Towards Open-Ended Discovery for Low-Resource NLP

Bonaventure F. P. Dossou[1,2,*], **Henri Aïdasso**[3,*]

[1]McGill University [2]Mila Quebec AI Institute, [3]École de technologie supérieure (ÉTS)
[*]Equal Contribution
bonaventure.dossou@mila.quebec, henri.aidasso@etsmtl.ca

Abstract

Natural Language Processing (NLP) for low-resource languages remains fundamentally constrained by the lack of textual corpora, standardized orthographies, and scalable annotation pipelines. While recent advances in large language models have improved cross-lingual transfer, they remain inaccessible to underrepresented communities due to their reliance on massive, pre-collected data and centralized infrastructure. In this position paper, we argue for a paradigm shift toward open-ended, interactive language discovery, where AI systems learn new languages dynamically through dialogue rather than static datasets. We contend that the future of language technology, particularly for low-resource and under-documented languages, must move beyond static data collection pipelines toward interactive, uncertainty-driven discovery, where learning emerges dynamically from human-machine collaboration instead of being limited to pre-existing datasets. We propose a framework grounded in joint human-machine uncertainty, combining epistemic uncertainty from the model with hesitation cues and confidence signals from human speakers to guide interaction, query selection, and memory retention. This paper is a call to action: we advocate a rethinking of how AI engages with human knowledge in under-documented languages, moving from extractive data collection toward participatory, co-adaptive learning processes that respect and empower communities while discovering and preserving the world's linguistic diversity. This vision aligns with principles of human-centered AI, emphasizing interactive, cooperative model building between AI systems and speakers.

1 Introduction

The recent progress in Natural Language Processing (NLP) has been largely shaped by a data-driven paradigm. Foundation models, built on large-scale internet corpora and empowered by scaling laws, have unlocked impressive generalization across tasks and languages (Kaplan et al., 2020; Brown et al., 2020; Scao et al., 2023). However, this trajectory has come at a cost: the assumption that performance improves with ever more data and compute has made cutting-edge research increasingly inaccessible, especially to researchers and communities in the Global South (Sambasivan et al., 2021; Schwartz et al., 2020).

Despite efforts to democratize NLP, a stark imbalance persists. African languages, which make up over 30% of the world's linguistic diversity, account for less than 1% of NLP research output (Joshi et al., 2020). These languages typically lack large-scale text corpora, parallel datasets, and standardized annotation practices. Transfer learning, active learning, self-supervised and semi-supervised learning, have all been proposed to address this data scarcity (Howard and Ruder, 2018; Devlin et al., 2019; Ein-Dor et al., 2020; Dossou et al., 2022; Dossou, 2025; Dossou et al., 2025), but even these methods depend on the availability of some unlabeled or previously seen language data. In environments where data is extremely scarce or non-digitized, such assumptions break down.

Moreover, while recent Large Language Models (LLMs) have demonstrated impressive cross-lingual abilities, their success is closely tied to data scale, computational resources, and increasingly centralized infrastructure. As scaling laws plateau and operational costs rise, the current paradigm risks becoming both *unsustainable* and *exclusive*, limiting participation from underrepresented communities and preventing scalable solutions for the languages that need them most (Strubell et al., 2019; Bender et al., 2021; Pouget et al., 2024).

We argue that NLP must now evolve beyond static, data-hungry training regimes. Inspired by recent work in open-ended discovery and self-improving AI (Hughes et al., 2024; Lu et al., 2025), we propose a shift toward *interactive, uncertainty-*

driven language learning. In our vision, AI systems learn languages not from vast corpora, but through *natural dialogue*, identifying gaps in their understanding, asking questions, and incorporating feedback in real time.

Imagine an AI system that only understands English, but receives a human input in Fon (Dossou and Emezue, 2020, 2021a; Dossou and Sabry, 2021; Dossou and Emezue, 2021b; Dossou et al., 2023). Instead of guessing or ignoring it, the system responds: ***"I do not recognize this language. Could you help me understand it?"*** *From this first exchange, it starts acquiring the new linguistic concepts interactively. Over time, through repeated exposure and correction, the system transitions from total ignorance to conversational fluency in the new language. This vision shifts the emphasis* ***from training on what we have to*** *learning from* ***what we do not yet understand****, as humans do.*

In this position paper, we explore the technical and conceptual foundations for such systems. We argue that *open-ended language learning*, grounded in *epistemic uncertainty*, *dialogue*, and *human-in-the-loop adaptation*, represents a scalable and inclusive path forward for low-resource NLP, especially in contexts where static data is not available, representative, or sufficient. We also outline a set of *open challenges* that arise from this vision, including the need for reliable uncertainty estimation, continuous learning mechanisms, and equitable access to interaction data. We discuss both the promise and the risks of this approach, including the question of whether such systems can acquire meaningful language competence without sufficient exposure or human feedback, and what architectures, incentives, or evaluation schemes would be required to support them.

2 Background and Related Work

2.1 Low-Resource Languages

Africa is one of the most linguistically diverse continents, home to over 3,000 indigenous languages (Epstein and Kole, 1998; Eberhard et al., 2024), which account for about one-third of the world's 7,159 living languages (Eberhard et al., 2024). In an increasingly digital world, where today's AI advancement such as LLMs offer unprecedented possibilities, the non-integration of these languages into the technological landscape not only exacerbates social inequalities but also poses a serious threat to the survival of entire linguistic cultures.

