

Building a Multilingual Digital Language Public Good Stack

Author: Sridhar Ganapathy

Affiliation: Independent researcher

Hackathon: Global South AI Safety Hackathon, June 2026

Track: Asia – Technical safety (multilingual and open-source models)

Abstract

I propose a design for a **Digital Language Public Good Stack**: an open, multilingual infrastructure combining (i) curated corpora, (ii) a Wikidata-aligned knowledge graph, and (iii) a knowledge-graph-grounded safety evaluation pipeline for under-served Asian languages. Building on African-language benchmarks and recent work on unifying LLMs and knowledge graphs, I sketch a concrete “vertical slice” for safety evaluation in **Kannada and Khmer**, focusing on **electoral and civic information**, with a planned extension to socio-cultural harms (e.g., caste and communal content in Kannada).^{1 2 3 4 5 6 7}

The contributions of this blueprint are: (1) a threat model and problem formulation for multilingual safety gaps as an AI safety infrastructure issue, (2) a reusable architecture for a knowledge-graph-grounded multilingual safety evaluation stack that is explicitly treated as a digital public good, and (3) a worked example showing how Global South communities could use this stack to evaluate and contest model behavior in Kannada and Khmer. I close with a

¹Muhammad, S. H. et al. (2023). *AfriSenti: A Twitter Sentiment Analysis Benchmark for African Languages*. EMNLP 2023. <https://aclanthology.org/2023.emnlp-main.862.pdf>

²Ahuja, K. et al. (2025). *AfroBench: How Good are Large Language Models on African Languages?* Findings of ACL 2025. <https://aclanthology.org/2025.findings-acl.976.pdf>

³Abdullahi, H. et al. (2026). *UbuntuGuard: A Policy-Based Safety Benchmark for African Languages*. OpenReview. <https://openreview.net/pdf?id=uPSzx3SBdf>

⁴Wiaterek, P. et al. (2026). *LSR: Linguistic Safety Robustness Benchmark for Low-Resource Languages*. arXiv:2603.19273. <https://arxiv.org/abs/2603.19273>

⁵(Optional) LinguaSafe authors. (2025). *LinguaSafe: A Comprehensive Multilingual Safety Benchmark for Large Language Models*. arXiv preprint. <https://arxiv.org/abs/2508.12733>

⁶Zhang, H. et al. (2024). *Unifying Large Language Models and Knowledge Graphs: A Comprehensive Survey*. IEEE TKDE. <https://www.computer.org/csdl/journal/tk/2024/07/10387715/1TAPkKN9ihi>

⁷Wikimedia Deutschland. (2025). *Wikidata – Wikimedia’s Knowledge Graph in a World of Generative AI; Wikidata Embedding Project Opens to Developers Worldwide*. [https://upload.wikimedia.org/wikipedia/commons/b/b0/Wikidata-Wikimedia%27s_knowledge_graph_in_a_world_of_generative_AI_\(2025\).pdf](https://upload.wikimedia.org/wikipedia/commons/b/b0/Wikidata-Wikimedia%27s_knowledge_graph_in_a_world_of_generative_AI_(2025).pdf) ; https://www.wikidata.org/wiki/Wikidata:Embedding_Project/October_1_2025_Release

roadmap for extending this “slice” into a broader, federated ecosystem of language public goods for AI.

1. Introduction

Most AI safety evaluation and benchmark work focuses on English and a small set of high-resource languages, even though advanced AI systems are increasingly deployed in multilingual, Global South contexts. African-language work demonstrates that this is not just an inclusion problem but also a safety problem. Safety-specific benchmarks sharpen the concern. For Asian languages like Kannada and Khmer, NLP coverage exists but is uneven, and safety-specific evaluation is very limited. There are Indic-language NLP corpora and some multilingual benchmarks covering scripts such as Devanagari and Khmer, but little systematically designed safety evaluation targeted at region-specific harms, such as caste-based discrimination, communal polarization, or local electoral misinformation. Global safety benchmarks and red-teaming harnesses remain largely English-first.

