

Article

Digital inclusion and sustainable language preservation: Examining Wikipedia as a tool for social equity through Setswana and Hà Nhì language communities

Shahid Minhas^{1,*}, Abiodun Salawu², Thu Vo Anh¹¹ School of Communications and Design, RMIT University, Hanoi 11110, Vietnam² Indigenous Language Media in Africa (ILMA), North-West University, Mahikeng Campus, Mahikeng 2777, South Africa* **Corresponding author:** Shahid Minhas, shahidihsas@gmail.com

CITATION

Minhas S, Salawu A, Anh TV.
Digital inclusion and sustainable
language preservation: Examining
Wikipedia as a tool for social equity
through Setswana and Hà Nhì
language communities. *Sustainable
Social Development*. 2026; 4(4):
8716.
<https://doi.org/10.23812/ssd8716>

ARTICLE INFO

Received: 28 May 2026
Revised: 6 July 2026
Accepted: 13 July 2026
Available online: 11 August 2026

COPYRIGHT

Copyright © 2026 by author(s).
Sustainable Social Development is
published by Asia Pacific Academy
of Science Pte. Ltd. This work is
licensed under the Creative
Commons Attribution (CC BY)
license.
[https://creativecommons.org/licenses/
by/4.0/](https://creativecommons.org/licenses/by/4.0/)

Abstract: This article investigates the role of Wikipedia as a mechanism for digital inclusion and sustainable language preservation among two minority language communities, Setswana speakers in South Africa and the Hà Nhì ethnic minority in Vietnam. A quanti-qualitative mixed-methods design combined semi-structured interviews with Wikipedia editors, language activists, digital literacy educators, and community practitioners with quantitative Wikipedia platform metrics, including article counts, active editor figures, and content growth trajectories, was employed to investigate this issue. The results show that both communities demonstrate emergent but constrained digital participation. Their barriers are structurally distinct but epistemically convergent, i.e., infrastructural inequity, editorial gatekeeping, procedural governance, and the systemic undervaluation of non-hegemonic knowledge all recur across cases. Cross-case comparison shows that Wikipedia's ostensibly neutral platform architecture reproduces existing knowledge hierarchies by privileging well-resourced language communities. The study introduces "platform-induced hermeneutical injustice" as a fifth dimension of digital exclusion, supplementing existing models of digital inclusion. It grounds Fricker's epistemic injustice framework in digital platform governance and offers evidence-based implications for the Wikimedia Foundation, national governments, educators, and language preservation practitioners across the Global South.

Keywords: digital inclusion; epistemic justice; Wikipedia; Setswana; Hà Nhì; linguistic sustainability; social equity

1. Introduction

The global linguistic landscape is experiencing an unprecedented crisis. At least 40 % of the approximately 7000 languages spoken around the world today are severely endangered. Linguists estimate that by the end of the twenty-first century, between 50 % and 90 % of existing languages may face extinction [1]. This trajectory signifies not only the destruction of communicative systems but also the irrevocable loss of accumulated cultural knowledge, civilization, and epistemological diversity, along with social identities and values. In this context, language death represents both a cultural and epistemic catastrophe, disproportionately impacting communities that are already marginalized within global political and economic hierarchies.

The emergence of digital platforms as a source of knowledge production and cultural preservation has opened new analytical and practical ground for scholars, activists, technologists, and policymakers concerned with linguistic sustainability. Wikipedia, the world's largest open-access, collaboratively edited encyclopedia, has

attracted the attention of scholars and practitioners as a potential instrument of digital social innovation for minority language communities. With over 60 million articles across more than 330 language editions as of 2023 [2,3], the platform offers, in principle, a democratic and democratizing infrastructure for the documentation, revitalization, and global dissemination of endangered and minority languages.

However, this optimistic framing has a deeper structural issue. The distribution of content and editorial participation on Wikipedia is profoundly unequal. Language editions associated with high-income, technologically advanced, and geographically central communities, i.e., English, German, French, and Spanish, command millions of articles and thousands of active editors. By contrast, editions in minority, indigenous, and politically marginalized languages remain characteristically sparse, fragmented, and editorially limited. This disparity is not incidental; it reflects and reproduces the broader digital divide, which van Dijk [4] has conceptualized as nine forms of power that comprise the framework: construction, coercion, domination, discipline, dependency, information, persuasion, and authority—and at all levels (individuals, organizations, and societies).

The present study focuses on two such underrepresented language communities: Setswana-speaking populations in South Africa, representing a Bantu language community embedded within the complex postcolonial linguistic politics of Southern Africa [1], and the Hà Nhi ethnic minority in Vietnam, whose language belongs to the Tibeto-Burman family and is spoken by approximately 17,000 individuals concentrated in the mountainous regions of northern Vietnam [5]. These two cases are analytically productive precisely because of their asymmetry. They differ in geographic location, colonial history, linguistic typology, state policy context, and platform engagement, yet they share a common epistemic predicament, the systematic marginalization of their languages within global digital knowledge systems.

This study aims to answer these four research questions. First, how do Setswana and Hà Nhi language communities perceive Wikipedia as a tool for digital inclusion and language sustainability? Second, what structural and epistemic barriers impede meaningful participation in these communities' Wikipedia-related digital practices? Third, what are the implications of these findings for theories of digital inclusion, epistemic justice, and sustainable cultural development? Fourth, what policy and institutional interventions are necessary to transform Wikipedia into a genuinely equitable instrument of linguistic sustainability?

2. Literature review

2.1. The digital divide and minority language representation

The concept of the digital divide has evolved considerably since its initial articulation as a binary distinction between those with and without internet access [6]. Contemporary scholarship has reframed the divide as a multidimensional and dynamic phenomenon encompassing access to infrastructure, digital literacy, content relevance, and participatory capacity [4,7]. Selwyn [8] has argued persuasively that the digital divide is not a technical problem amenable to technical solutions, but a social phenomenon inseparable from pre-existing patterns of inequality rooted in class, race, gender, and geographic marginalization. Van Dijk [9] further elaborates this position

through his four-barrier model, identifying motivational, material, skills-based, and usage dimensions as interdependent layers of exclusion that reinforce one another across social strata.

The digital divide has a unique aspect for minority language communities. It involves not just connectivity or device ownership, but also the representation of their languages within digital content ecosystems. Raihan and colleagues [10] have documented how deficit-oriented framings of marginalized communities, including language minorities, obscure the systemic dimensions of digital exclusion and redirect attention toward individual behavioral change rather than structural transformation. In the context of language preservation, this deficit framing is particularly damaging, as it positions community members as lacking the agency or technical competence to participate in digital knowledge production, rather than addressing the institutional and platform-level barriers that structurally constrain their participation.

Mair [11] argues that digital technologies simultaneously create new opportunities for minority language communities and impose new forms of cultural and linguistic pressure. The dominance of English, and, to a lesser extent, other major European and Asian languages, in digital content production constitutes what Canagarajah [12] has theorized as a form of translingual subordination, wherein minority language speakers are systematically financed to produce content in hegemonic languages at the expense of their own linguistic heritage. This dynamic operates not through explicit coercion but through the structural asymmetries of platform design and digital resource allocation.

