International Journal of Computer Science, Engineering and Applications (IJCSEA),
Vol. 15, No. 6, December 2025

NATURAL LANGUAGE PROCESSING PIPELINE FOR CO-DESIGNING CULTURALLY AWARE HEALTH CHATBOTS FROM USER STORIES TO SYSTEM SPECIFICATIONS

Oritsemeyiwa Gabriel Orugboh

United Kingdom

ABSTRACT

Designing culturally aware health chatbots for underserved communities is essential to advancing global health equity. A major challenge, however, is the conversion of contextual, often qualitative user requirements into technical specifications. In this paper, we present a natural language processing (NLP) pipeline for the co-design of culturally aware health chatbots. The pipeline systematically converts user stories into system specifications. By employing Large Language Models (LLMs), the pipeline automates the extraction of cultural and contextual requirements from user stories, which inherently counteracts biases from general AI models. By using the requirements engineering phase to direct attention to domain-specific and culturally relevant data, this approach ensures that the resulting specifications for chatbots are grounded in the realities of the communities they aim to serve. This approach does not only improve the efficiency of the design process, it also proactively embeds cultural sensitivity and inclusivity at the heart of AI-driven health solutions

KEYWORDS

Culturally Aware Chatbots, Health Natural Language Processing (Health NLP), Requirements Engineering, AI Bias Mitigation, Co-Design

1. INTRODUCTION

Natural language processing (NLP) is evolving at a steady pace, providing a range of new opportunities for its application, especially in the health domain. One such application is in the form of conversational agents (chatbots) to provide users with health information and support [1]. To be successful and to promote equity, conversational agents and their underlying AI technology need to be carefully designed to ensure contextual and cultural fit, particularly in the majority world [2]. This requires greater emphasis on building cultural awareness into the design and specification process [2, 3].

In the context of building culturally aware health chatbots for the majority world, this means incorporating an understanding of local health beliefs, practices, languages, and user expectations into the system design. This paper will, therefore, focus on using NLP techniques to translate user stories into system specifications for co-designing culturally aware health chatbots [3]. This will contribute to the emerging discussion on the need to address contextual awareness and bias mitigation in large language models [3].

The proposed pipeline will utilise state-of-the-art NLP techniques to extract and translate user stories into system specifications [6]. This process will focus on capturing the cultural and contextual nuances of the target users and translating them into specific, measurable, achievable, relevant, and time-bound (SMART) requirements for the chatbot [7]. This will involve using NLP techniques to identify and extract relevant information from user stories and map them to specific system requirements. The proposed approach will help to ensure that the resulting chatbot specifications are aligned with the cultural context of the users and the requirements engineering is based on domain-specific, high-quality data, which can help to mitigate some of the biases observed in large language models [7]. This approach is critical for the pursuit of health equity, as it allows the identification and mitigation of potential biases in AI development and deployment through the co-production process [4].

2. Background and Motivation

2.1. The Imperative for Cultural Awareness in Health AI

Healthcare NLP tools, while traditionally involving clinical subject matter experts to extract health information or aid interpretation, often have few if any community stakeholders with lived experience as part of their development process [1]. This top-down approach often produces healthcare NLP tools that don't have cultural considerations baked in. While there is understandable excitement for using conversational AI (CAI) for health in majority world contexts, in order for CAI to be truly useful, it must respond appropriately in culturally and linguistically diverse settings [2]. Mechanisms are needed to help contend with the fact that existing LLMs leave out lived experiences from around the world.

2.2. Challenges of Bias and Context in LLMs

The continuous growth of conversational and generative artificial intelligence (AI) has seen an increase in healthcare adoption of AI tools. Conversational AI tools promise both more efficient healthcare service delivery and an increase in access to healthcare, however, ethical, practical and inclusive concerns have been raised [4]. The latest Human-Computer Interaction (HCI) research has shown that there is still much work to be done in terms of context awareness and bias mitigation in large language models (LLMs) [3]. It is therefore critical that diversity and inclusion principles are baked into every stage of the AI lifecycle, particularly during the requirements engineering process, to build systems that truly meet the needs of diverse user groups and reflect ethical considerations [9].

3. Proposed NLP Pipeline for Co-Design

The approach outlined in this paper is a structured Natural Language Processing (NLP) pipeline that serves as an interface between bottom-up and locally grounded insights and a corresponding technical system design [3]. It translates qualitative data from participatory co-design workshops [2] into a format that can be used for software development.