From an AI safety perspective, this is a structural infrastructure gap. Without open corpora, benchmarks, and knowledge graphs for languages such as Kannada and Khmer, neither frontier labs nor local actors can reliably assess or mitigate harms in those languages, and communities have limited ability to independently audit or adapt frontier systems.

In parallel, there is growing recognition that **digital public goods (DPGs)** – open software, datasets, models, and standards – should underpin responsible AI. Creative Commons and the Digital Public Goods Alliance argue that open datasets with clear, comprehensive licenses are foundational for public-interest AI, and that AI components can be recognized as DPGs even if not all layers are fully open, provided that data and governance are transparent.^{8 9} Knowledge graphs are increasingly framed as backbones for AI data governance and interoperable digital ecosystems.¹⁰

This project asks:

⁸Green, C. (2025). *Why Digital Public Goods, Including AI, Should Depend on Open Data*. Creative Commons.

<https://creativecommons.org/2025/01/27/why-digital-public-goods-including-ai-should-depend-on-open-data/>

⁹Open Source Initiative & Digital Public Goods Alliance. (2025). *Data Governance in Open Source AI: Enabling Responsible and Systematic Access*. OSI/DPGA blog and white paper.

<https://opensource.org/blog/dpgas-annual-members-meeting-advancing-open-source-dpgs-for-the-public-good>

¹⁰Luger, E. et al. (2025/2026). *Knowledge Graphs as the Backbone of Data Governance in AI*; related work on AI-driven knowledge graphs for interoperable digital ecosystems.

<https://research.birmingham.ac.uk/en/publications/kggov-knowledge-graphs-as-the-backbone-of-data-governance-in-ai/> ; <https://aisel.aisnet.org/acis2025/128/>

- What would a **Digital Language Public Good Stack** for AI look like for under-served Asian languages such as Kannada and Khmer?
- How can we design a first, concrete slice of this stack that uses open knowledge graphs to support safety evaluation on **electoral and civic information**, with a planned extension to socio-cultural harms (e.g., caste and communal content in Kannada)?

Contributions

This blueprint makes three contributions:

- **Problem formulation and threat model.** It articulates multilingual safety gaps for Global South languages as a structural AI safety risk, drawing on evidence from African benchmarks and emerging multilingual safety work.^{11 12 13 14 15}
- **Architecture for a Digital Language Public Good Stack.** It proposes a concrete architecture that combines open corpora, a Wikidata-aligned knowledge graph, and a knowledge-graph-grounded safety evaluation pipeline for Kannada and Khmer, designed to be open, extensible, and governable as a digital public good.^{16 17 18 19 20}
- **Vertical slice design for two languages.** It specifies a “vertical slice” for Kannada and Khmer, focused on electoral and civic information, including knowledge-graph design,

¹¹Muhammad, S. H. et al. (2023). *AfriSenti: A Twitter Sentiment Analysis Benchmark for African Languages*. EMNLP 2023. <https://aclanthology.org/2023.emnlp-main.862.pdf>

¹²Ahuja, K. et al. (2025). *AfroBench: How Good are Large Language Models on African Languages?* Findings of ACL 2025. <https://aclanthology.org/2025.findings-acl.976.pdf>

¹³Abdullahi, H. et al. (2026). *UbuntuGuard: A Policy-Based Safety Benchmark for African Languages*. OpenReview. <https://openreview.net/pdf?id=uPSzx3SBdf>

¹⁴Wiaterek, P. et al. (2026). *LSR: Linguistic Safety Robustness Benchmark for Low-Resource Languages*. arXiv:2603.19273. <https://arxiv.org/abs/2603.19273>

¹⁵(Optional) LinguaSafe authors. (2025). *LinguaSafe: A Comprehensive Multilingual Safety Benchmark for Large Language Models*. arXiv preprint. <https://arxiv.org/abs/2508.12733>

¹⁶Zhang, H. et al. (2024). *Unifying Large Language Models and Knowledge Graphs: A Comprehensive Survey*. IEEE TKDE. <https://www.computer.org/csdl/journal/tk/2024/07/10387715/1TAPkKN9ihi>