2.2. Wikipedia and minority language communities

Wikipedia has been the subject of a substantial and growing body of research examining its capacity to serve as an instrument of equitable knowledge production. Minhas and Salawu [3] documented striking geographic and linguistic inequalities in Wikipedia's global coverage, especially focusing on Setswana and Punjabi languages, and Yin [13] demonstrated that North American and Western European topics are disproportionately represented relative to their demographic or cultural significance, a pattern the authors term digital divisions of labor. Ferran-Ferrer et al. [14] identified gender biases in Wikipedia's editorial community, while Rollin and Lages [15] documented cultural biases in the representation of national figures across language editions.

Studies particularly investigating Wikipedia's treatment of minority and indigenous languages, such as Nguyen et al. and colleagues [16] analyzed the structural characteristics of small Wikipedia language editions, identifying patterns of low article density, high content importation from larger editions, and limited editorial diversity as characteristic features of minority language Wikipedias. Ford [17] observed that while Wikipedia's founding philosophy proclaims universal editability, the actual distribution of participation privileges those with the social capital, time, technical skills, and linguistic confidence to navigate the platform's complex norms and bureaucratic architecture, a finding that resonates powerfully with Guillory's [18] concept of cultural capital.

Research specifically addressing African language Wikipedias remains limited

but is beginning to emerge. Sälevä and Lignos [19] have examined Swahili, Afrikaans, and Zulu Wikipedia editions, identifying the role of diaspora communities, university-based initiatives, and individual champions in sustaining modest levels of editorial activity. The Setswana Wikipedia (language code: TN), established in 2004, has followed a comparable trajectory, characterized by episodic growth driven by short-term institutional initiatives, notably WikiAfrica-sponsored editing sprints, and hampered by chronic shortages of consistent editorial engagement. Research on Southeast Asian minority language digital spaces is even more sparse. Do [20] has examined Vietnam's digital linguistic landscape, noting its domination by Vietnamese and the minimal institutional or platform support received by ethnic minority languages despite state-mandated commitments to ethnic cultural preservation.

2.3. Language preservation in the digital era

The discourse on digital language preservation has been shaped by two broadly contending perspectives. The first optimistic view holds that digital platforms offer unprecedented opportunities for minority language documentation, revitalization, and community building [21]. Proponents of this viewpoint point to cases of successful digital language activism, Welsh, Hawaiian, Māori, as evidence that digital tools can galvanize community engagement and generate new speakers [22]. The second, more critical perspective, articulated by scholars including Mühlhäusler [23] and Cunliffe and Herring [24], argues that digital environments, absent deliberate institutional intervention, tend to accelerate the erosion of linguistic diversity by reinforcing the communicative dominance of high-prestige languages.

The academic consensus has gravitated toward a more nuanced position that acknowledges both the emancipatory potential and the structural constraints of digital language preservation efforts. UNESCO's framework on language vitality and endangerment [25] has been particularly influential in establishing the principle that sustainable digital language preservation requires more than technology deployment: it demands the development of linguistic corpora and computational tools, the cultivation of literate and technically proficient editorial communities, the establishment of institutional partnerships, and the integration of language preservation within broader frameworks of cultural sustainability and social equity.

2.4. Knowledge production, the global south, and platform governance

The politics of knowledge production in the Global South constitute a crucial backdrop for understanding the limitations of platform-based language preservation. Fricker's [26] concept of epistemic injustice has been taken up by digital scholars including Birhane [27], who argues that algorithmic systems, including knowledge platforms, systematically reproduce and amplify the epistemic exclusions of the societies within which they are developed. Mignolo's [28] concept of the 'colonial matrix of power' within knowledge systems provides an additional theoretical lens, illuminating how Wikipedia's structural architecture, its reliance on text-based, citation-verifiable content, its bureaucratic dispute-resolution mechanisms, and its English-language dominance in meta-discussions and policy documentation reflect and reinforce longstanding hierarchies of epistemic authority.

Platform governance has emerged as a critical but underexamined dimension of this problematic. Benkler [29] has argued that the governance architecture of digital commons, including Wikipedia, is not politically neutral; it encodes assumptions about expertise, participation, and legitimacy that may systematically advantage certain communities over others. Spyridou [30] further developed this line of analysis by examining the institutional rules that govern knowledge commons, arguing that the sustainability of such commons depends critically on whether governance arrangements are perceived as legitimate by all participating communities. Graham et al. [31] extend this argument to the global digital sphere, demonstrating that patterns of digital knowledge production map closely onto pre-existing geopolitical and economic hierarchies, a finding that underscores the need for critical, structurally informed analysis of Wikipedia's role in minority language communities.

3. Theoretical framework

This study is grounded on an integrated and critically oriented theoretical framework that engages in a dialogue between digital inclusion theory and epistemic justice theory. This framework serves as mutually reinforcing analytical lenses, allowing for a nuanced examination of Wikipedia as a socio-technical system. This framework highlights the interaction between structural inequalities and platform governance, illuminating how digital participation is influenced by power dynamics, normative standards, and imbalances in knowledge production. This approach enables an analysis that transcends superficial participation metrics to explore how inclusion is negotiated, constrained, and transformed within global knowledge infrastructures.

3.1. Digital inclusion theory

Drawing on van Dijk's [4,9] multi-dimensional model of digital access, comprising motivational, material, skills, and usage dimensions, and Fratini's [32] emphasis on digital sovereignty, which involves the assertion of control over digital infrastructures to safeguard autonomy in the digital realm. This reconceptualization challenges reductionist accounts that equate access with connectivity, instead of foregrounding the conditions under which participation becomes meaningful, sustained, and agentic. Within the context of Wikipedia, this theoretical lens demands attention not only to disparities in infrastructure and digital literacy but also to the institutional norms and governance regimes that regulate content production.

Critically, the framework extends van Dijk's model by introducing a fifth dimension, epistemic architecture. This dimension captures the normative, procedural, and classificatory systems through which knowledge is validated, curated, and rendered visible [4,9]. It thus exposes how platform rules, such as verifiability and notability, may inadvertently privilege hegemonic knowledge systems while marginalizing oral, indigenous, or context-specific epistemologies. By incorporating epistemic architecture, the study shifts the analytical focus from access alone to the politics of recognition embedded in digital environments, thereby revealing the structural boundaries that delimit participation for Setswana and Hà Nhì communities.

3.2. Epistemic justice theory

Digital inclusion theory and epistemic justice theory are not parallel lenses applied to the same data; they are mutually constitutive within this framework. Van Dijk's access-based model explains who can reach the platform, but it cannot, on its own, explain why fluent, connected, digitally literate contributors still find their knowledge excluded once they arrive [4,9]. Epistemic justice theory supplies that missing mechanism: it identifies the normative architecture, verifiability, notability, citation-based legitimacy, through which inclusion at the access level can coexist with exclusion at the level of recognition. The epistemic architecture dimension introduced below is the hinge between the two traditions: it is simultaneously the fifth dimension of digital inclusion and the platform-specific site where Fricker's hermeneutical injustice becomes operative [26].

Fricker's [26] theory of epistemic injustice provides the critical core of the framework, offering a normative vocabulary to interrogate the marginalization of knowledge producers within digital platforms. The distinction between testimonial injustice, where credibility is unjustly denied, and hermeneutical injustice, where experiences are obscured by gaps in shared interpretive resources, illuminates how minority language contributors are systematically disadvantaged in Wikipedia's knowledge economy. Their contributions are delegitimized due to a perceived lack of authoritative sources, while their lived realities remain underrepresented within dominant epistemological schemas.

