

Simulating Empathic Interactions with Synthetic LLM-Generated Cancer Patient Personas

Rezaur RASHID^{a,*}, Saba KHEIRINEJAD^a, Brianna M WHITE^a, Soheil HASHTARKHANI^a, Parnian KHEIRKHAH RAHIMABAD^a, Fekede A KUMSA^a, Lokesh CHINTHALA^a, Janet A ZINK^a, Christopher L BRETT^b, Robert L DAVIS^a, David L SCHWARTZ^c, and Arash SHABAN-NEJAD^{a,§,1}

^aCenter for Biomedical Informatics, Department of Pediatrics, College of Medicine, University of Tennessee Health Science Center (UTHSC), Memphis, Tennessee, USA

^bThe University of Tennessee Graduate School of Medicine, Knoxville, Tennessee, USA

^cDepartment of Radiation Oncology, UTHSC, Memphis, TN, USA

ORCID ID: ^{*}<https://orcid.org/0000-0003-1343-5364>,

[§]<https://orcid.org/0000-0003-2047-4759>

Abstract. Unplanned interruptions in radiation therapy (RT) increase clinical risks, yet proactive, personalized psychosocial support remains limited. This study presents a proof-of-concept framework that simulates and evaluates Empathic AI-patient interactions using large language models (LLMs) and synthetic oncology patient personas. Leveraging a de-identified dataset of patient demographics, clinical features, and social determinants of health (SDoH), we created realistic personas that interact with an empathic AI assistant in simulated dialogues. The system uses dual LLMs, one for persona generation and another for empathic response, which engage in multi-turn dialogue pairs per persona. We evaluated the outputs using statistical similarity tests, quantitative metrics (BERTScore, SDoH relevance, empathy, persona distinctness), and qualitative human assessment. The results demonstrate the feasibility of scalable, secure, and context-aware dialogue for early-stage AI development. This HIPAA/GDPR compliant framework supports ethical testing of empathic clinical support tools and lays the groundwork for AI-driven interventions to improve RT adherence.

Keywords. Large Language Models, Radiation Therapy, Personas, Empathy, LLM

1. Introduction

Empathic communication is critical in oncology, particularly for patients undergoing radiation therapy (RT), where clinical, emotional, and social needs intersect. Large language models (LLMs) offer scalable solutions for personalized care, but face privacy, ethical testing, and contextualization challenges [1]. Recent studies have highlighted the potential of LLMs in oncology for patient education, emotional support, and clinical guidance, with chatbots increasingly simulating human-like conversations [2,3]. Digital

¹ Corresponding Author: Arash Shaban-Nejad, Center for Biomedical Informatics, Department of Pediatrics, College of Medicine, UTHSC, Memphis, TN, USA; E-mail: ashabann@uthsc.edu

personas can further enhance personalization and empathy by tailoring LLM-generated support to individual patient profiles [4]. However, producing authentically empathic and contextually relevant responses remains a key challenge [5].

We propose a novel proof-of-concept framework using dual LLMs: one to generate synthetic oncology patient personas from a de-identified patient dataset, and the other to simulate empathic dialogues addressing RT interruptions. The personas incorporate demographics, clinical features, and social determinants of health (SDoH), enabling realistic dialogue simulations that preserve patient anonymity. Our contributions include (i) realistic persona synthesis preserving SDoH complexity; (ii) context-sensitive empathic dialogue simulation; (iii) validation through statistical and qualitative metrics; and (iv) ethical, secure early-stage AI testing sandbox.

2. Methods

2.1. Dataset

The study utilized a de-identified institutional database of 2,525 RT patients treated at University of Tennessee Medical Center (UTMC) in Knoxville, containing 16 variables per RT course, anonymized per institutional review board guidelines, and compliant with HIPAA/GDPR. Key variables included: **Demographics** (Age, gender, race/ethnicity, geographic location); **Clinical Features** (ICD-10 cancer site, stage, prescribed dose, missed treatment schedule); **SDoH** (Insurance status, distance to facility, median income quintile).

2.2. Persona Generation

A transformer-based LLM (Llama3.1:8b, temperature 0, top p 0.9, repeat-penalty 1.1) was fine-tuned on the dataset to generate 10 synthetic patient personas, preserving statistical distributions of demographics, clinical features, and SDoH. The process involved:

- **Input Processing:** Structured prompts fed anonymized records into the LLM.
- **Feature Synthesis:** Generated 120-word vignettes with demographics, emotional state (e.g., frustration over missed sessions), and SDoH (e.g., transportation barriers), forming a 14-field persona card (Figure 1).
- **Validation:** Kolmogorov-Smirnov tests confirmed statistical fidelity to the dataset.

Example Prompt for Persona Generation. “Generate a realistic oncology patient persona going through radiation therapy treatment based on the following details: age, race, gender, cancer diagnosis type (ICD-10 category), SDoH, and treatment. Include a 120-word vignette with demographics, emotional state, social context, and structured attributes relevant to RT interruptions.”

