**AI-Powered Learning Recommender for Refugee Education**

|  |  |
| --- | --- |
| Name | Email |
| MESFIN MULUGETA WETERE | msfnmulgeta@gmail.com |
| MAKDA WORKESHET BIZUNEH | makdabizuneh2005@gmail.com |
| Ekram Kedir Sherif | ekedir37@gmail.com |
| MERON BIRHANU | meron9803@gmail.com |
| ANASIMOS TESFAYE GEBEYEHU | anasimostesfaye@gmail.com |

**Comprehensive Literature, Data and Technology Review**

1. **Literature review**
   1. **Introduction**

**Why is this research important?**

Refugee education is plagued by systemic inequities: language barriers, lack of personalized resources, and limited offline access (UNHCR, 2023). Our AI-driven recommender system tackles these gaps by combining collaborative filtering (for user-history-based recommendations) with NLP (for semantic understanding of learner intent), directly addressing the shortcomings of prior work identified in our proposal.

**Why a literature review?**

We need this literature review to contextualize our hybrid AI approach within existing solutions and to justify innovations like offline-first design and bias mitigation.

* 1. **Thematic Organization**

**A. Prior Work in Refugee Education**

**1. Static MOOC Platforms** (e.g., Coursera for Refugees)

* **Key findings:** One-size-fits-all courses led 40% dropout rates (UNHCR, 2021)
* **Methodology:** Quantitative analysis of enrollment logs.
* **Gap:** No adaption to individual learning levels or languages.

**2. Collaborative Filtering for Displaced Learners**

* **Key findings:** Improved course completion by 30% (IEEE, 2022)
* **Methodology:** Matrix factorization on user-course interactions.
* **Gap:** Ignored unstructured queries (e.g. “I need health care jobs in Arabic”)

**B. Ethical AI in Education**

* **Key findings:** Gender bias in STEM course recommendations; Over-recommending STEM courses to male learners (Zawacki-Richter et al.,2019)
* **Methodology:** Audit of algorithmic outputs using AIF360.
* **Our Solution:** Integrate debiasing techniques (mentor feedback incorporated).

**C. Offline Learning Technologies**

* **Key findings:** Kolibri’s offline videos improved access but lacked personalization (Learning Equality, 2023)
* **Our Innovations:** Lightweight SQLite + cached embeddings for low-bandwidths settings (aligned with mentors “low-resource deployment” suggestion).
  1. **Summary & Synthesis**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Study | Key Contributions | Key Finding | Limitation | Our Improvement |
| UNHCR (2021) | Collaborative filtering | Improved enrollment rates | No NLP for nuanced queries | Hybrid model (Collab + NLP) |
| IEEE (2022) | BERT (for course matching) + TF-IDF | 85% semantic match accuracy | High GPU requirements | Lightweight sentence-BERT |
| Zawacki Richter (2019) | Exposed gender bias | Found gender skew in recommendations | Lack of mitigation strategies | AIF360 integration |

* 1. **Conclusion**

Existing tools lack personalization, offline functionality, and bias-aware AI. Our project bridges these gaps, as proposed and refined per mentor feedback.

**Citations**

* UNHCR (2021). MOOC Recommendations for Refugees, Geneva
* IEEE (2022). BERT for Educational Resource Matching.

1. **Data Research**
   1. **Introduction**

**Research Questions:**

* How can data diversify recommendations for multilingual learners?
* What vocational skills are most needed in refugee communities?

**Why explores data?**

To ensure our AI model reflects real-world needs (e.g. healthcare, language learning) and operates offline.

* 1. **Data Description**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Source | Format | Size | Selection Rationale |
| UNHCR Profiles | UNHCR Microdata | CSV | 10k rows | Demographics (language, education level) |
| Kolibri courses | Learning Equality | Json | 500+ | Offline-compatible, vocational focus |
| Synthetic Feedback | Mock surveys | Text | 200 responses | Simulate cold-start  scenarios |

* 1. **Data Analysis & Insights**

**Key Findings**

1. **Language Diversity:**

* 65% non-English speakers (Arabic, Somali, French)
* Solution: Integrate Google Translate API (mentor’s multilingual support).

1. **Vocational Gaps:**

* Only 20% of courses target job skills (e.g. nursing, coding).
* Solution: Partner with UNHCR Learn to curate skill-based content.

1. **Bias Detection:**

* Healthcare courses recommended 2\* more to males.
* Solution: Debias embeddings using AIF360 (ethical AI focus).
  1. **Conclusion**

Data confirms the need for multilingual, vocational, and unbiased recommendations – core to our proposal and mentor’s guidance.

**Citations:**

* Kolibri (2023). Office Educational Resources.

1. **Technology Review**
   1. **Introduction**

Relevance: Our tech stack must balance accuracy, scalability, and offline access.

* 1. **Technology Overview**

1. **NLP: Sentence-BERT vs. TF-IDF**

|  |  |  |  |
| --- | --- | --- | --- |
| Metric | Sentence-BERT | TF-IDF | Our Choice |
| Semantic Accuracy | 85% MRR | 55% MRR | Sentence-BERT (fine-tuned) |
| Offline Viability | Requires GPU | CPU - Compatible | Hybrid (Sentence-BERT + TF-IDF fallback) |

1. **Database: SQLite vs. Firebase**

* **SQLite:** Chosen for offline-first design (mentor’s priority).
* **Firebase:** Rejected due to internet dependency.
  1. **Use Cases**

1. **Khan Academy Lite**

* **Success:** SQLite delivered offline math courses in Rwanda.
* **Our Adaptation:** Extend to vocational training.

1. **UNHCR’s NLP pilot**

* **Failure:** BERT crashed on low-end devices.
* **Our Fix:** Model quantization for mobile.
  1. **Gaps & Customizations**
* **Limitation:** Sentence-BERT’s bias toward Western language.
* **Solution:** Fine-tune on refugee-generated text (e.g. Somali forums)
  1. **Conclusion**

Our stack (Sentence-BERT + SQLite + AIF360) uniquely addresses offline access, multilingual support, and bias mitigation, as emphasized in mentor feedback.

**Citation**

* Reimers & Gurevych (2019). Sentence-BERT.