The LLM language divide: a critical examination of generative AI performance disparities in non-Anglophone journalism and the democratic and security implications for public information ecosystems

Andrei Nistor

Story Agents SRL / Press Cluster NGO, Hunedoara, Romania

andrei.nistor@storyagents.ro

Abstract

Objectives: This paper investigates the LLM Language Divide, the systematic performance disparity between AI journalism tools operating in English and those deployed in non-Anglophone and minority language contexts, arguing it constitutes a structural challenge for democratic information ecosystems and smart city cyber-resilience. **Prior work:** The study builds on computational linguistics research documenting Anglophone bias in major LLM training corpora, journalism studies on AI integration in newsrooms, Habermasian democratic theory, and smart city security scholarship published in SCRD Journal and TRUST conference proceedings, including contributions by Virtosu [33], Drobot [34], and Vangelov [35]. **Approach:** This is a theoretical synthesis paper employing systematic literature review across computational linguistics, natural language processing, journalism studies, and information security scholarship to map causes and consequences of cross-linguistic LLM performance inequality. **Results:** LLM performance inequality is rooted in training data asymmetry, English constitutes approximately 92% of major pre-training corpora, producing higher hallucination rates, cultural misrepresentation, and tokenisation inefficiencies in non-Anglophone journalism contexts. Cross-linguistic hallucination amplification creates compounded risks including fabricated attribution, institutional misrepresentation, and cultural distortion that disproportionately affect minority language communities with weaker verification infrastructure. **Implications:** The findings carry direct relevance for media policymakers, AI developers, EU regulatory authorities (EU AI Act, Digital Services Act, European Media Freedom Act), and smart city governance frameworks, calling for enforceable multilingual performance standards and community media resilience investment. **Value:** The paper introduces the original normative framework of Linguistic Media Equity (LME) and extends the analysis into smart city cyber-resilience, framing the LLM language divide as a structural information security vulnerability, a contribution at the intersection of AI governance, journalism studies, and the TRUST conference theme of building cyber-resilient intelligent cities.

Keywords: hallucination; linguistic media equity; smart cities; training data inequality; Central and Eastern Europe.

1. Introduction

Journalism occupies a unique position in democratic societies: it is at once a professional practice, a cultural institution, and a public service, the primary mechanism through which citizens access the information necessary to participate meaningfully in democratic life [1]. The rapid integration of generative artificial intelligence (GenAI), and specifically large language models (LLMs), into news production and distribution has therefore attracted substantial scholarly and policy attention. A comprehensive bibliometric analysis of AI and journalism research conducted by Sonni [2] documented a 172 percent surge in related scholarly publications in 2024 alone, driven primarily by the deployment of generative AI tools across the media industry. A 2024 Reuters Institute survey of editors, digital leaders, and media chief executives found that 78 percent believe AI investment will be critical to journalism's survival in competitive digital markets [3].

Yet the dominant scholarly and practitioner discourse on AI in journalism has largely assumed an implicitly English-language default. The empirical studies, case studies, and normative frameworks that populate this rapidly growing literature overwhelmingly draw on Anglophone newsroom contexts, principally those of the United States, United Kingdom, and Australia, while treating the experiences of non-English-speaking journalism communities as secondary, peripheral, or simply absent. This paper argues that this Anglophone default is not merely a gap in the literature: it reflects and reproduces a genuine structural asymmetry embedded in the architecture of the AI systems being studied.

This asymmetry, which this paper terms the LLM Language Divide, emerges directly from the composition of training data. The datasets used to pre-train major commercial and open-source LLMs are overwhelmingly constituted of English-language text. A 2025 survey of multilingual LLMs by Qin and colleagues [4] confirms that most major models remain fundamentally English-centric, limiting their applicability in a linguistically diverse world in which more than 80 percent of the global population does not use English as a primary language. In journalism, where precision, factual accuracy, cultural fidelity, and institutional accountability are democratic necessities, the performance gap between English and non-English LLM outputs generates risks that extend well beyond computational inconvenience.

This paper makes four primary contributions to the literature. First, it provides the most comprehensive theoretical account to date of the root causes and mechanisms of LLM language performance inequality in journalism contexts, synthesising evidence from computational linguistics, natural language processing, and journalism studies. Second, it documents, with reference to verified empirical studies, the specific, amplified risks that hallucination and cultural misrepresentation in lower-resource language LLMs present for

journalistic credibility and democratic information access. Third, it proposes the normative framework of Linguistic Media Equity (LME) as a new conceptual tool for evaluating the democratic legitimacy of AI systems deployed in multilingual journalism contexts. Fourth, it extends the analysis into smart city information security, demonstrating that the LLM language divide constitutes a structural cyber-resilience challenge for non-Anglophone intelligent urban environments, a contribution positioned directly within the thematic scope of the TRUST 2026 conference.

The paper proceeds as follows. Section 2 examines the structural causes of cross-linguistic LLM performance inequality. Section 3 analyses the specific and compounded hallucination risks in non-Anglophone journalism. Section 4 theorises the democratic implications for press freedom, media pluralism, and public trust. Section 5 introduces the LME framework and its policy implications. Section 6 identifies research gaps and proposes a future empirical agenda. Section 7 extends the analysis into smart city cyber-resilience, positioning the language divide as a structural information security vulnerability. Section 8 concludes.