2.3. Empathic AI Assistant

A second LLM (Gemma3:12B, temperature 0, top p 0.9, repeat-penalty 1.1) was supervised fine-tuned with a parameter-efficient LoRA adapter (rank = 8, α = 16) on ESCConv emotional-support dataset [6] to deliver 1–2 sentence empathic responses using

LangChain ConversationBufferMemory for context-aware questioning, affirmation, and suggestions. Responses incorporated RT jargon (e.g., “fractions”) and accessible language.

Example Prompt. “*You’re an empathic AI assistant helping name, facing SDoH issues, and feeling emotional state. Respond in 1–2 sentences with empathy, addressing RT interruption risks (if needed), and specific solutions (e.g., shuttle, financial aid where necessary).*”


| Linda Parker (name generated by LLM Model) | | |
|---|---|--|
|  | Age: 63 y/o Gender: Female Race: Black Profession: Teacher (retd.) | Description: Linda Parker is a 63-year-old black woman and a retired teacher residing in a suburban area about 30 miles from her treatment facility. She's battling gynecologic cancer but maintains high adherence to her radiation therapy regimen. Her lifestyle includes being a former smoker but abstains from alcohol and substances. She lives in a community characterized by modest income and low educational attainment. Linda relies on her insurance, Syntrix Liability 6YX, to cover medical expenses. She actively manages her health with regular medical check-ups and has family support due to her stable marital status. |

Figure 1. Synthetic oncology patient persona is produced by the persona-generation module.

2.4. Dialogue Simulation

A dual-agent loop (persona: Llama3.1:8b; assistant: Gemma3:12b) simulated 10–12 dialogue turns per persona using the open-source LLM API tool, Groq, and LangChain. Personas initiated queries (e.g., “I ain’t got no ride to my RT center, and I’m freaking out. What can I do?”), and the assistant responded empathetically, iterating naturally until completion.

2.5. Evaluation

Two board-certified radiation oncologists independently blind-rated a subset of persona-assistant dialogue pairs on three criteria: (1) Suggestion, evaluating the practicality of RT-specific advice (e.g., transportation solutions); (2) Consistency, ensuring alignment with personas’ SDoH and emotional states; and (3) Comforting Ability, assessing the assistant’s emotional support.

Automatic evaluations compared assistant responses against LLM-generated nurse navigator reference responses. The framework was evaluated on:

- **Persona Realism:** Kolmogorov-Smirnov tests for statistical similarity to the dataset.
- **Dialogue Quality (Quantitative):** Linguistic Metrics, BERTScore for semantic similarity, BLEU, and ROUGE-L against dataset-driven proxy reference dialogues (e.g., “I ain’t got no ride” → “Try the clinic’s shuttle”).
- **SDoH Relevance and Use of Empathy Cues:** Rule-based keyword scoring assessing whether assistant responses explicitly address SDoH barriers (e.g., transportation, financial concerns) and consistently include empathic phrasing (e.g., “I understand your concerns”).
- **Persona Distinctness:** HEXACO model [7] (Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, Openness) scored via rule-based keyword matching (e.g., “freaking out” for Emotionality), with diversity assessed via Euclidean distances.
- **Dialogue Quality (Qualitative):** Human evaluation by oncologists and social caregivers assessing suggestion quality, consistency, and comforting ability.

- **Security:** Verified no protected health information (PHI) exposure.

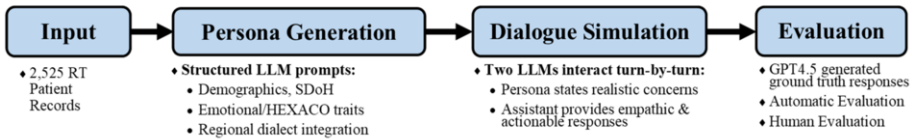


Figure 2. End-to-end workflow from patient data input through persona generation and dialogue simulation to evaluation for empathic RT-interruption support. Our evaluation consists of automatic linguistic metrics (e.g., BERTScore, BLEU-2, ROUGE-L), HEXACO trait consistency, and qualitative oncology expert assessments (Section 2.5). Additional clinically relevant evaluation metrics, such as sensitivity, specificity, PPV, and NPV, are planned as future enhancements.

3. Results

We generated 10 synthetic personas with persona-assistant conversations (≈ 110 turns), statistically similar to the 2,525-patient dataset (Kolmogorov-Smirnov $p = 0.05$ for all features). Table 1 summarizes mean scores. A mean BERTScore-F of 0.839 ± 0.03 indicates strong semantic similarity to the nurse-navigator reference (LLM-generated proxy reference) replies. Lexical overlap is naturally lower for short, paraphrastic counseling utterances (BLEU-2 = 0.047; ROUGE-L = 0.126). Task-specific indicators show that 62% of assistant turns explicitly addressed a SDoH barrier, and 94% contained an empathic cue. The mean HEXACO distance of 0.524 confirms the limited personality drift between persona specification and simulated dialogue.