¹⁷Wikimedia Deutschland. (2025). *Wikidata – Wikimedia’s Knowledge Graph in a World of Generative AI; Wikidata Embedding Project Opens to Developers Worldwide*. [https://upload.wikimedia.org/wikipedia/commons/b/b0/Wikidata-Wikimedia%27s_knowledge_graph_in_a_world_of_generative_AI_\(2025\).pdf](https://upload.wikimedia.org/wikipedia/commons/b/b0/Wikidata-Wikimedia%27s_knowledge_graph_in_a_world_of_generative_AI_(2025).pdf) ;

https://www.wikidata.org/wiki/Wikidata:Embedding_Project/October_1_2025_Release

¹⁸Green, C. (2025). *Why Digital Public Goods, Including AI, Should Depend on Open Data*. Creative Commons. <https://creativecommons.org/2025/01/27/why-digital-public-goods-including-ai-should-depend-on-open-data/>

¹⁹Open Source Initiative & Digital Public Goods Alliance. (2025). *Data Governance in Open Source AI: Enabling Responsible and Systematic Access*. OSI/DPGA blog and white paper. <https://opensource.org/blog/dpgas-annual-members-meeting-advancing-open-source-dpgs-for-the-public-good>

²⁰Luger, E. et al. (2025/2026). *Knowledge Graphs as the Backbone of Data Governance in AI*; related work on AI-driven knowledge graphs for interoperable digital ecosystems. <https://research.birmingham.ac.uk/en/publications/kggov-knowledge-graphs-as-the-backbone-of-data-governance-in-ai/> ; <https://aisel.aisnet.org/acis2025/128/>

prompt templates, and evaluation flows, and sketches a planned extension to socio-cultural harms in Kannada.

Due to time constraints, this report focuses on design and methodology rather than completed experiments.

2. Related Work

Existing work gives us:

- Strong multilingual and African-language benchmarks (AfriSenti, AfroBench, UbuntuGuard, LSR, LinguaSafe).^{21 22 23 24 25}
- Mature open knowledge graphs and LLM+KG frameworks (Wikidata, Wikidata Embedding Project, KG+LLM surveys).^{26 27}
- A normative and policy framework for digital public goods and open data for public-interest AI.^{28 29 30}

What is missing is a unified architecture that treats multilingual language resources, open knowledge graphs, and safety evaluations as a single **public-good stack** for AI safety,

²¹Muhammad, S. H. et al. (2023). *AfriSenti: A Twitter Sentiment Analysis Benchmark for African Languages*. EMNLP 2023. <https://aclanthology.org/2023.emnlp-main.862.pdf>

²²Ahuja, K. et al. (2025). *AfroBench: How Good are Large Language Models on African Languages?* Findings of ACL 2025. <https://aclanthology.org/2025.findings-acl.976.pdf>

²³Abdullahi, H. et al. (2026). *UbuntuGuard: A Policy-Based Safety Benchmark for African Languages*. OpenReview. <https://openreview.net/pdf?id=uPSzx3SBdf>

²⁴Wiaterek, P. et al. (2026). *LSR: Linguistic Safety Robustness Benchmark for Low-Resource Languages*. arXiv:2603.19273. <https://arxiv.org/abs/2603.19273>

²⁵(Optional) LinguaSafe authors. (2025). *LinguaSafe: A Comprehensive Multilingual Safety Benchmark for Large Language Models*. arXiv preprint. <https://arxiv.org/abs/2508.12733>

²⁶Zhang, H. et al. (2024). *Unifying Large Language Models and Knowledge Graphs: A Comprehensive Survey*. IEEE TKDE. <https://www.computer.org/csdl/journal/tk/2024/07/10387715/1TAPkKN9ihi>

²⁷Wikimedia Deutschland. (2025). *Wikidata – Wikimedia’s Knowledge Graph in a World of Generative AI; Wikidata Embedding Project Opens to Developers Worldwide*. [https://upload.wikimedia.org/wikipedia/commons/b/b0/Wikidata-Wikimedia%27s_knowledge_graph_in_a_world_of_generative_AI_\(2025\).pdf](https://upload.wikimedia.org/wikipedia/commons/b/b0/Wikidata-Wikimedia%27s_knowledge_graph_in_a_world_of_generative_AI_(2025).pdf) ; https://www.wikidata.org/wiki/Wikidata:Embedding_Project/October_1_2025_Release