2. The Structural Roots of the LLM Language Divide

2.1. Training Data Composition: The Anglophone Bias

The performance characteristics of any large language model are fundamentally determined by the composition of its pre-training corpus. This is not a contingent technological feature but a mathematical necessity: LLMs learn statistical distributions over the tokens present in their training data, and they perform best on tasks structurally similar to those distributions [5]. The systematic underrepresentation of non-English languages in LLM training data therefore produces not merely reduced performance on non-English tasks, but a qualitatively different relationship between the model and languages for which it has seen little data, one characterised by higher uncertainty, lower calibration, and a tendency to fall back on English-language patterns even when generating text in another language [6].

The empirical evidence for this compositional asymmetry is striking. Analysis of ChatGPT training data by Zhang and colleagues [6] found that English text accounts for approximately 92 percent of the total pre-training corpus, while languages such as Romanian, Slovak, Bulgarian, and Hungarian, each an official language of European Union member states spoken by millions of citizens, represent fractions of a single percent. Li and colleagues [7], who proposed a quantitative Language Ranker metric to measure multilingual abilities using LLM internal representations, confirmed a strong correlation between a language's proportional representation in pre-training corpora and subsequent model performance on that language, establishing definitively that the data composition gap translates directly into a capability gap. The mC4 dataset, one of the most widely used

multilingual corpora, nominally covers 108 languages; yet its token distribution reveals English dominating with approximately 2,733 billion tokens, while Romanian receives fewer than 35 billion and Slovak fewer than 5 billion [7].

Critically, this is not a problem that scales away with model size. Li and colleagues' Language Ranker study [7] demonstrates that LLMs of varying parameter sizes exhibit the same relative performance ordering across languages, indicating that the language divide is a feature of the data ecosystem rather than a limitation of specific models at a given developmental stage. Adding more English data or increasing model parameters does not automatically improve performance in languages that remain statistically rare in training data.

2.2. Cultural Context and the Limits of Cross-Lingual Transfer

Beyond raw token counts, LLMs trained predominantly on English data inherit and reflect the cultural values, political framings, and social assumptions embedded in English-language discourse. A landmark Stanford HAI white paper [8] documented that most major LLMs are not merely quantitatively underperforming in non-English languages but are qualitatively misaligned with the cultural contexts of non-Anglophone communities. This distinction matters enormously for journalism: the craft of news production is a culturally embedded practice of meaning-making that draws on shared social knowledge, political context, and community norms [9].

GPT models exhibit political and cultural biases aligning with Anglophone, and particularly United States-centric, value systems, what Motoki and colleagues [10] describe as having an 'American accent' regardless of the language in which the model is prompted. A 2025 study examining the multilingual capabilities of GPT across English, Korean, and Japanese [11] found that the model systematically generalises English-language interpretive preferences onto Korean-language outputs, where different linguistic norms would predict different responses. These findings suggest that cross-lingual transfer does not carry cultural context alongside linguistic form. For journalism, this is not trivial: a model generating news about Romanian institutional politics, Slovak election procedures, or Hungarian civil society may produce linguistically plausible text that is nevertheless culturally incoherent, politically misleading, or factually incongruent with local realities.

The concept of the 'curse of multilinguality,' documented by Conneau and colleagues and reviewed in recent multilingual LLM surveys [12], adds a further dimension. As the number of languages included in a training corpus increases, each individual language receives proportionally less attention, potentially degrading performance even for languages with moderate representation. This creates a difficult trade-off that has, in practice, consistently been resolved in favour of English and other high-resource languages, at the direct expense of lower-resource communities.

2.3. The Tokenisation Problem

A structural dimension of the LLM language divide that receives insufficient attention in journalism-adjacent scholarship is the problem of tokenisation, the way in which LLMs segment text into processable units during both training and inference. Most major LLMs, including GPT-4 and its successors, employ tokenisers optimised for English orthography. Romanian, Hungarian, Slovak, and Bulgarian, which feature morphological complexity, diacritical marks, and agglutinative constructions absent from English, are consequently tokenised far less efficiently, requiring substantially more tokens to represent equivalent semantic content [12]. This is not merely a computational efficiency problem: higher token-per-word ratios mean that the model has less contextual window space available for non-English inputs, that context is truncated more quickly in longer documents, and that generating extended non-English text is disproportionately costly. For news articles of several thousand words, this tokenisation inefficiency compounds at scale, creating a structural disadvantage for non-English news generation that English-language journalism does not face.

3. Hallucination in Non-Anglophone Journalism: Amplified Risks

3.1. Defining Hallucination in Journalistic Contexts

Hallucination in large language models refers to the generation of content that is syntactically and semantically fluent but factually incorrect, unsupported by evidence, or outright fabricated, produced with a surface-level confidence that gives no signal of the underlying inaccuracy [13]. A 2025 comprehensive survey catalogues two primary types: intrinsic hallucinations, in which the model generates content directly contradicting verifiable facts, and extrinsic hallucinations, in which unverifiable claims are introduced that cannot be grounded in evidence [13]. Both types are problematic in journalistic contexts; the latter is particularly insidious because it may resist automated detection while still introducing error into the public information record.

Anh-Hoang, Tran, and Nguyen [14], in a systematic empirical analysis of hallucination attribution across multiple LLMs, found that even state-of-the-art models including GPT-4 and LLaMA exhibit hallucination rates exceeding 15 percent when evaluated against standardised factual benchmarks. A systematic literature review covering 253 primary studies published between 2020 and 2025 [15] found that hallucination remains a persistent and unresolved challenge across all major LLM architectures, with detection and mitigation techniques remaining substantially below the reliability thresholds required for high-stakes professional applications. Hallucination is, in this sense, an inherent property of the probabilistic text generation architecture underlying current LLMs, not a bug that sufficiently advanced engineering will eliminate.