Therefore, the study advances the concept of platform-induced hermeneutical injustice, defined as the structural incapacity of digital knowledge architectures to accommodate non-hegemonic ways of knowing. This extension underscores the role of platform design and policy in reproducing epistemic exclusion, shifting the analytical focus from individual bias to systemic constraints. So, it positions Wikipedia not merely as a neutral repository of knowledge but as a contested epistemic space where inclusion is mediated by historically situated power relations. This critical perspective facilitates a more profound examination of how digital platforms can concurrently promote visibility and sustain forms of epistemic erasure, especially for linguistically marginalized communities.

4. Methodology

This study employs a quanti-qualitative mixed-methods comparative case study design [33]. Comparative case study enables the analytical identification of both convergent and divergent patterns across two carefully selected cases, facilitating theoretically generative inferences about the relationship between contextual factors and outcomes related to digital inclusion and linguistic sustainability. The quanti-qualitative design reflects the interpretive and socially constructed nature of the phenomena under investigation; barriers, perceptions, and equity implications cannot be adequately captured through quantitative metrics alone, while the integration of Wikipedia platform metrics provides an essential material-structural grounding for qualitative findings. Biddix and Bourke's [34] foundational criteria for case study rigor, construct validity, internal validity, external validity, and reliability informed all stages of design, data collection, and analysis.

4.1. Case selection and justification

The selection of Setswana (South Africa) and Hà Nhì (Vietnam) reflects a strategic reason of maximum variation along several theoretically relevant dimensions such as geographic location, colonial history, national language policy context, and Wikipedia platform engagement history. This is combined with minimal variation in the central explanatory variable, where both cases represent minority language communities in the Global South with emergent yet constrained digital language presence. This design allows comparative analysis to identify context-specific factors while also isolating cross-contextual patterns attributable to shared structural dynamics.

Setswana is a Bantu language spoken by approximately 4.6 million people in South Africa, where it holds official language status [35]. Its Wikipedia edition, established in 2004 (ISO code: TN), has grown modestly to approximately 4866 articles and 870 articles in Q3 2023 only (see **Table 1**), with a small but active editorial community. Hà Nhì is a Tibeto-Burman language spoken by approximately 17,000 individuals in northern Vietnam, concentrated in Lào Cai and Điện Biên provinces [5]. It currently lacks an independent Wikipedia edition, though limited Hà Nhì language content exists within the broader Vietnamese Wikipedia infrastructure and in external digital archives maintained by researchers and civil society organizations. While the two cases differ substantially in speaker population (4.6million vs 17,000), this maximum variation design provides analytical leverage for understanding how scale interacts with structural barriers. It allows us to examine both mid-sized official minority languages (Setswana) and very small, oral languages (Hà Nhì). This comparison highlights common platform-induced epistemic challenges across different magnitudes of marginalization rather than assuming strict equivalence.

Table 1. Comparative Wikipedia metrics: Setswana and Hà Nhì language communities.

Metric	Setswana Wikipedia (TN)	Hà Nhì (No dedicated edition)
Wikipedia Edition Status	Active (since 2004)	No dedicated edition
ISO Language Code	tn	,
Total Articles	870 (Q3 2023)	N/A
Registered Users	89	N/A
Active Editors (monthly avg.)	9	N/A
Total Accumulated Edits	14,320	N/A
Annual Growth Rate (2019–2023)	+12.4 %	N/A
Hà Nhì Content (External Digital Repositories)	N/A	~47 documents
Approximate Speaker Population	~4.6 million (SA + diaspora)	~17,000
Official Language Status	Yes (South Africa)	No
Community Digital Literacy Rate (est.)	~58 %	~21 %
Regional Internet Penetration (est.)	~72 % (national, SA)	~31 % (highland regions, VN)
Institutional Support for Digitization	Moderate (WikiAfrica, DHET)	Limited (scattered NGO)
Existence of Standardized Digital Script	Yes (Latin-based)	Partial (contested)

Note: Data extracted from stats.wikimedia.org and community digital audits, Q3 2025. Digital literacy and internet penetration estimates drawn from national and regional surveys [4,32]. N/A = not applicable given the absence of a dedicated edition.

4.2. Participants and sampling strategy

Twenty-four participants were recruited through a purposive-then-snowball sampling strategy [36], following Patton's [36] criterion of purposive selection based on information richness rather than statistical representativeness. Initial participants were identified through Wikimedia project participation lists, language advocacy networks, and academic referrals. Subsequent participants were identified through recommendation chains anchored in the initial purposive sample. Sampling continued until thematic saturation was achieved, operationalized as the point at which no new codes emerged from consecutive interviews. Participants were selected to represent diverse roles within each language community's digital preservation ecosystem, including active Wikipedia editors, language activists, educators, community leaders, NGO practitioners, and academic researchers.

4.3. Data collection procedures

4.3.1. Semi-structured interviews

Semi-structured interviews were conducted between March and October 2023. Interviews lasted between 45 and 90 min ($M = 69$ min), were conducted in English, Setswana, Vietnamese, and Hà Nhì with professional interpreter support where necessary, and were scripted and audio-recorded with participant consent. A semi-structured interview guide was developed and piloted with two expert reviewers, organized around four thematic domains: (i) awareness of and engagement with Wikipedia; (ii) experiences of participation or non-participation; (iii) perceived barriers and facilitators; and (iv) perceptions of language sustainability in digital environments. The guide allowed for flexibility and probing to accommodate participant-led narrative trajectories. All recordings were professionally transcribed and, where relevant, translated, with translation quality verified by bilingual reviewers.

4.3.2. Wikipedia platform metrics

Quantitative data were extracted from Wikipedia's publicly accessible statistics infrastructure (stats.wikimedia.org) and supplemented by manual content audits of the Setswana Wikipedia edition. Metrics collected included total article count, active editor figures (editors making five or more edits per month), total accumulated edits, content growth trajectory (2010–2023), and article deletion and retention rates. For Hà Nhì, given the absence of a dedicated edition, metrics related to Hà Nhì language content in the Vietnamese Wikipedia and external digital repositories were collected through manual assessment methods. **Table 1** presents comparative Wikipedia metrics for both cases.

4.4. Analytical approach

Automatic thematic analysis was applied to the collected data following the six-phase procedures outlined by Braun and Clarke [37]: (1) familiarization with data; (2) initial code generation; (3) theme construction; (4) theme review; (5) theme definition and naming; and (6) production of the analytic narrative. Coding was conducted using NVivo 14 software. Initial open coding generated 142 codes across the full dataset; axial coding consolidated these into 22 sub-themes; selective coding organized sub-themes into four superordinate analytical themes. A second coder independently coded

a 20 % subsample of transcripts, yielding a Cohen's kappa coefficient of 0.81, indicating strong inter-rater reliability.

Cross-case comparison was conducted as shown in **Table 2** using a structured matrix approach, systematically comparing Setswana and Hà Nhì findings for each analytical theme to identify convergent and divergent patterns. Quantitative metrics data were analyzed descriptively and integrated into the qualitative narrative at the discussion stage, following the 'connected' mixed-methods integration model [33]. **Table 3** presents the thematic coding summary. Ethical approval was obtained from (anonymized for review) in accordance with protocols governing research involving human subjects.