As inclusion and diversity gain global importance, commendable efforts have been made by researchers to identify available, albeit scarce, data sources (e.g., the Bible in Fon). Moreover, there are growing efforts for datasets creation (sometimes done manually and on a voluntary basis). These datasets have been used to create machine translation models that produce acceptable results (Dossou and Emezue, 2020; Adelani et al., 2022a). As a result, some of the very low-resource languages such as Fon, Ewe have been recently integrated into Google Translate,[1] for textual translations.

Despite these important advances, several major challenges persist that existing solutions do not, and arguably cannot address. In particular, current approaches still rely heavily on larger amounts of textual data (Adelani et al., 2022a; Dossou et al., 2022; Nekoto et al., 2020), resources that are extremely scarce or absent for many African languages and dialects (Nekoto et al., 2020; Joshi et al., 2020). Due to this reliance, existing solutions only cover a tiny fraction ($\approx$1%) of the languages, typically selected based on speaker population size or researchers' ties (Adelani et al., 2022a,b). These choices overlook the existing diversity and will ineluctably reinforce existing social inequalities and discrimination. For instance, Nigeria alone has over 500 indigenous languages (Eberhard et al., 2024), most of which severely lack written resources. Even more concerning is the practical impact of current solutions. In fact, most low-resource languages exist solely through oral traditions, meaning that the vast majority of native speakers can only speak them and struggle to read written versions, if such versions exist at all (Dossou and Emezue, 2021a; Olatunji et al., 2023b,a). Therefore, solutions that rely on textual translations are fundamentally misaligned with how these languages are actually used, making them ineffective for real-world communication needs.

2.2 Human Uncertainty Estimation

Incorporating human uncertainty into interactive learning frameworks has emerged as a critical complement to model uncertainty, as human feedback is often non-deterministic and can significantly shape model learning dynamics. Collins et al. (2023) explore concept-level interventions where humans provide feedback on intermediate concepts rather than final labels. They show that capturing the

[1] https://translate.google.com/?sl=en&tl=fon

confidence or uncertainty of these interventions, through soft labels or probabilistic feedback, improves model robustness and generalization.

Mendes et al. (2025) study the relationship between human-perceived and model-predicted uncertainties, finding only limited correlation between the two. This indicates that model uncertainty alone is insufficient to assess ambiguity in real-world settings. Explicitly modeling human uncertainty, for example, through elicited confidence scores or inter-annotator variance, can lead to more calibrated and reliable learning.

From a broader perspective, Bhatt et al. (2021) argue that exposing both human and model uncertainties enhances transparency and mutual understanding in human-AI collaboration. Similarly, collaborative annotation frameworks such as CoAnnotating (Li et al., 2023) leverage these uncertainty estimates to decide when to defer to human expertise or proceed autonomously, improving both efficiency and reliability in human-in-the-loop learning pipelines.

2.3 Model Uncertainty Estimation

In machine learning models, uncertainty estimation plays a crucial role in determining whether a model can respond confidently or should request clarification from the user. We denote by $f_\theta : \mathcal{X} \to \mathcal{Y}$ a parametric model with parameters θ, input $x \in \mathcal{X}$, and predictive distribution $p_\theta(y|x)$. $\mathcal{D}$ is the training dataset.

Kendall and Gal distinguish two types of uncertainty: aleatoric uncertainty (U_a) and epistemic uncertainty (U_e). Most literature works focus on U_e which is approximated by:

$$U_e(x) = \mathbb{V}_{p(\theta|\mathcal{D})}[\mathbb{E}_{p(y|x,\theta)}[y]]$$

This is U_e because directly tied to limited data or lack of model knowledge. The two most common ways of estimating U_e are the following:

With Bayesian Neural Networks BNNs (MacKay, 1992; Neal, 1996) define a posterior over weights:

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta),$$

and predictive uncertainty as:

$$p(y|x,\mathcal{D}) = \int p(y|x,\theta)p(\theta|\mathcal{D})\, d\theta.$$

In practice, this integral is untractable, and approximated using variational inference (Blundell et al., 2015) or Monte Carlo sampling (Gal and Ghahramani, 2016).

With Deep Ensembles Given M independently trained models $\{f_{\theta_m}\}_{m=1}^{M}$, predictive uncertainty is quantified via:

$$U_m(x) = \frac{1}{M}\sum_m H(f_{\theta_m}(y|x)),$$

where $H(\cdot)$ is Shannon entropy.

In summary, while advances in uncertainty estimation have improved model reliability (Kendall and Gal; Gal and Ghahramani, 2016; Kirsch et al., 2019) and recent work has explored uncertainty in human feedback (Collins et al., 2023; Mendes et al., 2025), current AI systems still learn predominantly from static datasets or treat user input as deterministic corrections. This creates two major limitations: (i) uncertainty from humans and models is rarely considered jointly, reducing the system's ability to assess when to seek clarification or defer decisions, and (ii) learning processes remain largely offline, without mechanisms to dynamically adapt to evolving user input.

To address these shortcomings, we introduce an **interactive learning system** that moves beyond passive, data-driven training. Instead of relying solely on pre-collected corpora, the system engages directly with users, identifies gaps in understanding, requests clarification when uncertainty is high, and incorporates feedback into its evolving knowledge state. This approach aims to fuse human and model uncertainties to guide the dialogue flow, enabling real-time, adaptive, and more sample-efficient language acquisition.