3.2. The Cross-Linguistic Amplification Effect

The research on cross-linguistic hallucination, the rate at which LLMs generate factually incorrect content in languages other than their primary training language, represents one of the most significant and least-examined dimensions of the LLM language divide. A dedicated study published at EMNLP 2025 [16] titled 'How Much Do LLMs Hallucinate across Languages?' provides the most direct available empirical evidence: hallucination rates are substantially and systematically higher when LLMs generate content in low-resource languages compared to English, and this differential is consistent across multiple model architectures and evaluation benchmarks. The mechanism is theoretically coherent: a model with insufficient training data in a given language lacks the statistical grounding to calibrate confidence in factual claims in that language, and therefore generates plausible-sounding text that may bear only a statistical, rather than factual, relationship to the world it purports to describe.

Research on LLM performance in African languages [17], which, despite having hundreds of millions of speakers, are similarly underrepresented in LLM training corpora to many CEE languages, provides compelling benchmark evidence for this cross-linguistic amplification effect. Studies using standardised question-answering and fact-verification benchmarks found that GPT-4 and comparable models perform substantially below human baselines in these language contexts, with performance gaps that cannot be closed through prompt engineering alone. The parallel to Romanian, Hungarian, Slovak, and Bulgarian is direct: these EU official languages, while better represented than many African languages, remain substantially underrepresented relative to English in all major commercial and open-source LLM training datasets.

The SemEval-2025 Mu-SHROOM dataset, a multilingual hallucination evaluation benchmark [18], confirmed that detection methods optimised for English hallucinations generalise poorly to low-resource language contexts, and that the best-performing detection systems for low-resource languages still substantially underperform compared to those for high-resource languages. This creates a double bind for non-Anglophone journalism: not only do LLMs hallucinate at higher rates in low-resource language contexts, but the detection tools available to identify and mitigate those hallucinations are themselves less capable in those same contexts.

3.3. Journalism-Specific Hallucination Risks

For journalism, the risks associated with cross-linguistic hallucination amplification manifest in specific failure modes that can directly damage public trust, harm individuals, and undermine democratic accountability. Three categories of journalism-specific hallucination risk are particularly significant.

Fabricated quotation attribution. The most direct and harmful failure mode involves the fabrication of quotes attributed to real individuals, politicians, public officials, institutional spokespersons, or private citizens. When an LLM generates a news article in Romanian and lacks sufficient contextual grounding in Romanian political discourse, it may plausibly construct a quote for a named politician that bears no relationship to anything that individual actually said. This failure mode is not hypothetical: verified cases of AI-generated journalism involving fictitious quotes attributed to real public figures have been documented in multiple countries [3]. The democratic consequence extends beyond individual reputational harm: it undermines the evidentiary basis of public accountability that the press is constitutionally expected to provide.

Geographic and institutional misrepresentation. A second category involves the confident misrepresentation of geographic, administrative, or institutional facts in local-language contexts. An LLM with limited exposure to Romanian administrative geography, Slovak institutional structures, or Bulgarian local government systems may generate articles that correctly name institutions in the target language while misrepresenting their legal powers, geographic jurisdiction, or relationship to other bodies. This type of hallucination is particularly difficult for readers to detect, the institutional names may be correct even as the attributed facts are invented, and particularly damaging in accountability journalism contexts where accurate characterisation of institutional behaviour is the central democratic function of the reporting.

Cultural and historical contextual distortion. A third category involves the imposition of English-language cultural framings onto non-Anglophone news contexts. Journalism is a culturally embedded practice: the selection, framing, and interpretation of news events draws on shared cultural codes that are language-specific and community-specific [9]. An LLM trained predominantly on English-language cultural references may generate news content in Romanian or Hungarian that, while grammatically correct, frames events through assumptions about political culture, civil society norms, or historical memory that do not correspond to Romanian or Hungarian experience. The result is news content that is factually plausible but culturally inauthentic, a subtle form of hallucination that automated fact-checking systems are entirely ill-equipped to detect.

4. Democratic Implications: Trust, Pluralism, and the Public Sphere

4.1. The Pre-Existing Trust Crisis in Journalism

The deployment of AI journalism tools in non-Anglophone contexts does not occur against a neutral backdrop. It occurs within a journalism ecosystem already experiencing a severe and accelerating crisis of public trust. The World Economic Forum [19] reports that only 40 percent of the global population trusts news, with concerns over disinformation cited as the primary driver of declining trust. A multinational survey found that 87 percent of

citizens across eight countries identified online misinformation as their top societal concern, and AI-generated misinformation has topped the World Economic Forum's Global Risks Report for two consecutive years, in 2024 and 2025 [20].

Within this context, the proliferation of AI-generated content at scale, including the approximately 1,200 AI-generated 'pink slime' sites identified by NewsGuard in the lead-up to the 2024 US election, which mimicked legitimate local news sources to spread disinformation [21], is placing unprecedented pressure on readers' capacity to distinguish credible from fabricated journalism. Von Sikorski and Hameleers [22] identify a recursive dynamic: as AI-generated misinformation proliferates, the awareness of AI's role in content production generates distrust even in accurate, human-authored journalism. The risk is not merely that individual AI-generated articles may be factually incorrect: it is that the systematic introduction of inaccuracy into minority language news ecosystems, where verification infrastructure is less developed and fact-checking organisations are fewer, could accelerate trust collapse in precisely those communities most dependent on local-language journalism for democratic information access.

