

From Quotes to Themes: Automating Qualitative Insights from IDI Transcripts in Mental Health and Psychological Wellbeing Help-Seeking

Abstract

We present an Explorer-Replicator pipeline that transforms in-depth interview (IDI) audio recordings into interpretable, auditable thematic summaries of mental-health help-seeking-without losing participants' voices.

From forty Polish IDIs (approx. 21,000 utterances, 40 hours of audio), we analysed 3,520 respondent answers after automated transcription, diarization and denoising. The Explorer stage applies Latent Dirichlet Allocation (LDA, $K=14$) to identify coherent, semantically grounded topics of psychological wellbeing and help-seeking. The Replicator stage aligns these topics with twelve human-coded categories using TF-IDF similarity mapping (character 3-5 n-grams) and data-driven thresholding, achieving approximately 70% coverage of all answers. The resulting 26 briefs (14 machine-derived, 12 human-coded) integrate both data-driven and interpretive perspectives. Agglomerative clustering of the briefs yields five higher-order meta-themes: professional roles and hierarchies, access and stigma, everyday coping, cultural routines, and values and spirituality. This hybrid, auditable approach separates discovery from confirmation, supports reproducible qualitative reporting, and provides a scalable pathway from raw narratives to defensible thematic insights in psychological wellbeing research.