Mapping Writer-AI Conversations for HTA: An Analysis of 12.500 Messages

HTA227

A. Wiehe, P. Čuk and F. Woeste.
Pharos Labs, Germany

Introduction

Health technology assessment (HTA) reports synthesize complex clinical and economic evidence to inform reimbursement decisions. The writing process is time-intensive, requiring writers to navigate diverse sources, synthesize findings, and maintain regulatory compliance. AI-powered writing assistants offer potential to accelerate evidence synthesis, but their real-world usage patterns remain poorly understood.

Objective

To identify tasks an agentic AI assistant supports during HTA writing and how conversational tone varies, informing interface and model refinements that could shorten evidence timelines.

Methods

Data Collection: Usage logs from an AI-powered HTA writing assistant captured 7,009 messages from 32 writers (February 1 – October 15, 2025) after excluding internal test users and low-activity accounts.

Task Classification: Messages were embedded using OpenAI's text-embedding-3-large model. We developed a hybrid LLM-embedding clustering approach: Gemini 2.5 Pro [2] analyzed 300 diverse messages to identify natural task categories along two dimensions (Action and Topic). Gemini 2.5 Flash [2] classified these samples, and high-confidence classifications were used to compute embedding centroids. All messages were then assigned to categories via cosine similarity to centroids.
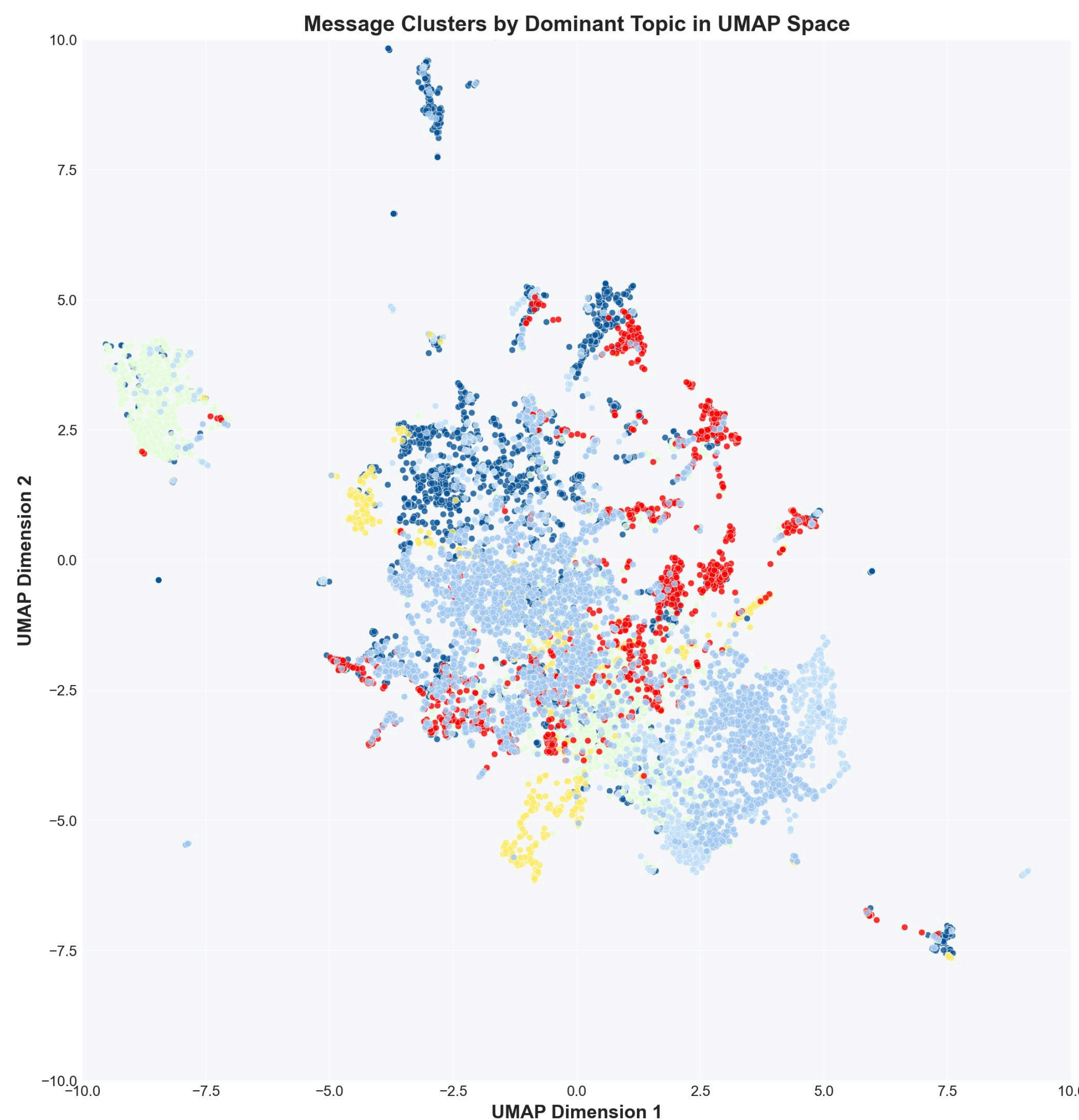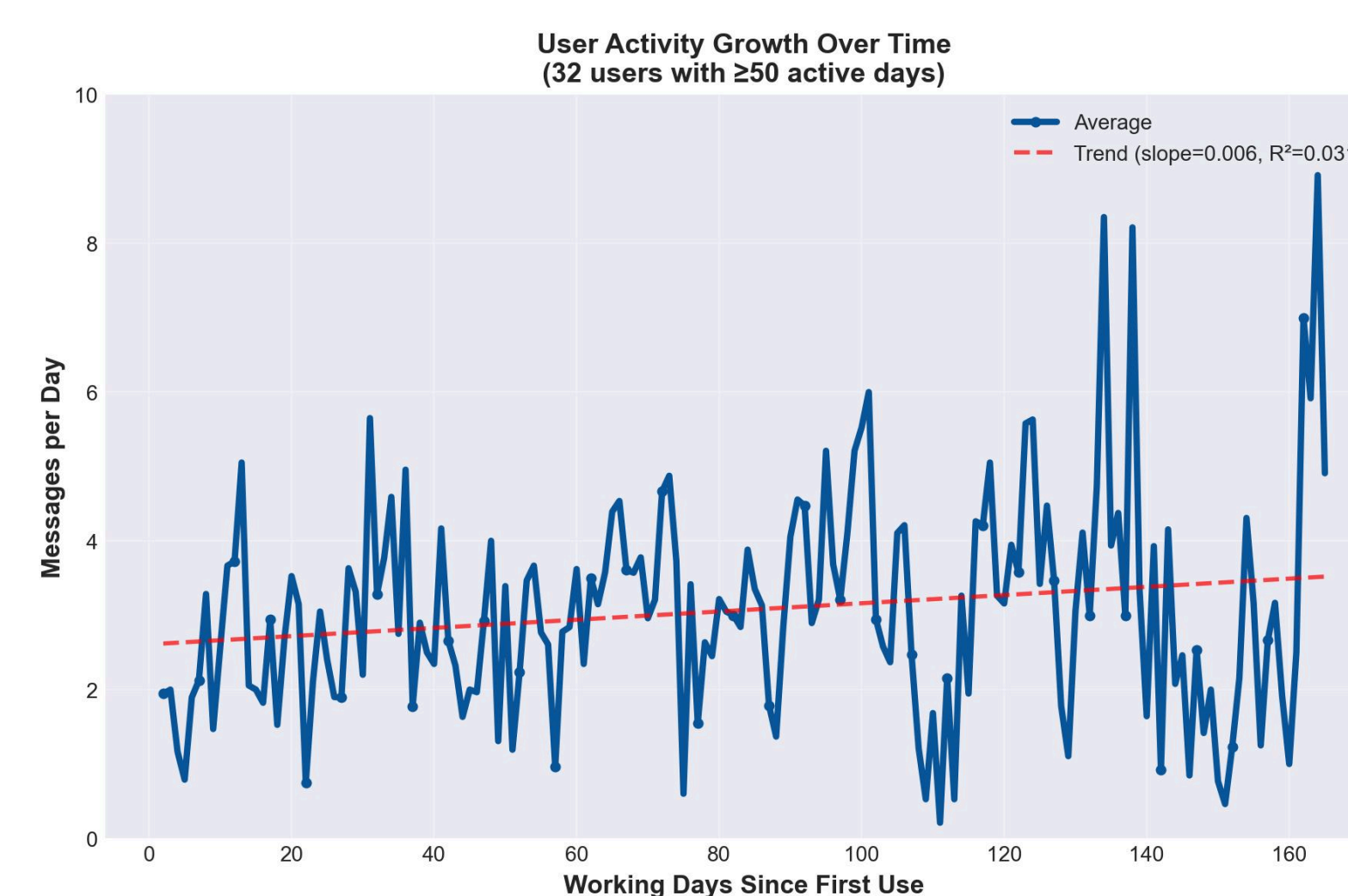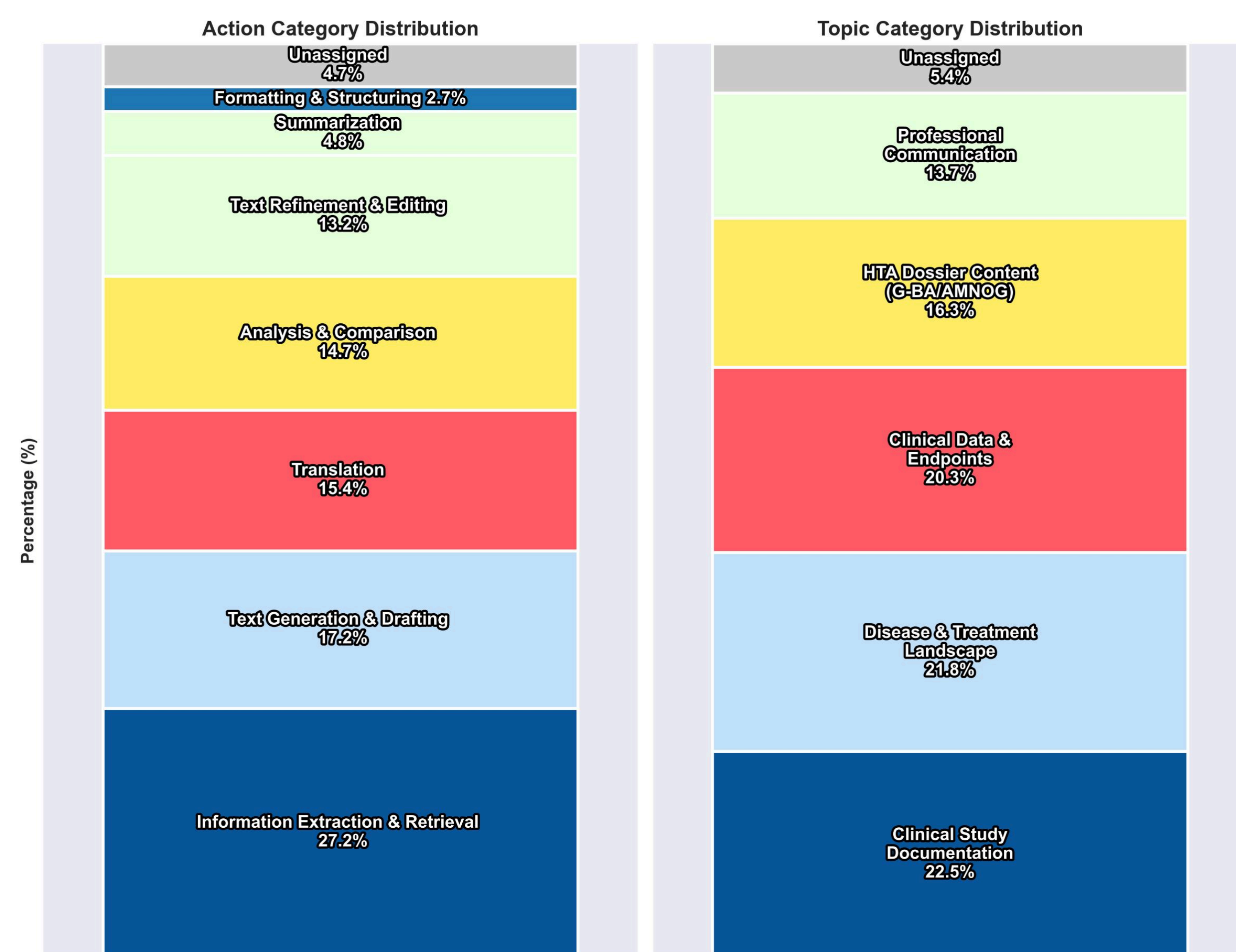
Visualization: UMAP [1] reduced embeddings to 2D for visualization. User adoption trends were assessed via linear regression.