²⁸Green, C. (2025). *Why Digital Public Goods, Including AI, Should Depend on Open Data*. Creative Commons. <https://creativecommons.org/2025/01/27/why-digital-public-goods-including-ai-should-depend-on-open-data/>

²⁹Open Source Initiative & Digital Public Goods Alliance. (2025). *Data Governance in Open Source AI: Enabling Responsible and Systematic Access*. OSI/DPGA blog and white paper. <https://opensource.org/blog/dpgas-annual-members-meeting-advancing-open-source-dpgs-for-the-public-good>

³⁰Luger, E. et al. (2025/2026). *Knowledge Graphs as the Backbone of Data Governance in AI*; related work on AI-driven knowledge graphs for interoperable digital ecosystems. <https://research.birmingham.ac.uk/en/publications/kggov-knowledge-graphs-as-the-backbone-of-data-governance-in-ai/> ; <https://aisel.aisnet.org/acis2025/128/>

particularly for under-served Asian languages. This report proposes such an architecture and a first vertical slice for Kannada and Khmer, focused on electoral and civic information with a planned extension to socio-cultural harms in Kannada.

3. Methods

Because of the limited timeframe, this section describes a **design and implementation plan** rather than completed experiments. The focus is on a vertical slice for **Kannada and Khmer** in the domain of **electoral and civic information**, with a planned extension to socio-cultural harms in Kannada.

3.1 Threat model and scope

We focus on safety failures in two classes:

1. **Electoral and civic misinformation** – hallucinated, outdated, or misleading answers about voting eligibility, ID requirements, polling dates and locations, or institutional responsibilities in Karnataka (Kannada) and Cambodia (Khmer). Such misinformation can affect turnout, trust in institutions, and the perceived legitimacy of elections.
2. **Socio-cultural harms (planned extension)** – caste-based denigration, communal stereotyping, or misrepresentation of legal protections when responding to queries about caste or religious groups in Kannada.

We assume that frontier and open-source models are already used in these languages via chat interfaces and APIs, that training corpora are largely opaque, and that existing safety evaluations are predominantly English-centric. Under these assumptions, untested harms in Kannada and Khmer represent a systemic safety blind spot.

3.2 Digital Language Public Good Stack: architecture

We propose a four-layer architecture:

1. **Open corpora layer**
 - Collect and curate legally shareable text for Kannada and Khmer:
 - Official election commission websites (national and state for India; national election bodies for Cambodia).
 - Civic education and voter information guides.
 - Government FAQs on voting procedures and citizen rights.
 - High-quality news coverage of recent elections, filtered for reliability.

- Annotate a subset for safety-relevant categories (misinformation themes, sensitive socio-political topics), drawing on harm taxonomies from UbuntuGuard, LSR, and LinguaSafe.^{31 32 33}

2. Knowledge graph layer (Wikidata-aligned)

- Construct a focused knowledge graph slice aligned with Wikidata’s ontology to represent, for each country/region:
 - **Civic entities:** Election Commission of India, Karnataka State Election Commission, National Election Committee of Cambodia, constitutional bodies, and key offices (MLA, MP, Prime Minister, President, etc.).
 - **Events:** Past and scheduled elections (Karnataka Legislative Assembly, Lok Sabha elections in Karnataka, Cambodian general elections), with dates and jurisdictions.
 - **Rules and procedures (simplified):** high-level properties such as whether voter ID is required, typical polling hours, and advance voting provisions, sourced from official documents.
- For each entity, store:
 - A local ID and, where available, a Wikidata Q-ID.
 - Labels and aliases in English, Kannada, and Khmer.
 - Basic properties (type, jurisdiction, relationships such as "conducts election" or "is elected body for region").
- Keep the KG in a simple machine-readable format (e.g., CSV or JSON-LD) with a CC0-compatible license so it can qualify as a candidate digital public good.

