# Menglin (Miley) Liu

 Seattle, WA ✉ menglinliu@cuhk.edu.cn ☎ 530-761-6139 🔗 https://menglinmileyliu.github.io/index.html

## Professional Summary

Applied research engineer and computational social scientist with Ph.D.-level training and 5+ years of experience building end-to-end NLP and survey analytics systems. Specialized in LLM-based text classification, interpretable modeling, and decision-oriented analytics for non-technical stakeholders. Strong track record translating ambiguous research and policy questions into scalable, production-ready pipelines and user-facing analytical tools.

## Experience

**Assistant Professor of Computational Social Science**                                    *2025 – Present*
*CUHK-Shenzhen*

- Led applied research and system development at the intersection of NLP, survey analytics, and social data.
- Designed and deployed LLM-based pipelines for text and survey analysis used in active research and product prototyping.
- Supervised applied research translating real-world questions into scalable analytical systems.

**Founder & Lead Research Scientist**                                    *2025 – Present*
*SurveyFluency*

- Founded and led development of an AI-driven survey intelligence platform for rapid, interpretable analysis of complex survey data.
- Designed system architecture supporting SPSS and longitudinal survey ingestion, demographic crosstabs, toplines, and visual analytics.
- Built natural-language interfaces for structured survey exploration, enabling users to identify subgroup differences and temporal patterns without code.
- Applied the platform to large-scale public datasets (e.g., General Social Survey cumulative file) to support early-stage exploratory analysis.
- Focused on interpretability, transparency, and researcher-facing workflows for applied social science and policy analysis.

**Data & Policy Analyst**                                    *June – Sept. 2023*
*Committee of 100 (Remote)*

- Created interactive geospatial dashboards visualizing property restriction laws and policy changes across U.S. states and cities.
- Built an end-to-end data pipeline from legislative data ingestion and cleaning to mapping-ready outputs and stakeholder-ready visual deliverables.

## Selected Applied Research Projects

**SurveyFluency: AI System for Survey Intelligence 🌐 — LLMs, NLP, Structured Data Reasoning**     *2024 – Present*

- Developed LLM-based pipelines for analyzing structured survey data, including topline generation, demographic crosstabs, and pattern discovery.
- Implemented natural-language querying over tabular survey data to support hypothesis-free exploration and rapid insight generation.
- Evaluated system outputs using real-world datasets (e.g., General Social Survey) to assess robustness across survey waves and sub-populations.
- Designed outputs to prioritize human interpretation, transparency, and researcher trust rather than automated decision-making.

**PoliPrompt: LLM-Based NLP Classification Toolkit** — Prompt Optimization, Transformers, PyTorch, Open Source     *Jun – Sep 2024*

- Developed a modular classification pipeline with prompt optimization and exemplar retrieval, achieving substantial accuracy gains with lower inference cost.
- Integrated Hugging Face Transformers and PyTorch for scalable training and evaluation across large text corpora.
- Released as an open-source PyPI package. 🔗🐍

**LLM-Based Analysis of Survey Open-Ended Responses** — Prompting, Topic Modeling, Interpretability     *May 2025 – Present*

- Built pipelines analyzing 50,000+ open-ended responses for effort, emotion, moral content, and ideological stance using multiple prompting strategies.
- Applied topic modeling and clustering to extract latent dimensions related to ambivalence, cross-pressure, and cognitive complexity.
- Validated outputs against human-coded benchmarks to ensure measurement reliability and interpretability.

**Data Processing for Housing Policy Debates** — ASR, Diarization, NLP Pipelines, CI/CD     *June 2025 – Present*

- Automated transcription and speaker diarization for 500+ hours of local government meeting audio; built reproducible pipelines for extraction and analysis.
- Developed CI/CD-ready processing workflows for cleaning, feature extraction, and downstream NLP tasks, reducing manual processing time substantially.
- Produced structured quantitative measures of policy discourse for non-technical research partners and decision-makers.

## Technical Skills

**Programming:** Python, R, SQL, Stata
**NLP / ML Libraries:** Hugging Face Transformers, PyTorch, scikit-learn, SpaCy, NLTK, Gensim, TensorFlow
**Data and Model Management:** DVC, Git, Weights & Biases, MLflow
**Cloud / Deployment:** AWS (EC2/S3), CI/CD workflows
**Other:** Causal inference, clustering, regression modeling, survey analysis