4.2. The Disproportionate Impact on Minority Language Communities

The LLM language divide compounds existing structural vulnerabilities in minority language news ecosystems. Research on AI adoption in Central European newsrooms [23] found that usage does not exceed 15 percent, with LLM factual unreliability in local languages cited as the primary barrier to wider adoption. Journalists working in Czech, Slovak, Hungarian, and Romanian contexts explicitly identified language-specific hallucination as making AI tools unsuitable for unsupervised content generation in their languages. This creates a paradoxical situation: the newsrooms that could most benefit from AI efficiency gains, small, resource-constrained local and regional newsrooms, are precisely those that cannot deploy available AI tools without unacceptable quality risk.

A 2025 study of AI adoption among 128 Romanian broadcasting journalists [24] confirmed that AI tools are perceived as helpful for efficiency but that quality concerns, specifically concerns about language accuracy and cultural fidelity in Romanian, constrain their deployment in editorial contexts. Romania's media system, characterised by significant economic precarity and political pressures on editorial independence [25], represents a context in which the quality of available AI journalism tools directly contributes to, or constrains, the capacity to produce public-interest journalism.

The Journal of Democracy [26] analysis of how AI threatens democracy provides the theoretical framework for understanding why this matters at the level of political systems, not merely media industries. Generative AI that floods information ecosystems with plausible but inaccurate content threatens democratic representation by making it impossible for citizens to identify and trust reliable information sources; threatens

democratic accountability by undermining the evidentiary basis of journalistic accountability work; and corrodes social trust in ways that exacerbate political polarisation. All three threats apply with particular force in minority language communities where AI-generated content quality is systematically lower and verification infrastructure is weaker.

4.3. The Habermasian Dimension: AI and the Public Sphere

Habermas's concept of the public sphere, the social space in which private individuals come together as a public to deliberate on matters of common concern, mediated by a free press [27], provides the foundational normative framework for understanding the democratic stakes of the LLM language divide. In post-communist Central and Eastern European democracies, the construction of functioning public spheres has been incomplete, contested, and fragile. The literature on CEE media systems [28] consistently identifies political parallelism, economic clientelism, and restricted editorial independence as structural constraints on the democratic function of journalism in the region.

Into this context, AI journalism tools that systematically produce lower-quality content in minority languages introduce a new structural constraint on the functioning of non-Anglophone public spheres. The Habermasian ideal requires that participants in public deliberation have access to accurate, culturally relevant information, the kind of information that quality journalism provides [29]. When the AI tools increasingly mediating news production in non-Anglophone contexts are structurally less capable of producing accurate, culturally coherent content in those languages, the informational foundation of democratic deliberation is structurally undermined. This is not a hypothetical risk: it is the direct implication of the empirical evidence reviewed in Sections 2 and 3.

5. Towards a Framework of Linguistic Media Equity

5.1. Defining Linguistic Media Equity

The evidence reviewed in this paper points towards the need for a new normative framework through which scholars, policymakers, and practitioners can evaluate the democratic legitimacy of AI journalism systems in multilingual contexts. This paper proposes Linguistic Media Equity (LME) as that framework. LME is defined as the principle that communities have a right to AI-assisted journalism that meets equivalent standards of factual accuracy, cultural authenticity, and editorial reliability regardless of the language in which that journalism is produced, and that persistent and systematic failure to meet those standards in non-Anglophone contexts constitutes a form of structural inequality in democratic information access.

LME draws on and extends three bodies of existing scholarship. First, it builds on digital equity theory [29], which frames access to digital tools and resources as a dimension of

social justice and democratic rights rather than merely a market provision question. Second, it draws on media pluralism scholarship, and particularly the Media Pluralism Monitor [30], which identifies cultural and geographic diversity in news provision as a core dimension of democratic media systems. Third, it extends AI fairness theory [31], which has developed robust conceptual and technical tools for identifying and mitigating systematic algorithmic discrimination, into the multilingual journalism domain where these tools have not previously been applied.

5.2. The Three Dimensions of LME

LME as a normative framework operates along three dimensions, each of which generates specific evaluative criteria and policy implications.

Factual equity requires that AI journalism tools deployed in non-Anglophone contexts meet equivalent hallucination rates and factual accuracy standards to those achievable in English-language contexts, or that their limitations are transparently disclosed and editorially managed. This is a minimum threshold, not an aspirational goal: it is the baseline below which AI journalism tools are actively harmful to democratic information ecosystems regardless of other efficiency benefits they may provide.

Cultural equity requires that AI journalism tools demonstrate cultural competence, the capacity to generate content that reflects the specific political cultures, social norms, institutional frameworks, and historical memories of the communities they serve. Cultural equity cannot be achieved through translation or cross-lingual transfer alone; it requires either substantive fine-tuning on community-specific data or robust human editorial oversight capable of identifying and correcting cultural misrepresentations.

Governance equity requires that communities have meaningful participation in the governance of AI systems that mediate their public information environments. This includes the right to audit AI journalism tool performance in their languages, the right to demand transparent disclosure of training data composition and known limitations, and the right to participate in decisions about the deployment of AI tools in their news ecosystems.

5.3. LME and EU Regulatory Frameworks

The LME framework is directly legible within the emerging European Union regulatory architecture for AI. The EU AI Act (2024), which classifies AI systems according to their potential impact on fundamental rights, provides the primary regulatory instrument through which LME concerns could be addressed in European contexts. The Act's fundamental rights impact assessment requirements, transparency obligations, and high-risk system classifications are all potentially applicable to AI journalism tools that systematically underperform in minority language contexts, where the fundamental rights engaged include

freedom of expression (Article 10 ECHR), the right to information (Article 11 ECHR), and the right to media pluralism (Article 11 EU Charter).

