WikiGap: Promoting Epistemic Equity by Surfacing Knowledge Gaps Between English Wikipedia and other Language Editions

ZINING WANG, University of British Columbia, Canada and Vector Institute for AI, Canada

YUXUAN ZHANG, University of British Columbia, Canada

DONGWOOK YOON, University of British Columbia, Canada

NICHOLAS VINCENT, Simon Fraser University, Canada

FARHAN SAMIR, University of British Columbia, Canada and University of Toronto, Canada

VERED SHWARTZ, University of British Columbia, Canada and Vector Institute for AI, Canada
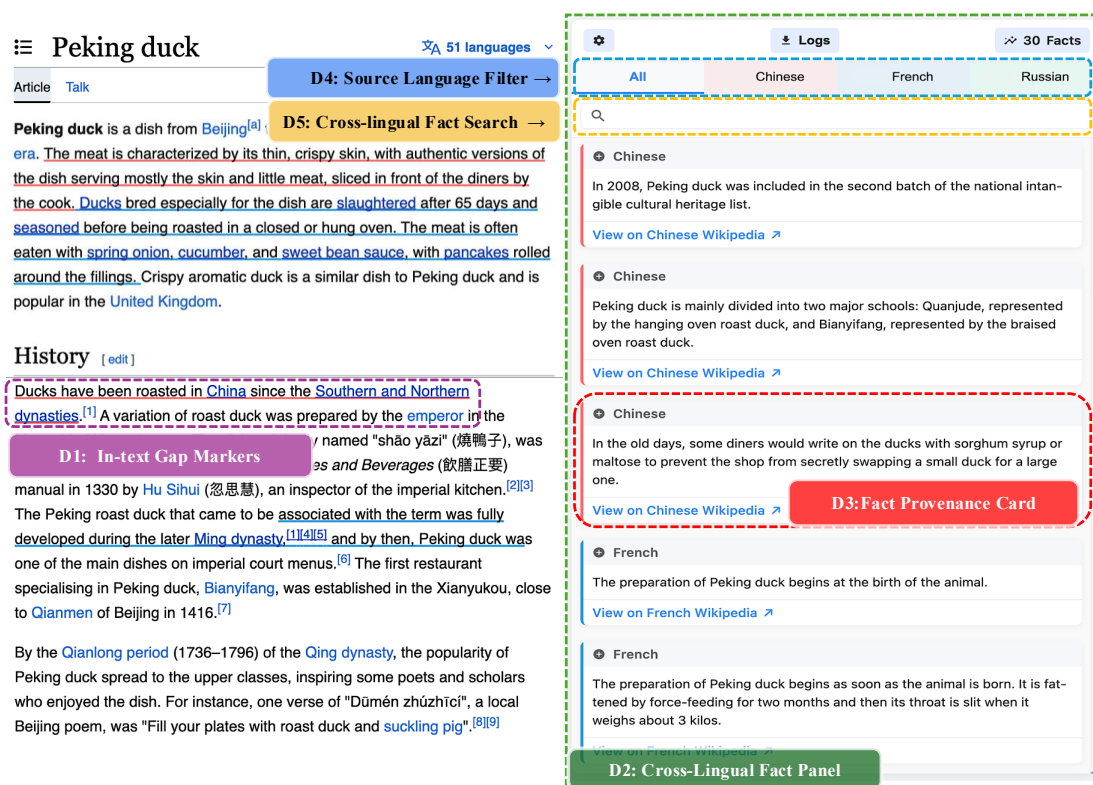
Fig. 1. The WIKIGAP interface embeds cross-lingual facts into English Wikipedia via five key design elements (D1–D5), supporting in-place access, traceability, and multilingual engagement.

With more than 11 times as many pageviews as the next largest edition, English Wikipedia dominates global knowledge access relative to other language editions. Readers are prone to assuming English Wikipedia as a superset of all language editions, leading many to prefer it even when their primary language is not English. Other language editions, however, comprise complementary facts rooted in their respective cultures and media environments, which are marginalized in English Wikipedia. While Wikipedia's user interface enables switching between language editions through its Interlanguage Link (ILL) system, it does not reveal to readers that other language editions contain valuable, complementary information. We present WikiGap, a system that surfaces complementary facts sourced from other Wikipedias within the English Wikipedia interface. Specifically, by combining a recent multilingual information-gap discovery method [52] with a user-centered design, WikiGap enables access to complementary information from French, Russian, and Chinese Wikipedia. In a mixed-methods study (n=21), WikiGap significantly improved fact-finding accuracy, reduced task time, and received a 32-point higher usability score relative to Wikipedia's current ILL-based navigation system. Participants reported increased awareness of the availability of complementary information in non-English editions and reconsidered the completeness of English Wikipedia. WikiGap thus paves the way for promoting epistemic equity across language editions. Taken together, these findings show how interface design can redistribute reader attention and engagement across Wikipedia's multilingual knowledge production communities, supporting epistemic equity.

1 Introduction

Wikipedia hosts more than 300 language editions, inviting contributions from many distinct linguistic and geographic regions. The investment in creating multiple language editions reflects Wikipedia's principles of democratizing both access to and contribution of information. These language editions are sociogeographically situated; it is generally well understood in Wikipedia scholarship that they are not mere translations of one another [26]. Instead, the language editions tend towards reflecting the knowledge that is contextually relevant, often based on cultural experience and localized media diets [4, 34].

However, this heterogeneity is not well understood by the broader public that engage with Wikipedia, whether readers or contributors [43]. Rather, the popular understanding of multilingual Wikipedia is that there is a global consensus on what information is reliable and noteworthy [26]. Moreover, English Wikipedia is by and large perceived as the most comprehensive and complete account of this consensus, containing as much as or more information than any other, making other languages the minority. This *English-as-superset* assumption is widespread, even among people who speak multiple languages proficiently; in a 2011 report, 93% of all editors surveyed attested to reading English Wikipedia, even though it was a primary language for only 52% of them [34, 60]. In the same vein, 20% of English Wikipedia's visitors were based in India, even more than readers from the United Kingdom. English Wikipedia is so widely understood to be an epistemic authority that is commonly treated as a source of neutral and complete information in and of itself by editors contributing to minority language editions [28].

This epistemic assumption is consequential not only for readers' understanding of Wikipedia but also for the organization of its peer production communities. Wikipedia is sustained by interconnected yet unevenly resourced

volunteer communities, each centered on different language editions [54]. Prior research has shown that volunteer engagement in Wikipedia depends on motivational factors such as enjoyment, opportunities for learning, and perceived recognition within a contributor community, beyond shared ideological commitments to free and open knowledge [3, 46]. When English Wikipedia is widely perceived as the most complete and authoritative source, reader attention and contributor effort become disproportionately concentrated in the English-language community. This concentration structurally amplifies the conditions that prior work has shown to support sustained participation, while making those conditions less accessible in minority language editions. One byproduct of this imbalance is a reduction in incentives for volunteer contributions to those communities [37], undermining Wikipedia's ethos as a globally and cooperatively produced multilingual knowledge infrastructure. In this work, we ask how interface design can redistribute attention and engagement across Wikipedia's multilingual peer production communities, promoting epistemic equity. [1]

Prior works have sought to incorporate contributions in minority language editions by consolidating information from all editions into a single unified knowledge base [e.g., 1, 14]. However, this conceptual formulation fails to *center* minority language editions and the unique volunteer contributions they contain. It continues to perpetuate the ideology of there being a global consensus on noteworthy and reliable information, a universal knowledge base. As such, it does little to challenge the status quo of English Wikipedia as an objective and complete epistemic authority.

In our work we take a social constructionist view of knowledge [22] in approaching the problem by improving engagement with minority language editions. That is, we center necessarily incomplete and sociogeographically positioned curation of knowledge. To highlight the positionality (rather than objectivity) of English Wikipedia, we aim to highlight its content gaps relative to other language editions. At the same time, we balance this intervention against being minimally disruptive to the current Wikipedia reading design [20]. We present WIKIGAP, an augmentation of the Wikipedia reading experience that surfaces sentence-level factual differences from English to 3 other language editions of Wikipedia (French, Russian, and Chinese). The system combines a computational backend (INFOGAP) that identifies multilingual content gaps [52] with a browser-based interface that presents these differences through in-page highlights and a sidebar display (Figure 1).

We evaluated WIKIGAP through a mixed-methods user study using a fact-finding task focused on culturally rich food articles – a domain where cross-lingual disparities are particularly salient [35]. Our results show that participants strongly preferred WIKIGAP over Wikipedia's default interlanguage links system (ILLs), with System Usability Scale (SUS) scores 32 points higher on average. In terms of performance, participants were significantly more accurate and faster at answering factual questions when using WIKIGAP, suggesting that surfacing multilingual information directly within the English interface not only improves user satisfaction but also enhances learning efficiency.

Through a Post-Study interview, we find that participants, prior to using WIKIGAP, were largely unaware of content gaps in English Wikipedia, demonstrating that the problematic *English-as-superset* epistemic assumption persists more than a decade after it was identified in seminal studies by Hecht [26]. Participants were overwhelmingly appreciative of WIKIGAP's capacity to surface these substantive information gaps during the reading of English Wikipedia pages, and expressed openness and enthusiasm about interacting with multilingual content in English Wikipedia. Moreover, a few participants with prior editing experience described how the system could support their editorial workflows by enabling source verification and facilitating the integration of multilingual content. These findings underscore WIKIGAP's broader value as a bridge between communities of editors and readers across languages.

---

[1]Throughout this paper, we use epistemic equity as a design lens, focusing specifically on how interface design can redistribute attention and engagement across Wikipedia's multilingual peer production communities, rather than addressing broader institutional or political dimensions of epistemic equity.

In Section 6, we reflect on how WikiGap challenges the English-as-superset assumption by redistributing reader attention and engagement across Wikipedia's multilingual peer production communities. We further interpret the system as a boundary object [56, 57] that supports articulation work between readers and editors across language editions. We contribute a validated system design for surfacing fact-level cross-lingual information gaps in Wikipedia. The system combines a computational pipeline that detects and integrates missing facts from non-English editions with an unobtrusive interface for English readers. A mixed-methods evaluation (n=21) shows that this design improves fact-finding accuracy and efficiency, outperforms Wikipedia's interlanguage-link system in usability, and increases engagement with multilingual knowledge—challenging assumptions about English Wikipedia's epistemic completeness.[2]

2 Background and Related Work

Wikipedia's multilingual architecture aspires to democratize knowledge access across languages, yet its design and community practices systematically privilege English content. In §2.1, we examine how this "English-as-superset" assumption [26] persists despite empirical evidence to the contrary, and how current Wikipedia interface reinforces epistemic hierarchies that undermine the platform's multilingual vision by concentrating attention and participation within a single language edition. In §2.2, we explain that cross-linguistic variation should be surfaced rather than obscured by drawing literature from prior CSCW research on asset-based design [23, 33, 63] and feminist epistemology [7, 22]. Then, we review prior technical interventions designed to surface cross-linguistic knowledge differences, evaluating them along two critical dimensions: the granularity of differences they expose and their integration with readers' existing browsing practices to support engagement across language communities. This analysis reveals persistent limitations that motivate our WikiGap design.

2.1 English-as-Superset and Epistemic Inequity in Multilingual Wikipedia

Wikipedia embodies a central tension in collaborative knowledge systems: although it is architecturally multilingual and aspires to provide knowledge to "every single person on the planet in their own language" [11], its design and community practices systematically privilege English content [16, 51]. This imbalance shapes how knowledge is organized and encountered on the platform, influencing both which perspectives become visible and how reader attention is distributed across language-specific peer production communities. At the interface level, this dynamic obscures cross-linguistic differences and reinforces what Hecht [26] term the "English-as-superset" assumption—the belief that English Wikipedia exhaustively represents the knowledge available across other language editions.