Results

Analysis of 32 writers with over 50 active days revealed a modest but statistically significant increase in engagement over time (slope=0.006, p=0.024). Average daily messages grew by 32.4% from the early (days 1-20) to the late period (days 50-165). However, high daily variability (SD > 6) suggests intermittent, project-based interaction rather than consistent daily use.

A hybrid LLM-embedding clustering approach identified the most frequent user actions as Information Extraction & Retrieval (27.2%) and Text Generation & Drafting (17.2%), predominantly applied to topics like Clinical Study Documentation (22.5%) and Disease & Treatment Landscape (21.8%). UMAP projection of message embeddings confirmed this structure, revealing distinct semantic clusters for procedural tasks like translation, separate from overlapping, content-focused query clusters.

User Activity Growth Over Time (32 users with ≥50 active days)

Category Distribution Overview — Action Category Distribution / Topic Category Distribution

Message Clusters by Dominant Topic in UMAP Space

Conclusions

AI assistants in HTA writing are primarily leveraged as a specialized 'toolbox' for high-value sub-tasks like information extraction and translation, rather than for general content creation. User engagement is intermittent, aligning with project-based workflows. The clear semantic clustering of tasks suggests that future AI tools should de-emphasize generic chat in favor of specialized interfaces optimized for these dominant use cases. By aligning the assistant's functionality with observed user behavior, we can better support HTA writers, streamline evidence synthesis, and ultimately help shorten regulatory evidence timelines.

References

1. McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction
2. Google DeepMind. (2024). Gemini: A Family of Highly Capable Multimodal Models. Technical Report.

The complete list of references can be found under pharos-labs.com/research

Contact

More information on the survey

Anton Orell Wiehe, Head of AI
✉ anton.wiehe@pharos-labs.com
in linkedin.com/in/antonwiehe
🌐 pharos-labs.com

Pharos Labs