5. Findings

5.1. Thematic overview

Analysis of interview data and platform metrics generated four analytically rich themes: (1) perceptions of Wikipedia as a language preservation instrument; (2) structural and epistemic barriers to platform participation; (3) social equity implications of digital language work; and (4) the perceived sustainability of community-driven digital language efforts. The Setswana Wikipedia article growth trajectory is presented in **Figure 1**. **Table 2** provides a structured cross-case comparison across all four themes. **Table 3** presents the thematic coding summary. The following sections present each theme, drawing on participant testimony and supported by quantitative evidence where relevant.

5.2. Theme 1: Perceptions of Wikipedia as a language preservation instrument

Across both cases, participants articulated a nuanced and internally differentiated set of perceptions regarding Wikipedia's value and limitations as a language preservation tool. A dominant pattern, present in all 22 participants who contributed to this theme, was what the researcher termed 'qualified optimism': A genuine belief in the platform's symbolic and practical potential, consistently tempered by awareness of structural constraints.

Among Setswana editors and practitioners, Wikipedia was predominantly perceived as a site of epistemic legitimation see **Table 2**, a space in which the language's existence, vitality, and intellectual capacity could be asserted against a global cultural backdrop that routinely marginalizes African languages. This perception aligns directly with Fricker's [26] concept of testimonial justice: the platform is experienced as a space in which epistemic credibility can be reclaimed.

"When I add an article in Setswana, I am not just giving information. I am saying: this language is alive. It can explain science. It can explain history. It is not primitive."
P01, Wikipedia Editor, Mahikeng.

Among participants associated with the Hà Nhì community, perceptions of Wikipedia were more ambivalent, reflecting the absence of a dedicated Hà Nhì edition and the community's limited digital infrastructure. Several participants articulated 'aspirational distance': A clear awareness of Wikipedia's existence and value

combined with a strong sense of its inaccessibility under current community conditions.

Table 2. Cross-case thematic comparison: Setswana and Hà Nhì communities.

Theme	Setswana community	Hà Nhì community	Convergences	Divergences
Perceptions of Wikipedia	Epistemic legitimization; qualified optimism; platform as symbolic assertion of language vitality; institutional exposure creates awareness	Aspirational distance; awareness without access; hope deferred by structural inaccessibility; no direct editorial engagement	Both communities' value Wikipedia's symbolic role; both identify gap between aspiration and actual participation capacity.	Setswana: Active though limited engagement; Hà Nhì: more fundamental physical and institutional distance from platform
Barriers to Participation	Verification norms excluding oral knowledge; skills gaps; time poverty; absence of Setswana-language Wikipedia interface support	No written standard (contested orthography); severe connectivity deficits; lack of platform interface in Hà Nhì; no editorial community	Epistemic architecture (platform norms) is the central structural barrier; temporal and economic constraints universal across cases.	Setswana: Barriers are navigable for a minority of users; Hà Nhì: barriers are foundational and largely pre-platform
Social Equity Implications	Post-apartheid epistemic marginalization; digital absence reinforces historical inferiority; economic and cultural visibility stakes	Educational equity (medium of instruction mismatch); generational knowledge loss; limited development visibility internationally	Digital absence perpetuates and deepens existing social marginalization; economic and cultural consequences in both cases.	Nature of equity stakes differs: political/historical in SA context; educational and developmental in Vietnamese highland context
Sustainability of Digital Efforts	Sprint-dependency; episodic article growth; fragile editorial community; no multi-year institutional funding mechanisms	Foundational institutional absence; dispersed preservation efforts across platforms; no central aggregation infrastructure	Both communities describe precarious, unsustainable, individually-driven efforts with no institutional long-term funding.	Setswana: Precarious but present platform infrastructure; Hà Nhì: pre-platform stage of digital language development

Note: Themes derived from reflexive thematic analysis of 24 semi-structured interviews. Comparative matrix constructed using structured cross-case analysis procedures.

“We know about Wikipedia. We know it is important. But to write there, you need good skills, you need support, you need to know how to write the language officially. We do not have all these things yet.” P16, Educator, Hanoi

A significant divergence between cases concerns the role of institutional backing. Setswana's status as an official South African language has historically generated modest levels of institutional support for language digitization. Multiple Setswana participants noted WikiAfrica-organized editing sprints as catalysts for article creation bursts, a pattern visible in the growth trajectory presented in **Figure 1**. This qualitative account is supported by quantitative data among the article creation spikes shown in **Figure 1**; three of the four largest align with documented WikiAfrica sprint dates. In contrast, the periods in between show almost no growth, reflecting the pattern of institutional dependency that participants described during the interviews.

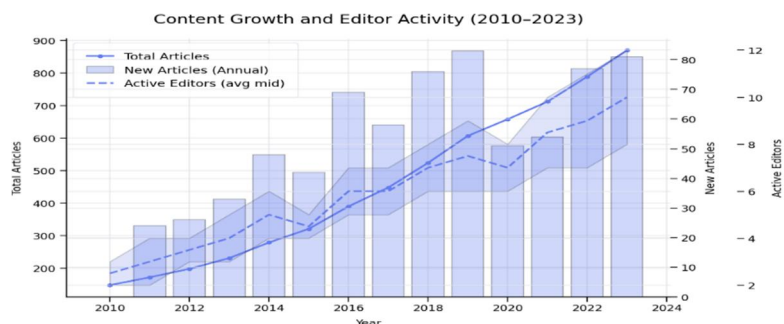


Figure 1. Setswana Wikipedia article growth trajectory, 2010–2023.

No comparable institutional infrastructure was identified in the Hà Nhì case, where digital language work remained almost entirely dependent on individual activism and community members.

5.3. Theme 2: Structural and epistemic barriers to participation

Where Theme 1 showed that both communities recognize Wikipedia's value, Theme 2 explains why recognition has not translated into proportionate participation. Both communities identified overlapping but contextually inflected barriers to Wikipedia participation. Six barrier types were identified across the data: (i) infrastructural barriers (connectivity, device access); (ii) skills-based barriers (digital literacy, Wikipedia-specific procedural knowledge); (iii) linguistic-technical barriers (absence of digital writing conventions, orthographic instability); (iv) epistemic barriers (knowledge validation norms, citation requirements); (v) temporal barriers (lack of time, competing economic demands); and (vi) motivational barriers (perception of futility or invisibility). All twenty-four participants contributed to this theme, the only theme achieving full sample representation.

The most analytically consequential findings concern epistemic barriers; those rooted not in technical capacity but in Wikipedia's normative and epistemological architecture. Participants from both communities reported experiences of 'platform-induced hermeneutical injustice', the rendering of community-held knowledge illegible within Wikipedia's verification requirements.

"Wikipedia says: you need a source. But who wrote about our kgotla in a published book? The knowledge is in the community. The elders know it. But Wikipedia cannot accept the elder's word." P09, Society Cultural Leader, NW.

This phenomenon, the systematic exclusion of orally transmitted community-held knowledge from a platform formally positioned as a repository of "the sum of all human knowledge", represents a structurally embedded form of epistemic injustice that cannot be resolved through skills training and connectivity improvements alone. It is a governance and architecture problem, not a capacity problem.

