README: Latent Traces – A Computational Framework for Assessing Hybrid Authorship in Human–AI Co-Creation

Author: Tsehaye Haidemariam

Affiliation: Norwegian Business School

Contact: tsehaye.haidemariam@bi.no

Date: 18.07.2025

Overview

This study introduces a novel computational framework for measuring hybrid authorship in text-to-text generative AI workflows. The approach integrates prompt engineering, cosine-based semantic similarity, originality scoring, and latent space projection (PCA and t-SNE) to quantify the extent and nature of human influence in AI-generated creative content.

Contents

This repository includes the following components:

- Manuscript File (.docx) – The full paper, including abstract, methodology, results, discussion, and conclusion.
- Figures (.png)– High-resolution visualizations of semantic similarity, originality scores, and latent space projections (Figures 1–8).
- Appendix– Prompt–output screenshots and raw texts organized by category for transparency and reproducibility.
- Baseline text– A csv file containing the five neutral baseline descriptions used to compute semantic deviation.
- Prompt_output_pairs.csv– A master file of all prompts–output pairs used in the study, organized by category (e.g., Poetic, Technical, Narrative, etc.).
- Semantic_similarity_scores.csv–Cosine similarity scores comparing prompts to outputs across all categories.
- Originality_scores_vs_baseline.csv – Cosine similarity scores comparing outputs to the baseline corpus to measure originality.
- Aggregated_similarity_by_category.csv– Aggregated average cosine similarity scores by prompt category (prompt vs. output).
- Semantic_drift_from_neutral.csv– Semantic drift values calculated by subtracting each category's similarity from the Neutral baseline.

- Pca_coordinates_prompts_outputs.csv– PCA-reduced TF-IDF vector coordinates of prompts and outputs for visualizing semantic space.
- Tsne_coordinates_prompts_outputs.csv–t-SNE-reduced TF-IDF vector coordinates of prompts and outputs for visualizing local semantic clustering.
- Pca_coordinates_baseline_outputs.csv– PCA projections comparing baseline corpus entries and outputs to illustrate divergence from conventional language.
- Tsne_coordinates_baseline_outputs.csv– t-SNE projections comparing baseline corpus entries and outputs.
- TF-IDF_Semantic_Similarity_Prompt_vs_Output_

Instructions for Reproduction

To replicate the analyses, users will require the following tools and packages:

- Python 3.10+
- scikit-learn (for TF-IDF vectorization, PCA, and cosine similarity)
- matplotlib / seaborn (for plotting)
- openTSNE or sklearn.manifold.TSNE (for t-SNE)
- SentenceTransformers (for advanced embedding models, optional)

Data files and code (if made available) will be structured in folders corresponding to each analysis step: preprocessing, vectorization, similarity scoring, visualization.

Contact

For questions regarding the methodology or use of this framework, please contact the corresponding author:
Tsehaye Haidemariam– tsehaye.haidemariam@bi.no

License and Acknowledgements

This research is intended for academic and non-commercial use. All datasets used (e.g., baseline corpus) were custom-curated and do not include copyrighted material.

Citation

If you use this dataset, framework, or visualizations in your work, please cite:

Haidemariam T. (2025). Latent Traces: A Computational Framework for Assessing Hybrid Authorship in Human–AI Co-Creation [Journal Name]