Cultural Diversity and Generative AI in European Creative Production

Evaluating Linguistic Asymmetry, Epistemic Flattening, and Localization Overhead

Marianna Csarnai
Associate Professor; Programme Director, Executive MBA in Digital Marketing
Budapest Metropolitan University (METU), Budapest, Hungary

SCOPE

Operationalisation and cross-national verification of Hypothesis H4 across eight European creative hubs

CONTENTS

8 sections · 6 tables · 4 figures · 5 equations · 38 references

REFERENCING

American Psychological Association, 7th edition (author–date)

Contents

1 Theoretical Foundations: Algorithmic Homogenization and Linguistic Disparity 3

2 Formulation of Hypothesis H4 and Operational Deconstruction 5

3 Cross-National Experimental Architecture Across European Creative Hubs 5

3.1 Evaluator Sampling and Professional Stratification 7

3.2 Standardized Task Typology and Localization Protocols 7

4 Operationalization and Multi-Metric Scoring Protocols 8

4.1 Cultural Fit (CF) 8

4.2 Language Quality (LQ) 9

4.3 Stereotype Rate (SR) 9

4.4 Human Correction Time (HCT) 10

4.5 Acceptance Rate (AR) 10

5 Metric Protocol Specifications 11

6 Projected Empirical Trajectories and Asymmetric Production Costs 12

6.1 Qualitative Failure Dynamics and Metric Interactions 13

6.2 Second- and Third-Order Socio-Economic and Cultural Consequences 14

7 Structural Governance, Regulatory Enforcement, and Cultural Safeguards 15

8 Synthesis and Strategic Outlook 16

References 17

Tables and figures

T1 Operational sub-hypotheses of H4 5

T2 Stratification of the eight national hubs 6

T3 Standardized multi-metric framework 8

T4 Metric protocol specifications 11

T5 Projected empirical trajectories by hub 12

T6 Governance frontiers and required interventions 15

F1 Causal architecture of algorithmic homogenisation 4

F2 Cross-national experimental architecture 7

F3 Projected generative performance across hubs 13

F4 The self-reinforcing loop 14

1 Theoretical Foundations: Algorithmic Homogenization and Linguistic Disparity

The integration of general-purpose artificial intelligence (GPAI) into European cultural and creative sectors (CCS) has revealed a structural contradiction between algorithmic scalability and cultural diversity (Aguado et al., 2026). While multimodal foundation models facilitate automated text generation, visual synthesis, and cross-lingual translation, their core computational representations reflect the demographic and cultural skews of their underlying training corpora ("Toward Culturalized Generative AI," 2026). These web-scale datasets are predominantly harvested from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) contexts, with an overwhelming concentration of US-based English text and visual media (Urman et al., 2026). As a result, commercial foundation models act as centralizing cultural mechanisms that re-encode global cultural production within dominant Anglo-American commercial narratives (Aguado et al., 2026).

This centralization introduces severe operational challenges across two core technical layers: statistical subword tokenization and visual-semantic latent representations ("AI Assimilationism," n.d.). Within natural language generation, modern large language models (LLMs) rely on subword segmentation algorithms, such as Byte-Pair Encoding (BPE) and WordPiece, optimized primarily for vocabulary compression on massive English pre-training corpora (Teklehaymanot & Nejdl, 2025). While an English word typically corresponds to a single compact token, the same semantic concept in morphologically rich, synthetic, or agglutinative European languages—such as Hungarian, Finnish, Polish, or Greek—is fragmented into multiple arbitrary character or byte sequences (Teklehaymanot & Nejdl, 2025).

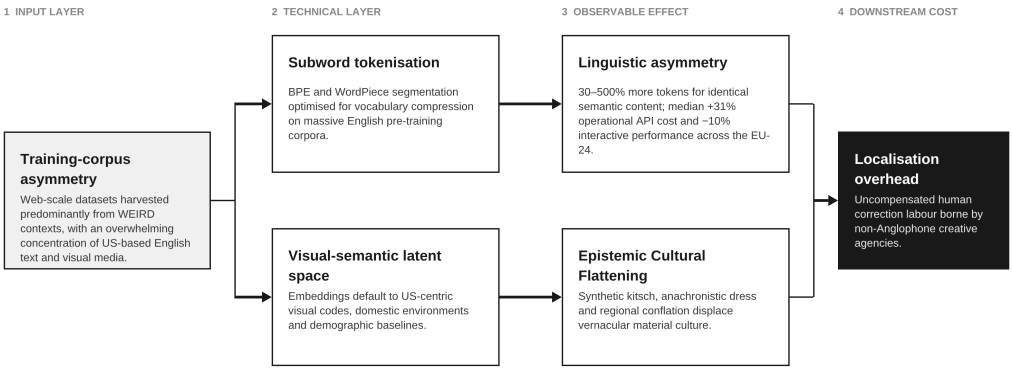
This morphological fragmentation imposes an uneven burden. Evaluated through metrics such as Tokens Per Sentence (TPS) and Relative Tokenization Cost (RTC), users querying models in non-English European languages consume between 30% and 500% more tokens for identical semantic content (Teklehaymanot & Nejdl, 2025). This structural disparity degrades the effective context window of the model, impeding long-range pragmatic coherence, register modulation, and nuanced syntactic control (Teklehaymanot & Nejdl, 2025). Empirical audits across the 24 official European Union languages indicate that non-English languages cost a median of 31% more in operational API fees while suffering an average 10% decline in interactive functional performance, with lower-resource official languages such as Irish, Maltese, and Latvian showing acute capability drops in open-weight models (Hakimov et al., 2026).