3 Proposed Approach

Our proposed framework enables AI systems to acquire language competence through open-ended, interactive learning. This process is illustrated in Figure 1 through interactions between a human and the AI system (agent). Rather than training on large static corpora, the system learns by engaging with users in real time, identifying gaps in its knowledge, soliciting clarification, and integrating feedback. The methodology consists of three core components: (1) modeling interactional uncertainty, (2) language acquisition via feedback, and (3) continual learning from dialogic exposure.

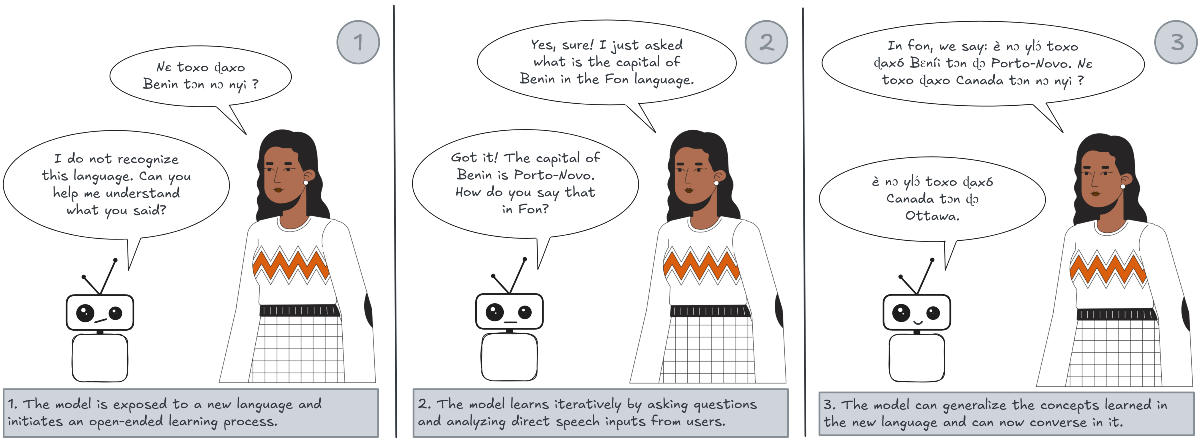
1
Nɛ toxo ɖaxo Benin tɔn nɔ nyi ?
I do not recognize this language. Can you help me understand what you said?
1. The model is exposed to a new language and initiates an open-ended learning process.
2
Yes, sure! I just asked what is the capital of Benin in the Fon language.
Got it! The capital of Benin is Porto-Novo. How do you say that in Fon?
2. The model learns iteratively by asking questions and analyzing direct speech inputs from users.
3
In fon, we say: è nɔ ylɔ́ toxo ɖaxɔ́ Bɛníì tɔn ɖɔ Porto-Novo. Nɛ toxo ɖaxo Canada tɔn nɔ nyi ?
è nɔ ylɔ́ toxo ɖaxɔ́ Canada tɔn ɖɔ Ottawa.
3. The model can generalize the concepts learned in the new language and can now converse in it.

Figure 1: Illustration of the proposed approach for open-ended learning of low-resource languages. It shows the voice conversation between a human and an agent who teaches the agent to recognize and respond to requests for the capital city of a country in the Fon language.

3.1 Modeling Interactional Uncertainty

At the heart of our approach is the notion of epistemic uncertainty, which refers to the system's awareness of what it does not know. In conventional NLP, model uncertainty is often used for tasks like active learning or confidence calibration (Kendall and Gal; Gal and Ghahramani, 2016; Houlsby et al., 2011; Guo et al., 2017). Here, we extend this principle to guide decision-making during interactive language learning.

We define a composite uncertainty signal combining both human and machine contributions:

$$\mathcal{U}_{\text{total}} = \alpha \cdot \mathcal{U}_{\text{human}} + (1 - \alpha) \cdot \mathcal{U}_{\text{model}}$$

where $\mathcal{U}_{\text{model}}$ is the model's epistemic uncertainty, estimated via entropy, ensemble disagreement, or Bayesian approximations (Kendall and Gal; Kirsch et al., 2019; Gal et al., 2017; Gal and Ghahramani, 2016; Kirsch et al., 2023), $\mathcal{U}_{\text{human}}$ reflects uncertainty inferred from hesitation cues, conflicting corrections, or prosodic markers, and α controls the relative influence of human versus machine uncertainty.

Given this signal, the system selects a query Q^* to ask the human speaker, optimizing:

$$Q^* = \arg\max_{Q} \frac{\mathbb{E}[\text{InfoGain}(Q)]}{\text{Cost}(Q, \mathcal{U}_{\text{human}})}$$

where

$$\text{Cost}(Q, \mathcal{U}_{\text{human}}) = c(Q)\big(1 + \lambda * \mathcal{U}_{\text{human}}\big)$$

with $c(Q)$ representing the baseline time or cognitive effort required for query type Q, and $\lambda \geq 0$ controlling how strongly human uncertainty increases perceived cost. This interaction cost reflects the human effort required to answer a query and the likelihood of confusion when the speaker is already uncertain. Scaling the cost by $(1 + \lambda * \mathcal{U}_{\text{human}})$ ensures the system avoids queries that are both expensive and likely to yield ambiguous responses. This improves efficiency and user experience, making learning cooperative rather than extractive.