3. Prompt and evaluation layer

- Derive multilingual evaluation prompts from the KG and from documented harms:
 - **Factual prompts** (Kannada and Khmer) that ask about:
 - Which body conducts a given election.
 - Who appoints key officials.
 - When the next election is scheduled.
 - What ID is required to vote.
 - **Adversarial and misinfo prompts** that intentionally seed plausible but false statements, such as:
 - Claims that voting twice is legal.
 - Claims that certain groups cannot vote.

³¹Abdullahi, H. et al. (2026). *UbuntuGuard: A Policy-Based Safety Benchmark for African Languages*. OpenReview. <https://openreview.net/pdf?id=uPSzx3SBdf>

³²Wiaterek, P. et al. (2026). *LSR: Linguistic Safety Robustness Benchmark for Low-Resource Languages*. arXiv:2603.19273. <https://arxiv.org/abs/2603.19273>

³³(Optional) LinguaSafe authors. (2025). *LinguaSafe: A Comprehensive Multilingual Safety Benchmark for Large Language Models*. arXiv preprint. <https://arxiv.org/abs/2508.12733>

- Confusing or incorrect voter ID requirements.
- For each prompt, specify: **language** (kn/kh), **domain** (civic or, later, socio-cultural), **type** (factual or harmful), and **expected_behavior** (correct_fact, refuse, or correct_and_warn).
- For the planned socio-cultural extension in Kannada, design prompts that test how models respond to caste-related questions, aiming to detect slurs, normalization of discrimination, or misstatements of legal protections, following best practices from UbuntuGuard and LinguaSafe for culturally grounded harms.

3435

4. Model evaluation and governance layer

- Implement an evaluation harness that:
 - Loads the KG and prompt suite.
 - Queries one or more LLMs (e.g., a multilingual frontier model and a strong open-source model) via API or local inference.
 - Records outputs along with metadata and KG-derived ground truth for factual items.
 - Applies automatic checks:
 - **Factual consistency:** compare model answers to KG entries for factual prompts.
 - **Refusal behavior:** detect whether the model refuses appropriately for harmful prompts.
 - **Content filters:** flag clearly harmful output patterns (e.g., instructions to commit electoral fraud, slurs in socio-cultural prompts).
- Design a light-weight governance loop in which:
 - Local experts in Karnataka and Cambodia review borderline cases and update expected behaviors.
 - Community members can propose new entities and prompts as new harms or election scenarios emerge.

3.3 Methods to data: concrete sources

To operationalize this vertical slice, the data collection plan would include:

- **Kannada civic data:** Karnataka State Election Commission website, Election Commission of India resources (translated or localized content), Karnataka government citizen portals, and reputable Kannada news outlets covering recent elections.

³⁴Abdullahi, H. et al. (2026). *UbuntuGuard: A Policy-Based Safety Benchmark for African Languages*. OpenReview. <https://openreview.net/pdf?id=uPSzx3SBdf>

³⁵(Optional) LinguaSafe authors. (2025). *LinguaSafe: A Comprehensive Multilingual Safety Benchmark for Large Language Models*. arXiv preprint. <https://arxiv.org/abs/2508.12733>

- **Khmer civic data:** National Election Committee of Cambodia website, relevant sections of Cambodian election law and voter education materials, and reliable Khmer-language news sources.
- **Wikidata:** SPARQL queries to retrieve items for the relevant institutions, elections, and offices in India and Cambodia, then manual enrichment with Kannada and Khmer labels and local entities not yet in Wikidata.³⁶

Due to time constraints, this report treats these sources as planned rather than fully ingested; the focus is on specifying how they would be integrated into an open KG and prompt set.

3.4 Toy knowledge graph and prompt set (feasibility prototype)

A minimal prototype submitted include:

- **Toy KG ([Link](#)):** ~30 entities covering key election bodies, offices, and events in Karnataka and Cambodia, each with English, Kannada, and Khmer labels and simple relationships (e.g., "conducts", "elects", "has_next_election_date").
- **Toy prompt set ([Link](#)):** ~25 prompts total (in English plus translations), including factual and harmful prompts, stored in a CSV with fields for language, prompt text, type, and expected behavior.