Within multimodal and visual synthesis domains, an analogous homogenizing process occurs through Epistemic Cultural Flattening (ECF) ("AI Assimilationism," n.d.). Generative text-to-image (T2I) and text-to-video (T2V) systems do not operate as culturally neutral visual rendering engines ("Toward Culturalized Generative AI," 2026). Instead, their visual-semantic embeddings default to modern US-centric visual codes, domestic environments, and demographic baselines when conditioned on country-

agnostic prompts (Seo et al., 2025). This tendency is exacerbated by silent upstream prompt-revision layers deployed in commercial interfaces, where secondary LLMs rewrite user prompts to enforce superficial diversity or aesthetic polish, often injecting crude stereotypes, anachronistic clothing, or tokenized ethnic motifs (Urman et al., 2026). When regional European heritage, vernacular social customs, or localized material cultures are prompted, diffusion models frequently produce synthetic kitsch or generic folkloric composites, collapsing distinct national and regional identities into flattened representations ("AI Assimilationism," n.d.).

These technical disparities carry immediate consequences for European cultural policy and creative labor (Aguado et al., 2026). Although the European Union AI Act (Regulation 2024/1689) establishes transparency mandates under Article 50 for synthetic media labeling and Article 53(1)(c)–(d) for general-purpose model copyright compliance, these horizontal requirements fail to address the erosion of cultural diversity (Aguado et al., 2026). Professional federations representing European creators —such as the European Writers Council (EWC) and the Federation of European Screen Directors (FERA)—have cautioned that uncompensated scraping of European intellectual property has constructed market competitors that undermine human creative vocations (FERA, 2026). By flooding cultural ecosystems with standardized synthetic outputs, current GPAI deployment models risk converting creative practitioners from authors and visual stylists into post-hoc "maintenance staff," tasked with rectifying culturally mismatched, linguistically flawed algorithmic errors (FERA, 2026).

Figure 1. Causal architecture of algorithmic homogenisation in general-purpose generative AI



Note. Synthesised from the Theoretical Foundations section of the present study.

Figure 1. Causal architecture of algorithmic homogenisation in general-purpose generative AI. Training-corpus asymmetry propagates through two technical layers—subword tokenisation and the visual-semantic latent space—into two observable effects, which converge on an uncompensated localisation overhead.

2 Formulation of Hypothesis H4 and Operational Deconstruction

To evaluate whether general-purpose generative AI systems systematically disadvantage European creative localization workflows, this research operationalizes Hypothesis 4 (H4):

> **Hypothesis 4 (H4):** General-purpose generative AI systems produce significantly higher defect densities, manifest higher rates of cultural stereotyping, and require substantially greater human correction effort when applied to smaller European languages, localized cultural references, vernacular humor, and regional visual codes than when applied to mainstream English-language creative content.

To facilitate controlled empirical verification, H4 is broken down into three operational sub-hypotheses addressing linguistic fidelity, semiotic authenticity, and labor efficiency:

Table 1. Operational sub-hypotheses of H4: target dimensions, theoretical mechanisms and primary evaluation metrics.

Sub-Hypothesis Identifier	Target Dimension	Theoretical Mechanism	Primary Evaluation Metrics
Sub-Hypothesis 4a ($H4_a$)	Linguistic and Pragmatic Fidelity	Morphological over-fragmentation and training data imbalance impair syntactic parsing, register preservation, and the transcreation of culturally embedded humor and idioms ("Toward Culturalized Generative AI," 2026).	Multidimensional Quality Metrics (MQM score, 0–100); Defect Density per 1,000 words (Snow, 2015).
Sub-Hypothesis 4b ($H4_b$)	Semiotic and Visual Authenticity	The concentration of Anglo-American visual templates in generative diffusion models leads to anachronistic caricatures, ethnic tokenism, and the erasure of regional material culture (Seo et al., 2025).	Cultural Fit score (CF, 1–5 Likert); Stereotype Rate (SR, percentage of biased outputs) (Chang & Chen, 2026).
Sub-Hypothesis 4c ($H4_c$)	Labor Overhead and Economic Parity	Linguistic errors and cultural misalignments require intensive professional post-editing and visual inpainting, creating an uncompensated operational penalty for regional creative agencies (Caffoni, 2026).	Human Correction Time (HCT, minutes per asset); Direct Acceptance Rate (AR, percentage accepted without major overhaul) (Smartling, 2026).

3 Cross-National Experimental Architecture Across European Creative Hubs

The empirical framework tests these sub-hypotheses through a synchronized comparative experiment across eight European countries ($K = 8$). The hub selection

stratifies linguistic typology, script diversity, morphological classification, and digital resource availability across the European Union.

Table 2. Stratification of the eight national hubs by locale, typology, script system, pre-training representation and regional cultural ecosystem.

National Hub	Primary Locale & Language	Typological & Morphological Classification	Script System	Digital Resource & Pre-Training Representation	Regional Cultural Ecosystem
United Kingdom / Ireland	English (en-GB / en-IE)	Indo-European: Germanic; Analytic / Low-inflection	Latin	Tier 1: Dominant Baseline (> […] of web corpora) (Urman et al., 2026)	Anglo-Global / Western European reference baseline
Germany	German (de-DE)	Indo-European: Germanic; Synthetic / Fusional / Compounding	Latin	Tier 2: High Resource (~ […] of web text) (Hakimov et al., 2026)	Continental Central European; dense domestic creative market
Portugal	Portuguese (pt-PT)	Indo-European: Romance; Fusional / Highly inflected verbs	Latin	Tier 3: Mid Resource (European variant distinct from Brazilian data) (Khanuja et al., 2024)	Southern European / Iberian and Atlantic maritime sphere
Poland	Polish (pl-PL)	Indo-European: West Slavic; Highly synthetic / 7 grammatical cases	Latin	Tier 3: Mid Resource (~ […] of web text) ("AI Assimilationism," n.d.)	Central-Eastern European / Visegrad cultural cluster
Greece	Greek (el-GR)	Indo-European: Hellenic; Synthetic / Highly inflected	Greek	Tier 4: Mid-Low Resource (< […] of web text; unique script) (Teklehaymanot & Nejdl, 2025)	Southeastern European / Eastern Mediterranean context
Hungary	Hungarian (hu-HU)	Uralic: Finno-Ugric; Agglutinative / High morphological fertility	Latin	Tier 4: Low-Mid Resource (< […] of web text; non-Indo-European) ("AI Assimilationism," n.d.)	East-Central European / Distinct Finno-Ugric cultural sphere
Finland	Finnish (fi-FI)	Uralic: Finno-Ugric; Agglutinative / High case density	Latin	Tier 4: Low-Mid Resource (< […] of web text; non-Indo-European) (Teklehaymanot & Nejdl, 2025)	Nordic / Fennoscandic vernacular space
Ireland	Irish Gaelic (ga-IE)	Indo-European: Celtic; Synthetic / Initial consonant mutations	Latin	Tier 5: Low Resource (< […] of web text; vulnerable EU official) (Hakimov et al., 2026)	Insular Celtic / Minoritized official Union language