The Digital Services Act (2022) adds a complementary layer of regulatory relevance, establishing transparency and accountability obligations for platforms that host and distribute news content. The European Media Freedom Act (2024) explicitly addresses AI in journalism in its provisions on editorial independence and algorithmic transparency, though it does not yet adequately address the language-specific dimension of AI journalism quality inequity. The Cohere AI Policy Brief on the AI Language Gap [32] provides additional regulatory context, arguing that closing the AI language gap will require coordinated action across policy, technology, and civil society domains. The LME framework provides the normative foundation from which these policy recommendations can be derived as requirements of democratic principle rather than merely aspirational industry standards.

6. Research Gaps and Future Empirical Agenda

6.1. The Measurement Gap

The most fundamental gap in the existing literature is the absence of systematic, peer-reviewed empirical benchmarking of LLM journalism content quality across non-Anglophone languages using journalism-specific evaluation criteria. The hallucination literature reviewed in Section 3 is technically sophisticated but domain-agnostic: it evaluates LLM performance on general factual tasks rather than on the specific challenges of news article generation, source attribution, institutional fact-checking, and cultural contextualisation that define journalistic quality. There is no validated, peer-reviewed instrument for measuring what this paper terms Journalistic Language Quality, the composite of factual accuracy, cultural fidelity, editorial appropriateness, and hallucination rate that determines whether AI-generated journalism serves its democratic function.

Future research should prioritise the development and validation of such an instrument, drawing on the emerging methodological literature on LLM evaluation benchmarks [18] and adapting it to the specific requirements of journalistic quality assessment. Such an instrument would need to be co-developed with professional journalists in the target languages, incorporate cultural expert evaluation alongside automated NLP metrics, and be validated against gold-standard human-authored news content across multiple journalistic genres and topics.

6.2. The CEE Newsroom Gap

A related research gap is the absence of ethnographic and organisational research on how non-Anglophone newsrooms, and specifically Central and Eastern European newsrooms,

are experiencing, negotiating, and resisting the integration of AI journalism tools that perform unevenly in their languages. The existing research [23] provides a starting point, but its coverage of Romanian, Slovak, and Bulgarian newsrooms is limited, and it does not examine the language-quality dimension of AI adoption decisions in sufficient depth. Understanding how editorial professionals in linguistically marginalised media markets make sense of AI tool limitations and develop adoption strategies is essential both for scholarly understanding and for the development of practically useful policy guidance.

6.3. The Governance Architecture Gap

Finally, there is a substantial gap in the governance literature: no comprehensive analysis yet exists of how EU AI Act, DSA, and EMFA provisions apply to the specific problem of language-specific AI journalism quality inequality, and no governance model has been proposed for ensuring multilingual performance standards in the AI journalism domain. The LME framework proposed in Section 5 provides the normative foundation for such a governance architecture, but the translation of normative principles into regulatory instruments, audit mechanisms, and enforcement procedures requires dedicated legal and policy research that the existing literature has not yet produced.

7. The LLM Language Divide as an Information Security Threat: Cyber-Resilient Intelligent Cities and the Multilingual Information Integrity Challenge

7.1. From Journalistic Quality to Urban Security: Reframing the Problem

The preceding sections have analysed the LLM language divide primarily through the frameworks of journalism studies, democratic theory, and media policy. This section extends the analysis into a domain that has received insufficient attention in the existing literature: the relationship between AI-generated multilingual misinformation and the security of intelligent urban information ecosystems. The TRUST conference proceedings (SCRD Digital Library, 2024–2025) have established Communications and Information Sabotage as a distinct domain of inquiry within the broader field of machine intelligence and smart city security, as evidenced by contributions including Virtosu [33], Drobot [34], and Vangelov [35]. This section argues that the LLM language divide constitutes a structural vulnerability within this domain, one that specifically threatens the information integrity of non-Anglophone smart cities and must be addressed within cyber-resilience frameworks.

Smart cities are, at their core, information-dependent systems. The governance of urban mobility, resource allocation, public health response, emergency services coordination, and participatory democracy in intelligent urban environments depends on the quality, accuracy, and trustworthiness of the information flows that animate these systems. Urban residents who cannot trust the journalistic and institutional information available in their

language are functionally excluded from the democratic governance of the smart city. Vangelov and Felisi [36] demonstrate that citizen engagement in smart urban environments depends critically on the quality and credibility of digital communication channels: when those channels are contaminated by unreliable AI-generated content, citizen participation atrophies and democratic legitimacy erodes.

7.2. AI-Generated Misinformation as Information Sabotage

The field of information security has traditionally conceptualised sabotage as a deliberate act by an adversarial actor, a cyberattack, a disinformation campaign, an influence operation. The LLM language divide reveals a more troubling category: structural information sabotage, in which harm to information ecosystems is produced not by deliberate adversarial action but by the systematic architectural inadequacies of AI systems deployed in contexts for which they were not adequately designed. Whether the source of inaccurate AI-generated journalism in Romanian, Slovak, or Bulgarian is a malicious actor exploiting LLM limitations or simply the unintended consequences of Anglophone-biased training data, the outcome for the affected information ecosystem is functionally identical: reduced information quality, increased hallucination rates, and erosion of public trust.

Virtosu's [33] analysis of the dual-use nature of AI in cyber security, which identifies how the same technologies that enhance defensive capabilities are increasingly exploited by adversaries to conduct more sophisticated, targeted, and evasive cyberattacks, applies with direct force to the multilingual journalism domain. AI systems capable of generating plausible minority-language news content can serve either legitimate journalistic purposes or deliberate disinformation campaigns. Crucially, the LLM language divide means that in non-Anglophone contexts, both legitimate and malicious AI-generated content will exhibit higher hallucination rates and lower cultural fidelity, making it more difficult for readers to distinguish reliable from unreliable information, and thereby amplifying the vulnerability of minority language information ecosystems to exploitation by malicious actors.