3.1. From User Stories to System Specifications

The pipeline is centered on the analysis of user stories, with a focus on systematically identifying and operationalizing culturally bound needs and preferences as parameters for system design [8]. Through the translation of narrative, qualitative user input into concrete technical specifications, the pipeline aims to facilitate a requirements elicitation and analysis process that is well-suited for developing culturally aware and user-centered AI, especially in sensitive fields such as health [4].

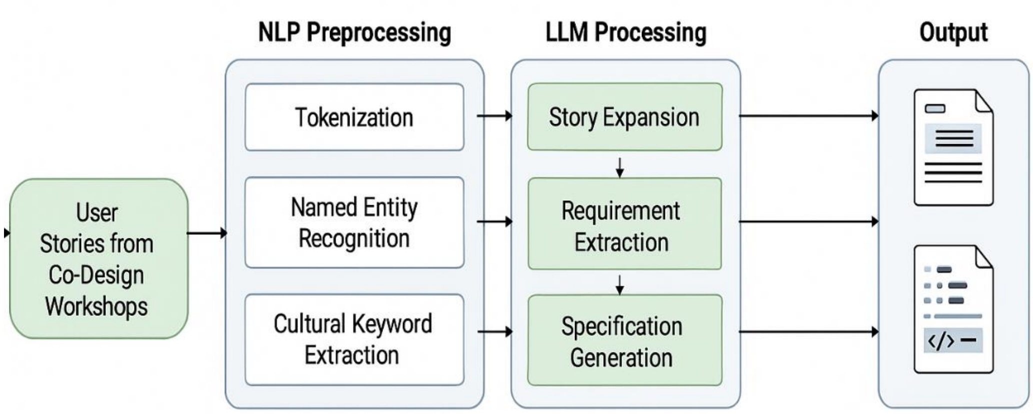

Figure 1. Overview of the proposed NLP pipeline translating user stories into system specifications for culturally aware health chatbots.

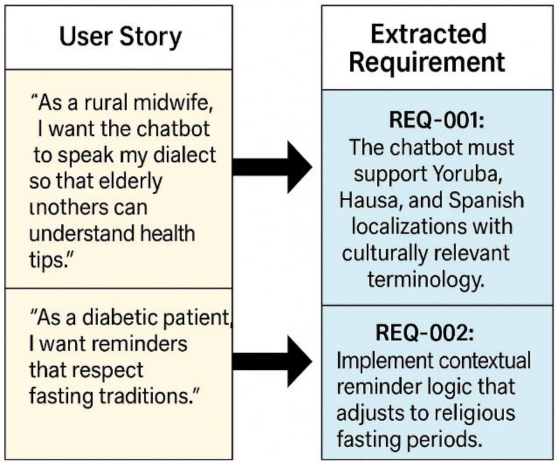

Figure 2. Example of the NLP pipeline output translating user narratives into technical requirements.

3.2. The Critical Role of Large Language Models

Large Language Models form the core technical component of the pipeline for two key reasons: The ability to generate and refine user stories automatically and to transform them into system specifications [10,11]. This makes use of the potential of these models to understand, process and produce human language in order to automate part of the otherwise manual and thus, time-intensive task of requirements elicitation and analysis [10,12]. In addition, their potential to work with unstructured text data allows them to process the cultural and contextual information contained in user stories and translate it into actionable insights [7].

This, in turn, allows for identifying subjective yet relevant information such as social and behavioral determinants of health which are often only found in unstructured patient notes and are key in designing tailored healthcare for marginalized groups [13]. It also allows for much more efficient requirements engineering, as parts of the elicitation and specification generation process which are especially prone to human error, can be automated [10].

4. LITERATURE REVIEW

The emergence of large language models (LLMs) and their recent progress have shown that they could be a significant asset in a wide range of software engineering applications [14]. In the field of software development lifecycle (SDLC), LLMs have been widely used in various automated tasks such as requirements engineering [15], code generation and testing.

Requirements engineering (RE) is an essential phase in the SDLC, which involves requirements elicitation, analysis, specification, and change. As an important yet challenging part of the software development process, RE has been a subject of constant research over decades, with various problems like communication, early stage uncertainties, and estimation of the resources required for accurate development [7]. In this empirical study, we explore the efficiency of LLMs for automating requirements analysis tasks.