CSCW scholarship on epistemic equity [2, 8] helps explain why this assumption matters: collaborative systems should support the visibility and engagement of diverse epistemic standpoints rather than subsuming them under a dominant perspective. In Wikipedia's multilingual context, this means recognizing that language communities produce knowledge shaped by distinct cultural, geographic, and historical conditions that are not uniformly visible across language editions

2.1.1 English Wikipedia Is Treated as the Default, Yet Knowledge Gaps Persist. Given English's status as a global lingua franca, the English edition of Wikipedia has far outpaced other language editions in size and coverage. As a result, it is often implicitly treated as a superset of their knowledge. Hecht [26] identified this "English-as-superset" viewpoint, which, though rarely stated explicitly, remains prevalent—even among researchers who rely on multilingual Wikipedia to train text-generation and text-embedding models. However, substantial knowledge gaps persist between English

---

[2]Code and data will be made publicly available upon publication.

and other language editions. Hecht and Gergle [25] showed that English Wikipedia lacks coverage on many topics documented in other languages, while more recent work demonstrates that even when topics overlap, English articles often omit facts present elsewhere [52].

2.1.2 *From Knowledge Gaps to Epistemic Invisibility.*

From an epistemic equity perspective, such variation is not an error or deficiency, but an expected outcome of situated knowledge production. One interpretation of Wikipedia's multilingual vision assumes a language-agnostic set of facts that can be uniformly translated across languages—a view that treats language as a neutral vessel for meaning, abstracted from socio-geographic context [18, Chapter 4]. Yet much of Wikipedia's content concerns culturally and geographically situated histories, practices, and identities [61]. Accordingly, language editions are shaped by the communities that produce them [30] and exhibit a well-documented "self-focus bias" toward regionally relevant topics [9, 24, 52]. This emphasis on positionality echoes feminist epistemological arguments that knowledge is always produced from particular social, cultural, and linguistic standpoints [22]. When such epistemic differences are rendered invisible, they not only obscure situated knowledge but also limit opportunities for readers and editors to encounter and engage with the communities that produce it.

The "English-as-superset" assumption can thus contribute to cycles of epistemic invisibility [2], in which visibility and legitimacy are recursively reinforced within collaborative systems. On Wikipedia, these dynamics manifest in both editorial and readership practices. Bilingual editors often consult English Wikipedia to determine what content belongs in other language editions, treating it as a reference point [28]. Wikipedia's emphasis on verifiability further privileges English-language sources, creating structural barriers for knowledge rooted in other linguistic or cultural contexts [54]. Although multilingual editors can bridge across editions, they constitute only a small fraction of Wikipedia's editor base [21].

Readers similarly assume that different language editions contain equivalent information (Section 5). As engagement with smaller editions declines, editors in those communities receive less visibility and recognition, weakening key motivations for participation [3, 46]. This feedback loop reduces incentives for contribution and undermines Wikipedia's vision as a collaboratively produced multilingual knowledge infrastructure [37].

2.1.3 *Interface Design and the Obscuring of Epistemic Difference.*

These epistemic hierarchies are not only reflected in community norms and editorial practices, but are also enacted through interface design. Wikipedia primarily supports cross-lingual navigation through Interlanguage Links (ILLs), presented as a drop-down menu on each article. While ILLs enable access to other language editions, the interface provides no cues indicating that articles may differ substantially in content. By framing language switching as a matter of translation rather than epistemic difference, the design implicitly suggests a shared underlying knowledge base, limiting opportunities for readers to engage with the work of other language communities. Yet prior work shows that many unique facts remain distributed across non-local language editions, even for regionally associated topics [52]. As a result, Wikipedia's current interface renders epistemic differences structurally invisible to readers, making it difficult to engage with other language editions.

Together, this body of work motivates a focus on design interventions that challenge the "English-as-superset" assumption by making cross-linguistic differences visible and supporting engagement across Wikipedia's multilingual peer production communities.

2.2 Prior Work Targeting Cross-Lingual Knowledge Dissemination in Wikipedia

2.2.1 *Surfacing vs. Consolidating Cross-Linguistic Differences.*

Prior work on cross-linguistic variation in Wikipedia has generally followed two contrasting orientations. One line of work treats differences across language editions primarily

as obstacles to information access, motivating efforts to consolidate multilingual knowledge into a single, unified representation [1, 14]. While such approaches aim to improve coverage and navigability, they implicitly assume that cross-linguistic variation should be minimized, and risk obscuring the distinct perspectives embedded in distributed knowledge production by enforcing uniform representations [53].

In contrast, a second orientation emphasizes surfacing cross-linguistic differences as meaningful expressions of positionality [22]. From this perspective, variation across language editions is not a shortcoming to be resolved, but an expected and valuable outcome of situated knowledge production. This framing aligns closely with our earlier discussion of epistemic equity, which calls for collaborative systems to recognize and sustain diverse ways of knowing rather than subsume them under a dominant perspective.

When translated into design terms, this orientation corresponds to an asset-based approach [23, 33, 63]. Rather than treating heterogeneity as a problem to be corrected, asset-based design views existing differences as resources for sense-making and learning. Applied to multilingual Wikipedia, this suggests that interfaces should help readers perceive and engage with cross-linguistic differences, instead of rendering them invisible through consolidation or assumed equivalence. Building on this perspective, our work extends prior efforts by exploring how interface design can surface positionality within everyday Wikipedia browsing practices.

2.2.2 Prior Multilingual Interface Designs on Surfacing Cross-lingual Knowledge Differences. Several systems have attempted to move beyond Wikipedia's default ILLs by providing interfaces that surface content from multiple language editions. Manypedia [39] displays two language editions side-by-side using machine translation, but it does not explicitly highlight factual inconsistencies or information gaps between them. WikiCompare [50] takes a step further by identifying topic-level differences. When a user reads an article in one language (e.g., Hindi), the interface highlights subtopics that are discussed in another language edition (e.g., English) but missing in the current version; clicking a highlight redirects the user to the corresponding English section. While informative, WikiCompare still operates at the topic level, rather than identifying specific factual discrepancies across languages. Moreover, neither Manypedia nor WikiCompare formally evaluates how such cross-lingual surfacing affects readers' comprehension, behavior, or information-gathering efficacy.

The most ambitious multilingual interface to date is Omnipedia [6], which visualizes how different language communities conceptualize a topic by aggregating anchor-link structures across editions. This design allows simultaneous comparison of multiple languages but comes at the cost of requiring users to adopt an entirely new interaction paradigm. Users must access Omnipedia through its dedicated search interface, which contrasts with how most people encounter Wikipedia—via search engines and in-page browsing [59]. As a result, Omnipedia's workflow is difficult to integrate unobtrusively into readers' existing habits on Wikipedia.

Complementing such reader-centered systems, the WIKIPEDIA DIVERSITY OBSERVATORY (WDO) [44] offers editor-facing dashboards that quantify cultural and geographic content gaps across language editions and suggest missing but culturally relevant articles. However, WDO also operates outside Wikipedia's native reading and editing interfaces and focuses only on article-level differences—an even coarser granularity than tools such as WikiCompare.

In summary, prior systems demonstrate the value of surfacing cross-lingual heterogeneity, yet they share several infrastructural limitations that restrict their practical impact: (1) they primarily expose topic-level rather than sentence-level differences; (2) they rely on workflows external to the natural reading experience, limiting their ability to integrate multilingual knowledge unobtrusively, As Grudin's classic analysis of groupware failures highlights [20], systems that impose new routines or additional work misaligned with existing practices frequently struggle to achieve sustained

adoption; and (3) with the exception of Omnipedia, they lack empirical evaluation of how such designs actually affect readers.

2.2.3 Advances in multilingual factual knowledge retrieval. One of the challenges faced by prior systems that work at the article or topic-level was lack of access to computational tools that can recognize finer-grained differences between articles. Recent advances in multilingual NLP enable automatic detection of sentence-level factual gaps, which we leverage for developing our user-centric WIKIGAP design. Specifically, we rely on INFOGAP [52, Sec. 3.2]. The INFOGAP paper corroborated through a large-scale analysis that English Wikipedia lacks facts from other language editions, including facts that may have a broader appeal. For example, Apple CEO Tim Cook's association with the Russia-Ukraine war is only mentioned on Russian Wikipedia, but is conceivably noteworthy to the wider array of readers that engage with English Wikipedia. Building on INFOGAP, our WIKIGAP system embeds those unique facts within Wikipedia's native interface, overcoming limitations from the aforementioned interfaces.

3 WikiGap System

The goal of WIKIGAP is to challenge the English-as-superset assumption by making complementary knowledge from other language editions visible to readers during everyday Wikipedia use. Rather than treating multilingual differences as gaps to be corrected, WIKIGAP is designed to surface them in context, allowing readers to encounter cross-lingual knowledge while browsing English Wikipedia pages. We combine a computational pipeline that identifies sentence-level knowledge gaps across languages with a reader-facing interface that presents these differences through in-situ annotations and structured fact cards.

We begin with outlining the design requirements informed by user needs and foundational theories that shaped the WikiGap interface (§3.1). We then step behind the scenes to introduce INFOGAP [52], which we use to detect language-exclusive facts lacking in English Wikipedia (§3.2). Furthermore, we describe how we extended and adapted INFOGAP to detect and deliver language-exclusive facts, including how we filter, translate, and bind them to UI components at runtime (§3.3). We conclude this section by outlining the system implementation details (§3.4).

3.1 Designing WIKIGAP

We describe the exploratory interview we conducted to understand users' needs, from which we developed the design requirements (§3.1.1). Then, we describe the core UI elements that we derived from the requirements (§3.1.2).

3.1.1 Design Requirements Informed by User Needs and Theory. Guided by an asset-based perspective that shifts the focus from correcting gaps to making differences visible and interpretable, we combined findings from our exploratory interviews with established theories of reading and information seeking to define the design requirements for WIKIGAP. These requirements share an overarching goal: enabling readers to access cross-lingual knowledge while preserving the familiar experience of browsing English Wikipedia. The findings from the exploratory interview, requirements, and design elements are summarized in Figure 2 and detailed below.

Exploratory Interviews on Multilingual Wikipedia Use. We conducted exploratory interviews with four Computer Science graduate students (2 women, 2 men), all regular Wikipedia users; three spoke an additional language (Korean, Russian, or Hindi). Each 30-minute session followed an unstructured, conversational format guided by open-ended questions about their multilingual Wikipedia usage, perceptions of the Wikipedia ILL system, and ideas for improving cross-lingual content presentation. Participants reviewed an English Wikipedia article on *mooncake*–a topic unfamiliar

to them–and reflected on how multilingual facts could be surfaced in this context. They proposed interface features to better support multilingual reading habits. We took notes during the sessions and synthesized them using thematic coding to derive the following design requirements for WikiGap.
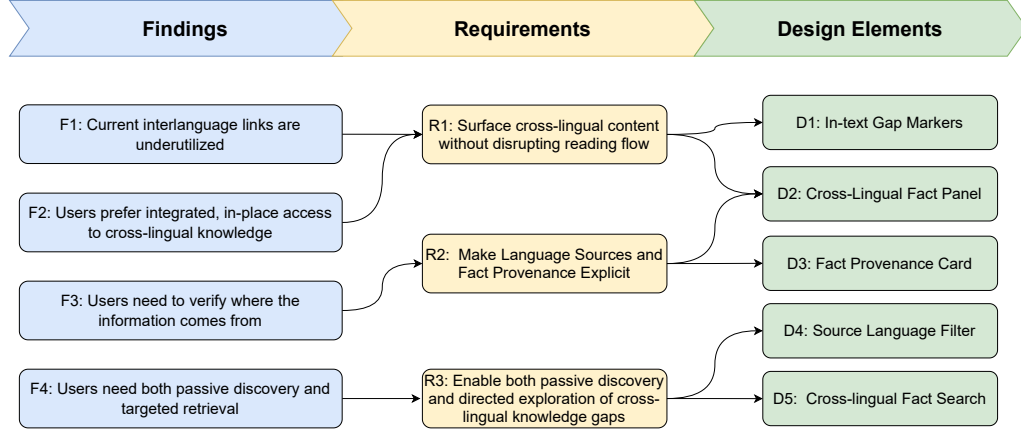

Fig. 2. How the findings from our preliminary interviews (F1-F4) informed WikiGap's design requirements (R1-R3) and core design elements (D1-D5).

