

Multimodal Training Data for Dementia Care

A Foundational Resource for the AI Community

Team Noyce | University of California, Irvine

Submission to the 2025 AlzInsights AI Prize Committee

The Field is Missing Its Most Essential Tool: Data

0

Publicly Available Datasets

of natural conversations during dementia-related agitation

- **Current Landscape:** Existing resources like DementiaBank are limited to structured clinical interviews, not real-world interactions.
- **The Unmet Need:** There are **zero** publicly available datasets of natural conversations during dementia-related agitation.
- **The Bottleneck:** Without this data, developing truly empathetic and effective AI for dementia care is impossible.

Building the World's First High-Fidelity Dementia Conversation Corpus

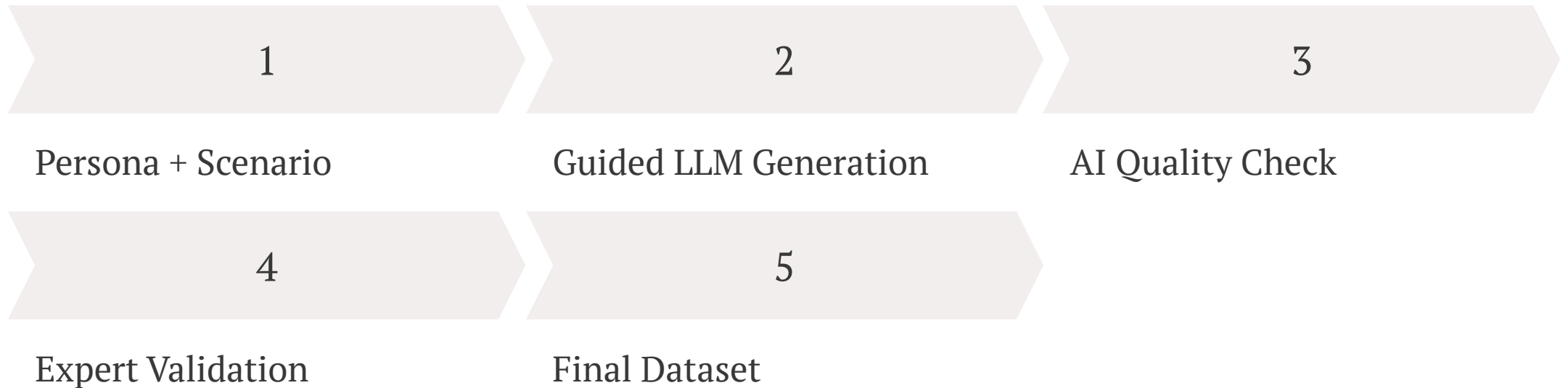
Our systematic framework generates a rich, diverse, and clinically-relevant dataset.

- 9 Expert-Designed Personas
- 35 Clinically-Validated Scenarios
- 315+ Base Combinations

Target: ~2,000 High-Quality Conversations

Ensuring Clinical Authenticity at Every Step

Our process combines AI generation with layers of human expertise to ensure the highest quality.



Live Demonstration

Proving Our Quality Assurance Process

Project Timeline

December 2025: Text Dataset Release

Public release of **2,000+ validated text-based conversations** for the global research community.

1

2

March 2026: Full Multimodal Release

Release of the **complete multimodal dataset**, adding prosody-tuned audio and behavioral annotations.

Unlocking Progress for the Entire Field

We are not just building another dataset. We are building the **foundational infrastructure** that will enable and accelerate innovation for every dementia AI team worldwide.

Thank you

Team Noyce | University of California, Irvine

Contact: cheonkaj@hs.uci.edu