The semiotic analysis offered by Vangelov [35] of social engineering and phishing provides a complementary analytical lens. Phishing attacks succeed by exploiting cognitive shortcuts, the tendency of users to interpret familiar signs (official logos, institutional language, authority cues) as reliable indicators of trustworthiness. AI-generated journalism in minority languages exploits a directly analogous mechanism: content that exhibits the surface features of professional news production (journalistic genre conventions, institutional naming, official-sounding attribution) is processed by readers as credible even when its factual content is unreliable. The LLM language divide means that this surface-credibility / factual-unreliability gap is systematically wider in non-Anglophone contexts, creating a structural vulnerability to what might be termed 'semiotic exploitation', the

weaponisation of familiar journalistic signs to lend false credibility to hallucinated or manipulated content.

7.3. Misinformation Propagation Dynamics in Non-Anglophone Smart City Contexts

Research on misinformation propagation in technology-related domains provides important empirical grounding for understanding how AI-generated inaccuracies in minority language journalism are likely to spread within urban information ecosystems. Drobot's [34] analysis of fake news about 5G technology in smart cities demonstrates that misinformation about AI and digital infrastructure follows specific propagation dynamics: it exploits pre-existing institutional distrust, spreads rapidly through social networks within affected communities, and generates concrete downstream consequences including infrastructure vandalism, delays in digital development, and erosion of investment confidence. These dynamics are not idiosyncratic to 5G: they represent a generalisable pattern of how technologically uncertain but institutionally consequential misinformation spreads in urban contexts.

Applied to AI-generated journalism, these dynamics predict that factually inaccurate AI-generated content in non-Anglophone languages will propagate through the same channels and at comparable speed to other forms of urban misinformation. However, the LLM language divide introduces a specific amplifier: because hallucination in minority language contexts is structural rather than episodic, the supply of potentially inaccurate AI-generated content in non-Anglophone languages is continuous and persistent, not limited to specific events or actors. This chronic misinformation pressure on non-Anglophone urban information ecosystems represents a qualitatively different threat to smart city information integrity than the event-driven disinformation campaigns that cyber-resilience frameworks have primarily been designed to address.

7.4. Governance Principles for Multilingual Information Integrity in Cyber-Resilient Cities

Building cyber-resilient intelligent cities, the overarching theme of the TRUST 2026 conference, requires confronting the multilingual information integrity dimension of urban cyber resilience as a first-order security concern. The following governance principles, derived from the LME framework developed in Section 5 and extended into the information security domain, provide an initial architecture for this confrontation.

Principle 1: Multilingual information integrity auditing. Smart city governance frameworks should establish regular, independently conducted audits of AI-generated content quality in all languages used by the urban population. These audits should employ journalism-specific quality metrics, including hallucination rate, cultural fidelity, and institutional accuracy, and their results should be publicly disclosed. Audit findings should

inform procurement decisions about AI journalism tools, ensuring that tools with demonstrated language-specific quality deficits are not deployed in minority language journalistic contexts without adequate editorial oversight.

Principle 2: Minimum performance standards for AI journalism tools. Regulatory frameworks governing AI deployment in smart city information ecosystems, including the EU AI Act, the Digital Services Act, and the European Media Freedom Act, should establish minimum performance standards for AI-assisted news generation across all EU official languages. These standards should require AI journalism tool providers to disclose training data composition, conduct independent multilingual performance evaluations, and demonstrate compliance with minimum hallucination rate thresholds before deployment in news production contexts.

Principle 3: Community media resilience investment. Cyber-resilient intelligent cities must invest in the human and institutional infrastructure of minority language journalism as a dimension of information security, not merely as a cultural or media policy concern. This means sustained public funding for professional journalism in minority languages, support for community media organisations that provide editorial oversight of AI-generated content, and investment in AI literacy programmes that equip citizens with the critical evaluation skills necessary to identify and respond to AI-generated inaccuracy in their language communities.

Principle 4: Cross-border information security cooperation. The LLM language divide affects multiple EU member states simultaneously, Romanian is an official language in both Romania and the Republic of Moldova; Hungarian is co-official in Hungary, Slovakia, and spoken by significant communities in Romania; Bulgarian is used by communities across the Balkans. Effective governance of multilingual information integrity in intelligent cities therefore requires transnational coordination mechanisms, including shared hallucination detection infrastructure, cross-border journalistic fact-checking networks, and coordinated regulatory enforcement across the EU single market for AI journalism tools.

The integration of these principles into smart city cyber-resilience frameworks would represent a significant expansion of the information security agenda, from its current focus on network security, data protection, and adversarial attack mitigation to include the structural quality of the AI-generated information on which urban governance and democratic participation increasingly depend. A cyber-resilient intelligent city that cannot guarantee the information integrity of its minority language public sphere is not, in any meaningful sense, resilient at all.

8. Conclusion

This paper has argued that the LLM language divide, the systematic performance gap between AI journalism tools operating in English and those operating in non-Anglophone and minority language contexts, represents a structural challenge with direct and serious consequences for democratic information ecosystems. The root causes of this divide lie in training data composition, cultural misalignment, and tokenisation inefficiencies that systematically disadvantage languages beyond the Anglophone core of major LLM training corpora. The consequences for journalism are qualitatively distinct from the general multilingual performance challenges documented in the NLP literature, because journalism is not merely a linguistic exercise but a democratic public service whose quality directly determines the informational foundation of democratic deliberation.