R1: Surface cross-lingual content without disrupting reading flow. This requirement directly addresses participants' tendency to overlook multilingual content due to the high cost of switching language editions. Participants expressed a strong desire to access multilingual information without breaking their reading experience. They reported rarely using Wikipedia's ILLs (**F1**), noting the disorienting nature of full-page switches: *"Even if I click the other language, I don't know where to look for the thing I care about. It's like starting over."* Such page jumps impose high navigation costs [49], requiring users to reorient to a new layout and language, which discourages exploration.

Instead, users preferred lightweight, non-intrusive cues – similar to annotations in Grammarly or Google Docs – that subtly signal the presence of additional information (**F2**). This aligns with Anchored Annotation Theory [38], which finds that tying annotations to specific text locations improves comprehension and recall. Our design anchors cross-lingual facts directly to the English article using click-triggered underlines and a collapsible panel, allowing users to access additional facts only when interested. This approach follows the Spatial Contiguity Principle [29], which emphasizes placing related information close together. Collectively, these design choices preserve reading flow while enabling low-friction, in-place access to multilingual content.

R2: Ensure clear language attribution and traceability of multilingual facts to support trust and credibility. This requirement reflects readers' need to assess the provenance of multilingual facts, especially when such information is absent from the English article. In our interviews, a recurring concern among participants was the trustworthiness of information surfaced from other language editions–particularly when such facts were not present in the English version. Participants emphasized that they wanted to know *"where the information comes from,"* including which language edition it originated from and the ability to read it in its original context (**F3**), because it is crucial for accessing information's

credibility. This requirement highlights the importance of traceability and attribution in multilingual information access. Readers must be able to understand both the linguistic source and the original context of each surfaced fact. This entails more than just showing a translation, but rather also requires visual and interactive mechanisms that reinforce provenance and credibility.

To meet this need, we present each fact with an explicit, color-coded language label and allow readers to trace each fact back to its source by opening the corresponding language edition and viewing the sentence in context. These strategies support transparency and accountability, giving readers confidence in the surfaced content and encouraging deeper multilingual exploration.

R3: Support both passive discovery and active exploration of cross-lingual knowledge gaps (customizable exploration). This requirement reflects the need to support different levels of engagement, recognizing that readers vary in how much cross-lingual information they want to encounter. User feedback revealed a range of interaction preferences: users want flexible access modes–both passive discovery and targeted retrieval (**F4**). Different participants had different expectations for how cross-lingual facts should be surfaced. Some favored passive discovery, where the system automatically highlights interesting differences without requiring user input. Others preferred more active control, such as being able to search for specific content or filter information by language or topic. They discussed concern over potential information overload if too many facts were presented at once.

This requirement reflects that readers should be able to modulate how much cross-lingual content they see and which languages they want to include in their exploration. From a cognitive standpoint, this helps reduce information overload and supports a more goal-directed behavior. It necessitates features like search, filtering, and toggling between languages – functions that empower users without requiring them to dig through entire articles in other languages.

3.1.2 Core Design Elements. Based on our design process and the identified requirements (R1-R3), we developed five core design elements for WIKIGAP, as illustrated in Figure 1.

D1: In-text Gap Markers. This design addresses R1 (non-disruptive presentation) and R2 (traceability) to make multilingual differences visible at the moment of reading without interrupting readers' flow, we implemented a subtle underlining system that signals the presence of additional information from other language editions. Each marker is country-color coded and appears as a patterned underline within the English article. When clicked, the side panel retrieves the fact relevant to that sentence, where the extension identifies complementary or missing information from another language edition. The underline then becomes more prominent, providing interaction feedback and inviting further exploration. Each language is associated with a distinct color – red for Chinese, blue for French, and green for Russian. This color coding is applied across both in-text highlights and corresponding elements in the side panel, reinforcing source attribution as emphasized in R2. As we elaborate in Sec. 3.3.2, only significant knowledge gaps identified by INFOGAP are highlighted, rather than minor variations or differences in wording. This selective highlighting prevents information overload and ensures that users' attention is drawn to substantive cross-lingual differences.

D2: Cross-Lingual Fact Panel. To fulfill R1 and R2, we also developed a margin-anchored sidebar that displays the facts, translated to English and organized by their source language. When a user interacts with an in-text highlight, the corresponding fact card is revealed in the sidebar, which appears on the right side of the article. The sidebar is collapsible and can be pinned for extended browsing or hidden entirely to maintain a minimalist reading view. Each language section presents its facts in dedicated groups. Each fact is shown in English, accompanied by attribution

information, including the source language tag and link to the original article. The close spatial relationship between in-text cues and sidebar content supports R1's emphasis on minimizing reading disruption.

D3: Fact Provenance Card. At the heart of WIKIGAP's cross-lingual surfacing system is the fact card interface, which presents individual facts from other language editions in a compact, standardized format. Each fact card includes: (1) the target fact, translated into English, (2) a color-coded badge indicating the source language, and (3) a hyperlink labeled "View on [Language] Wikipedia" that opens the source article in a new tab with the original sentence highlighted. This design primarily addresses R2 (transparent attribution) by balancing at-a-glance comprehension with direct access to original context.

D4: Source Language Filter. To support R3 (customizable exploration), we introduced a language filter function located at the top of the sidebar. These filters allow users to specify which language(s) they want to view content from, enabling personalization based on their linguistic background and interests. The filters use both language names and colors to maximize clarity. This functionality directly supports both passive and active interaction modes (R3).

D5: Cross-lingual Fact Search. To further support R3, specifically the directed exploration function, we added a search box feature that allows users to perform keyword-based queries across multilingual facts. When a user enters a search term, relevant fact cards from the selected languages are retrieved and displayed in the sidebar. This feature caters to users with goal-oriented information-seeking tasks, such as verifying a specific claim or comparing facts across language editions. While D1 and D2 emphasize passive and contextual discovery, the search box fulfills a complementary role by enabling intentional, focused retrieval. Together, these elements provide a spectrum of access modes that reflect users' varying preferences for how and when to engage with multilingual content.

3.2   The INFOGAP Pipeline

While the interface makes multilingual differences visible to readers, INFOGAP provides the backend capability that enables such differences to be identified at sentence-level granularity. It was a previously developed computational pipeline introduced by Samir et al. [52], which detects factual misalignments between Wikipedia articles across languages. This subsection introduces the core components of the original INFOGAP system, which we use to identify sentence-level knowledge gaps between English and other language editions. In the following section, we will describe how we extended this pipeline to support additional languages and adapted its outputs for integration into the WIKIGAP interface.

INFOGAP is a state-of-the-art LLM-based pipeline that takes an article L_s in the *source language* and the respective article L_t in the *target language* and returns the sets of common facts and facts that are exclusive to one article. As illustrated in Figure 3, given a Wikipedia article in the source language (L_s, e.g., English) and the respective article in another language (L_t, e.g., French), INFOGAP identifies three categories of facts: those shared across editions, those unique to the source, and those unique to the target (i.e., language-exclusive facts). For example, "*Oolong is a semi-oxidized Chinese tea*" is shared across languages, while "*It is served in US restaurants*" may appear only in English.

The pipeline leverages a multilingual LLM and sentence embeddings to perform cross-lingual comparison in three stages, as illustrated in Figure 4:

(1) **Fact decomposition**. Each paragraph in L_s and L_t is decomposed into atomic factual statements using prompts issued to a multilingual LLM. The output includes both the fact and its paragraph index for downstream alignment.
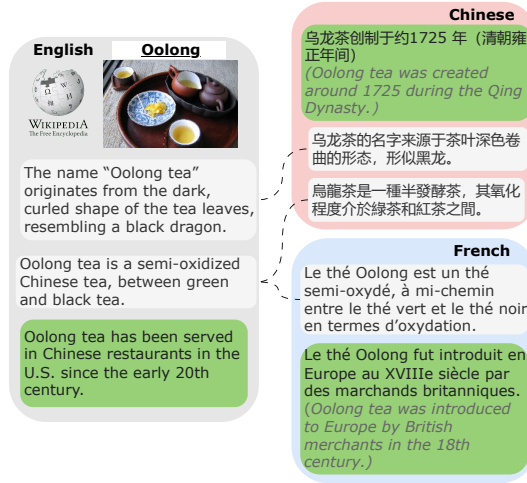
Fig. 3. Knowledge differences in the Wikipedia coverage of *Oolong* identified by INFOGAP. Connecting lines mark overlapping facts; green boxes highlight facts unique to each language edition. English sentences in *italics* represent translations of facts from Chinese or French Wikipedia, while non-italicized English sentences are from the original English Wikipedia article.

(2) **Multilingual alignment**. Each fact is encoded using a multilingual sentence encoder. For every fact in L_t, the top three nearest neighbors in L_s are retrieved based on cosine similarity. This reduces noise and narrows the search space when looking for equivalent facts.

(3) **Alignment verification**. For each fact in L_t, the LLM is prompted to determine whether it is inferable from any of its nearest neighbors in L_s. If a matching fact is found in L_s, these facts are considered aligned, otherwise, the fact is labeled as a *knowledge gap* in L_s.

3.3 Extending and Integrating INFOGAP for WIKIGAP

In this section, we outline the conceptual modifications and integration steps that enable INFOGAP to serve as the backend for WIKIGAP. We adapted and extended the existing INFOGAP pipeline [52] in two key ways: (1) by enabling new language (Chinese) support beyond the original study, and (2) by building an integration layer that transforms INFOGAP's output into interactive interface components usable within our system.

3.3.1 Rationale for Language Selection. In this study, we intentionally focused on English and three additional language editions, namely Russian, French, and Chinese. We chose the first two because the original INFOGAP pipeline was developed for detecting cross-lingual gaps in biographical articles between English and either Russian or French. We included them because of they have been validated in the research to ensure accuracy. To generalize the method to a new domain and an additional language, we incorporated support for Chinese into the fact decomposition and verification modules, as some co-authors were proficient in Chinese. This involved custom preprocessing for tokenization, crafting prompts, and conducting human evaluation on Chinese texts to confirm that INFOGAP performs comparably to its performance on the original languages studied by Samir et al. [52].
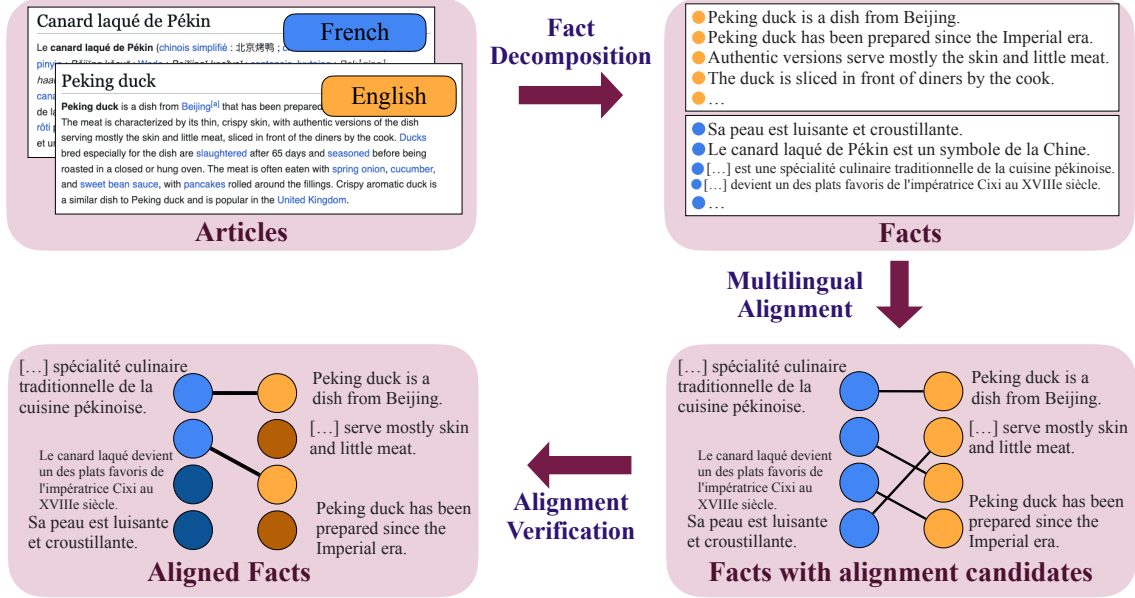
Fig. 4. Overview of the INFOGAP backend pipeline for cross-lingual fact alignment, reproduced from Samir et al. [52] and adapted. For additional technical details, see the original paper.

