bitter is an art historian, cultural studies scholar, whose interdisciplinary career bridges academic research, digital innovation, and heritage practice. As Principal Investigator of the project "*Impact of AI on preserving the cultural heritage of low-resource languages*", she leads a pioneering research project advancing cultural and linguistic equity in image-generating AI systems with special regard to the Visegrad countries. Holding a PhD in Film, Media, and Cultural Studies and MAs in Law and Political Science as well as in Art History, she is uniquely positioned to address the ethical, representational, and legal dimensions of cultural data and artificial intelligence.

Tibor Bacsi is a senior software engineer with more than twenty years of professional experience. His work supports computational research, generative AI experimentation, and the technical development of visual datasets. In this project, he contributed to the generative experimental setup, prompt strategies, image-database curation, and resource management.

Szilárd Szakács is a senior UX/UI expert, software engineer, and data-analysis specialist with more than twenty years of professional experience. His work focuses on automated analytical systems, statistical pipelines, and quantitative research infrastructures. In this project, he contributed to the survey design, analytical software, statistical pipeline, formal data analysis, and data visualisation.

Kateřina Marková is a PhD candidate at the Centre for Ethics as Study in Human Value (CE) at the University of Pardubice. She is a PhD level researcher at the Centre for Environmental and Technology Ethics – Prague (CETE–P), where, as part of the project "Human-centered AI for a Sustainable and Adaptive Society," she researches the impact of digital content—curated through algorithms guided by the principle of profitability—on the autonomy of vulnerable individuals and their communities. Kateřina holds an MA in multimedia from the Faculty of Engineering of the University of Porto. She has over 15 years of experience as a designer in the tech industry. Her prior research focused on the management of personal data in connection with death, from ethical, environmental, and economic perspectives.

Anna Keszeg is an Associate Professor at the Moholy-Nagy University of Art and Design, where she also leads the Heritage in Motion Lab. With a background in the humanities (literature, philosophy, history), her research focuses on visual culture and fashion studies, particularly through the lens of popular geopolitics. Her most recent book, published in Hungarian, explores the transmedia strategies of the fashion industry. She has contributed to a range of international journals, such as European Review, Journal of European Popular Culture, Fashion Highlights Journal, etc.

Jiří Philippe Janda is a Czech visual artist, architect, and doctoral researcher at the Faculty of Art and Design at Jan Evangelista Purkyně University in Ústí nad Labem. His work focuses on generative AI as a cultural and aesthetic system, with a particular interest in synthetic images, regional visibility, and AI-driven moving image. Through practice-based research, he develops AI-generated films and speculative visual projects exploring the relationship between human and machine creativity.

David Kořínek is an artist and university professor. Since 2022 he has been working at the Department of Art History and Theory at the Faculty of Art and Design, Jan Evangelista Purkyně University in Ústí nad Labem, where he was appointed full professor in 2024. Since 2023 he has served as Vice-Dean for External Relations and Internationalization at FUD UJEP. Until 2022 he was Head of the Centre for Audiovisual Studies at FAMU in Prague. He also teaches at UMPRUM Prague and Scholastika, Prague. In 2008 he co-founded the Supermedia Studio at the Academy of Arts, Architecture and Design in Prague together with Federico Díaz and led it until 2018. At the Department of Media Studies at Masaryk University he established the Digital Media programme and headed the Media Lab. He has worked as a dramaturg at Czech Television, with which he has long collaborated as a director. Since 2007 he has been a member of the artist group Rafani, which regularly exhibits in European galleries and institutions. The group's work spans a wide range of media, from gallery installations and public-space projects to videos and feature-length documentaries. Rafani has received numerous awards, and its works are held in both institutional and private collections. David Kořínek focuses on the theory of the moving image in relation to visual art and is the author of scholarly texts published in art journals and edited volumes. In his academic practice he concentrates on contemporary art in relation to audiovisual media (video, film, new media, post-internet art, performance, etc.).

Albín Kuchta (also known as Alžbeta Kuchtová) is a researcher at the Slovak Academy of Sciences, Institute of Philosophy in Bratislava, Slovakia. He focuses on French Philosophy and Phenomenology, Environmental Philosophy, Posthumanism, Philosophy of Technology and Post-soviet Feminism. He works on translations (E. Levinas, J. Derrida) from French to Slovak. He is the author of *The Ungraspable as a Philosophical Problem* (2024) and the co-editor of *Repenser la logique du vivant après Jacques Derrida* (2024). His latest publications include "La radicalité du manger" in *Journal of French and Francophone Philosophy* (2024), "The Incalculability of Generated Text" in *Philosophy & Technology* (2024), and "Humanity as a New Image of the Divine Absoluteness," a chapter in *The Modern Experience of the Religious* (2023).

Visegrad Fund

THE PROJECT IS CO-FINANCED BY THE GOVERNMENT CZECHIA, HUNGARY, POLAND AND SLOVAKIA, THROUGH VISEGRAD GRANT FROM THE INTERNATIONAL VISEGRAD FUND. THE MISSION OF THE FUND IS TO ADVANCE IDEAS FOR SUSTAINABLE REGIONAL COOPERATION IN CENTRAL EUROPE.

Project ID #/Title: 22510385