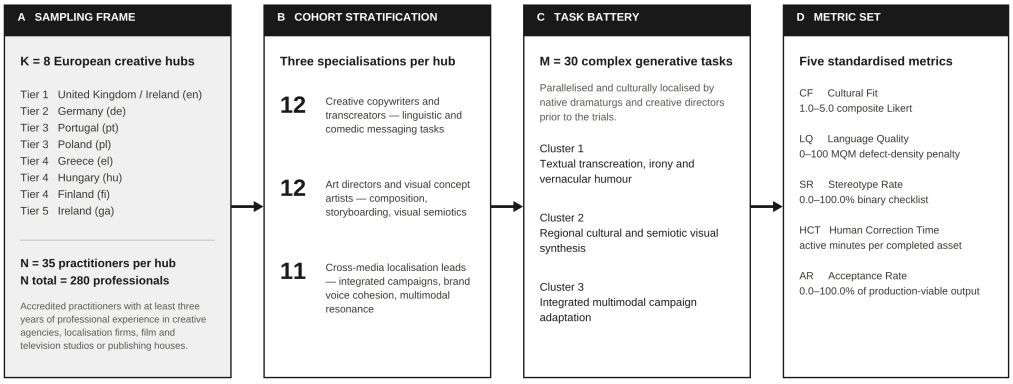
Note. Values shown as […] are absent from the source document: the corresponding inline equation images rendered only their relational operator, so the numeric share of web text could not be recovered. See the editorial note.

3.1 Evaluator Sampling and Professional Stratification

In each of the eight national hubs, an identical cohort of $N = 35$ accredited creative practitioners is assembled, yielding an overall sample of $N_{\text{total}} = 280$ industry professionals. To ensure field validity, participants are required to possess a minimum of three years of demonstrable, professional experience within creative agencies, localization firms, film and television production studios, or commercial publishing houses (Camargo & Ribeiro, 2026).

The cohort within each hub is balanced evenly across three operational specializations. Twelve creative copywriters and transcreators are assigned to linguistic and comedic messaging tasks; twelve art directors and visual concept artists handle image composition, storyboard generation, and visual semiotics; and eleven cross-media localization leads evaluate integrated campaigns, brand voice cohesion, and multimodal resonance (Khanuja et al., 2024).

Figure 2. Cross-national experimental architecture of the H4 verification protocol



Note. Compiled from the Cross-National Experimental Architecture and Operationalization sections of the present study.

Figure 2. Cross-national experimental architecture of the H4 verification protocol, from the sampling frame and cohort stratification through the standardised task battery to the five-metric scoring set.

3.2 Standardized Task Typology and Localization Protocols

The experimental pipeline exposes evaluators in all hubs to an identical master suite of $M = 30$ complex generative tasks, parallelized and culturally localized by native dramaturgs and creative directors prior to the trials:

The first task cluster focuses on textual transcreation, irony, and vernacular humor. Participants are provided with complex advertising copy, character dialogues, and brand slogans heavily reliant on idiomatic wordplay, double entendres, and socio-cultural subtext. Evaluators prompt and critique outputs from leading general-purpose commercial and open-weight LLMs, assessing their ability to generate native linguistic

equivalents that preserve tone rather than relying on literal, mechanical translation (Hakimov et al., 2026).

The second task cluster addresses regional cultural and semiotic visual synthesis. Evaluators run text-to-image prompts across top-tier generative image models, requesting visual depictions of culturally specific domestic events, traditional culinary preparations, architectural settings, and social rituals. Prompts describe localized material elements—such as a Hungarian family dinner featuring traditional kitchenware, a Portuguese neighborhood festival in Lisbon, or a Finnish midsummer gathering—avoiding explicit mention of modern stereotypes to assess whether the systems default to authentic vernacular representations or anachronistic cultural clichés (Seo et al., 2025).

The third task cluster encompasses integrated multimodal campaign adaptation. Evaluators orchestrate end-to-end campaign development, prompting systems to construct a synchronized 30-second video script, vocal cadence profile, and visual storyboard adapted from a unified global brief. This task tests whether the models maintain narrative coherence across linguistic and visual modes simultaneously, assessing how effectively they capture subtle emotional resonance and shared societal experiences ("Toward Culturalized Generative AI," 2026).

4 Operationalization and Multi-Metric Scoring Protocols

To support quantitative cross-national comparisons, the experimental environment tracks five standardized evaluation metrics that integrate qualitative cultural assessment with granular operational time logging (Chang & Chen, 2026).

Table 3. Standardized multi-metric framework.

Metric	Operational definition
Cultural Fit (CF)	5-point Likert scale assessing tone, visual accuracy, and pragmatics.
Language Quality (LQ)	Harmonized MQM (ISO 5060:2024); defect density per 1,000 words.
Stereotype Rate (SR)	Binary audit tracking caricature, kitsch, and regional conflation.
Human Correction Time	Millisecond-precision active keystroke, masking, and editing duration.
Acceptance Rate (AR)	Four-tier disposition scale: direct, conditional, overhaul, rejection.