Among Hà Nhì participants, epistemic barriers were compounded by the absence of a standardized written form of the language. The Hà Nhì writing system, based on a modified Latin script, was developed relatively recently and remains contested within the community, creating significant uncertainty about orthographic conventions. P19 (Sociolinguist) observed that this orthographic instability creates a situation where even those who wish to produce digital content cannot agree on how to write what they know, a form of structural precondition that must be addressed before platform-level barriers can become the primary focus.

5.4. Theme 3: Social equity implications

These barriers are not experienced as a neutral inconvenience; participants consistently read them as continuous with older, pre-digital patterns of marginalization, where participants across both communities articulated clear and analytically rich connections between language and broader social equity, though the character of these connections differed meaningfully across cases. In the Setswana context, the relative absence of the language from global digital knowledge systems was perceived as part

of a broader pattern of post-apartheid epistemic marginalization. The digital exclusion of Setswana was understood not as a technical deficit but as a structural continuation of colonial language hierarchies.

“In apartheid, they told us our language was not for education, not for science. Today, when our language is not on the internet, when it is not on Wikipedia, it continues the same message. It says this language is not important.” P14, Cultural Organization representative, Johannesburg.

In the Hà Nhì context, social equity concerns were more tightly focused on educational access and generational knowledge transmission. Multiple participants identified the absence of Hà Nhì language digital content as a concrete barrier to educational equity: children from Hà Nhì communities must navigate their formal education entirely through Vietnamese, without access to bilingual digital resources that might bridge the communicative gap. This finding resonates powerfully with SDG 4’s emphasis on culturally and linguistically inclusive education.

Both communities also raised concerns about the economic dimensions of the absence of digital language. Practitioners noted that community members working in tourism, cultural production, and heritage sectors identified the absence of authoritative online language documentation as a constraint on economic development opportunities. International research collaborators, NGOs, and cultural institutions routinely rely on digital sources; the absence of minority language content further marginalizes these communities within global information networks and funding landscapes.

5.5. Theme 4: Sustainability of digital language efforts

Even where these equity concerns are acknowledged and addressed in the short term, a further question remains: can the resulting gains be sustained? This final theme suggests not, in either case, without structural change. One of the most prominent findings across both cases relates to the perceived sustainability or, more accurately, the structural precariousness of community-driven digital language preservation initiatives. Participants consistently depicted a scenario defined by strong individual commitment, minimal institutional backing, structural vulnerabilities, and restricted pathways to scalability.

In the Setswana case, the platform metrics data corroborated participant testimony about precariousness. As shown in **Figure 1**, the Setswana Wikipedia’s article growth curve shows episodic rather than continuous improvement, periods of relative stagnation punctuated by short-term article creation drives. One participant, a Botswana-based digital humanities researcher, described this pattern as ‘sprint-dependency’: The edition’s growth is disproportionately reliant on short-term project-driven interventions rather than sustainable community practice. The active-editor figures in **Table 1** reinforce this reading that the contributor base is too thin to sustain organic growth between sprints, which is precisely the structural vulnerability participants like P05 described.

“We always have these big events. Two days, everyone edits. Fifty articles created. Then the event is over. People go back to their jobs. Six months later, we do

it again. But no community keeps growing in between.” P05, Digital Humanities Researcher, Polokwane.

For Hà Nhì, the sustainability challenge is more foundational, as without a dedicated Wikipedia edition and no written script, community-generated content is mostly in Vietnamese and other languages and is dispersed across YouTube, Facebook, personal blogs, and academic archives, with no centralized infrastructure for aggregation, quality control, and long-term preservation. This dispersal depicts the community’s digital language work as epistemically invisible at the global scale even when locally significant digital language activity occurs.

6. Discussion

6.1. Extending digital inclusion theory

The findings of this study extend digital inclusion theory in several significant directions. Most substantively, the results demonstrate that van Dijk’s [4,9] multi-level model of digital access requires supplementation with an epistemic dimension when applied to minority language knowledge production contexts. The challenges highlighted in this study, particularly the “platform-induced hermeneutical injustice” experienced by participants in both communities, do not fit effectively into the motivational, material, skills, or usage categories of van Dijk’s model [4,9]. The results constitute a distinct, fifth dimension of digital exclusion, like the structural incapacity of dominant platform architectures to accommodate non-hegemonic forms of knowing.

This theoretical expansion has implications beyond the Wikipedia context. It suggests that digital inclusion scholarship needs to move beyond its prevailing focus on access and skills to engage more rigorously with questions of epistemic architecture. These normative and procedural frameworks determine whose knowledge counts, how it must be validated, and for whom the platform was fundamentally designed. Fratini’s [32] digital sovereignty and control over digital infrastructures to safeguard autonomy in the digital realm anticipate elements of this argument, but the present findings provide a more specific, empirically grounded specification of the epistemic dimension than his framework articulates.

Second, the findings complicate the prevailing optimism of digital social innovation literature regarding Wikipedia’s transformative potential. Murray et al. [38] characterize digital social innovation as inherently emancipatory, driven by the democratizing affordances of networked digital platforms. The present findings reveal a more complex picture: while Wikipedia does create genuine, if constrained, opportunities for minority language communities, its structural architecture systematically disadvantages these communities through the very mechanisms that are supposed to ensure quality and neutrality. Platform design and governance reforms are prerequisites for meaningful digital language inclusion, not consequences of it.

6.2. Platform-induced hermeneutical injustice: A fifth dimension, operationalized

The concept introduced in Section 3.2 can now be specified more precisely against the findings. Platform-induced hermeneutical injustice differs from Fricker's original formulation in three respects relevant to future empirical work [26]. First, its locus is not an individual hearer's prejudice but a codified ruleset, Wikipedia's verifiability and notability policies, meaning the injustice persists even where individual editors act in good faith, as the data from both communities suggest they generally did. Second, its mechanism is exclusion-by-design rather than exclusion-by-omission: oral testimony is not merely under-represented on the platform; it is procedurally inadmissible as a primary source, regardless of how many community members could attest to it. Third, its distribution is uneven across communities in ways that track prior incorporation into literate, citation-producing institutions, which is why Setswana contributors, despite facing the same formal verifiability rule, experience a less totalizing version of the injustice than Hà Nhì contributors do. This specification suggests an operational path for future research: the concept could be measured comparatively via a "verifiability gap" index, the proportion of a community's claimed cultural or historical knowledge that exists in a form citable under platform policy, set against the proportion that exists only orally or communally. A wide gap would predict susceptibility to platform-induced hermeneutical injustice independent of a community's connectivity, literacy, or digital skill levels, separating this dimension empirically from the four access-based dimensions in van Dijk's model [4,9].

6.3. Epistemic injustice in digital knowledge systems

The most theoretically significant contribution of this study concerns its empirical grounding of epistemic injustice theory within digital platform contexts. Fricker's [26] theoretical framework was developed primarily with reference to interpersonal communicative contexts, but the present findings demonstrate that its core insights apply, with important modifications, to algorithmically mediated and institutionally governed digital knowledge platforms.

Wikipedia's citation and verifiability norms function as structural hermeneutical gaps that render oral, communal, and intergenerational knowledge systematically inexpressible within the platform's architecture. This is not a consequence of individual editorial bias, though such bias clearly exists, but of the platform's foundational epistemological commitments, which privilege text-based, institutionally validated, and citation-verifiable forms of knowledge. The structural analogy with Fricker's [26] concept of hermeneutical injustice is precise: the platform lacks the conceptual and procedural resources to render certain forms of knowledge legible, not because those forms of knowledge are epistemically inferior, but because the platform was designed with a particular epistemic tradition as its norm.