The expected information gain from a query Q is defined as the anticipated reduction in predictive uncertainty:

$$\begin{aligned}\text{InfoGain}(Q) = {} & \mathbb{H}[Y \mid x, \mathcal{D}] \\ & - \mathbb{E}_{A \sim p(A|Q)}\big[\mathbb{H}[Y \mid x, \mathcal{D}, Q, A]\big]\end{aligned}$$

where $\mathbb{H}[\cdot]$ denotes Shannon entropy, $\mathcal{D}$ is the current learner state, and A denotes a human response sampled from $p(A \mid Q)$. This term quantifies how much uncertainty the query is expected to resolve. We define $p(A|Q)$ as the conditional distribution over possible human responses given a query Q. This distribution models the variability and uncertainty in human feedback due to ambiguity in meaning, hesitation or noise in responses, and contextual variability across speakers.

The selected query Q^* and the anticipated distribution of human responses $p(A|Q)$ provide the necessary context for the next stage, where human feedback is integrated into the model.

3.2 Language Acquisition via Human Feedback

Once a query Q^* has been selected based on the joint uncertainty signal, the AI system receives a feedback signal A from the human speaker. In this stage, the goal is to integrate the new information into the model's knowledge while accounting for both human and model uncertainty.

A targeted query Q is designed to elicit clarifying information about input x, such as asking ***"What does this word mean?"*** or ***"How would you say this sentence?"***. The response is denoted as $A \sim p(A|Q)$, sampled from a conditional distribution over possible answers. This distribution reflects that feedback may vary or include ambiguity, such as multiple possible translations or uncertain corrections.

We denote $p_\theta(\cdot|x)$ as the model's current predictive distribution over possible meanings or utterances for input x, parameterized by θ. The human feedback is represented as y_{human}, a meaning distribution derived from the response A. It can be sharp, corresponding to a single unambiguous answer, or soft, capturing several plausible meanings with associated probabilities. Finally, we introduce a reliability weight $w_f = 1 - \mathcal{U}_{\text{human}}$, which downscales the influence of uncertain human feedback. When human uncertainty is high, the system places less emphasis on the feedback to avoid reinforcing potentially misleading signals.

Using these definitions, the system constructs a new target distribution that combines its own prior predictions with the received feedback:

$$\tilde{y} = w_f \cdot y_{\text{human}} + (1 - w_f) \cdot p_\theta(\cdot|x)$$

This weighted target guides the parameter update:

$$\theta' = \theta - \eta \nabla_\theta \mathcal{L}(p_\theta(\cdot|x), \tilde{y})$$

where η is the learning rate and $\mathcal{L}$ is a loss function. KL Divergence can be used to align the model's predicted distribution with human-provided meaning probabilities in a continuous space, making it well-suited for uncertain or soft feedback. Contrastive Loss distinguishes correct meanings from alternative ones in an embedding space, supporting open-ended discovery where meanings are not predefined. Categorical Cross-Entropy works when the system has a finite set of candidate meanings, though it is less ideal for open-ended language learning since it assumes predefined categories.

This approach allows the system to integrate human feedback incrementally and proportionally to its reliability, while still preserving useful prior knowledge from its own predictions. In the future, more appropriate loss functions could be designed specifically for dialogic, open-ended learning scenarios to better reflect the uncertainty and flexibility inherent in human language interactions.

3.3 Continual Learning from Dialogic Exposure

Language acquisition is not a single-step process. Over multiple interactions, the system must consolidate knowledge, refine uncertain examples, and adapt to evolving feedback. To achieve this, every interaction is stored in a memory bank:

$$\mathcal{M} = \{(x_i, A_i, w_i)\}$$

where each element consists of the input x_i, the human feedback A_i, and an associated weight:

$$w_i = (1 - \mathcal{U}^{(i)}_{\text{human}})(1 - \mathcal{U}^{(i)}_{\text{model}})$$

This weight captures the combined confidence of both the human and the model for a given interaction.

The memory bank $\mathcal{M}$ acts as a growing repository of past interactions with human speakers, each stored alongside a weight indicating reliability. Periodically, the system revisits stored samples to reinforce reliable information and re-query ambiguous examples. Past interactions are used to improve the model through uncertainty-aware gradient updates:

$$\theta \leftarrow \theta - \eta \sum_i w_i \nabla_\theta \mathcal{L}(p_\theta(\cdot|x_i), A_i)$$

$p_\theta(\cdot|x_i)$ refers to the same predictive model introduced in Section 3.2, now updated iteratively using both immediate feedback and stored memory samples. We reuse the notation to emphasize that the model evolves over time through repeated uncertainty-guided interactions.

Low-weight samples contribute less to the update, preventing uncertain or noisy feedback from degrading the learned representation. They are not discarded but flagged for future re-querying when opportunities arise. This creates a closed interactive loop where the system encounters new input x, computes $\mathcal{U}_{\text{total}}$ and selects an optimal query Q^*, collects human feedback A and updates parameters incrementally, stores the interaction in $\mathcal{M}$ with weight w_i, and periodically revisits uncertain cases to refine or validate earlier knowledge, looping back when necessary.