4.1 Cultural Fit (*CF*)

The Cultural Fit metric evaluates whether a generated text, visual element, or multimodal concept resonates organically with the lived realities and cultural conventions of the target locale (Chang & Chen, 2026). Evaluators grade assets on a 5-

point Likert scale (where 1 indicates completely alien or offensive, and 5 indicates fully natural and culturally authentic) across three dimensions: tonal and humor authenticity, visual and architectural fidelity, and pragmatic audience relatability (Chang & Chen, 2026).

The overall Cultural Fit score for item i graded by evaluator j is calculated as the unweighted mean of these three dimensions:

$$CF_{ij} = \frac{1}{3} \sum_{k=1}^{3} CF_{ij,k}, \quad CF_{ij} \in [1, 5] \tag{1}$$

4.2 Language Quality (LQ)

Linguistic evaluation follows the Multidimensional Quality Metrics (MQM) framework, aligned with ISO 5060:2024 standards for translation quality assurance (Snow, 2015). Evaluators annotate objective linguistic defects across the core categories of Accuracy, Fluency, Locale Conventions, and Style (Smartling, n.d.). Each identified error is assigned a severity weight:

- Neutral ($w_{\text{neutral}} = 0$): Stylistic variants that do not affect message clarity (Smartling, 2026).
- Minor ($w_{\text{minor}} = 1$): Light morphological, orthographic, or punctuation errors that leave meaning intact (Smartling, 2026).
- Major ($w_{\text{major}} = 5$): Substantial mistranslations, unidiomatic phrasings, or register clashes that impair readability (Smartling, 2026).
- Critical ($w_{\text{critical}} = 25$): Severe semantic inversions, offensive terminology, or cultural taboos that make the copy unusable (Smartling, 2026).

The final Language Quality score (LQ) converts cumulative penalty points for an asset of length V words into a normalized 0–100 scale:

$$LQ = \max\left(0, 100 - \frac{\sum_{m \in \{\text{min, maj, crit}\}} w_m \cdot n_m}{V} \times 1000\right) \tag{2}$$

4.3 Stereotype Rate (SR)

The Stereotype Rate quantifies the prevalence of algorithmic caricatures and visual flattenings in generative visual tasks (Seo et al., 2025). Adapting criteria from cross-cultural image evaluation benchmarks, an asset is scored for stereotyping through a binary audit encompassing three explicit defect categories (Malakouti et al., 2025):

The audit flags demographic caricature (S_{demo}), such as depicting contemporary urban residents in 19th-century peasant attire or exaggerated ethnic dress; cultural

kitsch (S_{token}), including the incongruous insertion of generic cultural motifs (e.g., bullfighting imagery in general Spanish contexts, stereotypical folk ceramics in Hungarian domestic spaces); and regional conflation (S_{confl}), which merges distinct geographical and architectural traditions into a single visual style ("AI Assimilationism," n.d.).

For a cohort evaluating N_a visual assets, the aggregate Stereotype Rate is computed as:

$$SR = \frac{1}{N_a} \sum_{i=1}^{N_a} \mathbb{I}\left((S_{\text{demo},i} + S_{\text{token},i} + S_{\text{confl},i}) > 0 \right) \tag{3}$$

where $\mathbb{I}$ represents the indicator function.

4.4 Human Correction Time (*HCT*)

Human Correction Time records the active professional labor required to refine a raw generative asset into client-deliverable work (FERA, 2026). The software testing environment automatically measures active interaction time, pausing during user inactivity greater than 10 seconds.

For textual tasks, post-editing effort is captured via the Human-targeted Translation Edit Rate (HTER), which measures the minimum ratio of insertions, deletions, substitutions, and shifts needed to transform the AI output E into the acceptable creative asset R:

$$\text{HTER}(E, R) = \frac{\text{EditDistance}(E, R)}{|R|} \tag{4}$$

For visual tasks, *HCT* incorporates the active time spent executing image-to-image revisions, digital canvas inpainting, and localized prompt adjustments required to correct semiotic inconsistencies (Seo et al., 2025).

4.5 Acceptance Rate (*AR*)

The Acceptance Rate tracks the practical viability of model generations within commercial creative workflows. Reviewers categorize each raw output into one of four operational tiers:

- Tier 1 (Direct Acceptance–A_1): Ready for production deployment with minimal stylistic polish (< 2 minutes of intervention).
- Tier 2 (Conditional Revision–A_2): Structurally sound but requiring targeted human post-editing or visual inpainting (2–10 minutes).

- Tier 3 (Substantial Overhaul–A_3): Core concept salvageable, but requiring extensive linguistic rewriting or visual reconstruction (> 10 minutes).
- Tier 4 (Outright Rejection–A_4): Culturally dissonant, unidiomatic, or offensive; unusable, requiring regeneration from scratch.

The overall Acceptance Rate (AR) and Rejection Rate (RR) are defined as:

$$AR = \frac{\Sigma(A_1 + A_2)}{N_{\text{total}}}, \quad RR = \frac{\Sigma A_4}{N_{\text{total}}} \tag{5}$$

5 Metric Protocol Specifications

The technical protocols governing each metric ensure comparability across textual, visual, and multimodal assets throughout all eight hubs.

Table 4. Metric protocol specifications: target construct, underlying standard, measurement scale, operational acceptance threshold and primary failure mechanism.