The hallucination risks associated with cross-linguistic LLM deployment in journalism, including fabricated attribution, institutional misrepresentation, and cultural distortion, are amplified in non-Anglophone contexts precisely because the statistical grounding that constrains hallucination in English is absent for lower-resource languages. As established in Section 7, and consistent with Virtosu's [33] analysis of AI's dual-use nature in cyber security, the same generative capabilities that enable legitimate AI-assisted journalism in minority languages also lower the technical barrier for deliberate information sabotage targeting those communities. In an information environment already characterised by declining public trust and proliferating AI-generated misinformation, the systematic introduction of higher-hallucination-rate AI tools into minority language journalism ecosystems represents an active threat to democratic resilience and the information security of intelligent urban environments.

The normative framework of Linguistic Media Equity (LME) proposed in Section 5 provides a basis for evaluating AI journalism systems not merely on technical performance criteria but on the democratic requirements of factual equity, cultural equity, and governance equity. Translated into the smart city domain through Section 7's four governance principles, LME offers a concrete architecture for ensuring multilingual information integrity as a dimension of urban cyber resilience, directly relevant to the TRUST 2026 conference theme of Building Cyber-Resilient Intelligent Cities. The EU's emerging regulatory architecture, the AI Act, Digital Services Act, and European Media Freedom Act, provides the legislative context within which these principles could be translated into enforceable standards, though significant legal and policy work remains.

The scholarly agenda this paper identifies encompasses three priority research directions: systematic empirical benchmarking of journalistic language quality across non-Anglophone contexts using a validated Journalistic Language Quality instrument; ethnographic research on CEE newsroom AI adoption decisions and their language-

specific dimensions; and the development of a multilingual AI journalism governance architecture that operationalises LME within existing EU regulatory frameworks. This programme is of both academic importance and urgent democratic relevance, and, as Section 7 has demonstrated, of growing information security significance as well.

At a moment when journalism's role as the informational foundation of democratic life is under unprecedented pressure from declining trust, proliferating AI-generated content, and the erosion of local and minority language newsrooms, ensuring that the AI tools transforming news production serve all language communities equitably is not merely a technical challenge. It is a democratic imperative, and, in the age of the intelligent city, a security imperative too.

Declarations

Author Contributions: The sole author conceived the research question, conducted the literature review, developed the theoretical framework (LME), designed the smart city governance principles, and wrote the manuscript in its entirety.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest: The author declares no conflicts of interest. The author is affiliated with Story Agents SRL, the only Romanian newsroom in the Google CEE.AI Lab, which may have a commercial interest in AI journalism tools; this affiliation did not influence the research findings.

Data Availability: This paper is a theoretical review article. No primary data were generated. All sources cited are peer-reviewed publications or publicly accessible institutional reports, fully cited in the reference list below.

Ethical Approval: Not applicable.

References

[1] J. Habermas, *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society.* Cambridge, MA: MIT Press, 1989.

[2] A. F. Sonni, "Digital transformation in journalism: mini review on the impact of AI on journalistic practices," *Front. Commun.,* vol. 10, art. 1535156, Feb. 2025. doi: 10.3389/fcomm.2025.1535156

[3] N. Newman *et al.*, *Reuters Institute Digital News Report 2024,* Reuters Institute for the Study of Journalism, University of Oxford, Oxford, UK, Tech. Rep., 2024.

[4] X. Qin *et al.*, "A Survey on Multilingual Large Language Models: Corpora, Alignment, and Bias," arXiv preprint arXiv:2404.00929v2, Apr. 2024. doi: 10.48550/arXiv.2404.00929

[5] L. Huang *et al.*, "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions," arXiv preprint arXiv:2311.05232v2, Nov. 2024. doi: 10.48550/arXiv.2311.05232

[6] X. Zhang, S. Li, B. Hauer, N. Shi, and G. Kondrak, "Don't trust ChatGPT when your question is not in English: A study of multilingual abilities and types of LLMs," in *Proc. 2023 Conf. on Empirical Methods in Natural Language Processing (EMNLP),* pp. 7915–7927, 2023. doi: 10.18653/v1/2023.emnlp-main.491

[7] Z. Li *et al.*, "Quantifying Multilingual Performance of Large Language Models Across Languages," arXiv preprint arXiv:2404.11553, Apr. 2024. doi: 10.48550/arXiv.2404.11553

[8] C. Meinhardt, S. Nong, G. Webster, T. Hashimoto, and C. Manning, "Mind the Language Gap: Mapping the Challenges of LLM Development in Low-Resource Language Contexts," Stanford HAI White Paper, Stanford University, Stanford, CA, 2024. Available: https://hai.stanford.edu/policy/mind-the-language-gap

[9] M. Schudson, *Why Democracies Need an Unlovable Press.* Cambridge, UK: Polity Press, 2008.

[10] F. Motoki, V. P. Neto, and V. Rodrigues, "More human than human: Measuring ChatGPT political bias," *Public Choice,* vol. 198, pp. 3–23, 2024. doi: 10.1007/s11127-023-01097-2

[11] S. H. Amanda Tan *et al.*, "Multilingual capabilities of GPT: A study of structural ambiguity," *PLOS ONE,* 2025. doi: 10.1371/journal.pone.0319766

[12] A. Conneau *et al.*, "Unsupervised Cross-lingual Representation Learning at Scale," in *Proc. 58th Annual Meeting of the Association for Computational Linguistics,* pp. 8440–8451, 2020. doi: 10.18653/v1/2020.acl-main.747.