3.3.2 Language attribution and fact selection. Our focus is on *knowledge gaps* – facts that appear in the L_t version (e.g., French, Russian, or Chinese) of an article but *not* in L_s (English). We therefore filter the INFOGAP output to retain only gap facts. To ensure interface responsiveness, we use a precomputed output from INFOGAP for each article topic. For every topic, we prepare a unified dataset that includes exclusive facts from three target languages (French, Russian, and Chinese). These facts are grouped by language code (fr, zh, ru) to support language-specific interaction features such as language filtering (D4). To reduce cognitive overload, we limit the number of displayed facts to **10** per language, proportionally sampling based on section-level gap counts. If fewer than 10 gaps exist for a language, we display all available facts.

3.3.3 Translations and traceability. Because WIKIGAP displays cross-lingual content within the English Wikipedia page to readers who we do not assume read other languages, we translate the gap facts from their original language (L_t) into English for presentation in the fact cards. To support source traceability and deeper exploration, we create a direct *link-to-highlight* for each fact. An encoded version of the original sentence in L_t is appended to the target article's URL, sending readers to the exact sentence in context when they click on the "View on [L_t] Wikipedia" in the card (D3).

3.3.4 In-text highlights. For each gap fact, INFOGAP outputs the most semantically related English sentence from the source article, even when the fact itself is absent in English. We use these sentences as anchors for subtle in-text highlights (D1), which allow readers to preview cross-lingual differences inline without disrupting their reading flow. Figure 5 illustrates how such outputs map onto WIKIGAP's interface components. For example, the French Wikipedia article for *Peking Duck* mentions that the dish became a favorite of Empress Cixi in the 18th century – a detail missing from the English article. INFOGAP aligns this fact to the closest English sentence "The Peking roast duck that came to be

associated with the term was fully developed during the later Ming dynasty", which provides a semantically-relevant anchor for surfacing the French-exclusive fact in context.
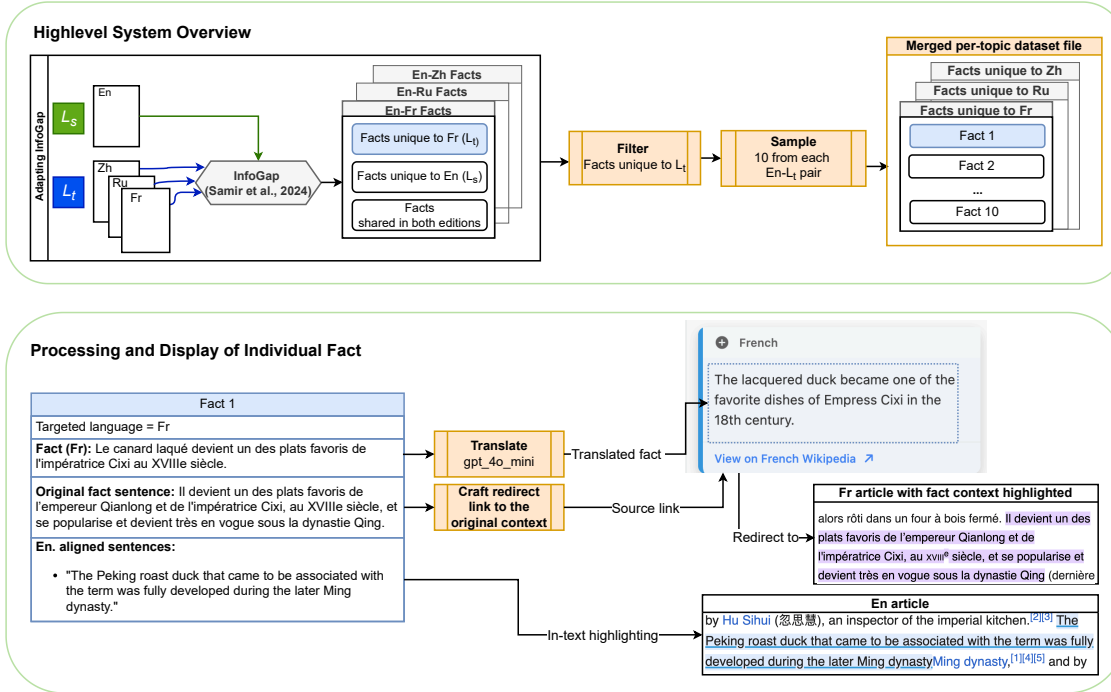

Fig. 5. **Overview of system implementation and data processing pipeline.** *Top:* A high-level overview of the data stream in the WIKIGAP system. We adapted the INFOGAP pipeline to support Chinese-language input alongside existing language pairs, followed by post-processing steps to standardize and merge datasets by topic. Orange process blocks indicate components we developed to enable proper integration and display of multilingual facts in the UI. *Bottom:* The data structure and rendering flow for an individual fact. This illustrates how each multilingual fact is transformed – through translation, alignment, tagging, and contextual linking – into an interactive component in the WIKIGAP interface.

3.4 System Implementation

We implemented WIKIGAP as a browser-based system that overlays multilingual content directly onto English Wikipedia articles. The system interface was developed as a Chrome extension using standard web technologies – HTML, CSS, and vanilla JavaScript – while the underlying data pipeline generates and serves precomputed JSON files derived from the INFOGAP framework. When running INFOGAP, we used GPT-4o for fact decomposition and fact alignment verification, and computed sentence embeddings using the multilingual LaBSE model [15, 48]. For each English Wikipedia topic, INFOGAP produces three separate datasets, each containing factual gaps between English and one of three target languages (fr, zh, ru). We then merge these files by topic into a unified, standardized JSON file named after the English article. This consolidation supports runtime efficiency: when a user visits an English topic page, the extension loads the corresponding file based on the topic title. Each file contains multilingual facts translated into English which are generated using GPT-4o-mini [47]. The overall system workflow is illustrated in Figure 5.

Table 1. Description of study metrics and corresponding data collection and analysis. All metrics were collected at the individual participant level and compared across WɪᴋɪGᴀᴘ and control conditions.

Metrics	Data Collected	Data Analysis
Performance		
Quiz Accuracy	Percentage of correct answers in each condition	Paired t-test comparing WɪᴋɪGᴀᴘ vs. control conditions
Completion Time	Time (in minutes) to complete each quiz session	
Preference		
Usability	System Usability Scale (SUS) score in each condition	Paired t-test comparing SUS scores between conditions
User Feedback	Open-ended responses and interview transcripts	Thematic coding to extract perceived usability, utility, and challenges

The extension interacts with the live English Wikipedia page, dynamically injecting highlights and sidebar content based on the current topic the user is browsing (the corresponding dataset is selected for rendering the data on the UI based on that topic). For each fact card, we generate external links that redirect readers to the original sentence in the source-language Wikipedia article. This is done by appending an encoded version of the original sentence to the article's base URL, allowing users to jump directly to the fact in context. Additionally, the search function is implemented by scanning both the fact body and language label for matches with the user's input and dynamically hiding non-matching entries.

4 Evaluation Methods

To examine whether surfacing cross-lingual knowledge gaps within the English Wikipedia interface can meaningfully support multilingual engagement, we conducted a mixed-methods user study combining quantitative performance measures with qualitative feedback. The study assesses usability, knowledge acquisition, and users' perceptions of the system and the cross-lingual content it surfaces. Table 1 summarizes the study metrics and analysis.[3]

4.1 Participants

Twenty-one people participated in the study. Of these, eight were student volunteers from the HCI course where this project originated, and the remaining thirteen participants were recruited through Upwork, and were compensated at a rate of 20 USD for the user study. Participants ranged in age from 18 to 44 and represented diverse cultural and ethnic backgrounds (see Table 2). All participants reported that they typically use the English version of Wikipedia. Twelve had previously contributed to Wikipedia to varying degrees, with two considering themselves frequent editors. Participants were required to meet the following inclusion criteria: (i) regular use of English Wikipedia, (ii) comfort with using a Chrome browser, and (iii) willingness to contribute user experience feedback.

The study followed a within-subject design, in which each participant was assigned two different topics out of the total five topics (Sec. 4.3). Each participant read one article using the WɪᴋɪGᴀᴘ extension and another article using the default Wikipedia interface (with ILLs) as a control. We selected interlanguage links (ILLs) as the control to reflect the default, in-context Wikipedia experience and to simulate an extension built on top of existing knowledge-seeking workflows; moreover, prior multilingual interface designs are not publicly available for deployment. The assignment of treatment and control conditions was randomized for each participant.

---

[3]The study protocol received approval from our institutional REB.

Table 2. Participant self-reported cultural background.

Cultural / Ethnic background	% of participants
East Asian	24%
White or European	19%
South Asian	14%
Hispanic or Latin American	14%
Black or African	10%
Middle Eastern or North Southeast Asian	10%
African	5%
Other	5%

4.2 Tasks

Participants completed an open-book reading quiz consisting of 10 multiple-choice questions in each condition. For each quiz, participants were provided with four Wikipedia articles on the same topic—specifically, the English article and its corresponding versions in French, Russian, and Chinese. They were instructed to rely solely on these four articles to answer the quiz questions and were not allowed to use external resources such as Google Search. Each participant completed the quiz twice: once using the standard Wikipedia interface (control condition) and once using the WikiGap extension (treatment condition). For every question, participants were also asked to indicate the language edition in which they found the answer.

Quiz Construction. We constructed one quiz per article topic using facts extracted by InfoGap and surfaced through the WikiGap interface. From the 30 multilingual facts identified for each article, we manually selected a subset of 10, ensuring a roughly equal number of facts from each of the three language editions—French (fr), Russian (ru), and Chinese (zh)—to maintain balanced representation. For a fair comparison, we ensured that all quiz questions are answerable from InfoGap facts and by extension from the respective articles in French, Russian, and Chinese.

We then prompted a large language model GPT-4o [48] to generate a multiple-choice question with four answer choices from each fact. The questions were manually reviewed and edited to ensure clarity, factual correctness, and alignment with the presented content. In addition to selecting the correct answer, participants were asked to indicate the language edition in which they found the supporting fact. Although all quiz facts were sourced from non-English Wikipedia editions, we note that some participants may have inferred the correct answer from contextual clues in the English article. Nonetheless, the quizzes were designed to evaluate whether participants engaged with multilingual content. If participants had guessed entirely at random, the expected accuracy would be 25%, given the four-choice format. Example questions are included in the Appendix A.

Performance Metrics. We recorded each participant's task completion time and quiz accuracy score under both conditions. Quiz accuracy captured how much multilingual knowledge participants were able to obtain, while completion time reflected how quickly they could locate and interpret the relevant information. Taken together, these two measures provide an indication of learning efficiency and allowed us to quantitatively compare the effectiveness of WikiGap with the standard Wikipedia interface.

Table 3. Number of knowledge gap facts discovered per food topic and language by INFOGAP, and the total number of facts shown in the WikiGap interface. For each topic, up to 10 facts per language were selected to ensure balanced representation and reduce cognitive overload. *Injera* had only 8 gaps in Chinese, resulting in 28 total facts shown (*).