Through these mechanisms, uncertainty evolves from a static confidence score into an active principle governing when to trust, query, defer, or memorize. This continual process ensures that learning is incremental, reliable, and co-adaptive. It enables the system to refine its internal representations over time, progressively improving its understanding of a new language while remaining sensitive to the reliability of past and future feedback. Together, these three stages establish a self-reinforcing loop for interactive language discovery, where uncertainty not only shapes individual interactions but also drives long-term, co-adaptive learning.

4 Opportunities and Challenges

Our proposed framework for open-ended language discovery leverages joint human-machine uncertainty to guide interaction, query selection, and memory retention. While the approach introduces a novel paradigm for low-resource language acquisition, its success and limitations stem directly from the mechanisms we designed. Unlike conventional NLP pipelines that rely on static, curated datasets and post-hoc analysis, this framework is designed for real-time, adaptive interaction. It emphasizes uncertainty-driven decision-making, enabling language acquisition to progress even when large corpora, standardized orthographies, or expert annotators are unavailable.

4.1 Why This Could Work

The framework builds on several principles that make it uniquely suited for interactive, low-resource settings. By explicitly modeling epistemic uncertainty, the system learns what it does not know and can focus queries on areas of high information gain rather than engaging in blind memorization. This targeted querying mechanism has the potential to accelerate language acquisition compared to static corpus-based training approaches. Incorporating $\mathcal{U}_{\text{human}}$ allows the system to defer or prioritize information based on human confidence, ensuring that reliable feedback from fluent speakers directly shapes the learned representation and reduces noise in the earliest stages of learning. Over time, dynamic weighting (α) adapts reliance on each contributor according to their observed consistency and reliability, making the system robust to heterogeneous or occasional feedback. Furthermore, confidence-weighted memory retention enables iterative refinement of knowledge: high-certainty information consolidates quickly, while ambiguous examples remain open for re-querying, progressively building a stable and trustworthy knowledge base. Together, these mechanisms enable data-efficient learning that can bootstrap language understanding from a small number of high-value interactions, making it feasible in settings where large corpora are unavailable. These properties suggest that joint human–machine uncertainty could form the backbone of scalable, respectful, and data-efficient language acquisition, where conventional supervised NLP pipelines cannot operate.

4.1.1 In the Context of Low-Resource African Languages

Low-resource African languages often face a unique combination of challenges that make standard NLP pipelines ineffective: severe data scarcity, highly variable orthographies, oral traditions without standardized writing systems, and limited availability of expert annotators. The proposed framework is particularly well-suited to this context because it does not rely on pre-existing corpora or formal linguistic resources. Instead, it learns interactively from small, high-value exchanges, asking only those questions that are most informative given its current uncertainty. This targeted learning process minimizes the burden on speakers, who may have limited time or literacy in standardized orthography, while still allowing the system to rapidly form hypotheses about grammar, semantics, and phonology.

Moreover, the joint modeling of human and machine uncertainty makes the framework robust to the realities of field data collection in African settings, where contributors may have varying degrees of fluency, confidence, or even differing dialects of the same language. By adapting reliance on each contributor through dynamically learned weighting (α), the framework can filter noise while still capturing dialectal richness. Its ability to defer uncertain information and revisit ambiguous examples ensures that rare or culturally significant linguistic forms are not prematurely discarded. These properties make it a promising approach for preserving, documenting, and learning African languages where the cost of traditional data collection is prohibitive and where respectful, participatory collaboration with speakers is essential. This approach not only addresses data scarcity but also reframes language technology development as a collaborative process between AI systems and speakers. By moving away from extractive data collection toward live, adaptive interaction, it offers a pathway for NLP to support language documentation and revitalization efforts. Particularly in marginalized communities, this paradigm empowers speakers to co-create technology aligned with their linguistic and cultural realities, potentially reshaping how AI contributes to the preservation and expansion of global linguistic diversity.

4.1.2 In the Context of Human-Centered AI and Human-Computer Interactions

The proposed framework embodies principles of human-centered artificial intelligence by placing speakers at the center of the learning process. Rather than treating them as static annotators or sources of labels, it engages in a cooperative interaction where both human and machine uncertainty guide the flow of information exchange. This fosters transparency and trust, as speakers can see that the system acknowledges its own uncertainty, adapts to their confidence levels, and defers decisions when information is unclear.

From a Human-Computer Interaction (HCI) standpoint, the framework reduces the cognitive and emotional burden on contributors by focusing only on high-value, contextually relevant questions instead of overwhelming them with repetitive or trivial requests. It can adapt the pace and style of interaction based on hesitation cues, feedback latency, or non-verbal indicators of uncertainty, making it more accessible to non-expert participants. Additionally, the iterative refinement of memory ensures that early mistakes can be revisited and corrected collaboratively, giving speakers a sense of agency and ownership in shaping the emerging language model. This paradigm transforms data collection from a one-way, extractive process into a participatory dialogue, contributing to the development of AI systems that are not only technically effective but also socially aligned and respectful toward the communities they aim to serve. In doing so, it demonstrates a path toward genuinely human-centered AI, where computational methods adapt to people, rather than asking people to adapt to technology. This vision is aligned with participatory and co-design approaches explored in HCI research (Liao and Vaughan, 2023; Birhane et al., 2022; Delgado et al., 2023), which emphasize collaborative model building, transparency, and community agency in shaping AI behavior.