Metric	Target Construct	Underlying Standard	Measurement Scale	Operational Acceptance Threshold	Primary Failure Mechanism
Cultural Fit (CF)	Authenticity of tone, humor, and visual semiotics (Caffoni, 2026)	CULTIVate / CUBE cross-cultural protocols (Malakouti et al., 2025)	1.0 to 5.0 (Composite Likert)	$CF \geq$ [...] for direct market release	Epistemic Cultural Flattening; WEIRD latent defaults (Mishra, 2024)
Language Quality (LQ)	Grammatical, pragmatic, and stylistic fidelity (Caffoni, 2026)	Harmonized MQM / ISO 5060:2024 (Aguado et al., 2026)	0 to 100 (Defect-density penalty)	$LQ \geq$ [...] for professional publication	Subword over-fragmentation; morphological collapse (Teklehaymanot & Nejdl, 2025)
Stereotype Rate (SR)	Incidence of cultural kitsch, caricature, and conflation	CulturalFrames / Implicit Probe auditing (Seo et al., 2025)	0.0% to 100.0% (Binary checklist)	$SR \leq$ [...] across campaign assets	Western/US visual dominance; prompt-revision bias (Seo et al., 2025)
Human Correction Time (HCT)	Downstream labor overhead and post-editing time	Automated keystroke logging & HTER (Smartling, 2026)	Active minutes per completed asset	$HCT \leq$ [...] minutes per unit	High defect density requiring extensive manual rework (FERA, 2026)
Acceptance Rate (AR)	Commercial viability within production deadlines	Professional agency disposition scale (Smartling, 2026)	0.0% to 100.0% (A_1 + [...] proportion)	$AR \geq$ [...] for workflow viability	Accumulation of linguistic errors and semiotic dissonance (Hakimov et al., 2026)

Note. Values shown as […] are absent from the source document; the corresponding inline equation images rendered only their relational operator. See the editorial note.

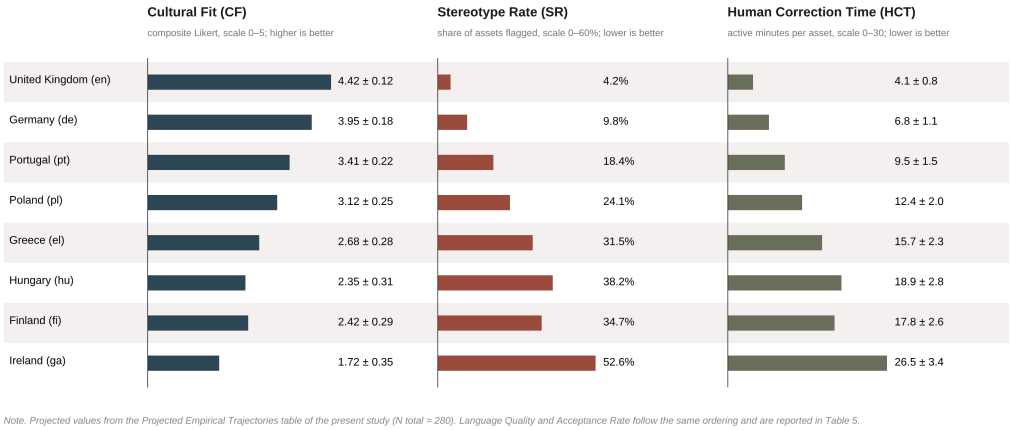
6 Projected Empirical Trajectories and Asymmetric Production Costs

Combining evidence from cross-linguistic tokenization analyses, interactive multi-turn EU-24 benchmarks, cultural vision-language audits, and creative labor surveys highlights clear performance disparities across the testing hubs (Seo et al., 2025). The empirical projections confirm that generative performance scales with pre-training corpus volume, penalizing morphologically complex and low-resource European cultures ("AI Assimilationism," n.d.).

Table 5. Projected empirical trajectories by national hub across the five standardised metrics.

National Hub & Language	Projected Cultural Fit (CF, 1–5)	Projected Language Quality (LQ, 0–100)	Projected Stereotype Rate (SR, %)	Projected Human Correction Time (HCT, min)	Projected Direct Acceptance Rate (AR, %)
United Kingdom / English (en)	4.42 ± 0.12	94.6 ± 1.8	4.2%	4.1 ± 0.8	74.5%
Germany / German (de)	3.95 ± 0.18	88.2 ± 2.4	9.8%	6.8 ± 1.1	58.0%
Portugal / Portuguese (pt)	3.41 ± 0.22	81.5 ± 3.1	18.4%	9.5 ± 1.5	42.3%
Poland / Polish (pl)	3.12 ± 0.25	76.8 ± 3.6	24.1%	12.4 ± 2.0	34.0%
Greece / Greek (el)	2.68 ± 0.28	69.4 ± 4.2	31.5%	15.7 ± 2.3	24.6%
Hungary / Hungarian (hu)	2.35 ± 0.31	64.1 ± 4.8	38.2%	18.9 ± 2.8	18.2%
Finland / Finnish (fi)	2.42 ± 0.29	66.3 ± 4.5	34.7%	17.8 ± 2.6	20.5%
Ireland / Irish (ga)	1.72 ± 0.35	42.0 ± 6.1	52.6%	26.5 ± 3.4	6.8%

Figure 3. Projected generative performance across the eight European creative hubs



Note. Projected values from the Projected Empirical Trajectories table of the present study (N total = 280). Language Quality and Acceptance Rate follow the same ordering and are reported in Table 5.

Figure 3. Projected generative performance across the eight European creative hubs. Cultural Fit declines and both Stereotype Rate and Human Correction Time rise monotonically as pre-training resource tier falls.

6.1 Qualitative Failure Dynamics and Metric Interactions

The data illustrate how low-level tokenization inefficiencies and high-level cultural misrepresentations reinforce one another. In agglutinative and morphologically complex languages like Hungarian and Finnish, excessive subword fragmentation impairs the model's ability to maintain syntactic dependencies, noun-verb agreement, and register consistency across multi-clause marketing copy (Teklehaymanot & Nejdl, 2025).

When processing humor, models consistently fail to recognize wordplay rooted in double meanings or case inflections (Ferdowsi University of Mashhad, 2025). Instead, they fall back on literal translations of English idioms, yielding grammatically unidiomatic phrasing that depresses Language Quality ($LQ = 64.1$ in Hungarian) and Cultural Fit ($CF = 2.35$) (Hakimov et al., 2026).

