Hiring Team

Zonda (formerly Hanley Wood | Meyers Research)

Glasgow, United Kingdom

Dear Hiring Team,

I am excited to apply for the Machine Learning Engineer II (NLP) position at Zonda. With a strong foundation in natural language processing, machine learning systems development, and interdisciplinary problem-solving, I am eager to contribute to your team's mission of building scalable, impactful NLP products that balance innovation with efficiency and real-world usability.

As an ESRC-funded PhD researcher at the University of York, I have spent the past several years designing and deploying end-to-end NLP pipelines that combine rigorous research with practical engineering. My work bridges social science and computer science—giving me a unique perspective on how language models can be developed not just to perform well, but to serve users meaningfully.

I bring hands-on experience across the full ML lifecycle, from dataset curation and cleaning to model deployment and monitoring. For example:

- I prepared and released the Hansard OCR Dataset, a cleaned corpus of historical UK parliamentary debates, designed for training and evaluating OCR models like TrOCR—showcasing my ability to structure datasets for real NLP applications.
- In my OCR Accuracy Evaluation project, I benchmarked multiple OCR engines (Tesseract, PaddleOCR, EasyOCR) augmented with LLMs (Gemini, Qwen, Deepseek, ChatGPT-4o), measuring WER and CER improvements using `jiwer`. This involved batch processing logic to manage API rate limits—a practical solution for cost- and time-efficient inference.
- I developed a modular literature review toolkit using RAG and FAISS, where semantic search over document collections enables dynamic knowledge retrieval—directly applicable to agentic workflows.

I have extensive experience working with agentic frameworks and LLM orchestration:

- Built a historical debate simulator using LangChain and Google Gemini, where two RAG-powered agents simulate dialogue based on embedded historical texts. The system uses FAISS for retrieval and Gradio for interactivity, and is deployed as a live demo on Hugging Face Spaces.
- Designed a chatbot for university staff using internal survey data to explain EDI policy changes—demonstrating how NLP systems can be tailored to organizational knowledge and user needs.

My technical stack includes Python, PyTorch, Hugging Face, scikit-learn, spaCy, Docker, AWS, and PostgreSQL, and I have built CLI tools and APIs for efficient model training and data processing. I am proficient in vector databases and retrieval systems, having used FAISS extensively across multiple projects for semantic search and topic exploration.

I also prioritize model evaluation and maintenance:

- In my published work, *"Holistic Evaluations of Topic Models"* (arXiv, 2025), I introduced novel metrics such as Gini-based topic inequality, Pairwise Uniqueness Values (PUV), and corpus coverage analysis to assess trade-offs in BERTopic models—skills directly transferable to monitoring model drift and performance degradation in production.
- I implemented randomized parameter searches across 60+ model runs, logging configurations and outputs systematically—ensuring reproducibility and traceability, key aspects of technical documentation and model governance.

What sets me apart is my ability to communicate complex ML concepts clearly—a skill honed through teaching roles across primary, secondary, and university levels, as well as through public-facing writing and open-source documentation. I've led seminars, managed teams, and secured funding for community initiatives, all of which have strengthened my collaboration and stakeholder engagement abilities.

I am particularly drawn to Zonda's focus on meaningful, product-driven ML development and your commitment to inclusion and continuous learning. I would welcome the opportunity to contribute my expertise in NLP, agentic workflows, and responsible model design to your innovative team.

Thank you for considering my application. I look forward to the possibility of discussing how I can help advance your NLP product goals.

Sincerely,

Thomas Compton