While these properties highlight the potential of our framework to enable scalable, and data-efficient language learning, realizing this vision in practice is far from trivial. Uncertainty-guided discovery introduces its own vulnerabilities, and deploying such systems in real-world low-resource environments presents additional technical and sociotechnical barriers, that must be addressed. The following section discusses these open challenges.

4.2 Challenges

Several challenges could undermine the effectiveness of the proposed framework in practice. A first concern lies in the reliability of uncertainty estimation. Because the system operates on highly out-of-distribution data such as new languages, unseen constructs, and unpredictable input patterns, its uncertainty signals may not be well calibrated. Miscalibration could lead to redundant or unnecessary queries, or conversely, to missed opportunities to acquire valuable information early on.

Human uncertainty signals introduce another layer of complexity. Hesitation cues, conflicting answers, or silence are not always reliable indicators of a speaker's true confidence. Cultural norms and individual communication styles can further distort these signals, leading the system to over-trust uncertain information or defer excessively even when a speaker would have provided correct input. This unreliability in feedback interpretation can propagate downstream errors in learning.

Errors may also arise in the adaptive weighting mechanism. Because α must be learned online from sparse observations, early interactions can dominate future weighting, allowing biases from the first few contributors to persist unchecked. In heterogeneous communities where speaker reliability varies widely, it becomes difficult to estimate contributor trustworthiness accurately, which risks amplifying noise and reducing the value of human input. This interacts closely with query selection: without stable reliability or cost estimates, the system may waste interactions on poorly chosen clarifications, frustrating users and slowing overall progress.

The memory component presents its own risks. Confidence-weighted retention is designed to consolidate reliable information quickly, but if misinterpreted feedback is assigned high confidence, early errors risk becoming fossilized in the learned representation. Conversely, rare linguistic forms may repeatedly receive low-confidence scores, preventing their integration and leaving parts of the language undocumented or misunderstood. This challenge is compounded in what we term a "double-uncertainty deadlock," where both the model and the human contributors remain uncertain for extended periods. In such cases, the system may repeatedly defer decisions, becoming overly cautious and failing to test hypotheses that could break the cycle of uncertainty.

Finally, practical constraints in real-world deployment cannot be ignored. Reliable uncertainty estimation, adaptive weighting, and dynamic query selection all introduce computational overhead that may be infeasible on low-cost, battery-limited, or offline devices. Connectivity issues, limited processing power, and fragile hardware environments could hinder the ability of the framework to operate effectively in the very settings it aims to serve.

4.3 Future Directions

Addressing these challenges requires progress on several fronts. Improving epistemic uncertainty estimation in open-ended, out-of-distribution language input is a priority, as more reliable measures would reduce unnecessary queries and strengthen the system's ability to make informative decisions early on. Equally important is the development of context-aware and culturally adaptive models of human uncertainty, since hesitation and confidence cues vary widely across individuals and communities. Advancing methods for learning α from sparse interactions will also be key to mitigating early biases, ensuring that the system adapts fairly and dynamically to multiple contributors over time.

Meta-learning approaches offer a promising path toward improving α estimation. By transferring priors on speaker reliability from related language acquisition sessions or typologically similar languages, the system could begin with more informed weighting strategies, reducing the risk of overfitting to a handful of early interactions. This would make adaptation faster and more stable, even in diverse or previously unseen linguistic settings.

Developing multi-agent exploration policies could further enhance query selection. Instead of treating human interactions in isolation, coordinated strategies could balance information gain, contributor reliability, and annotation cost across multiple speakers. Such strategies might deliberately diversify queries to capture rare linguistic forms, seek cross-validation from independent sources to resolve ambiguities, and avoid overloading single contributors, making learning more efficient and collaborative.

Breaking double-uncertainty deadlocks will require exploration mechanisms that take calculated risks when both human and model uncertainty remain high. Periodic re-querying, rediscovery routines, and targeted hypothesis testing could help overcome conservativeness and expand the system's knowledge base over time. Finally, lightweight, offline-capable implementations of the framework are necessary for real-world deployment. Achieving efficient uncertainty estimation, adaptive query selection, and meta-learning-based weighting on low-power devices would make the approach scalable and practical for under-resourced communities that lack access to high-compute infrastructure.

If these research directions are pursued, joint human-machine uncertainty could unlock scalable, interactive, and respectful language learning systems capable of discovering and documenting under-resourced languages without relying on large curated datasets. Ultimately, this line of research bridges technical innovation and participatory design, opening opportunities for AI systems that learn with people, not just from data.

5 Conclusion

This paper outlines a vision for open-ended language discovery based on joint human-machine uncertainty. We argue that future NLP systems, particularly for low-resource languages, must move beyond static data pipelines and toward interactive, participatory approaches that adapt to sparse, uncertain, and heterogeneous feedback. While many technical and sociotechnical challenges remain, this is not merely a research proposal but an ideological stance: language technology should be co-created with speakers. We position this work as a challenge to current practices that treat language as extractable data, advocating instead for AI systems that become collaborative participants in language preservation and revitalization. We call on the NLP and HCI research communities to develop methods, tools, and evaluation practices that support co-adaptive language learning systems, opening new pathways for linguistic documentation, preservation, and empowerment in the digital age. This position paper advocates for a paradigm shift: from building models that passively learn from existing data to designing systems that actively learn with people in real time, fostering respectful, human-centered AI for linguistic diversity.