Food Topic	Russian (ru)	French (fr)	Chinese (zh)	Facts Shown in WikiGap
Wiener schnitzel	65	20	62	30
Peking duck	13	23	69	30
Paella	28	62	52	30
Philippine adobo	21	15	84	30
Injera	10	14	8	28*

4.3 Materials and Topic Assignment

Selected topics. We focus on food articles because food serves as a culturally rich lens through which to explore knowledge gaps across language editions in Wikipedia. Culinary practices are deeply embedded in regional customs, social histories, and national identities, making food a meaningful proxy for cultural knowledge [36, 62]. Analyzing food allows researchers to capture both shared and divergent cultural representations, such as differences in dish names, preparation styles, and regional associations [62]. This cultural complexity makes food an ideal domain for surfacing factual asymmetries. We selected five culturally specific dishes: *Injera* (Ethiopia), *Paella* (Spain), *Philippine Adobo* (Philippines), *Peking Duck* (China), and *Wiener Schnitzel* (Austria). Each of these dishes has a dedicated Wikipedia article and distinct cultural heritage. Table 3 summarizes the number of language-exclusive facts detected per topic and the total number shown in the WIKIGAP interface.

All five dishes satisfy the following practical constraints: (i) they originate outside the English-speaking world, providing a comparable degree of cultural distance from the English source language, (ii) their English Wikipedia article is moderately sized (1,000–2,000 words), and (iii) the WIKIGAP extension can surface ~30 multilingual facts for each.[4]

Topic assignment. Each participant is assigned two distinct topics, one for each condition (control and WIKIGAP). Topic assignment was designed to minimize cultural confounds and familiarity-driven performance differences. Specifically, the food topic chosen for each participant could not be strongly linked to the participant's self-reported cultural or ethnic background. To reduce individual-performance bias, each topic was assigned in every condition to **exactly four** different participants.

4.4 Procedure

The lead investigator conducted the study remotely over a 1–1.5 hour recorded Zoom call. At the start, participants received a zip file containing the WIKIGAP extension and were instructed to install it on their Google Chrome browser. A short tutorial was provided by the researcher for both the treatment and control conditions. During the treatment condition, we began the round by introducing the WIKIGAP's functionality and allowed participants to explore the extension on their own until they felt ready to proceed. In the control condition, we demonstrated how to access language versions on Wikipedia using ILLs, and how to use Chrome's Google Translate function to read the content in English. At the beginning of each round, one of the five topics was assigned to the participant according to the topic assignment rules. After completing each round, the participant completed a System Usability Scale (SUS) after finishing the quiz. We used the SUS to measure ease-of-use and users' preference on with or without the WIKIGAP

---

[4]*Oolong tea* was included in the pilot with the first two participants but was dropped because its cultural context overlapped substantially with *Peking Duck*. It was therefore replaced with *Philippine Adobo* to preserve cultural diversity in topic representation.
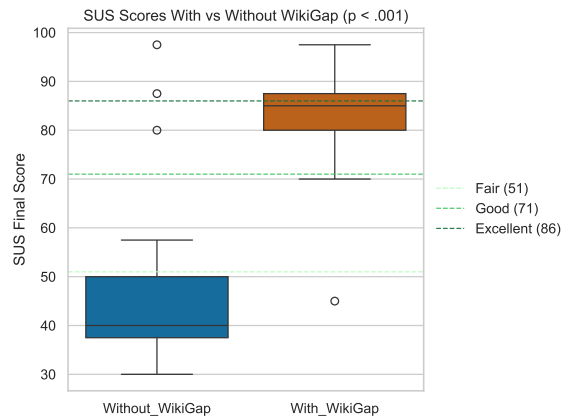
Fig. 6. Box plot showing the System Usability Scale (SUS) scores for each condition. Blue represents the control condition (no WikiGap), and orange represents the treatment condition (with WikiGap). Horizontal dashed lines represent standard usability benchmarks in varying shades of green: light green for *Fair* usability (SUS > 51), medium green for *Good* usability (SUS > 71), and dark green for *Excellent* usability (SUS > 86).

extension. To conclude, the first twelve participants completed an open-ended questionnaire, while the remaining nine took part in a semi-structured interview. Both instruments covered the same six areas: overall impressions of the system, presentation of multilingual information, perceived cultural impact, trust in the augmented facts, perceptions of Wikipedia's reliability after using WikiGap, and future inclination to explore multilingual content (see Appendix B for the full list of questions).

5 Findings

We report findings from our mixed-method evaluation of WikiGap. Together, these findings show how making multilingual knowledge differences visible within the English Wikipedia interface affects users' performance, preferences, and engagement with non-English content. Across quantitative measures and qualitative feedback, participants consistently described shifts in how they accessed information, perceived Wikipedia's completeness, and engaged with multilingual knowledge during routine fact-finding.

1. Making Multilingual Differences Visible Improves User Preference and Reading Experience (D1, D2, D3). We evaluated user preference using the System Usability Scale (SUS). The average SUS scores for each condition are presented in Figure 6, along with general SUS cutoff scores for fair, good, and excellent usability [5]. Overall, WikiGap achieved a substantially higher usability score compared to the control. WikiGap falls within the "excellent" usability category, while the default Wikipedia page with ILL scored below the "fair" threshold. A paired-samples *t*-test revealed a significant difference in scores between the WikiGap condition ($M = 82.5$, $SD = 10.9$) and the control condition ($M = 48.2$, $SD = 18.6$), $t(20) = 6.60$, $p < .001$. (Figure 6).

Participants attributed their strong preference for WikiGap not only to its ability to surface knowledge disparities, but also to the various design elements (D1-D3) that supported the quick access and ease of reading. Many highlighted how having access to multilingual content directly within the English article – through translated facts in a sidebar (D2, D3) – lowered the barrier to engaging with other language editions. It helped users "absorb information without the language difference obstacle" (P8). Participants appreciated the sidebar's structure and organization by language,

which kept all cross-lingual facts "apparent within a single page" (P5), eliminating the need to navigate away and reducing friction in multilingual exploration. In contrast to ILL's full-page switch model, participants noted that the **sentence-level fact card interface (D3)** made the information easier to process. As P11 remarked, *"[WikiGap] presents sentences instead of long paragraphs from the Wikipedia page."*

Other visual elements also shaped participants' perceptions of usability. Several users found the **in-text gap markers (D1)** helpful in surfacing areas of missing information in the English article. However, reactions to the color scheme used for language cues were mixed. While some found the colors intuitive, others, like P19, felt that the red-green coding could unintentionally suggest *"right or wrong"*.

Despite minor confusion about the color scheme, most users described the overall experience as more enjoyable and cognitively less demanding than their usual experience with Wikipedia's default ILL interface. P9 summarized this sentiment: *"Sometimes even when I open my Wikipedia, I would not read that much because of laziness, but WikiGap did a good job extracting information and makes the whole experience much more pleasant."*

2. Visibility of Cross-Lingual Differences Improves Fact-Finding Performance and Supports Cross-Cultural Learning (D4, D5). To evaluate the impact of WikiGap on users' ability to retrieve and retain cross-lingual information, we measured task performance through a fact-finding quiz, using both accuracy and completion time as metrics. As shown in Figure 7, participants performed significantly better when using WikiGap, achieving higher quiz accuracy and completing the task more quickly than with the default Wikipedia interface. A paired-samples t-test showed that accuracy was significantly higher in the WikiGap condition ($M = 0.91$, $SD = 0.09$) than in the control condition ($M = 0.73$, $SD = 0.19$), $t(20) = 4.75$, $p < .001$. Completion time also improved significantly, with users finishing the quiz faster using WikiGap ($M = 12.39$ minutes, $SD = 5.78$) compared to the control ($M = 20.84$ minutes, $SD = 8.68$), $t(20) = -5.52$, $p < .001$.[5] These results suggest that WikiGap not only facilitated access to relevant information but also improved users' efficiency in locating and interpreting cross-lingual facts. WikiGap consistently improved user performance across all five culturally diverse topics. For a topic-by-topic breakdown of quiz accuracy and completion time, see Appendix Figure 8.

These performance gains were reinforced by participants' qualitative feedback, which highlighted the value of both passive discovery and active exploration features. **The fact search (D5)** enabled users to directly retrieve relevant facts, supporting goal-oriented behaviors and speeding up information retrieval. As P16 explained, *"I don't need to go to multiple pages. I can search in the WikiGap extension and find any information quickly. Using the WikiGap extension saved me a lot of time."* This targeted retrieval helped users answer questions more efficiently, contributing to the improved fact-finding performance observed in the quiz metrics. Complementing this, the **language filter controls (D4)** allowed users to personalize their exploration by prioritizing facts from languages culturally aligned with the topic. Several participants reported that they first filtered to the language they deemed most relevant, as they *"will trust more when the language is directly aligned with the original country of the food."* (P1). Similarly, P7 remarked, *"I would love to see how the Tagalog page talks about Philippine adobo. In fact, diverse cultural perspectives on food is more linked to the cultural ecosystem where the food is grounded.".* Together, D4 and D5 supported efficient, goal-driven access to cross-lingual content.

Beyond performance gains, the extension also facilitated passive discovery and cross-cultural learning – a central goal of our design. Several users remarked that they encountered information they would not have actively sought

---

[5]Faster completion times in the WikiGap condition is partly expected, as the quiz questions were constructed from facts extracted by InfoGap. In real-world use, not all facts would be presented by WikiGap, and completion time gains may therefore be more modest.
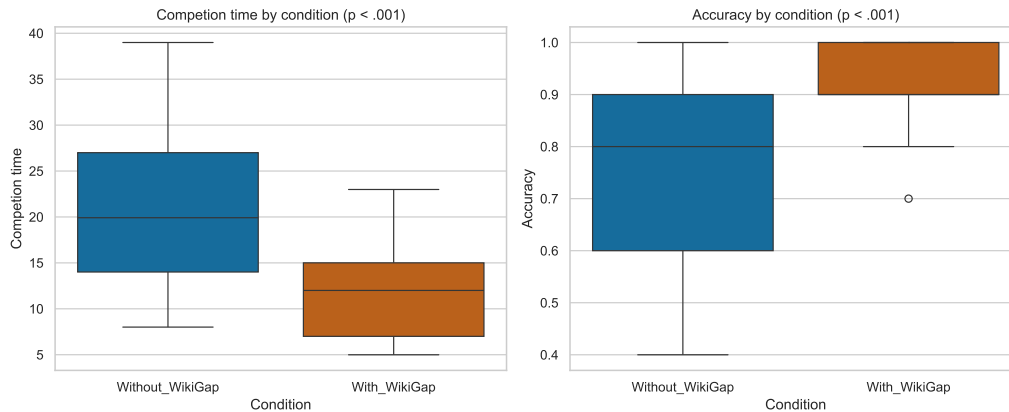
Fig. 7. **Average completion time** (left) and **quiz accuracy** (right) across conditions. Blue represents the control condition (no WikiGap), and orange represents the treatment condition (with WikiGap). Using WikiGap demonstrates a statistically significant improvement in user performance on the quiz.

out on their own. As P1 reflected, *"Without [WikiGap], I wouldn't have thought about looking into Wikipedia pages in different languages."* Others described moments of unexpected cultural insight, such as P5, who noted, *"I would have completely missed out on the historical aspect if I used Wikipedia normally (only in English)."* Participants also pointed to surprising culture-specific facts, like the culinary use of Sprite in a traditional dish (P13). These examples illustrate that making multilingual differences visible not only improved task efficiency, but also broadened users' exposure to culturally situated knowledge.

3. WikiGap Raises Awareness of Multilingual Gaps and Motivates Exploration. Participants consistently credited WikiGap for revealing factual disparities they were previously unaware of. Several users described being surprised that different language versions did not simply mirror each other. As P20 reflected, *"Before WikiGap, I always assumed that everything presented in English would also be visible in another language."* Similarly, P18 noted that they *"didn't know there could be so many differences in how the content was presented in various languages."*

This realization often prompted frustration as participants became aware of the limits of English Wikipedia. As P5 expressed, *"I'm frustrated that this doesn't already exist! I feel like I've been missing out on information from other languages because I just assumed other pages would have the same information translated."* Other participants had deeper reflections on how WikiGap changed their perception about knowledge: *"Knowledge is not plain, knowledge has different dimensions waiting to be discovered, and this tool offers this opportune moment."* (P10).