The finding that Setswana participants experienced Wikipedia as a site of epistemic legitimation (see **Table 3**), a space in which the language's intellectual capacity could be asserted, also resonates with Fricker's concept of testimonial justice [26]. Participants were not merely consuming information; they were engaging in what

the data suggest we might term ‘epistemic activism’: the deliberate use of the platform to reclaim and assert epistemic credibility against a global cultural backdrop that routinely discounts the knowledge of African language communities. This dimension of Wikipedia participation, its role in processes of epistemic self-determination, has received insufficient attention in existing literature. Epistemic activism, in this sense, differs from conventional accounts of Wikipedia contribution as either altruistic knowledge-sharing or career-building credentialism. It is better understood as a form of what might be called recognitional labor: editing not primarily to inform an external reader, but to perform, and thereby make durable, a claim about the editor’s own language and people that the offline information environment routinely denies them. This reframes Wikipedia participation, for communities in this position, as continuous with a longer repertoire of decolonial knowledge-assertion strategies rather than as novel digital-era behavior. It also generates a testable expectation for future work: epistemic activism should be most visible among editors from communities with a documented history of externally imposed epistemic delegitimizing, and comparatively muted among editors from minority-but-not-historically-delegitimized communities, a distinction the Setswana/Hà Nhì contrast in this study cannot itself resolve, since the historical trajectories of delegitimizing differ in kind, not just degree, between the two cases.

Table 3. Thematic coding summary.

Theme	Sub-themes	Representative codes	Frequency (N = 24)	% Participants
Perceptions of Wikipedia	Epistemic legitimation; Aspirational distance; Symbolic value	platform as proof of language vitality; awareness without access; Wikipedia as cultural archive	22	91.7 %
Barriers to Participation	Epistemic architecture; Skills deficits; Connectivity; Orthographic instability	oral knowledge exclusion; citation norms as gatekeeping; device access; no written standard	24	100 %
Social Equity Implications	Post-colonial epistemic marginalization; Educational inequity; Economic visibility	digital absence = continued erasure; medium-of-instruction mismatch; development invisibility	21	87.5 %
Sustainability of Efforts	Institutional precarity; Sprint-dependency; Individual burden	no long-term funding; episodic growth; one or two committed individuals carry the effort	23	95.8 %

Note: Frequency refers to number of participants whose data contributed codes to each theme. % Participants = percentage of total sample (N = 24). Themes derived from reflexive thematic analysis [33].

6.4. Challenging platform neutrality

This study challenges the notion of Wikipedia’s neutrality. The Wikimedia Foundation’s Neutral Point of View policy positions the platform as a space of impartial knowledge aggregation, free from ideological and political bias. The present findings reveal, however, that Neutral Point of View functions as an ideologically loaded epistemological commitment that systematically privileges Western, advanced, literate, and institutionally embedded knowledge traditions, a form of what Bourdieu [39] would recognize as symbolic violence, exercised through the naturalization of a particular epistemic habitus as the universal standard of legitimate knowledge.

The comparative case analysis (see **Table 1**) reveals that this challenge is not experienced uniformly. Setswana's official language status provides certain editorial community members with access to the institutional resources, academic publications, and governmental documentation necessary to satisfy Wikipedia's verification requirements, even if access to these resources is unequally distributed within the community. The Hà Nhi community, lacking both official language status and a substantial textual tradition, faces a more fundamental incompatibility between their knowledge traditions and Wikipedia's normative architecture. This asymmetry illustrates how structural features of the platform governance environment interact with contextual factors, specifically, the degree to which a language community has already been incorporated into dominant textual knowledge traditions, to produce differential outcomes.

6.5. Cultural sustainability and linguistic equity

The findings highlight important conflicts within the sustainable development framework as applied to linguistic diversity. While SDGs 4, 10, and 11 formally endorse cultural inclusivity and linguistic rights, the institutional mechanisms and digital infrastructure through which these goals might be pursued remain significantly underdeveloped [40]. The present study suggests that sustainable language preservation in digital environments requires the deployment of digital tools as well as the transformation of platform governance to ensure epistemic equity. This goal simultaneously involves the Wikimedia Foundation, governments, educational institutions, and civil society organizations.

Uddin's [41] concept of biocultural diversity enriches this analysis by locating linguistic sustainability within a broader ecology of cultural and biological vitality. From this perspective, the systematic exclusion of minority language communities from global digital knowledge platforms constitutes not merely a linguistic injustice but a form of biocultural impoverishment with implications for the long-term adaptive capacity of human social systems. This framing has significant implications for the integration of linguistic equity concerns into mainstream sustainability frameworks, specifically, the need to recognize digital knowledge access as a dimension of cultural sustainability no less important than physical heritage preservation or biodiversity conservation.

6.6. Generative AI and platform-induced hermeneutical injustice

The rise of large language models highlights a pressing issue with knowledge-sharing systems. Recent reports on small Wikipedia editions show that well-meaning editors, who do not speak minority languages, use AI translation tools like Google Translate to create articles [42]. This often results in large amounts of content that are grammatically incorrect or hard to understand. The languages most affected by this issue are those with fewer resources, as less available text means translation models can generate misleading output that seems plausible compared to the limited authentic content already present. For a language like Hà Nhi, which currently has no dedicated Wikipedia edition, there is an important concern. Rushing to create a Hà Nhi Wikipedia with machine translations could embed errors and inauthentic content

before the community can establish its own editorial standards. This could worsen the existing knowledge-sharing problem, as the inaccuracies would be created by machines rather than people. This shows that AI tools do not solve issues of knowledge justice; they can speed up and conceal these problems. Any response to this situation must link the quality control of AI-generated content with the proposed reforms to improve how information is verified.

The foregoing analysis highlighted that generative AI systems contribute to platform-induced hermeneutical injustice by shaping whose knowledge is recognized, represented, and rendered intelligible. Situating these findings within Digital Inclusion Theory provides a broader framework for understanding how such epistemic exclusions undermine meaningful digital participation and inclusion.

7. Policy and practical implications

The findings of this study generate several strong, evidence-based implications for multiple stakeholders. These implications are contextually grounded recommendations derived directly from the critical findings.

For policymakers, the most pressing implication is the effective integration of linguistic digital equity into national digital governance frameworks. South Africa's electronic communications and language policies have thus far operated largely in isolation from one another; their integration, and the creation of targeted funding mechanisms for minority and indigenous languages digital content production, would address several of the structural challenges identified in this study. The Department of Higher Education and Training's existing language development initiatives represent a potential institutional bridge for this integration. Vietnam's Ethnic Minority Policy framework [20] similarly requires a substantive digital language provision component, supported by infrastructure investment in the remote highland areas where Hà Nhi communities are concentrated and where internet penetration remains significantly below the national average.