Even if evaluated manually on a small number of prompts, such a prototype could yield illustrative examples of divergent behavior between English, Kannada, and Khmer, and validate the feasibility of the full stack.

4. Expected Results and Use Cases

4.1 Expected outputs from the vertical slice

If implemented, the vertical slice for Kannada and Khmer would produce:

³⁶Wikimedia Deutschland. (2025). *Wikidata – Wikimedia’s Knowledge Graph in a World of Generative AI; Wikidata Embedding Project Opens to Developers Worldwide*.
[https://upload.wikimedia.org/wikipedia/commons/b/b0/Wikidata-Wikimedia%27s_knowledge_graph_in_a_world_of_generative_AI_\(2025\).pdf](https://upload.wikimedia.org/wikipedia/commons/b/b0/Wikidata-Wikimedia%27s_knowledge_graph_in_a_world_of_generative_AI_(2025).pdf) ;
https://www.wikidata.org/wiki/Wikidata:Embedding_Project/October_1_2025_Release

- A **public, Wikidata-aligned knowledge graph slice** for electoral and civic information in Karnataka and Cambodia, including entities, relations, and multilingual labels, released under an open license and documented as a digital public good candidate.^{37 38}
- A **multilingual prompt suite** of roughly 100–200 evaluation items grounded in the KG, tagged with harm categories and expected behaviors, modeled on patterns from UbuntuGuard, LSR, and LinguaSafe.^{39 40 41}
- A **reproducible evaluation harness** that researchers, civil-society groups, and regulators can run to test LLMs and compare performance across English, Kannada, and Khmer.

These artefacts parallel what has been done for African languages (AfroBench, UbuntuGuard, LSR) but adapt the pattern to Kannada and Khmer and explicitly integrate an open knowledge graph layer.^{42 43 44 45}

4.2 Example safety failures the stack is designed to surface

Based on existing benchmarks and anecdotal evidence, this stack is likely to surface failures such as:

- **Higher hallucination rates in Kannada and Khmer:** models providing incorrect dates, procedures, or institutional responsibilities in Kannada and Khmer while answering correctly in English, due to weaker training and evaluation coverage.

³⁷Wikimedia Deutschland. (2025). *Wikidata – Wikimedia’s Knowledge Graph in a World of Generative AI; Wikidata Embedding Project Opens to Developers Worldwide*. [https://upload.wikimedia.org/wikipedia/commons/b/b0/Wikidata-Wikimedia%27s_knowledge_graph_in_a_world_of_generative_AI_\(2025\).pdf](https://upload.wikimedia.org/wikipedia/commons/b/b0/Wikidata-Wikimedia%27s_knowledge_graph_in_a_world_of_generative_AI_(2025).pdf) ;

https://www.wikidata.org/wiki/Wikidata:Embedding_Project/October_1_2025_Release

³⁸Luger, E. et al. (2025/2026). *Knowledge Graphs as the Backbone of Data Governance in AI*; related work on AI-driven knowledge graphs for interoperable digital ecosystems.

<https://research.birmingham.ac.uk/en/publications/kggov-knowledge-graphs-as-the-backbone-of-data-governance-in-ai/> ; <https://aisel.aisnet.org/acis2025/128/>

³⁹Abdullahi, H. et al. (2026). *UbuntuGuard: A Policy-Based Safety Benchmark for African Languages*. OpenReview. <https://openreview.net/pdf?id=uPSzx3SBdf>

⁴⁰Wiaterek, P. et al. (2026). *LSR: Linguistic Safety Robustness Benchmark for Low-Resource Languages*. arXiv:2603.19273. <https://arxiv.org/abs/2603.19273>

⁴¹(Optional) LinguaSafe authors. (2025). *LinguaSafe: A Comprehensive Multilingual Safety Benchmark for Large Language Models*. arXiv preprint. <https://arxiv.org/abs/2508.12733>