In addition to this, to collect requirements in agile software development, developers record them in user stories, focusing on the user's point of view, their needs, and the value of each software feature [8]. They are required to write shorter and high-quality user stories to aid the development process. In the following, we present our attempt to use LLMs for this purpose.

5. METHODOLOGY

5.1. Participatory Data Collection

Our aim was to build a nuanced and human-focused understanding of: (a) areas of cultural misalignment in digital health; (b) local understandings of chatbots for health and (c) recommendations for designing culturally appropriate CAI [2], with the under-researched area of Latin America as our focus [2]. Our findings suggest that academic definitions of the concept of culture are devoid of meaning at the grassroots level, and that technologies will have to be positioned within a wider paradigm; one which integrates the ways in which economics, politics, geography and local logistics are intertwined with cultural experience [2].

5.2. LLM-Driven Requirements Elicitation and Analysis

We developed a multi-agent system to produce agents using AI models to generate user stories from initial requirements, evaluate and enhance quality, and prioritize stories using selected technique [7].

Table 1. Comparative performance and characteristics of language models used in requirements analysis.

Model	Parameter Size	Task Focus	Semantic Similarity (%)	API Latency (ms)	Notes
GPT-3.5	175 B	Requirement Generation	87	450	Balanced accuracy & speed
GPT-4 Omni	1 T+	Full Pipeline Automation	93	720	Best contextual reasoning
LLaMA3-70	70 B	Domain Fine-Tuned	88	610	Efficient for cultural keywords
Mixtral-8B	8 B Mixture-of-Experts	Rapid Prototyping	82	390	Lightweight deployment

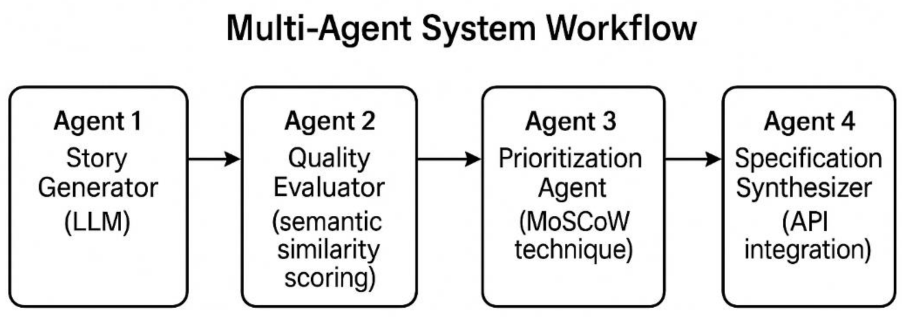

Figure 3. Workflow of the multi-agent architecture for automated requirements elicitation and prioritization.

In our approach, we deployed four models (GPT-3.5, GPT-4 Omni, Llama3-70, and Mixtral-8B) and executed experiments to perform requirements analysis on four real-world projects. We assess the results based on the semantic similarity and API performance of different models, as well as their effectiveness and efficiency in requirements analysis and gathered users' feedback about their experience [7].

6. Discussion and Implications

6.1. Towards Equitable and Effective Health Chatbots

Achieving health equity with conversational AI requires a roadmap for the inclusive design and implementation of chatbots in healthcare. A 10-stage roadmap was identified to reflect activities related to phases of equitable conversational AI design and implementation: 1) Conception and planning, 2) Diversity and collaboration, 3) Preliminary research, 4) Co-production, 5) Safety measures, 6) Preliminary testing, 7) Healthcare integration, 8) Service evaluation and auditing, 9) Maintenance and 10) Termination [4]. These stages highlight the significance of a team-based approach and patient group involvement in addressing the rapidly changing landscape of conversational AI.

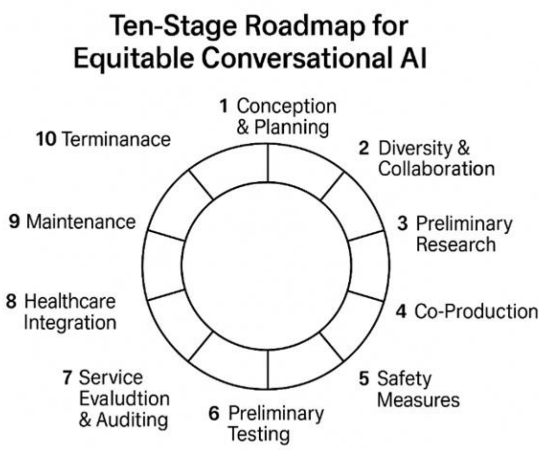

Figure 4. Ten-stage roadmap for designing equitable and inclusive health chatbots.