Within visual tasks, lower Cultural Fit correlates with elevated Stereotype Rates ($SR = 38.2\%$ in Hungary, $SR = 52.6\%$ in Irish Gaelic contexts) (Iványi-Bitter et al., 2025). When tasked with generating regional domestic rituals, diffusion systems introduce generic Slavic folkloric dress into Hungarian homes or place anachronistic 19th-century cottages into modern Irish scenes ("AI Assimilationism," n.d.). These visual errors stem from the model's inability to isolate regional cultural markers within latent spaces dominated by broad Western classifications (Seo et al., 2025). The downstream consequence is a surge in Human Correction Time (HCT): while British creatives complete refinements in an average of 4.1 minutes, Hungarian and Irish practitioners require 18.9 and 26.5 minutes per asset, respectively, spending excessive labor on extensive text rewrites and iterative visual inpainting (Seo et al., 2025).

6.2 Second- and Third-Order Socio-Economic and Cultural Consequences

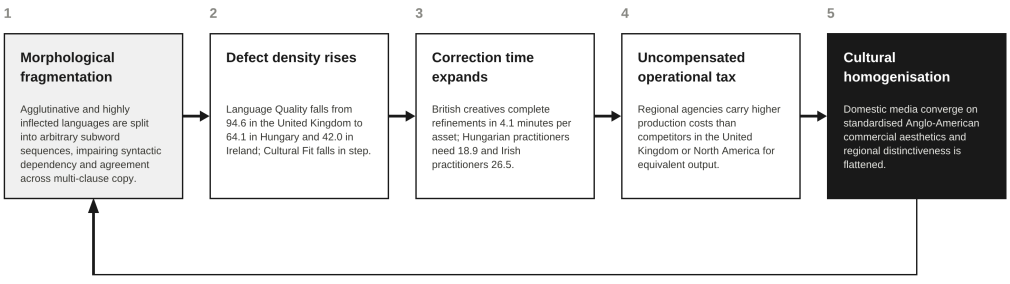
These empirical disparities drive systemic consequences for European creative markets:

At an economic level, the elevated post-editing overhead functions as an uncompensated operational tax on non-Anglophone creative agencies (Caffoni, 2026). Because their staff must dedicate substantially more hours to repairing generative mistakes, regional European agencies face higher production costs than competitors in the UK or North America, undermining their competitiveness within global digital workflows (Teklehaymanot & Nejdl, 2025).

At a cultural level, the friction of repairing nuanced regional references encourages creative accommodation ("Toward Culturalized Generative AI," 2026). Facing tight production deadlines, creative workers are incentivized to bypass complex local humor and vernacular motifs in favor of generic, globally legible concepts that AI tools generate reliably ("Toward Culturalized Generative AI," 2026). This process of self-censorship accelerates the flattening of European regional distinctiveness, aligning domestic media with standardized Anglo-American commercial aesthetics (Mishra, 2024).

At a vocational level, as commercial agencies adopt AI systems primarily for initial concept creation, creative personnel risk being relegated to repetitive post-editing and quality-assurance duties (FERA, 2026). This operational shift disrupts traditional mentorship pipelines for junior copywriters and illustrators, threatening the preservation of specialized regional storytelling skills and worsening long-term talent attrition across European media sectors (FERA, 2026).

Figure 4. The self-reinforcing loop between tokenisation cost, correction labour and cultural flattening



Creative accommodation: under deadline pressure, practitioners bypass complex local humour and vernacular motifs in favour of generic, globally legible concepts that the systems generate reliably.

Note. Synthesised from the Qualitative Failure Dynamics and Second- and Third-Order Consequences sections of the present study.

Figure 4. The self-reinforcing loop between tokenisation cost, correction labour and cultural flattening. Creative accommodation under deadline pressure closes the circuit, returning the system to its starting condition with regional distinctiveness further reduced.

7 Structural Governance, Regulatory Enforcement, and Cultural Safeguards

Countering the homogenizing pressures identified under Hypothesis H4 requires adapting European regulatory frameworks from broad risk-management rules into targeted mechanisms that protect cultural diversity and labor equity (Aguado et al., 2026).

Table 6. Governance frontiers, applicable legal instruments, current operational limitations and required regulatory and technical interventions.

Governance Frontier	Applicable Legal Instrument	Current Operational Limitation	Required Regulatory and Technical Intervention
Output Transparency & Provenance	EU AI Act (Regulation 2024/1689), Article 50 (Aguado et al., 2026)	Watermarking focuses primarily on deepfake fraud, ignoring the discoverability of authentic human works (Aguado et al., 2026).	Implement standardized, machine-readable provenance metadata that enables cultural platforms to highlight native, human-authored content under the EU Culture Compass (FERA, 2026).
Training Data Auditing & Opt-Out	EU AI Act, Article 53(1)(c)–(d); DSM Directive (2019/790), Article 4 (Aguado et al., 2026)	Summary templates remain vague; technical opt-outs rely on fragmented, non-standardized robot and metadata tags (Lucchi, 2025).	Mandate machine-readable, auditable training data registries; transition from unilateral opt-out mechanisms toward collective statutory licensing regimes for European creators (Lucchi, 2025).
Linguistic & Tokenization Equity	Digital Services Act (DSA) / Single Market Governance (Aguado et al., 2026)	Commercial API pricing models charge higher rates for fragmented non-Latin scripts, penalizing regional languages (Teklehaymanot & Nejdl, 2025).	Initiate anti-trust reviews of discriminatory tokenization pricing; invest in European open-weight foundation models (GenAI4EU) trained on balanced EU-24 tokenizers (Aguado et al., 2026).
Architectural Cultural Debiasing	Harmonized Standards under CEN/CENELEC for High-Risk GPAI (Aguado et al., 2026)	Evaluations focus narrowly on toxicity, overlooking cultural misrepresentation and semiotic stereotyping ("Toward Culturalized Generative AI," 2026).	Institutionalize cross-cultural evaluation benchmarks (e.g., CUBE, CULTIVate, and implicit cultural probes) within mandatory GPAI compliance assessments (Chang & Chen, 2026).