For the Wikimedia Foundation and platform governance stakeholders, the findings support a substantive reformulation of verification and citability norms for minority language editions. The "no original research" and "verifiability" policies, while defensible in the context of well-documented majority languages, constitute epistemic barriers for communities whose knowledge traditions are primarily oral. Alternative verification mechanisms, community endorsement panels, collaborative oral-to-text documentation protocols, and institutional partnerships with indigenous knowledge archives required serious and urgent consideration. The Wikimedia Foundation should also expand programs such as WikiAfrica with sustained, multi-year funding models, explicitly designed to transcend the sprint-based interventions that currently characterize institutional support for minority language editions.

This study emphasizes the need for comprehensive digital literacy programs that not only teach Wikipedia-specific editing skills but also promote orthographic standardization, provide guidance on the approval process, help establish a workforce, and critically engage participants with the epistemic aspects of using the platform. Both communities involved highlighted the lack of such integrated and contextually relevant training as a significant gap. This is particularly important for educators and

language preservation practitioners. Partnerships between universities and communities can serve as a powerful institutional bridge to deliver this support, as demonstrated by their success in similar contexts, but this has not yet been systematically applied to either the Setswana and Hà Nhi cases.

The findings underscore the importance of treating digital knowledge production as an integral component of a linguistic revitalization program, not a complementary or aspirational feature, but a core equity concern with direct implications for SDG achievement. Funding bodies and development organizations operating in language communities with limited digital presence should include Wikipedia edition establishment and capacity-building as explicit programmatic objectives.

Funding bodies and development organizations occupy a distinct structural position in this ecosystem: unlike the Wikimedia Foundation, they can underwrite the multi-year, non-sprint-based editorial capacity that this study identifies as the central sustainability gap in both cases. Three mechanisms merit explicit consideration. First, dedicated multi-year grant lines rather than one-off project funding for minority-language Wikipedia editor cohorts, modeled on the kind of sustained institutional investment that has underwritten successful cases elsewhere (Section 6.7), would directly address the “sprint-dependency” pattern documented in Section 5.5. Second, South–South cooperation funding instruments are particularly well-suited to this problem: a joint Setswana–Hà Nhi capacity-building partnership, for instance, would allow the more developed Setswana editorial community to mentor the as-yet-nonexistent Hà Nhi one, while giving funders a single instrument addressing two otherwise disconnected cases.

8. Conclusion

This study has examined Wikipedia as a source of digital inclusion and linguistic sustainability for two indigenous language communities, Setswana in South Africa and Hà Nhi in Vietnam, through an integrated theoretical framework combining digital inclusion theory, epistemic justice theory, cultural sustainability frameworks, and digital social innovation literature. The findings indicate that, although Wikipedia provides authentic, albeit structurally limited, opportunities for the preservation of digital languages, its platform architecture, particularly its epistemological norms, editorial governance, and verification standards and processes, systematically disadvantages minority language communities whose knowledge traditions do not align with the platform’s literate and citation-based epistemological commitments.

Comparative case analysis reveals that, although these challenges (see **Table 1**) are influenced by the specific histories of colonial language policies and state recognition, they share a common epistemic foundation. The inherent limitation of dominant digital knowledge platforms to encompass non-hegemonic forms of knowing reflects a form of digitally mediated epistemic injustice. This insight extends Fricker’s theoretical framework into the realm of digital platform governance and carries substantial implications for theories concerning digital inclusion, cultural sustainability, and social equity in the Global South [26].

The study’s primary theoretical contribution lies in the identification of a fifth dimension of digital exclusion, epistemic architecture, that supplements van Dijk’s

foundational multi-level model and reorients digital inclusion scholarship toward questions of knowledge system design, platform governance, and epistemic equity [4,9]. Its empirical contribution lies in the generation of detailed, theoretically organized comparative case findings that illuminate both the contextual specificity and structural generality of digital language exclusion in Global South minority communities. Its practical contribution lies in the generation of evidence-based policy and design recommendations that are theoretically coherent and empirically grounded, offering a pathway toward genuinely inclusive and linguistically sustainable digital knowledge ecosystems.

Looking ahead, it is unlikely that the trajectory of this issue will remain unchanged. The increasing influence of generative AI in producing content in minority languages, the gradual emergence of alternative knowledge commons platforms outside the Wikimedia ecosystem, and evolving international funding priorities concerning linguistic and digital sustainability will all reshape the landscape outlined in this study. Whether these developments will narrow or widen the identified epistemic gap will depend less on technological capabilities and more on whether platform governance, funding structures, and community institution-building are approached as an integrated issue rather than as three separate challenges, a perspective that the findings of this study suggest is currently more of an exception than the norm.

9. Limitations and future research

This study has several limitations. First, the relatively small interview sample ($n = 24$), while appropriate for a qualitative-dominant approach and sufficient for thematic saturation as operationalized in the present study, constrains the scope of theoretical generalization beyond the focal cases. Future research employing larger, more geographically diverse samples across multiple Global South minority language communities would strengthen the evidential basis for the study's theoretical claims and test the cross-contextual robustness of the epistemic architecture framework.

Second, the absence of a dedicated Hà Nội Wikipedia edition creates a structural asymmetry in the comparative design, as the metrics-based analysis was necessarily conducted at different levels of analysis for each case. A longitudinal study that tracks the establishment and early development of a Hà Nội Wikipedia edition, should one be created because of advocacy inspired in part by research findings such as these, would provide a powerful experiment for testing the barriers identified in the present study.

Third, while the study acknowledges gender as a dimension of editorial exclusion in passing, a more dedicated investigation of gendered dimensions of Wikipedia participation in minority language communities is necessary. Minhas and Salawu's [3] findings regarding gender disparities in Wikipedia editing suggest that gender intersects with language and ethnicity in ways that the present study has not fully explored. Future research should explicitly incorporate gender as an analytical variable and engage with feminist digital scholarship on epistemic exclusion.

Fourth, the theoretical framework does not engage with recent, rapid developments in machine learning and automated content generation, including large

language models, as potential tools for creating minority-language Wikipedia content. This rapidly evolving technological landscape represents an important and urgent frontier for future research on digital language preservation, raising questions about the quality, authenticity, and community acceptance of algorithmically generated minority language content.

Longitudinal case studies tracking the development of newly established minority-language editions on Wikipedia; experimental studies of alternative verification mechanisms for minority-language content; and multi-stakeholder governance analyses of the Wikimedia Foundation's language diversity policies and their implementation could be future research directions. Also, examining the relationship between formal language recognition policies and digital language vitality across Global South contexts and comparative studies that extend the present framework to indigenous language communities in Latin America, Oceania, and South Asia could be researched.

Author contributions: Conceptualization, SM and TVA; methodology, SM; software, TVA; validation, SM, AS and TVA; formal analysis, SM; investigation, TVA; resources, AS; data curation, SM; writing—original draft preparation, SM; writing—review and editing, TVA; visualization, SM; supervision, AS; project administration, SM; funding acquisition, SM. All authors have read and agreed to the published version of the manuscript.

Funding: None.

Ethical approval: This study was reviewed and approved by RMIT University (approval number: 2025-29852-30655).

Informed consent statement: Informed consent was obtained from all participants involved in the study.

Acknowledgement: Grammarly is used to improve grammar, clarity, and tone of the article.

Conflict of interest: The authors declare no conflict of interest.