6.2. Mitigating Bias through Localized Co-Design

LLMs in the public domain, which were mostly trained in global north contexts, were tested for bias in a real-world clinical setting. When the LLMs were fine-tuned to the local EHR data the performance of the model improved [13]. These findings highlighted the utility of no-code platforms for large scale deployment, and the balance required between automation and human in the loop. The research proposed a context-aware model to accelerate AI adoption in the domain of higher education, connecting global technology developments with Africa's socio-technical landscape, and prioritising ethical considerations and student-centred design [6].

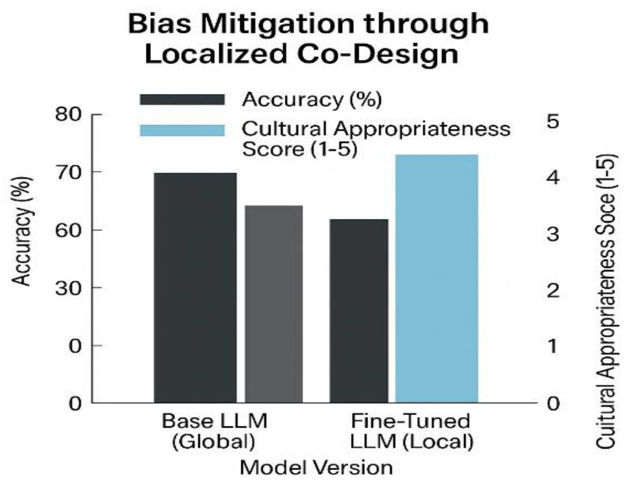

Figure 5. Improvement in performance and cultural relevance after fine-tuning models with local datasets.

7. CONCLUSION AND FUTURE WORK

This paper presented an NLP pipeline for co-designing culturally aware health chatbots, providing a step-by-step guide from user stories to system specifications using LLMs for generating technical code. In using such a participatory approach in leveraging LLMs to the point of directly producing system code, the pipeline could help address the common AI health risks of cultural bias and context irrelevance in automated decision support systems (ADSS) for deployment in majority world contexts. This could help ensure the systems built are appropriate, equitable, and truly serve the needs of local communities.

Future work can be done to implement the entire pipeline end-to-end for deployment in a specific culture of use, such as Latin American or South Asian contexts identified in the review [2, 13]. Further studies are also needed to measure the effects of each activity recommended in the pipeline in improving fairness of chatbots and reducing health disparities [4]. Ultimately, working together, clinical and technical experts can make sure responsible and unbiased powerful tech can be accessible to and used by under-resourced, overburdened settings in ways that are safe, fair, and helpful to all [13].

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to the community members, design participants, and subject-matter experts whose insights informed this research. Their contributions were invaluable in shaping the culturally grounded user stories and requirements that guided this work. I also appreciate the support and constructive feedback provided by colleagues and mentors who reviewed earlier versions of this study. Finally, I acknowledge the broader research communities in NLP, human-centered design, and global health whose foundational work made this project possible.

REFERENCES

[1] Pillai, M., Griffin, A. C., Kronk, C., & McCall, T. (2023). Toward Community-Based Natural Language Processing (CBNLP): Cocreating With Communities. *Journal of Medical Internet Research*, 25, e48498. https://doi.org/10.2196/48498

[2] Peters, D., Espinoza, F., da Re, M., Ivetta, G., Benotti, L., & Calvo, R. A. (2025). Towards culturally-appropriate conversational AI for health in the majority world: An exploratory study with citizens and professionals in Latin America. *arXiv*. https://arxiv.org/abs/2507.01719

[3] Dev, R., Ramani, D., Divate, T., Jalota, S., & Ismail, A. (2025). "Kya family planning after marriage hoti hai?": Integrating Cultural Sensitivity in an LLM Chatbot for Reproductive Health. *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '25)*. https://doi.org/10.1145/3706598.3713362