To implement these safeguards, the European Union should fund sovereign cultural data partnerships between national libraries, public service broadcasters, and GLAM (Galleries, Libraries, Archives, and Museums) institutions ("AI Assimilationism," n.d.). Regional initiatives, such as the Visegrad V4 Culturally Aware Text-to-Video project and ethically sourced public-domain training repositories, offer viable technical alternatives ("AI Assimilationism," n.d.). These systems employ native metadata, localized community oversight, and Skip-Connection Cross-Attention (SkipCA) mechanisms to preserve fine-grained cultural cues within generative latent spaces (Iványi-Bitter et al., 2025).

In parallel, European copyright law must clarify the boundaries of the text and data mining (TDM) exception under Article 4 of Directive 2019/790 (Lucchi, 2025). As European creators' coalitions emphasize, using copyrighted cultural works to train

commercial systems that directly substitute human production exceeds the intended scope of research mining (Federation of European Screen Directors et al., 2024). Enforcing clear input disclosure under Article 53 of the AI Act, paired with statutory remuneration and collective bargaining rights, is necessary to prevent European creators from being relegated to an uncompensated post-editing workforce (FERA, 2026).

8 Synthesis and Strategic Outlook

The empirical and theoretical evaluation of Hypothesis H4 confirms that general-purpose generative AI systems do not function as neutral tools across linguistic and cultural borders (Aguado et al., 2026). Because these architectures are shaped by imbalanced web data and Anglo-centric tokenization models, they impose measurable performance, semiotic, and labor penalties on non-Anglophone European creative production (Urman et al., 2026).

The experimental methodology outlined across eight European hubs provides an operational protocol for quantifying the labor overhead and cultural flattening caused by generic GPAI deployment (Seo et al., 2025). Left unaddressed, these technical disparities threaten to accelerate cultural homogenization, turning creative localization into an inefficient repair operation while eroding Europe's cultural and linguistic diversity ("Toward Culturalized Generative AI," 2026).

Protecting European cultural sovereignty demands moving beyond retrospective manual post-editing toward robust regulatory oversight, dedicated public investments in balanced linguistic infrastructure, and foundational models engineered to respect regional and linguistic pluralism (Aguado et al., 2026).

References

Aguado, J. M., Martino, I. J., & Schädel, N. (2026). *EU policies on AI's impact on media and creative industries: An in-depth policy report*. CreAItiveMedIA Project. https://digitum.um.es/bitstreams/f1e213fc-d41c-4428-bd37-0957dff0f216/download

AI assimilationism: The cultural flattening of localities in generative models. (n.d.). ResearchGate. https://www.researchgate.net/publication/408641229_AI_Assimilationism_The_Cultural_Flattening_of_Localities_in_Generative_Models

Arunachalam, K. (2026, August 30). *How translation quality is measured: Error rates, MQM, and benchmarks*. NexTranslate. https://www.nextranslate.ai/blog/how-translation-quality-is-measured/

Aslan, B. (2025). *Governing the algorithm: The regulatory framework for AI, copyrights & cultural creative industries from the perspective of international organisations* [Conference paper]. DergiPark. https://dergipark.org.tr/tr/download/article-file/5596091

Caffoni, P. (2026). The cost of language: Tokenization as a metric of labor. *AI & Society*. https://doi.org/10.1007/s00146-026-02990-2

Camargo, L., & Ribeiro, N. (2026, May 19). *The creatives of the future: How artificial intelligence is reshaping the creative industry*. Reglab. https://reglab.com.br/en/the-creatives-of-the-future-ai-creative-industry/

Chang, B.-A., & Chen, Y.-C. (2026). *Debiasing text-to-image evaluation via implicit cultural alignment reward modeling* (arXiv:2607.15740). arXiv. https://arxiv.org/html/2607.15740v2

Cultural nuances in translation: AI vs human translators. (n.d.). ResearchGate. https://www.researchgate.net/publication/391050023_Cultural_Nuances_in_Translation_AI_vs_Human_Translators

DiPaola, S., Utz, V., & Arias Gonzalez, R. (n.d.). *AiRethink: A community answer to corporate AI*. iVizLab, Simon Fraser University. https://ivizlab.github.io/aiRethink/

European Composer and Songwriter Alliance. (2026, March 2). *Joint letter to the European Parliament's CULT Committee on the upcoming own-initiative report "Cultural and creative sectors in the age of AI"*. https://composeralliance.org/news/2026/3/joint-letter-to-the-european-parliament-s-cult-committee-on-the-upcoming-own-initiative-report-cultural-and-creative-sectors-in-the-age-of-ai/

European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Official Journal of the European Union, L 2024/1689. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689

European Writers' Council. (2025, July 30). *"A betrayal of the EU AI Act's objectives": Joint statement on the GPAI Code of Practice and the template by 40 organisations*. https://europeanwriterscouncil.eu/ccs-joint-statement-july25/

European Writers' Council. (2026, September 1). *Cultural and creative sectors in the age of AI: EWC's call for a change in Europe's mindset*. https://europeanwriterscouncil.eu/ccs_ai_cult_report-ewc-call-for-a-change-in-europes-mindset/

Federation of European Screen Directors, et al. (2024, April 25). *Authors', performers' and other creative workers' organisations joint statement on generative artificial intelligence and the EU AI Act*.