Importantly, increased awareness also motivated future engagement. Participants described being more inclined to explore content outside their primary language now that differences were visible and easy to access.

> *"I was never on the pages of Wikipedia where the language is not what I speak, because I often assume that all pages under different languages are the same if they are of the same topic, so sometimes I struggled to find some information in the Wikipedia. But if I have WikiGap, I think I will definitely explore more."* — P9

Several emphasized that centralizing multilingual content within a single interface reduced the effort required to engage, making exploration more likely. Now they can *"see different viewpoints from different language pages"* (P17), they are more inclined to explore content outside their primary language in the future. Several users described feeling

more motivated to engage with multilingual content now that they understood its value. As P1 noted, "*If there's a tool like WikiGap that can help me easily get various information from other pages, I would like to explore. If not, I wouldn't navigate multiple pages myself.*"

4. Respondents' Views on LLMs and Wikipedia: Perceived Trust and Tensions. Awareness of cross-lingual disparities also led participants to reflect on the trustworthiness of Wikipedia and how it compares to LLM-based knowledge sources. We observed a wide range of perspectives. For some, WikiGap surfaced surprising inconsistencies between language editions. Those differences made them question the completeness or neutrality of English Wikipedia. Others interpreted these variations not as flaws, but as a feature of Wikipedia's multilingualism and editorial diversity. These divergent views led participants to compare Wikipedia to emerging alternatives such as LLMs, positioning WikiGap as a catalyst for rethinking how people evaluate and engage with knowledge sources.

A minority (n=2) of the participants reported a decline in trust in Wikipedia after using WikiGap. They described how uncovering missing or differing facts between language editions made them feel less confident in the platform's ability to represent comprehensive or balanced information. This, in turn, increased their inclination to rely on LLMs as knowledge sources:

> "*I feel that after using WikiGap, I would be inclined to use Wikipedia even less. Before this, I assumed that Wikipedia would be a reliable and complete source of knowledge. However, I now know that the English version misses several key details. This leads me to be more reliant on LLM-based tools that have all the information in one place.*" — P2

For this participant and others, WikiGap altered their perception of Wikipedia as a trustworthy authority. The perceived comprehensiveness of LLMs, combined with their ability to synthesize content across sources, positioned them as a more efficient alternative, even if they lacked transparency.[6]

However, not all participants responded this way. Others voiced strong concerns about the lack of attribution and provenance in LLM-generated responses, and emphasized that WikiGap made them appreciate Wikipedia's editorial structure and citation practices even more. For these users, the ability to trace facts back to their original source, including reading the sentence in the original language edition, was critical to their sense of trust:

> "*I don't trust ChatGPT that much because sometimes it just makes things up, and it doesn't provide sources. With WikiGap, I can click on, say, the French Wikipedia page, read it, and see the sources. It feels more reliable. I also appreciate that Wikipedia cites everything, so there's some quality control, even if it's not always perfect.*" — P13

These contrasting responses underscore an important dynamic in how users engage with knowledge platforms: while LLMs offer convenience and perceived comprehensiveness, they can lack the transparency and editorial accountability that are central to Wikipedia's epistemic model. For some users, WikiGap undermined their confidence in Wikipedia. For others, it reaffirmed the value of its provenance-driven structure, even in the face of inconsistency. We return to the socio-technical implications in Sec. 6.

5. WikiGap Suggests Opportunities for Supporting Cross-Lingual Editing. Although WikiGap was primarily designed to support readers by surfacing factual cross-lingual content, a few participants, particularly those with prior

---

[6]In practice, the assumption that multilingual LLMs such as ChatGPT and Gemini can synthesize information learned from web text in different languages is not entirely true; they often fail to retrieve facts learned in one language when prompted in another language [17].

editing experience, reflected on how it could assist with editorial tasks. These insights point to an emergent opportunity for WIKIGAP to support not only knowledge consumption but also knowledge curation on Wikipedia.

Some users noted that the ability to trace and verify information from other language editions made the tool useful for improving article quality. As one frequent editor explained, *"Sometimes an article in another language has crucial information that's missing in English. WIKIGAP helps me find and translate it, then cite it properly."* (P16). Another participant elaborated on how the tool fits into their existing editing workflow: *"I'd check the source WIKIGAP shows me, then cite the same source... WIKIGAP helps me discover facts and references."* (P17). While these reflections were not prompted by editor-specific questions, they point to an emergent opportunity: making multilingual gaps visible to readers may also lower barriers to cross-lingual knowledge integration, extending WIKIGAP's impact beyond consumption to potential curation.

6 Discussion

We interpret our findings through the lens of *boundary objects*—shared artifacts that enable coordination across heterogeneous communities without requiring consensus or uniform interpretation [56, 57]. We argue that WIKIGAP functions as a boundary object within Wikipedia's multilingual ecosystem by making cross-lingual knowledge differences legible, actionable, and traceable within everyday reading practices. In this role, WIKIGAP enables three intertwined forms of epistemic and coordinative work. First, it reframes readers' epistemic assumptions by exposing the incompleteness of English Wikipedia and foregrounding knowledge produced by other language communities (§6.1). Second, it provides a shared reference point that supports articulation work between readers and editors across language editions, enabling verification and coordination without enforcing uniformity (§6.2). Third, it gestures toward a broader infrastructural role by modeling how pluralistic, provenance-aware interfaces can counter knowledge consolidation in emerging AI-mediated systems (§6.3).

6.1 Promoting Epistemic Equity by Challenging the "English-as-Superset" Assumption

Our user study revisits prior findings that the dominance of English Wikipedia leads readers to treat it as a comprehensive superset of knowledge across language editions, rather than as a culturally situated and socio-geographically positioned construction [34]. Participants explicitly articulated this belief. Although more than 60% of participants reported a non-English primary language, all identified English as their primary source on Wikipedia, largely because English Wikipedia covers more topics and is updated more frequently. Despite being familiar with Wikipedia's ILLs, participants predominantly consumed content in English, reflecting how infrastructural dominance concentrates reader attention within a single language edition and reinforces the invisibility of knowledge produced by other language communities. At the interface level, Wikipedia's current design renders other language editions technically accessible but epistemically invisible: ILLs signal the existence of alternative versions, yet do not communicate that English content is neither exhaustive nor neutral. This gap was evident in participants' assumptions; as one remarked, *"Before WikiGap, I always assumed that everything presented in English would also be visible in another language."*

Consistent with this perception, participants strongly preferred WIKIGAP over the default ILLs interface and performed significantly better on fact-finding tasks, indicating that access alone—without cues about epistemic differences, provenance or community context—is insufficient to support meaningful cross-lingual engagement. WIKIGAP challenges the "English-as-superset" assumption by both raising awareness of cross-linguistic knowledge gaps and redistributing

reader attention toward knowledge produced by other language communities. Adopting an asset-based design orientation, the system foregrounds non-English content not as supplementary or corrective, but as a valuable product of distinct peer production efforts.

By embedding sentence-level complementary facts in context and pairing them with explicit provenance, WikiGap makes the limits of English Wikipedia visible and simultaneously inviting engagement with the work of other language editions. This design encourages readers to recognize Wikipedia as a collection of interconnected yet distinct knowledge production communities, rather than a single monolithic source of global consensus. In doing so, WikiGap promotes epistemic equity in a scoped, design-operational sense: by redistributing attention and engagement across Wikipedia's multilingual peer production communities and supporting cross-community interaction as part of everyday readership.

6.2 WikiGap as a Boundary Object Supporting Articulation Work Across Reader and Editor Communities

Wikipedia's multilingual knowledge production spans heterogeneous communities with distinct linguistic, cultural, and editorial norms. Our findings show that WikiGap operates as a boundary object in practice, providing a shared point of reference through which readers and editors from different language communities can interpret, evaluate, and act on cross-lingual knowledge differences without requiring consensus. By surfacing sentence-level complementary facts with clear provenance, WikiGap offers a way to support shared understanding across communities without requiring consensus or uniformity. Building on these findings, we interpret WikiGap as enabling an asset-based stance: it helps readers and multilingual contributors recognize non-English facts as usable resources by making them legible and actionable through provenance and in-context integration.

6.2.1 For Readers of English Wikipedia. WikiGap benefits readers by supporting fact-finding, cross-cultural learning, and interpretive flexibility. Participants described discovering content they "would have otherwise missed" in English and expressed surprise at how much meaningful information existed exclusively in other language editions (Findings 1–2). Several noted that the tool increased their curiosity and willingness to visit other language pages, shifting their sense of Wikipedia from a monolithic source of global consensus to a collection of situated, culturally inflected knowledge communities.

Through a CSCW lens, this suggests that WikiGap enhances the *legibility* of multilingual knowledge work. It exposes facts that English readers typically never encounter, foregrounds the situated perspectives of marginalized linguistic communities, and supports interpretive flexibility—one of the core affordances of boundary objects. Rather than collapsing perspectives into a single authoritative narrative, WikiGap allows readers to compare, contextualize, and reinterpret information based on their own background knowledge and cultural standpoint.

6.2.2 For (Potential) Editors Across Languages. Although our study focused on readers, participants with editing experience highlighted how WikiGap could support multilingual editorial workflows, they pointed to using the provenance link to see the original facts from the other languages to verify and potentially transfer content across editions. Maintaining consistency across language editions requires editors to reconcile differing source traditions, cultural emphases, and update rhythms. This process is inherently distributed and involves articulation work [53]: interpreting discrepancies, assessing their relevance to local communities, coordinating with other editors, and making decisions about whether and how to incorporate missing knowledge.

Participants noted that WikiGap reduces some of this articulation burden by automatically surfacing contextualized discrepancies, highlighting their provenance, and linking directly to original sentences. These cues help editors evaluate the origins and significance of differences, understand when they reflect meaningful cultural divergence versus simple

Manuscript submitted to ACM

incompleteness, and coordinate potential updates. Rather than automating agreement or enforcing uniformity, WIKIGAP supports the human work of interpretation and negotiation that underlies Wikipedia's collaborative ecosystem. By enabling multiple situated viewpoints to coexist while still facilitating coordination across them, the system exemplifies how boundary objects can scaffold the complex articulation work required for distributed knowledge production.

In this way, WIKIGAP renders the often invisible coordinative labor of multilingual maintenance more visible and manageable by making cross-lingual differences legible and actionable through provenance and in-context integration, reducing coordination friction and supporting editor engagement.

6.3 Positionality, LLMs, and the Future Role of WikiGap as Boundary Infrastructure

Across both quantitative preferences and qualitative interviews, participants expressed strong appreciation for WIKI-GAP's explicit display of provenance and its framing of multilingual knowledge as situated rather than objective. This recognition of positionality became particularly salient when contrasted—often spontaneously—against the increasingly common practice of using Large Language Models (LLMs) as ostensibly objective, consolidated knowledge sources.

A small subset of participants (n=2) expressed a desire for consolidated knowledge and therefore perceived gaps highlighted by WIKIGAP as evidence of English Wikipedia's "incompleteness." One participant noted that they would prefer to turn to LLMs, which present a single synthesized output. While LLM outputs are often perceived as authoritative or persuasive [10], they rarely foreground multiple cultural or linguistic perspectives and frequently exhibit English-centric biases [27, 41, 58]. They also obscure provenance [55], making it difficult for users to evaluate where information originates. LLMs thus risk presenting a "view from nowhere," [22] even though their outputs reflect specific training data and design choices.

Although WIKIGAP and LLMs appear to sit at opposite ends of the spectrum—positionality versus objectivity—they need not be mutually exclusive. Our findings suggest that asset-based design principles foregrounding positionality can inform future LLM interfaces by encouraging users to see responses as situated rather than uniformly objective. Concretely, interfaces could surface multiple culturally grounded viewpoints instead of a single synthesized answer, along with lightweight provenance cues indicating which language communities or traditions each perspective reflects. Such designs would counter the tendency of current systems to amplify majority perspectives and obscure minority ones, helping users explore diverse interpretations more intentionally. They would also make latent cultural knowledge in LLMs more accessible to researchers and end users alike. In this sense, embedding WIKIGAP-like scaffolding into LLM interfaces represents a promising direction for foregrounding epistemic positionality within increasingly consolidated AI ecosystems. This extends our system's asset-based design principles to cross-lingual and multilingual LLM interfaces, encouraging users to engage with knowledge as situated, grounded in provenance, and produced by distinct language communities.