References

1. Minhas S, Salawu A. Preserving and promoting indigenous languages: Social media analysis of Punjabi and Setswana languages. *Journal of Asian and African Studies*. 2025; 60(7): 4109–4126. doi: 10.1177/00219096241243061
2. Ashrafimoghari V. Detecting cross-lingual information gaps in Wikipedia. *Companion Proceedings of the ACM Web Conference 2023*. 2023. pp. 581–585. doi: 10.1145/3543873.3587539
3. Minhas S, Salawu A. Wikipedia and indigenous language preservation: Analysis of Setswana and Punjabi languages. *Frontiers in Communication*. 2025; 29(10): 1442935. doi: 10.3389/fcomm.2025.1442935
4. Van Dijk J. *Power and technology: A theory of social, technical and natural power*. John Wiley & Sons; 2024.
5. Edmondson JA, Gregerson KJ. The languages of Vietnam: Mosaics and expansions. *Language and Linguistics Compass*. 2007; 1(6): 727–749. doi: 10.1111/j.1749-818X.2007.00033.x
6. Norris P. *Digital divide: Civic engagement, information poverty, and the internet worldwide*. Cambridge University Press; 2001. doi: 10.1017/CBO9781139164887
7. Warschauer M. *Technology and social inclusion: Rethinking the digital divide*. MIT Press; 2004. doi: 10.7551/mitpress/6699.001.0001

8. Selwyn N. Reconsidering political and popular understandings of the digital divide. *New Media & Society*. 2004; 6(3): 341–62. doi: 10.1177/1461444804042519
9. Van Dijk JA. *The deepening divide: Inequality in the information society*. Sage Publications; 2005. doi: 10.4135/9781452229812
10. Raihan MM, Subroto S, Chowdhury N, et al. Dimensions and barriers for digital (in) equity and digital divide: A systematic integrative review. *Digital Transformation and Society*. 2025; 4(2): 111–127. doi: 10.1108/DTS-04-2024-0054
11. Mair C. World languages in competition: the digital factor. *Linguaculture*. 2025; 16: 13–27. doi: 10.47743/linu-2025-si-0435
12. Canagarajah S. *Translingual Practice: Global Englishes and Cosmopolitan Relations*. Routledge; 2012. doi: 10.4324/9780203073889
13. Yin Y, Thebault-Spieker J. Beyond access: Contextualizing the benefits of broadband through contributor dynamics on Wikipedia. In: *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*; 2026. pp. 1–2. doi: 10.1145/3772318.3790903
14. Ferran-Ferrer N, Centelles M, Fernández L. Wikipedia’s front page ten years evolution: Gender gap and implicit bias in biographies and editorial policies. *Online Journal of Communication and Media Technologies*. 2025; 15(4): e202535. doi: 10.30935/ojcm/17483
15. Rollin G, Lages J. The most influential philosophers in Wikipedia: A multicultural analysis. *EPJ Data Science*. 2025; 14(1): 69. doi: 10.1140/epjds/s13688-025-00570-w
16. Nguyen N, Takeda H, Karunathilake L. Empowering weak languages through cross-language hyperlink recommendation. *Information*. 2025; 16(9): 749. doi: 10.3390/info16090749
17. Ford H. *Writing the revolution: Wikipedia and the survival of facts in the digital age*. MIT Press; 2022.
18. Guillory J. *Cultural capital: the problem of literary canon formation*. University of Chicago Press; 2023.
19. Sälevä J, Lignos C. Mining Wikidata for name resources for African languages. *ArXiv Preprint ArXiv:2104.00558*. 2021. doi: 10.48550/arXiv.2104.00558
20. Do H. Language ideology and its educational impact: insights from Vietnamese community language schools. In: *Vietnamese Language, Education and Change in and Outside Vietnam*. Springer Nature Singapore; 2024. pp. 83–105. doi: 10.1007/978-981-99-9093-1_5
21. Patrick E. Language revitalization and new technologies: Cultures of electronic mediation and the refiguring of communities. *Annual Review Anthropology*. 2004; 33: 21–45. doi: 10.1146/annurev.anthro.33.070203.143900
22. Valijärvi RL, Kahn L. The role of new media in minority- and endangered-language communities. In: *Endangered Languages in the 21st Century*. Routledge; 2023. pp. 139–157. doi: 10.4324/9781003260288-12
23. Mühlhäusler P. *Language of environment, environment of language: A course in ecolinguistics*. London, UK: Battlebridge; 2003.
24. Cunliffe D, Herring SC. Introduction to minority languages, multimedia and the web. *New Review of Hypermedia and Multimedia*. 2005; 11(2): 131–137. doi: 10.1080/13614560512331392186
25. UNESCO. *Language vitality and endangerment: Document submitted to the International Expert Meeting on UNESCO Programme Safeguarding of Endangered Languages*. UNESCO; 2003.
26. Fricker M. *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press; 2007. doi: 10.1093/acprof:oso/9780198237907.001.0001
27. Birhane A. Algorithmic injustice: A relational ethics approach. *Patterns*. 2021; 12; 2(2). doi: 10.1016/j.patter.2021.100205
28. Mignolo WD. Epistemic disobedience, independent thought and decolonial freedom. *Theory, Culture & Society*. 2009; 26(7–8): 159–181. doi: 10.1177/0263276409349275
29. Benkler Y. *The wealth of networks: How social production transforms markets and freedom*. Yale University Press; 2026.
30. Spyridou P. Normative commitments and platform logics: Understanding journalism’s adaptive resilience through coverage of democratic innovations. *Media and Communication*. 2026; 14: 1–19. doi: 10.17645/mac.11560
31. Graham M, Ojanperä S, Dittus M. Internet geographies: Data shadows and digital divisions of labor. In: Graham M, Dutton WH, eds. *Society and the Internet: How Networks of Information and Communication Are Changing Our Lives*. Oxford University Press; 2019. doi: 10.1093/oso/9780198843498.003.0004
32. Fratini S. The sociotechnical politics of digital sovereignty: Frictional infrastructures and the alignment of privacy and geopolitics. *Big Data & Society*. 2025; 12(4): 20539517251400729. doi: 10.1177/20539517251400729

33. Creswell JW, Clark VL. Designing and conducting mixed methods research. Sage Publications; 2017.
34. Biddix JP, Bourke B. Using formative assessment approaches to enhance rigor and quality in research. *Quality & Quantity*. 2026; 60(2): 5387–5406. doi: 10.1007/s11135-025-02461-8
35. Mesthrie R. Language in South Africa. Cambridge: Cambridge University Press; 2002. doi: 10.1017/CBO9780511486692
36. Patton MQ. Qualitative research & evaluation methods: Integrating theory and practice. Sage Publications; 2014.
37. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology*. 2006; 3(2): 77–101. doi: 10.1191/1478088706qp063oa
38. Murray R, Caulier-Grice J, Mulgan G. The open book of social innovation. London: Nesta; 2010.
39. Bourdieu P. Theory of symbolic power. *Culture/power/history: A reader in contemporary social theory*. Liverpool University Press, 1994;155.
40. United Nations. Transforming our world: The 2030 agenda for sustainable development (Resolution A/RES/70/1). United Nations; 2015.
41. Uddin MA. Linguistic, cultural, and biological diversity: Patterns, processes, and prospects. *Canadian Journal of Educational and Social Studies*. 2025; 5(5): 190–210. doi: 10.53103/cjess.v5i5.406
42. Judah J. How AI and Wikipedia have sent vulnerable languages into a doom spiral. *MIT Technology Review*. 2025.