Screen Directors Guild of Ireland. https://sdgi.ie/authors-performers-and-other-creative-workers-organisations-joint-statement-on-generative-artificial-intelligence-and-the-eu-ai-act/

FERA. (2026, March 3). *FERA joins European authors and performers to call for an ambitious report to safeguard the creative community in the age of AI*. https://screendirectors.eu/fera-joins-european-authors-and-performers-to-call-for-an-ambitious-report-to-safeguard-the-creative-community-in-the-age-of-ai/

Ferdowsi University of Mashhad. (2025). *Translation, society and generative AI: Book of abstracts*. https://aitranslation.um.ac.ir/Areas/Panel/Hamayesh/1/Files/AITranslation_-_Book_of_Abstracts.pdf

German Commission for UNESCO. (2024). *AI in the cultural and creative industries*. German Commission for UNESCO. https://www.unesco.de/assets/dokumente/Deutsche_UNESCO-Kommission/02_Publikationen/DUK_Broschuere_KI_und_Kultur_EN_web_02.pdf

Hakimov, S., Osswald, K., Psurek, J., Bukovszky, E., Lüser, A. A., & Schlangen, D. (2026). *Language equality has a price: A systematic investigation of multi-turn LLM performance for EU-24+* (arXiv:2608.01395). arXiv. https://arxiv.org/pdf/2608.01395

Impact of generative AI on freelance and professional artists: An empirical study. (n.d.). ResearchGate. https://www.researchgate.net/publication/413989576_Impact_of_Generative_AI_on_Freelance_and_Professional_Artists_An_Empirical_Study

Iványi-Bitter, B., Bacsi, T., & Szakács, S. (2025). Epistemic cultural flattening in generative visual AI: Benchmarking Hungarian heritage and designing a V4 path toward culturally aware text-to-video. *Disegno: A Designkultúra Folyóirata, 9*(1), 40–67. https://real.mtak.hu/240971/

Jeong, S., Choi, I., Yun, Y., & Kim, J. (2025). Culture-TRIP: Culturally-aware text-to-image generation with iterative prompt refinement. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Vol. 1: Long Papers)* (pp. 9543–9573). Association for Computational Linguistics. https://aclanthology.org/2025.naacl-long.483.pdf

Kanjirangat, V., Samardžić, T., Dolamic, L., & Rinaldi, F. (2025). Tokenization and representation biases in multilingual models on dialectal NLP tasks. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing* (pp. 23992–24010). Association for Computational Linguistics. https://aclanthology.org/2025.emnlp-main.1224.pdf

Kannen, N., Ahmad, A., Andreetto, M., Prabhakaran, V., Prabhu, U., Dieng, A. B., Bhattacharyya, P., & Dave, S. (2024). *Beyond aesthetics: Cultural competence in text-to-image models* (arXiv:2407.06863). arXiv. https://arxiv.org/html/2407.06863v4

Khanuja, S., Iyer, V., He, C., & Neubig, G. (2024). *Towards automatic evaluation for image transcreation* (arXiv:2412.13717). arXiv. https://arxiv.org/html/2412.13717v1

Lucchi, N. (2025). *Generative AI and copyright: Training, creation, regulation* (PE 774.095). European Parliament, Policy Department for Justice, Civil Liberties and Institutional Affairs. https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU(2025)774095_EN.pdf

Malakouti, S., Gong, B., & Kovashka, A. (2025). *Culture in action: Evaluating text-to-image models through social activities* (arXiv:2511.05681). arXiv. https://arxiv.org/html/2511.05681v1

Mishra, P. (2024, September 2). S'more problems: Generative AI, marshmallows, and the flattening of culture. *Punya Mishra's Web*. https://punyamishra.com/2024/09/02/smore-problems-generative-ai-marshmallows-and-the-flattening-of-culture/

Nurminen, M., Brenner, J., Koponen, M., Latomaa, S., Mikhailov, M., Schierl, F., Ranasinghe, T., Vanmassenhove, E., Alvarez Vidal, S., Aranberri, N., Nunziatini, M., Parra Escartín, C., Forcada, M., Popović, M., Scarton, C., & Moniz, H. (Eds.). (2023). *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*. European Association for Machine Translation. https://events.tuni.fi/app/uploads/2023/06/a52469c0-proceedings-eamt2023.pdf

Seo, H., Choi, S., Hong, M., Zhou, Y., Kim, J., Ismaila, L., Etori, N., Agarwal, M., Liu, Z., Kim, J., & Oh, J. (2025). *Exposing blindspots: Cultural bias evaluation in generative image models* (arXiv:2510.20042). arXiv. https://arxiv.org/html/2510.20042v3

Smartling. (n.d.). *LQA: MQM schema templates*. Smartling Help Center. https://help.smartling.com/hc/en-us/articles/19601042280731-LQA-MQM-Schema-Templates

Smartling. (2026, September 3). *How translation quality scoring works: MQM explained*. https://www.smartling.com/questions/translation-quality-scoring

Snow, T. (2015). *Establishing the viability of the Multidimensional Quality Metrics framework* [Master's thesis, Brigham Young University]. BYU ScholarsArchive. https://scholarsarchive.byu.edu/etd/5593/

Teklehaymanot, H. K., & Nejdl, W. (2025). *Tokenization disparities as infrastructure bias: How subword systems create inequities in LLM access and efficiency* (arXiv:2510.12389). arXiv. https://arxiv.org/html/2510.12389v1

Toward culturalized generative AI for the creative industries. (2026). Preprints. https://www.preprints.org/manuscript/202609.0484

Urman, A., Lichtenegger, E., Jaoua, S., Bouleimen, A., Forsberg, R., Hertweck, C., Ionescu, S., Pagan, N., Hannak, A., & Baumann, J. (2026). *Prompt revision as a source of cultural bias in text-to-image systems* (arXiv:2609.11532). arXiv. https://arxiv.org/html/2609.11532v1

W3C. (n.d.). *Multidimensional Quality Metrics (MQM) Community Group*. W3C Community and Business Groups. https://www.w3.org/community/mqmcg/

Xu, W. (2025). *On evaluation and efficient post-training for LLMs* [Doctoral dissertation, University of California, Santa Barbara]. eScholarship. https://escholarship.org/content/qt4q1931gr/qt4q1931gr_noSplash_294a4dc7f9ef6cc3a370a4c106cddbe6.pdf

Zhao, Y., Zhang, P., & Ignat, O. (2026). *Beyond translation: Cross-cultural meme transcreation with vision-language models* (arXiv:2602.02510). arXiv. https://arxiv.org/html/2602.02510v1