6.4 Limitations and Future Work

Language Coverage. Our study focuses on English and three target languages (French, Russian, and Chinese). While these languages span different linguistic families and large Wikipedia communities, they do not capture the full diversity of Wikipedia's language ecosystem. Several participants expressed a desire for broader language coverage, particularly for pages in their language of origin. For example, one participant noted, *"I would love to see how the Tagalog page could complement it. In fact, diverse cultural perspectives on food are more linked to the cultural ecosystem where the food is grounded"* (P7). Expanding to more languages will enrich cross-cultural perspectives and align with users' expectations in future work. The underlying INFOGAP pipeline is language-agnostic: in principle, it can operate on *any* language that

GPT-4 supports, though quality may vary, especially for low-resource languages. This is important for future research as smaller language editions remain largely understudied and present additional challenges in their own right [32, 45].

WikiGap content selection. We selected a random subset of facts to present from other language editions (Sec. 3.3.2 for further detail). This selection could be improved for greater localization through applying a geoprovenance classifier [54], thereby emphasizing those facts that are sourced from a geographical region relevant to the language edition. This is an important step for future work as our respondents expressed the desire for *localized* Volunteered Geographic Information.

Capacity for misinformation contagion. Language editions vary considerably in their susceptibility to organized disinformation campaigns [31]. Thus, there is a risk of spreading misinformation from other language editions into English Wikipedia through WikiGap. WikiGap may thus inadvertently make smaller language editions more attractive for actors carrying out disinformation campaigns, as audience size has previously been proposed as a motivating factor for bad-faith contributions [31]. Care should be thus taken in deploying WikiGap, with attention paid to the language editions and topics that are selected for the extension.

System Architecture. WikiGap demonstrates a socio-technical intervention that aims to reshape how people access, relate to, and interpret multilingual knowledge, which is beyond a purely technical innovation. We have an overwhelmingly positive result from the evaluation. It's important to acknowledge that these results come from a relatively controlled setting. While the findings are encouraging, future work will need to address the technical and interaction challenges of deploying a live, scalable version of the system in more variable real-world conditions.

Future Work. Building on our discussion of pluralistic LLM interfaces, future work could extend WikiGap's design principles to other multilingual information-sharing platforms. Prior work shows that platforms such as Twitter and Flickr also exhibit substantial cross-lingual divergence in how events, entities, and cultural topics are discussed [12, 13, 19, 40]. Applying WikiGap could help users interpret how narratives differ across languages and communities beyond Wikipedia. Extending this approach across platforms would also inform the design of more culturally aware LLM interfaces, reinforcing our broader argument that pluralistic system design can mitigate the homogenizing effects of consolidated AI infrastructures.

7 Conclusion

WikiGap contends with longstanding structural inequalities in Wikipedia by enabling readers to access cross-lingual facts directly within the English interface. Through a novel combination of fact-centric text comparison [InfoGap; 52] based on LLMs [42] and user-centered design, the system unobtrusively displays knowledge missing due to linguistic and editorial asymmetries. Our evaluation shows that WikiGap not only improves usability and fact-finding efficiency, but also fosters greater awareness of cultural variation and the limitations of assuming English Wikipedia as a default knowledge base. By surfacing multilingual content in context, WikiGap connects readers with perspectives on topics that are situated outside of the Anglosphere, connections that were previously occluded by the hegemonic dominance of English Wikipedia [34, Chapter 3]. This work demonstrates how augmentative tools can shift reader practices and challenge epistemic assumptions, paving the way for more inclusive, transparent, and culturally grounded knowledge infrastructures. This paper examines how interface design can challenge the English-as-superset assumption by making multilingual knowledge differences visible to readers during everyday Wikipedia use.

8 Acknowledgments

References

[1] Eytan Adar, Michael Skinner, and Daniel S Weld. 2009. Information arbitrage across multi-lingual Wikipedia. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*. 94–103.

[2] Leah Hope Ajmani, Jasmine C. Foriest, Jordan Taylor, Kyle Pittman, Sarah Gilbert, and Michael Ann DeVito. 2024. Whose Knowledge is Valued? Epistemic Injustice in CSCW Applications. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2, Article 523 (Nov. 2024), 28 pages. doi:10.1145/3687062

[3] Ofer Arazy, Hila Liifshitz-Assaf, Oded Nov, Johannes Daxenberger, Martina Balestra, and Coye Cheshire. 2017. On the "How" and "Why" of Emergent Role Behaviors in Wikipedia. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 2039–2051. doi:10.1145/2998181.2998317

[4] Bunty Avieson and Suruchi Mazumdar. 2025. Wiki warriors: language editors counter knowledge hierarchies during the pandemic. *Asian Journal of Communication* (2025), 1–19.

[5] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: adding an adjective rating scale. *J. Usability Studies* 4, 3 (May 2009), 114–123.

[6] Patti Bao, Brent Hecht, Samuel Carton, Mahmood Quaderi, Michael Horn, and Darren Gergle. 2012. Omnipedia: bridging the wikipedia language gap. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1075–1084.

[7] Shaowen Bardzell. 2010. Feminist HCI: taking stock and outlining an agenda for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1301–1310. doi:10.1145/1753326.1753521

[8] Ruha Benjamin. 2023. Race after technology. In *Social Theory Re-Wired*. Routledge, 405–415.

[9] Ewa S Callahan and Susan C Herring. 2011. Cultural bias in Wikipedia content on famous persons. *Journal of the American society for information science and technology* 62, 10 (2011), 1899–1915.

[10] Shruthi Chockkalingam, Seyed Hossein Alavi, Raymond T. Ng, and Vered Shwartz. 2025. Should I go vegan: Evaluating the Persuasiveness of LLMs in Persona-Grounded Dialogues. In *Proceedings of the Third Workshop on Social Influence in Conversations (SICon 2025)*, James Hale, Brian Deuksin Kwon, and Ritam Dutt (Eds.). Association for Computational Linguistics, Vienna, Austria, 65–72. doi:10.18653/v1/2025.sicon-1.4

[11] Noam Cohen. 2008. Open-Source Troubles in Wiki World. New York Times. https://www.nytimes.com/2008/03/17/technology/17wikipedia.html

[12] Wei Dong and Wai-Tat Fu. 2010. Cultural difference in image tagging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 981–984. doi:10.1145/1753326.1753472

[13] Amir Dotan and Panayiotis Zaphiris. 2010. A cross-cultural analysis of Flickr users from Peru, Israel, Iran, Taiwan and the UK. *Int. J. Web Based Communities* 6, 3 (June 2010), 284–302. doi:10.1504/IJWBC.2010.033753

[14] Kevin Duh, Ching-Man Au Yeung, Tomoharu Iwata, and Masaaki Nagata. 2013. Managing information disparity in multilingual document collections. *ACM Transactions on Speech and Language Processing (TSLP)* 10, 1 (2013), 1–28.

[15] Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2020. Language-agnostic BERT Sentence Embedding. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 879–894. https://arxiv.org/abs/2007.01852

[16] Michel Foucault. 2013. *Archaeology of knowledge.* routledge.

[17] Omer Goldman, Uri Shaham, Dan Malkin, Sivan Eiger, Avinatan Hassidim, Yossi Matias, Joshua Maynez, Adi Mayrav Gilady, Jason Riesa, Shruti Rijhwani, Laura Rimell, Idan Szpektor, Reut Tsarfaty, and Matan Eyal. 2025. ECLeKTic: a Novel Challenge Set for Evaluation of Cross-Lingual Knowledge Transfer. arXiv:2502.21228 [cs.CL] https://arxiv.org/abs/2502.21228

[18] David Golumbia. 2009. *The cultural logic of computation.* Harvard University Press.

[19] Mark Graham, Scott A. Hale, and Devin Gaffney. 2014. Where in the World Are You? Geolocation and Language Identification in Twitter. *The Professional Geographer* 66, 4 (2014), 568–578. doi:10.1080/00330124.2014.907699 arXiv:https://doi.org/10.1080/00330124.2014.907699

[20] Jonathan Grudin. 1988. Why CSCW applications fail: problems in the design and evaluationof organizational interfaces. In *Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work* (Portland, Oregon, USA) *(CSCW '88)*. Association for Computing Machinery, New York, NY, USA, 85–93. doi:10.1145/62266.62273

[21] Scott A Hale. 2014. Multilinguals and Wikipedia editing. In *Proceedings of the 2014 ACM conference on Web science*. 99–108.

[22] Donna Haraway. 2013. Situated knowledges: The science question in feminism and the privilege of partial perspective 1. In *Women, science, and technology*. Routledge, 455–472.

[23] Jean Hardy and Jacob Thebault-Spieker. 2024. A Turn to Assets in Community-Based Computing Research: Tradeoffs, Deficits, and Neoliberalism in Technological Development. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 14 (April 2024), 20 pages. doi:10.1145/3637291

[24] Brent Hecht and Darren Gergle. 2009. Measuring self-focus bias in community-maintained knowledge repositories. In *Proceedings of the fourth international conference on communities and technologies*. 11–20.

[25] Brent Hecht and Darren Gergle. 2010. The tower of Babel meets web 2.0: user-generated content and its applications in a multilingual context. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 291–300.

[26] Brent Jaron Hecht. 2013. *The mining and application of diverse cultural perspectives in user-generated content.* Ph. D. Dissertation. Northwestern University.

[27] Daniel Hershcovich, Stella Frank, Heather Lent, Miryam de Lhoneux, Mostafa Abdou, Stephanie Brandl, Emanuele Bugliarello, Laura Cabello Piqueras, Ilias Chalkidis, Ruixiang Cui, Constanza Fierro, Katerina Margatina, Phillip Rust, and Anders Søgaard. 2022. Challenges and Strategies in Cross-Cultural NLP. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, Dublin, Ireland, 6997–7013. doi:10.18653/v1/2022.acl-long.482

[28] Molly G Hickman, Viral Pasad, Harsh Kamalesh Sanghavi, Jacob Thebault-Spieker, and Sang Won Lee. 2021. Understanding wikipedia practices through hindi, urdu, and english takes on an evolving regional conflict. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–31.

[29] Cheryl I Johnson and Richard E Mayer. 2012. An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied* 18, 2 (2012), 178.

[30] Isaac Johnson and Emily Lescak. 2022. Considerations for multilingual wikipedia research. *arXiv preprint arXiv:2204.02483* (2022).

[31] Zarine Kharazian, Kate Starbird, and Benjamin Mako Hill. 2024. Governance capture in a self-governing community: A qualitative comparison of the croatian, serbian, bosnian, and serbo-croatian wikipedias. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–26.

[32] Sejal Khatri, Aaron Shaw, Sayamindu Dasgupta, and Benjamin Mako Hill. 2022. The social embeddedness of peer production: A comparative qualitative analysis of three Indian language Wikipedia editions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18.

[33] John P. Kretzman and John McKnight. 1993. Building Communities from the Inside Out: A Path Toward Finding and Mobilizing a Community's Assets. https://api.semanticscholar.org/CorpusID:109453662

[34] Sangeet Kumar. 2021. *The digital frontier: Infrastructures of control on the global web.* Indiana University Press.

[35] Paul Laufer, Claudia Wagner, Fabian Flöck, and Markus Strohmaier. 2015. Mining cross-cultural relations from Wikipedia: a study of 31 European food cultures. In *Proceedings of the ACM web science conference*. 1–10.