Human Evaluation: Expert A rated seven of the 10 dialogues using a structured rubric across three specified criteria (Section 2.5), while Expert B annotated four overlapping dialogues, yielding comprehensive qualitative and quantitative assessments. The dialogues were generally empathetic and SDoH-relevant; however, experts identified several areas for improvement, including robotic tone, vague or nonspecific advice (e.g., lack of reference to local clinic resources), insufficient distinction between personas resulting in uniform responses, and a limited incorporation of region-specific linguistic features.

4. Discussion

Preliminary observations showed rural personas often prompted practical transportation solutions (e.g., “clinic shuttle”), whereas urban personas tended to generate discussions around work-life balance. Although limited scale, the 10-personas set provided an initial proof-of-concept balancing scalability and diversity. Quantitative metrics indicate promising dialogue quality, and the open-source LLM API enables efficient, secure persona generation without PHI exposure, ensuring HIPAA/GDPR compliance.

However, these preliminary findings should be interpreted with caution due to limitations, including the relatively small scale, proxy references lacking nuanced regional stereotypes, and human evaluation predominantly conducted by oncologists, potentially overlooking important subtleties that social caregivers and real patients might provide. Future work will expand significantly (100+ personas), integrate real patient

narratives, and refine LLM training with detailed RT-specific jargon and realistic clinical solutions, maintaining ethical testing via synthetic personas.

Table 1 Automatic metrics used to assess ten persona realism and patient-assistant dialogue quality. (Higher values denote better performance except for HEXACO distance, where lower values indicate better.)

| Criteria | What it Measures | Method | Results (mean±std) |
|---------------------|---|---------------------------|--------------------|
| Semantic Similarity | Semantic coherence of assistant replies (contextual embeddings) | BERTScore | 0.839 ± 0.006 |
| Lexical Precision | N-gram overlap precision | BLEU-2 | 0.047 ± 0.025 |
| Lexical Recall | Sequence overlap recall | ROUGE-L | 0.126 ± 0.012 |
| SDoH Relevance | Proportion of explicit addressing of SDoH barriers | Keyword-based Scoring (%) | 62.42% ± 28.3% |
| Empathy Cue Usage | Presence of explicit empathic phrasing | Keyword-based Scoring (%) | 93.94% ± 10.8% |
| Trait Consistency | Personality trait consistency across generated personas | HEXACO Euclidean Distance | 0.524 ± 0.090 |

5. Conclusion

This study presents a proof-of-concept model that explores the potential of a dual LLM system to generate synthetic oncology patient personas and simulate empathic dialogues for addressing RT interruptions in clinician-support interface testing. Preliminary evaluations indicate the feasibility and promise of our approach, while further rigorous validation is required before broader clinical application.

Acknowledgment. This study was supported by a grant from the Tennessee Department of Health.

References

- [1] Shaban-Nejad A, Michalowski M, Bianco S. Creative and generative artificial intelligence for personalized medicine and healthcare: Hype, reality, or hyperreality? *Experimental Biology and Medicine (Maywood)*, 248(24); 2023 Dec; pp.2497–2499. doi: 10.1177/15353702241226801.
- [2] Zitu MM, Le TD, Duong T, Haddadan S, et al. Large language models in cancer: potentials, risks, and safeguards. *BJR Artificial Intelligence*, 2(1); 2025 Jan; p.ubae019. doi: 10.1093/bjrai/ubae019
- [3] Sorin V, Brin D, Barash Y, et al. Large Language Models and Empathy: Systematic Review. *Journal of Medical Internet Research*, 26; 2024 Dec; p.e52597. doi: 10.2196/52597.
- [4] Wu S, Deng Y, Zhu Y, Hsu W, Lee ML. From personas to talks: Revisiting the impact of personas on llm-synthesized emotional support conversations. *arXiv*; 2025 Feb; preprint arXiv:2502.11451.
- [5] Cho S, Lee M, Yu J, et al. Leveraging Large Language Models for Improved Understanding of Communications With Patients With Cancer in a Call Center Setting: Proof-of-Concept Study. *Journal of Medical Internet Research*, 26; 2024 Dec; p.e63892. doi: 10.2196/63892
- [6] Liu S, Zheng C, Demasi O, et al. Towards emotional support dialog systems. *arXiv*; In *Proc. of the 59th Annual Meeting of the Association for Computational Linguistics, Vol. 1*, pages 3469–3483, 2021. doi: 10.18653/v1/2021.acl-long.269
- [7] Lee K, and Ashton MC. Psychometric properties of the HEXACO personality inventory. *Multivariate behavioral research*, 39(2); 2004 Apr; pp.329-358. doi: 10.1207/s15327906mbr39028