[4] Nadarzynski, T., Knights, N., Husbands, D., Graham, C. A., Llewellyn, C., Buchanan, T., Montgomery, I., & Ridge, D. (2024). Achieving health equity through conversational AI: A roadmap for design and implementation of inclusive chatbots in healthcare. *PLOS Digital Health*, 3(5), e0000492. https://doi.org/10.1371/journal.pdig.0000492

[5] Denecke, K. (2025). Bridging the Gap: Inclusive Artificial Intelligence for Patient-Oriented Language Processing in Conversational Agents in Healthcare. *Applied Clinical Informatics*. https://doi.org/10.18653/v1/2025.cl4health-1.42

[6] Bamurange, D., & Jonathan, K. N. (2025). Designing a Hybrid AI Chatbot Framework for Student Support: Integrating NLP and Human Oversight in African Universities. *Journal of Information and Technology*, 5(4), 41-52. https://doi.org/10.70619/vol5iss4pp41-52

[7] Sami, M., Wasim, M., Zhang, Z., Rasheed, Z., & Systä, K. (2024). AI based Multiagent Approach for Requirements Elicitation and Analysis. *arXiv*. https://arxiv.org/abs/2409.00038

[8] Sami, M., Wasim, M., Zhang, Z., Rasheed, Z., Systä, K., & Abrahamsson, P. (2024). Early Results of an AI Multiagent System for Requirements Elicitation and Analysis. In *Lecture Notes in Computer Science: Vol. 15488*. Springer, Cham. https://doi.org/10.1007/978-3-031-78386-9_20

[9] Bano, M., Zowghi, D., Gervasi, V., & Shams, R. A. (2023). AI for All: Operationalising Diversity and Inclusion Requirements for AI Systems. *arXiv*. https://arxiv.org/abs/2311.14695

[10] Askari, A., Sharbaf, M., Kolahdouz-Rahimi, S., Lano, K., & Tehrani, S. Y. (2025). Research directions for using LLM in software requirement engineering: a systematic review. *Frontiers in Computer Science*, 7. https://doi.org/10.3389/fcomp.2025.1519437

[11] Yamani, A., Baslyman, M., & Ahmed, M. (2025). Leveraging LLMs for User Stories in AI Systems: UStAI Dataset. *Proceedings of the 32nd IEEE International Requirements Engineering Conference (RE'25)*. https://doi.org/10.1145/3727582.3728689

[12] Ronanki, K., Berger, C., & Horkoff, J. (2023). Investigating ChatGPT's Potential to Assist in Requirements Elicitation Processes. *arXiv*. https://arxiv.org/abs/2307.07381

[13] Hashemian, S. A. R., Jamal, F., Azal, M. U., Khan, A. H., Abbas, S., Cheema, H. N., Shabbir, M. U., Hameed, I., Ayub, M., Masood, H., Jafar, A., Khan, A. M., Nazir, M. A., Jamil, M. A., Sultan, F., & Khalid, S. (2024). Assessing equitable use of large language models for clinical decision support in real-world settings: fine-tuning and internal-external validation using electronic health records from South Asia. *medRxiv*. https://doi.org/10.1101/2024.06.05.24308365

[14] Malik, A., Rasheed, Z., Wasim, M., Zhang, Z., Tomas, H., & Abrahamsson, P. (2024). A Tool for Test Case Scenarios Generation Using Large Language Models. *arXiv*. https://arxiv.org/abs/2406.07021

[15] Ali, J., Qayyum, S., & Dar, H. S. (2025). Large Language Model for Requirements Engineering: A Systematic Literature Review. *Requirements Engineering*. https://doi.org/10.21203/rs.3.rs-5589929/v1

AUTHOR

Orugboh Oritsemeyiwa Gabriel is a multidisciplinary researcher and professional whose work bridges computing, sociology, and social welfare. He holds an MSc in Computing from Cardiff University, Wales, an MSc in Sociology from the University of Ibadan, and a BSc in Sociology and Anthropology from the University of Benin, Nigeria. His research explores the intersection of technology, data analytics, and social systems, focusing on how computational sociology and big data can inform evidence-based social policy. He has published on social dynamics, migration, and digital inclusion in peer-reviewed journals. Oritsemeyiwa is a member of the International Sociological Association (ISA), the British Computer Society (BCS), and the IEEE Systems Council. Email: meyiwagab@gmail.com