Hans William Alexander Hanley 汉斯·汉隶

Semantic Modeling • Natural Language Processing • LLM Safety • Multilingual Representation Learning

hhanley@cs.stanford.edu hanshanley.com themarginoferror.com
hans-hanley-0694a180 github.com/hanshanley Google Scholar

I build scalable methods for interpreting and evaluating large language models, including multilingual representation-learning systems, semantic modeling pipelines, and LLM-based evaluators and agents for high-fidelity reasoning and content analysis. My PhD research uncovered latent structure in massive text ecosystems, aligned LLM behavior across languages and domains, and supported safer, more reliable deployment of model-driven systems.

Education

Stanford University **PhD, Computer Science, 2025 | MS, Computer Science, 2024**
GPA: 4.02; Advisor: Zakir Durumeric; Meta Research PhD Fellow; NSF GRFP; Stanford EDGE Doctoral Fellow

University of Oxford **MSc, Statistical Science, 2020 | MSc, Computer Science, 2019**
Distinction (Highest Honors); Daniel M. Sachs Class of 1960 Graduating Scholar at Worcester College

Princeton University **BSE, Electrical Engineering, 2018**
Highest Honors, GPA: 3.982; Information Security & Privacy Concentration; Phi Beta Kappa; Tau Beta Pi; Sigma Xi

Professional Experience

Microsoft AI, Member of Technical Staff **April 2026–Present**
- Research and build large-scale language-model systems for AI safety and truthfulness, spanning multilingual representation learning, semantic modeling, and evaluation of model-generated content.
- Develop LLM-based evaluators and measurement pipelines to assess factuality, reasoning quality, and safety and integrity risks in deployed model-driven products.

Meta, Research Scientist, CAS / Central Applied Sciences **March 2025–April 2026**
- Researched, built, and deployed large-scale multilingual and multimodal LLM pipelines to model content semantics, identify creator verticals, and understand user interaction behavior across Instagram (over 3 billion users) for downstream recommender systems.
- Developed LLM-based evaluation methods and production metrics to assess political content quality, reply/comment impact, safety and integrity risks, and downstream effects on user experience.

Google Research, Health AI Student Researcher **Summer 2024**
- Designed LLM-based agents for structured querying and high-fidelity reasoning over large clinical datasets; implemented verification layers to detect reasoning failures under distribution shifts.
- Built evaluation frameworks measuring accuracy, safety, and adversarial robustness for LLM-mediated medical data interactions.

Research Highlights & Selected Publications

Tracking the Takes and Trajectories of News Narratives **USENIX Security 2025**
- Built a large-scale narrative-tracking system combining specially fine-tuned (e.g., LoRA) encoder-LLM semantic similarity models and zero-shot stance detection to track 146K narratives across 4K+ news sites and understand narrative pathways (e.g., anti-vaccine and anti-Ukraine influence campaigns).

Across the Firewall: Foreign Media's Influence on Chinese Social Media **PNAS 2025**
- Quantified foreign-media penetration into Chinese platforms using cross-lingual alignment pipelines to detect narrative borrowing on the Russo–Ukrainian War.

Hierarchical Level-Wise News Article Clustering via Multilingual Matryoshka Embeddings **ACL 2025**
- Proposed multilingual Matryoshka embeddings enabling controllable semantic granularity (coarse-to-fine); achieved SOTA on SemEval 2022 Task 8 and built scalable hierarchical clustering pipelines surfacing stories, narratives, and macro-themes.

TATA: Stance Detection via Topic-Agnostic and Topic-Aware Embeddings **EMNLP 2023**
- Developed contrastive embedding models for topic-agnostic and topic-aware stance detection; achieved SOTA zero-shot performance (0.771 F1 on VAST) and improved cross-topic generalization in multi-domain news corpora.

Skills

Programming	Python, Java, C, C++, C#, Go, SQL, R, MATLAB; Git; PyTorch, TensorFlow, Hugging Face
ML/LLM	Transformers; LLM fine-tuning; multilingual modeling; interpretability; feature attribution; contrastive learning; LLM-based evaluation; safety & robustness analysis
Languages	English (native); Mandarin Chinese (HSK Level 5)

Selected Awards & Honors

- CSCW 2025 Best Paper Honorable Mention (Top 4%).
- Rising Stars in Data Science (Stanford/UCSD/UChicago); Meta Research PhD Fellow; NSF Graduate Research Fellowship (GRFP); Stanford EDGE Doctoral Fellow; University of Oxford Daniel M. Sachs Class of 1960 Graduating Scholar.
- Charles Ira Young Memorial Tablet & Medal; SEAS Hayes–Palmer Prize; George B. Wood Legacy Prize.