[36] Yiwei Luo, Kristina Gligoric, and Dan Jurafsky. 2023. Othering and Low Status Framing of Immigrant Cuisines in US Restaurant Reviews and Large Language Models. In *International Conference on Web and Social Media*. https://api.semanticscholar.org/CorpusID:259936860

[37] Emanuel Maiberg. 2025. Wikipedia Says AI Is Causing a Dangerous Decline in Human Visitors. *404 Media* (2025). https://www.404media.co/wikipedia-says-ai-is-causing-a-dangerous-decline-in-human-visitors/ Published online 16 October 2025; accessed 2025-12-03.

[38] Catherine C Marshall. 1997. Annotation: from paper books to the digital library. In *Proceedings of the second ACM international conference on Digital libraries*. 131–140.

[39] Paolo Massa and Federico Scrinzi. 2012. Manypedia: Comparing language points of view of Wikipedia communities. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*. 1–9.

[40] Julia Mendelsohn, Sayan Ghosh, David Jurgens, and Ceren Budak. 2023. Bridging nations: quantifying the role of multilinguals in communication on social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 17. 626–637.

[41] Rada Mihalcea, Oana Ignat, Longju Bai, Angana Borah, Luis Chiruzzo, Zhijing Jin, Claude Kwizera, Joan Nwatu, Soujanya Poria, and Thamar Solorio. 2025. Why AI Is WEIRD and Shouldn't Be This Way: Towards AI for Everyone, with Everyone, by Everyone. *Proceedings of the AAAI Conference on Artificial Intelligence* 39, 27 (Apr. 2025), 28657–28670. doi:10.1609/aaai.v39i27.35092

[42] Sewon Min, Kalpesh Krishna, Xinxi Lyu, Mike Lewis, Wen-tau Yih, Pang Wei Koh, Mohit Iyyer, Luke Zettlemoyer, and Hannaneh Hajishirzi. 2023. Factscore: Fine-grained atomic evaluation of factual precision in long form text generation. *arXiv preprint arXiv:2305.14251* (2023).

[43] Marc Miquel-Ribé and David Laniado. 2018. Wikipedia culture gap: quantifying content imbalances across 40 language editions. *Frontiers in physics* 6 (2018), 54.

[44] Marc Miquel-Ribé and David Laniado. 2020. The Wikipedia Diversity Observatory: A Project to Identify and Bridge Content Gaps in Wikipedia. In *Proceedings of the 16th International Symposium on Open Collaboration* (Virtual conference, Spain) *(OpenSym '20)*. Association for Computing Machinery, New York, NY, USA, Article 14, 4 pages. doi:10.1145/3412569.3412866

[45] Hellina Hailu Nigatu, John Canny, and Sarah E Chasins. 2024. Low-Resourced Languages and Online Knowledge Repositories: A Need-Finding Study.. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–21.

[46] Oded Nov. 2007. What motivates Wikipedians? *Commun. ACM* 50, 11 (Nov. 2007), 60–64. doi:10.1145/1297797.1297798

[47] OpenAI. 2024. GPT-4o mini: Advancing Cost-Efficient Intelligence. https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence. Large language model, July 18 version.

[48] OpenAI. 2024. GPT-4o System Card. *arXiv preprint arXiv:2410.21276* (2024). https://arxiv.org/abs/2410.21276

[49] Peter Pirolli and Stuart Card. 1999. Information foraging. *Psychological review* 106, 4 (1999), 643.

[50] Dwaipayan Roy, Sumit Bhatia, and Prateek Jain. 2022. Information asymmetry in Wikipedia across different languages: A statistical analysis. *Journal of the Association for Information Science and Technology* 73, 3 (2022), 347–361.

[51] Edward W Said. 1977. Orientalism. *The Georgia Review* 31, 1 (1977), 162–206.

[52] Farhan Samir, Chan Young Park, Anjalie Field, Vered Shwartz, and Yulia Tsvetkov. 2024. Locating Information Gaps and Narrative Inconsistencies Across Languages: A Case Study of LGBT People Portrayals on Wikipedia. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen (Eds.). Association for Computational Linguistics, Miami, Florida, USA, 6747–6762. doi:10.18653/v1/2024.emnlp-main.384

[53] Kjeld Schmidt and Liam Bannon. 1992. Taking CSCW seriously. 1, 1 (1992), 7–40. doi:10.1007/BF00752449

[54] Shilad W Sen, Heather Ford, David R Musicant, Mark Graham, OS Keyes, and Brent Hecht. 2015. Barriers to the localness of volunteered geographic information. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 197–206.

[55] Chirag Shah and Emily M Bender. 2022. Situating search. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*. 221–232.

[56] Susan Leigh Star. 2010. This is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology, & Human Values* 35, 5 (2010), 601–617. doi:10.1177/0162243910377624 arXiv:https://doi.org/10.1177/0162243910377624

[57] Susan Leigh Star and James R. Griesemer. 1989. Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science* 19, 3 (1989), 387–420. doi:10.1177/030631289019003001 arXiv:https://doi.org/10.1177/030631289019003001

[58] Yan Tao, Olga Viberg, Ryan S Baker, and René F Kizilcec. 2024. Cultural bias and cultural alignment of large language models. *PNAS nexus* 3, 9 (2024), pgae346.

[59] Nicholas Vincent and Brent Hecht. 2021. A deeper investigation of the importance of Wikipedia links to search engine results. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–15.

[60] Wikimedia Foundation. 2011. *Editor Survey Report – April 2011*. Technical Report. Wikimedia Foundation. https://upload.wikimedia.org/wikipedia/commons/7/76/Editor_Survey_Report_-_April_2011.pdf Accessed.

[61] Wikimedia Foundation. 2024. Announcing English Wikipedia's most popular articles of 2024. https://wikimediafoundation.org/news/2024/12/03/announcing-english-wikipedias-most-popular-articles-of-2024/ Accessed: 2025-09-09.

[62] Genta Indra Winata, Frederikus Hudi, Patrick Amadeus Irawan, David Anugraha, Rifki Afina Putri, Wang Yutong, Adam Nohejl, Ubaidillah Ariq Prathama, Nedjma Ousidhoum, Afifa Amriani, Anar Rzayev, Anirban Das, Ashmari Pramodya, Aulia Adila, Bryan Wilie, Candy Olivia Mawalim, Cheng Ching Lam, Daud Abolade, Emmanuele Chersoni, Enrico Santus, Fariz Ikhwantri, Garry Kuwanto, Hanyang Zhao, Haryo Akbarianto Wibowo, Holy Lovenia, Jan Christian Blaise Cruz, Jan Wira Gotama Putra, Junho Myung, Lucky Susanto, Maria Angelica Riera Machin, Marina Zhukova, Michael Anugraha, Muhammad Farid Adilazuarda, Natasha Christabelle Santosa, Peerat Limkonchotiwat, Raj Dabre, Rio Alexander Audino, Samuel Cahyawijaya, Shi-Xiong Zhang, Stephanie Yulia Salim, Yi Zhou, Yinxuan Gui, David Ifeoluwa Adelani, En-Shiun Annie Lee, Shogo Okada, Ayu Purwarianti, Alham Fikri Aji, Taro Watanabe, Derry Tanti Wijaya, Alice Oh, and Chong-Wah Ngo. 2025. WorldCuisines: A Massive-Scale Benchmark for Multilingual and Multicultural Visual Question Answering on Global Cuisines. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, Luis Chiruzzo, Alan Ritter, and Lu Wang (Eds.). Association for Computational Linguistics, Albuquerque, New Mexico, 3242–3264. https://aclanthology.org/2025.naacl-long.167/

[63] Marisol Wong-Villacres, Aakash Gautam, Wendy Roldan, Lucy Pei, Jessa Dickinson, Azra Ismail, Betsy DiSalvo, Neha Kumar, Tammy Clegg, Sheena Erete, Emily Roden, Nithya Sambasivan, and Jason Yip. 2020. From Needs to Strengths: Operationalizing an Assets-Based Design of Technology. In *Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing* (Virtual Event, USA) *(CSCW '20 Companion)*. Association for Computing Machinery, New York, NY, USA, 527–535. doi:10.1145/3406865.3418594

A Sample Quiz Questions

We present two example questions from each of the five topics used in our study. These examples illustrate the format and cultural specificity of the quiz items. All questions were drawn from the 30 language-exclusive facts surfaced by WikiGap, and were manually reviewed for clarity and accuracy. The correct answers are bolded in Table 4.

Table 4. Example Quiz Questions Across Topics. Correct answers are bolded.

Topic	Question 1	Question 2
Peking Duck	*What unique preparation step is done to the duck before roasting it for Peking duck?* A. It is marinated in soy sauce **B. It is inflated with air under the skin** C. It is stuffed with rice D. It is boiled in water	*What is served after the meat in a traditional Peking duck meal?* A. A fruit platter B. Rice **C. Chinese cabbage soup** D. A cup of tea
Wiener Schnitzel	*During which historical period was Wiener schnitzel brought to Italy and then to Austria?* A. World War I B. French Revolution **C. Napoleonic Wars** D. Renaissance	*Which city of China has a type of Western cuisine similar to Vienna schnitzel?* A. Beijing B. Nanjing C. Guangdong **D. Shanghai**
Paella	*What festival is Spanish paella associated with?* A. La Tomatina B. San Fermín **C. Falles** D. Semana Santa	*How was paella traditionally eaten?* A. On plates B. In bowls **C. Straight from the cooking pan** D. Consumed with bread
Philippine Adobo	*Which ingredient is sometimes added to adobo to replace palm or coconut sugar and help tenderize the meat?* A. Honey B. Molasses **C. Sprite** D. Maple syrup	*Which of the following is a traditional method used in adobo to keep meat fresh in tropical climates?* A. Drying the meat **B. Frying with vinegar** C. Smoking the meat D. Freezing the meat
Injera	*How large is traditional injera typically?* A. 30 centimeters B. Half a meter **C. About 1 meter** D. 2 meters	*Injera closely resembles which Middle Eastern pancake variant?* A. Naan **B. Lahoh** C. Pita D. Lavash

B Interview and Questionnaire Protocol

Participants were asked the following questions during either the semi-structured interview or the open-ended questionnaire.

- Describe your experience using WikiGap. What stood out to you the most, and why?
- When you enabled WikiGap to see multilingual facts, how did you feel about the way information was presented? What aspects of the presentation worked well or could be improved?
- Can you describe a specific fact you discovered through WikiGap that surprised you or changed your understanding of the topic? How did this discovery influence your perception or understanding of diverse cultural viewpoints?
- When using WikiGap, how did you decide which information to trust? Please explain your reasoning.
- After using WikiGap, do you see Wikipedia as more reliable or complete, especially for finding information from other languages? Please explain your answer with examples if possible.
- In what ways, if any, did WikiGap change your perception or understanding of diverse cultural viewpoints related to the topics you explored?
- Did using WikiGap make you more inclined to explore content outside your primary language in the future? Why or why not?

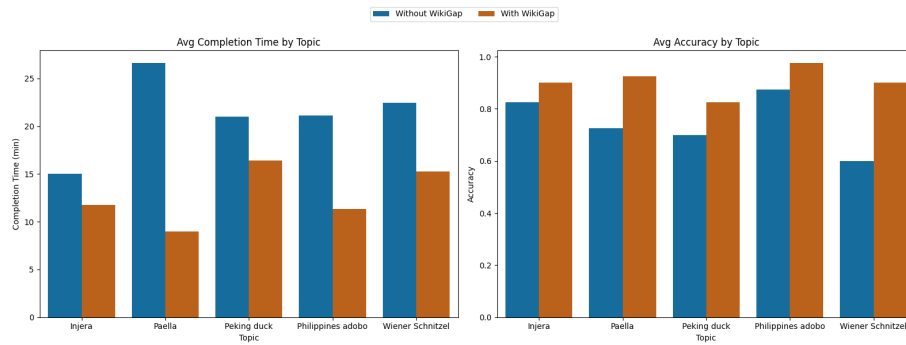## C   Average Completion Time and Accuracy across Different Topics

Fig. 8. Side-by-side bar charts comparing **average completion time** and **accuracy** across different topics, with and without the WikiGap extension. Blue represents the control condition (no WikiGap), and orange represent the treatment condition (with WikiGap).