References

David Ifeoluwa Adelani, Jesujoba Oluwadara Alabi, Angela Fan, Julia Kreutzer, Xiaoyu Shen, Machel Reid, Dana Ruiter, Dietrich Klakow, Peter Nabende, Ernie Chang, Tajuddeen Gwadabe, Freshia Sackey, Bonaventure F. P. Dossou, Chris Emezue, Colin Leong, Michael Beukman, Shamsuddeen H. Muhammad, Guyo D. Jarso, Oreen Yousuf, and 26 others. 2022a. A few thousand translations go a long way! leveraging pre-trained models for African news translation. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3053–3070, Seattle, United States. Association for Computational Linguistics.

David Ifeoluwa Adelani, Graham Neubig, Sebastian Ruder, Shruti Rijhwani, Michael Beukman, Chester Palen-Michel, Constantine Lignos, Jesujoba O. Alabi, Shamsuddeen H. Muhammad, Peter Nabende, Cheikh M. Bamba Dione, Andiswa Bukula, Rooweither Mabuya, Bonaventure F. P. Dossou, Blessing Sibanda, Happy Buzaaba, Jonathan Mukiibi, Godson Kalipe, Derguene Mbaye, and 26 others. 2022b. MasakhaNER 2.0: Africa-centric transfer learning for named entity recognition. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 4488–4508, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*.

Umang Bhatt, Javier Antorán, Yunfeng Zhang, Q. Vera Liao, Prasanna Sattigeri, Riccardo Fogliato, Gabrielle Melançon, Ranganath Krishnan, Jason Stanley, Omesh Tickoo, Lama Nachman, Rumi Chunara, Madhulika Srikumar, Adrian Weller, and Alice Xiang. 2021. Uncertainty as a form of transparency: Measuring, communicating, and using uncertainty. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '21, page 401–413, New York, NY, USA. Association for Computing Machinery.

Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the people? opportunities and challenges for participatory ai. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, pages 1–8.

Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. 2015. Weight uncertainty in neural networks. In *International Conference on Machine Learning*, pages 1613–1622.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, and 12 others. 2020. Language models are few-shot learners. *Advances in Neural Information Processing Systems*.

Katherine Maeve Collins, Matthew Barker, Mateo Espinosa Zarlenga, Naveen Raman, Umang Bhatt, Mateja Jamnik, Ilia Sucholutsky, Adrian Weller, and Krishnamurthy Dvijotham. 2023. Human uncertainty in concept-based ai systems. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23, page 869–889, New York, NY, USA. Association for Computing Machinery.

Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2023. The participatory turn in ai design: Theoretical foundations and the current state of practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '23, New York, NY, USA. Association for Computing Machinery.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Bonaventure F. P. Dossou. 2025. Advancing african-accented english speech recognition: Epistemic uncertainty-driven data selection for generalizable asr models. In *Proceedings of the ACL Student Research Workshop*.

Bonaventure F. P. Dossou, Ines Arous, and Jackie CK Cheung. 2025. Rethinking full finetuning from pretraining checkpoints in active learning for african languages. In *Proceedings of the ACL Student Research Workshop*.

Bonaventure F. P. Dossou and Chris Chinenye Emezue. 2021a. OkwuGbé: End-to-end speech recognition for Fon and Igbo. In *Proceedings of the Fifth Workshop on Widening Natural Language Processing*, pages 1–4, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Bonaventure F. P. Dossou, Atnafu Lampebo Tonja, Oreen Yousuf, Salomey Osei, Abigail Oppong, Iyanuoluwa Shode, Oluwabusayo Olufunke Awoyomi, and Chris Emezue. 2022. Afrolm: A self-active learning-based multilingual pretrained language model for 23 african languages. In *Proceedings of the Third Workshop on Data-Centric AI (SustainNLP)*.

Bonaventure FP Dossou and Chris C Emezue. 2021b. Crowdsourced phrase-based tokenization for low-resourced neural machine translation: The case of fon language. *arXiv preprint arXiv:2103.08052*.

Bonaventure FP Dossou, Iffanice Houndayi, Pamely Zantou, and Gilles Hacheme. 2023. Fonmtl: Towards multitask learning for the fon language. *arXiv preprint arXiv:2308.14280*.

Bonaventure FP Dossou and Mohammed Sabry. 2021. Afrivec: Word embedding models for african languages. case study of fon and nobiin. *arXiv preprint arXiv:2103.05132*.

Femi Pancrace Bonaventure Dossou and Chris Chinenye Emezue. 2020. FFR v1.1: Fon-French neural machine translation. In *Proceedings of the Fourth Widening Natural Language Processing Workshop*, pages 83–87, Seattle, USA. Association for Computational Linguistics.

David M. Eberhard, Gary F. Simons, and Charles D. Fennig. 2024. *Ethnologue: Languages of the World*, 27 edition. SIL International, Dallas, Texas.

Liat Ein-Dor, Alon Halfon, Ariel Gera, Eyal Shnarch, Lena Dankin, Leshem Choshen, Marina Danilevsky, Ranit Aharonov, Yoav Katz, and Noam Slonim. 2020. Active Learning for BERT: An Empirical Study. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7949–7962, Online. Association for Computational Linguistics.