[13] A. Alansari *et al.*, "Large Language Models Hallucination: A Comprehensive Survey," arXiv preprint arXiv:2510.06265v2, Oct. 2025. doi: 10.48550/arXiv.2510.06265

[14] P. Anh-Hoang, N. Tran, and T. Nguyen, "Survey and analysis of hallucinations in large language models: attribution to prompting strategies or model behavior," *Front. Artif. Intell.,* vol. 8, art. 1622292, 2025. doi: 10.3389/frai.2025.1622292

[15] "The rise of hallucination in large language models: systematic reviews, performance analysis and challenges," *Cluster Comput.,* Springer, 2025. doi: 10.1007/s10586-025-05891-z

[16] K. Guerreiro *et al.*, "How Much Do LLMs Hallucinate across Languages?" in *Proc. 2025 Conf. on Empirical Methods in Natural Language Processing (EMNLP),* 2025. doi: 10.18653/v1/2025.emnlp-main.1481

[17] "Bridging the Gap: Enhancing LLM Performance for Low-Resource African Languages with New Benchmarks, Fine-Tuning, and Cultural Adjustments," arXiv preprint arXiv:2412.12417, Dec. 2024. doi: 10.48550/arXiv.2412.12417

[18] P. Vásquez *et al.*, "SemEval-2025 Task 3: Mu-SHROOM, the Multilingual Shared-task on Hallucinations and Related Observable Overgeneration Mistakes," Association for Computational Linguistics, 2025.

[19] World Economic Forum, "Building trustworthy media ecosystems in the age of AI and declining trust," *WEF Insight Report,* Davos, Switzerland, 2023. Available: https://www.weforum.org

[20] W. Ejaz, R. Fletcher, R. K. Nielsen, and S. McGregor, "Misinformation and AI-Generated Content: A Multinational Survey of Public Attitudes," Reuters Institute Working Paper, University of Oxford, 2024.

[21] K. Ilagan and F. A. Rainis, "Chaos and Credibility: A Snapshot of How AI Is Impacting Press Freedom and Investigative Journalism," Global Investigative Journalism Network, Jan. 2025. Available: https://gijn.org

[22] C. von Sikorski and M. Hameleers, "Disinformation in the Age of Artificial Intelligence (AI): Implications for Journalism and Mass Communication," *Journalism & Mass Communication Quarterly,* 2025. doi: 10.1177/10776990251375097

[23] Thomson Foundation and Media and Journalism Research Center, "AI and Media in Central Europe: Between Fears and Hopes," Thomson Foundation Research Report, London, 2024. Available: https://journalift.org

[24] C. Mungiu-Pippidi *et al.*, "The Role of Artificial Intelligence in Romanian Broadcasting: Opportunities and Challenges," *Journal. Media,* vol. 6, no. 1, art. 22, Feb. 2025. doi: 10.3390/journalmedia6010022

[25] P. Gross, *The Cultural Core of Media Systems: The Romanian Case.* Lanham, MD: Rowman & Littlefield, 2023.

[26] A. Dafoe and S. Russell, "How AI Threatens Democracy," *Journal of Democracy,* vol. 36, no. 1, pp. 5–20, 2025. doi: 10.1353/jod.2025.a951337

[27] J. Habermas, *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy.* Cambridge, MA: MIT Press, 1996.

[28] D. Hallin and P. Mancini, *Comparing Media Systems: Three Models of Media and Politics.* Cambridge, UK: Cambridge University Press, 2004.

[29] "Digital media, democracy and civil society in Central and Eastern Europe," *J. Inf. Technol. Politics,* vol. 21, no. 1, 2024. doi: 10.1080/19331681.2023.2257012

[30] E. Brogi *et al.*, *Media Pluralism Monitor 2022.* Florence: Centre for Media Pluralism and Media Freedom, European University Institute, 2022.

[31] S. Barocas, M. Hardt, and A. Narayanan, *Fairness and Machine Learning: Limitations and Opportunities.* Cambridge, MA: MIT Press, 2023.

[32] Cohere AI Research, "The AI Language Gap: Policy Primer," Cohere Policy Brief, Toronto, Canada, 2024. Available: https://cohere.com/research/papers/the-ai-language-gap.pdf

[33] I. Virtosu, "Guardians or threats? The double-edged role of artificial intelligence in cyber security," *Int. Conf. Mach. Intell. Secur. Smart Cities (TRUST) Proc.,* vol. 2, pp. 61–73, Pro Universitaria, Dec. 2025. ISSN Print: 3061-2330. Available: https://www.scrd.eu/index.php/trust/article/view/744

[34] I.-A. Drobot, "Fake news about 5G in smart cities: Reasons and consequences," *Int. Conf. Mach. Intell. Secur. Smart Cities (TRUST) Proc.,* vol. 2, pp. 165–170, Pro Universitaria, Dec. 2025. Available: https://www.scrd.eu/index.php/trust/article/view/752

[35] N. Vangelov, "Social engineering and phishing: A semiotic perspective," *Int. Conf. Mach. Intell. Secur. Smart Cities (TRUST) Proc.,* vol. 2, pp. 109–118, Pro Universitaria, Dec. 2025. Available: https://www.scrd.eu/index.php/trust/article/view/747

[36] N. Vangelov and S. Felisi, "Advertising for the people: Digital media and citizen engagement in the smart urban environment," *Smart Cities Reg. Dev. (SCRD) J.,* vol. 10, no. 1, pp. 129–137, Feb. 2026. doi: 10.25019/gesn4v44.