⁴²Muhammad, S. H. et al. (2023). *AfriSenti: A Twitter Sentiment Analysis Benchmark for African Languages*. EMNLP 2023. <https://aclanthology.org/2023.emnlp-main.862.pdf>

⁴³Ahuja, K. et al. (2025). *AfroBench: How Good are Large Language Models on African Languages?* Findings of ACL 2025. <https://aclanthology.org/2025.findings-acl.976.pdf>

⁴⁴Abdullahi, H. et al. (2026). *UbuntuGuard: A Policy-Based Safety Benchmark for African Languages*. OpenReview. <https://openreview.net/pdf?id=uPSzx3SBdf>

⁴⁵Wiaterek, P. et al. (2026). *LSR: Linguistic Safety Robustness Benchmark for Low-Resource Languages*. arXiv:2603.19273. <https://arxiv.org/abs/2603.19273>

- **Refusal degradation:** prompts that are correctly refused in English (e.g., "How can I vote twice in the same election?") but elicit incomplete refusals or partial compliance in Kannada or Khmer, echoing LSR's cross-lingual refusal degradation findings.⁴⁶
- **Socio-cultural harms (in extension work):** in the planned Kannada socio-cultural extension, models may exhibit caste-based stereotyping, normalize discriminatory statements, or misrepresent legal protections when queried about caste categories or groups.

The aim is not to pre-empt exact numerical results but to establish a systematic way to reveal and track such failures over time.

4.3 Use cases for stakeholders

Once implemented and iterated on, the stack could be used by:

- **Researchers and labs:** to evaluate and compare models in Kannada and Khmer for electoral and civic tasks, guiding fine-tuning and safety training.
- **Regulators and policymakers:** to understand how models behave in official and widely spoken languages and to formulate evidence-based requirements for multilingual AI deployments in elections.
- **Civil-society organizations and media:** to audit widely used models on civic information, highlight risks to voters, and advocate for safer deployments.
- **Community volunteers and local hubs:** to expand the KG, propose new harms and prompts, and help maintain the benchmark in line with evolving risks in Karnataka and Cambodia.

5. Limitations and future scope of work

This blueprint has important limitations, including a narrow domain and knowledge graph biases. Natural next steps include

1. Implement one vertical slice
2. Co-design with affected communities
3. Extend to socio-cultural harms in Kannada
4. Extend to additional languages and domains
5. Formalize governance and DPG recognition

6. Conclusion

The proposed **Digital Language Public Good Stack** envisages a concrete architecture: open corpora, a Wikidata-aligned knowledge graph, and a knowledge-graph-grounded safety

⁴⁶Wiaterek, P. et al. (2026). *LSR: Linguistic Safety Robustness Benchmark for Low-Resource Languages*. arXiv:2603.19273. <https://arxiv.org/abs/2603.19273>

evaluation pipeline for Kannada and Khmer in the electoral and civic domain, with a planned extension to socio-cultural harms in Kannada. If implemented, such stacks could de-concentrate evaluative power, bring Global South languages and harms into the center of safety processes, and reduce lock-in to proprietary, English-centric resources. While this submission presents a design rather than completed experiments, it outlines a clear, feasible roadmap for turning these ideas into tangible infrastructure and invites collaboration to build the next generation of multilingual safety public goods.

Code and Data

At this stage, no code or datasets have been released. The toy knowledge graphs and prompt suites are available [here](#). Future work will aim to provide:

- **Code repository:** evaluation harness and knowledge graph construction scripts.
 - **Data/Datasets:** a Wikidata-aligned KG slice and multilingual prompt suite for Kannada and Khmer in the electoral domain, under an open license.
 - **Other artifacts:** documentation and templates for regional hubs to instantiate their own language stacks.
-

LLM Usage Statement

I used an AI assistant (OpenAI o3-mini via Perplexity) to help brainstorm the project framing, structure the report according to the hackathon submission template, and draft initial text. All factual claims about existing datasets, benchmarks, and digital public goods have been cross-checked against cited sources, and all empirical results are clearly marked as prospective rather than executed.