Edmund L. Epstein and Robert Kole. 1998. *The Language of African Literature*. Africa World Press. Google-Books-ID: XkkrDH27jmIC.

Yarin Gal and Zoubin Ghahramani. 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1050–1059, New York, New York, USA. PMLR.

Yarin Gal, Riashat Islam, and Zoubin Ghahramani. 2017. Deep Bayesian active learning with image data. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1183–1192. PMLR.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. 2017. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1321–1330. PMLR.

Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. 2011. Bayesian active learning for classification and preference learning. In *Proceedings of the 29th International Conference on Machine Learning*.

Jeremy Howard and Sebastian Ruder. 2018. Universal language model fine-tuning for text classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*.

Edward Hughes, Michael Dennis, Jack Parker-Holder, Feryal Behbahani, Aditi Mavalankar, Yuge Shi, Tom Schaul, and Tim Rocktäschel. 2024. Open-endedness is essential for artificial superhuman intelligence. *arXiv preprint arXiv:2406.04268*.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the nlp world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in Neural Information Processing Systems*, 30:5574–5584.

Andreas Kirsch, Joost van Amersfoort, and Yarin Gal. 2019. *BatchBALD: efficient and diverse batch acquisition for deep Bayesian active learning*. Curran Associates Inc., Red Hook, NY, USA.

Andreas Kirsch, Sebastian Farquhar, Parmida Atighehchian, Andrew Jesson, Frederic Branchaud-Charron, and Yarin Gal. 2023. Stochastic batch acquisition: A simple baseline for deep active learning. *Preprint*, arXiv:2106.12059.

Minzhi Li, Taiwei Shi, Caleb Ziems, Min-Yen Kan, Nancy Chen, Zhengyuan Liu, and Diyi Yang. 2023. CoAnnotating: Uncertainty-guided work allocation between human and large language models for data annotation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 1487–1505, Singapore. Association for Computational Linguistics.

Q. Vera Liao and Jennifer Wortman Vaughan. 2023. Ai transparency in the age of llms: A human-centered research roadmap. *Preprint*, arXiv:2306.01941.

Cong Lu, Shengran Hu, and Jeff Clune. 2025. Automated capability discovery via foundation model self-exploration. *Preprint*, arXiv:2502.07577.

David J. C. MacKay. 1992. A practical bayesian framework for backpropagation networks. *Neural Computation*, 4(3):448–472.

Pedro Mendes, Paolo Romano, and David Garlan. 2025. Uncertainty estimation by human perception versus neural models. *Preprint*, arXiv:2506.15850.

Radford M Neal. 1996. *Bayesian learning for neural networks*. Ph.D. thesis, University of Toronto.

Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Taiwo Fagbohungbe, Solomon Oluwole Akinola, Shamsuddeen Muhammad, Salomon Kabongo Kabenamualu, Salomey Osei, Freshia Sackey, Rubungo Andre Niyongabo, Ricky Macharm, Perez Ogayo, Orevaoghene Ahia, Musie Meressa Berhe, Mofetoluwa Adeyemi, Masabata Mokgesi-Selinga, Lawrence Okegbemi, Laura Martinus, and 28 others. 2020. Participatory research for low-resourced machine translation: A case study in African languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2144–2160, Online. Association for Computational Linguistics.

Tobi Olatunji, Tejumade Afonja, Bonaventure F. P. Dossou, Atnafu Lambebo Tonja, Chris Chinenye Emezue, Amina Mardiyyah Rufai, and Sahib Singh. 2023a. Afrinames: Most asr models "butcher" african names. In *Interspeech 2023*, pages 5077–5081.

Tobi Olatunji, Tejumade Afonja, Aditya Yadavalli, Chris Chinenye Emezue, Sahib Singh, Bonaventure F. P. Dossou, Joanne Osuchukwu, Salomey Osei, Atnafu Lambebo Tonja, Naome Etori, and Clinton Mbataku. 2023b. Afrispeech-200: Pan-african accented speech dataset for clinical and general domain asr. *Transactions of the Association for Computational Linguistics*, 11:1669–1685.

Angéline Pouget, Lucas Beyer, Emanuele Bugliarello, Xiao Wang, Andreas Peter Steiner, Xiaohua Zhai, and Ibrahim Alabdulmohsin. 2024. No filter: Cultural and socioeconomic diversity in contrastive vision-language models. *Preprint*, arXiv:2405.13777.

Nithya Sambasivan, Erin Arnesen, Ben Hutchinson, Tulsee Doshi, and Vinodkumar Prabhakaran. 2021. Re-imagining algorithmic fairness in india and beyond. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, page 315–328, New York, NY, USA. Association for Computing Machinery.

Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, Matthias Gallé, Jonathan Tow, Alexander M. Rush, Stella Biderman, Albert Webson, Pawan Sasanka Ammanamanchi, Thomas Wang, Benoît Sagot, Niklas Muennighoff, Albert Villanova del Moral, and 373 others. 2023. Bloom: A 176b-parameter open-access multilingual language model. *Preprint*, arXiv:2211.05100.

Roy Schwartz, Jesse Dodge, Noah A. Smith, and Oren Etzioni. 2020. Green ai. *Commun. ACM*, 63(12):54–63.

Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2019. Energy and policy considerations for deep learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3645–3650, Florence, Italy. Association for Computational Linguistics.