



Understanding Help-Seeking Behaviors in the U.S. Submarine Force: Comparing Non-AI and AI Augmented Qualitative Analytic Methods

Sruthi Chintalacharuvu, BS^{1,2}, D. Angus Clark, PhD^{1,2}, Justin Handy, PhD², & Dominica Hernandez, PhD²

¹Leidos Inc, ²Naval Submarine Medical Research Laboratory



NMF and LLM Show High Thematic Agreement Across Response Categories

	Perceived institutional barriers	Practical constraints	Peer / leadership support	Help-seeking facilitators
Confidentiality concerns	0.78	0.12	0.08	0.05
Career impact fears	0.82	0.09	0.11	0.04
Social / peer judgment	0.71	0.10	0.38	0.06
Time constraints	0.14	0.84	0.07	0.09
Access / availability	0.18	0.79	0.10	0.12
Peer / leadership support	0.11	0.08	0.88	0.22
Help-seeking facilitators	0.06	0.11	0.24	0.67

Rows = NMF subthemes | Columns = LLM broad themes | Values = Jaccard similarity index (0-1)

Low Agreement High Agreement

Results

Theme Convergence

- NMF and LLM demonstrated substantial thematic convergence – overall mean Jaccard similarity = 0.70

Thematic Granularity

- NMF produced finer-grained subthemes than LLM.
- For example, LLM grouped help-seeking facilitators with peer/leadership support.

Efficiency & Feasibility

- LLM required less runtime (~2 min vs. ~12 min), representing an 83% reduction in processing time.
- LLM analysis required only prompt engineering skills, whereas NMF required Python/R proficiency.

Conclusion & Implications

LLM and NMF produced comparable results, indicating promise for AI-augmented NLP in thematic analysis. Comparisons with human-centric inductive thematic analysis are recommended.

For exploratory analysis, LLM provides sufficient fidelity; for intervention design requiring subtheme precision, NMF remains preferable or should be used in combination.

Future work should examine output optimization through additional prompt engineering, multi-rater validation of LLM outputs, and applicability across broader military health domains.

Sponsor Acknowledgement

The authors would like to acknowledge the Department of Navy Office of Force Resiliency (DON OFR) for supporting this work.

Introduction

Military health research frequently relies on qualitative data to understand the “why” behind complex health behaviors.

Traditional qualitative analysis is labor-intensive and difficult to scale. Although Artificial Intelligence (AI) and Natural Language Processing (NLP) are increasingly prevalent, their use for qualitative analysis requires validation.

This study evaluated whether AI-assisted qualitative methods produces findings comparable to traditional analytic approaches and supports integration into military health research workflows.

Illustrative Use Case

Qualitative data on how stigma influences relationships between leadership/command climate, mental health, and treatment-seeking among submariners.

Methods

Sample: Secondary survey data from 413 active-duty U.S. submariners.

Data Characteristics: Open-ended responses from submariners detailing reasons why they would (or would not) use services/programs to address mental health difficulties.

Analytic Approach: Parallel qualitative analyses comparing results of **Non-Negative Matrix Factorization (NMF)** with a **Large Language Model (LLM)-based** approach (Gemini 3.1 Pro accessed on DoW GenAI).

LLM Reproduces NMF-Identified Themes While Substantially Reducing Analysis Time

