

Kavyanjali Agnihotri

📍 Tübingen, Germany 📩 kavyagnihotri17@gmail.com 💬 kavyagnihotri 💬 kavyagnihotri

M.Sc. Machine Learning at Uni Tübingen — ELIZA Scholar

Education

Eberhard Karls University of Tübingen

M.Sc. Machine Learning

Tübingen, Germany

Oct 2024 – Feb 2027 (Expected)

- **Honors:** ELIZA Scholar (Zuse School in AI) — 24-month stipend.

- **Relevant Coursework:** Probabilistic ML, Statistical ML, NLP, Reinforcement Learning, Advanced Info Retrieval.

BITS Pilani

B.E. Computer Science (German Grade: 1.8 / CGPA: 8.49)

Hyderabad, India

Nov 2020 – July 2024

- **Honors:** Merit scholarship for performance after 2nd semester.

- **Relevant Coursework:** Software Engineering, Information Retrieval, Machine Learning, Design and Analysis of Algorithms, Data Structure and Algorithms, Database Systems.

Skills

Languages: English (Native/C2), German (A2 - Basic), Hindi (Native)

Programming: Python, C++, SQL (PostgreSQL/MySQL), Bash, Java, JavaScript

Machine Learning: Pandas, NumPy, PyTorch, Scikit-learn, Transformers, (HuggingFace), NLTK

Web & Frameworks: MERN Stack (MongoDB, Express, React, Node.js), REST APIs

Tools & DevOps: Git, Docker, Jenkins, CI/CD Pipelines

Experience

University of Tübingen | Scholar Inbox

Tübingen, Germany

Student Research Assistant (HiWi)

Dec 2024 – Present

- Implementing additional functionality to Scholar Maps feature and digest system to visualize multi-cluster research interests, improving user discovery of cross-disciplinary academic papers.
- Architecting PaperAtlas features by leveraging LLMs to generate personalized author recommendations and conference summaries based on high-dimensional user interest embeddings from the recommender system.
- Spearheaded the platform redesign in Figma and implemented modular frontend components using React.js, ensuring a seamless transition from design mockups to the Python backend.
- Refactored feature-toggle architecture using bit manipulation to store user preferences in a single bitmap column; eliminated an entire auxiliary table, significantly reducing database storage overhead and join complexity.
- Enhanced the reliability of the Scholar Inbox platform by resolving critical bugs and optimizing the sharing infrastructure for the system featured in: (arXiv:2504.08385).

Cisco

Bengaluru, India

Technical Intern

Jan 2024 – June 2024, June 2023 – July 2023

- Engineered a real-time diagnostic overlay for Secure Network Analytics, bridging Python logic with Bash-driven log parsing to surface runtime errors directly within the UI, reducing MTTR (Mean Time to Repair).
- Modernized 20+ microservices by upgrading 15+ legacy Java libraries, resolving critical dependency vulnerabilities and improving system stability across the codebase.
- Streamlined deployment by containerizing 8+ repositories using Docker and architecting Jenkins CI/CD pipelines for automated error analysis.
- Developed and fine-tuned a BERT-based QA model to automate internal product support, enabling instant retrieval of module update specifications and technical documentation.
- Achieved and maintained 95% unit test coverage across 25+ core files, ensuring high-availability and reliability for security-critical network analytics components.

Palmtree Infotech

Chennai, India

Technical Intern (NLP)

May 2022 – July 2022

- Evaluated 5+ LLMs, including GPT-J-6B and GPT-NeoX, to determine optimal performance for Medical Named Entity Recognition (NER) across complex clinical datasets.
- Experimented with Zero-Shot and Few-Shot learning workflows to extract multi-entity medical data, bypassing the need for extensive fine-tuning while maintaining high extraction accuracy.
- Cleaned and filtered clinical datasets by removing non-ASCII characters and irrelevant metadata, restructuring the text into document-level JSONL files to align with the data schema used by the GPT-NeoX library.

Selected Projects

Reinforcement Learning for Air-Hockey | *Python, TD3, PyTorch, hockey-env* Jan 2026 – Present

- Developed a Reinforcement Learning agent using the TD3 (Twin Delayed DDPG) algorithm, achieving an 80% win rate against the environment's built-in strong heuristic opponent.
- Built custom implementations of Prioritized Experience Replay (PER) and Pink Noise from scratch to investigate their effects on exploration and sample efficiency in continuous action spaces.
- Conducted comparative analysis of agent performance, evaluating the trade-offs between standard Gaussian noise and correlated Pink Noise regarding training stability and convergence.
- Implementing a self-play curriculum to increase agent robustness by training the model against adversarial versions of itself to handle increasingly complex defensive strategies.

Compositional Understanding in Stable Diffusion | *Stable Diffusion 3.5* September 2025 – Present

- Conducting a systematic evaluation of Stable Diffusion 3.5 Large regarding spatial reasoning; identified circa 60% failure rate across relational prepositions like "above," "below," and "to the right of."
- Designed a controlled benchmarking suite using Self-Bench objects and spatial modifiers (e.g., "on", "under", "to the right of") to isolate linguistic patterns that trigger compositional failures.
- Engineered an automated VLM evaluation pipeline using Qwen3-VL-30B (MoE); implemented Chain-of-Thought (CoT) prompting which improved alignment with human ground truth.
- Investigating the impact of text encoder architectures on image synthesis, specifically researching the transition from standard CLIP-based encoders to CLIC-inspired contrastive learning models to improve attribute binding.

Generation of Scientific Lecture Notes | *Python, BERT, BeautifulSoup, LDA* Jan 2023 – Mar 2024

- Automated the extraction of multimodal scientific content from Wikipedia using BeautifulSoup, curating a rich dataset of technical definitions, LaTeX formulas, and code snippets.
- Architected a Knowledge Graph to represent conceptual dependencies between scientific topics, enabling the structured and hierarchical generation of academic lecture notes.
- Utilized BERT embeddings for contextual multimodal element placement, ensuring that scraped images and formulas were semantically aligned with the corresponding text sections.
- Applied Latent Dirichlet Allocation (LDA) for topic segmentation to refine content organization and minimize semantic bias in automatically synthesized scientific definitions.

Achievements

AI Incubator (Batch 6, 2025): Selected for an intensive startup development program; engineered a AI prototype and developed a comprehensive product roadmap through workshops on investor pitching and UI/UX design.

Google Girl Hackathon 2023: Placed in the **Top 2.5%** of candidates nationwide among thousands of participants.

JPMC Code for Good 2023: Selected as a **Finalist** out of 10+ teams for building a full-stack web solution to streamline operational workflows for an NGO.

JEE Advanced 2020: Achieved an All India Rank (AIR) of **4860** (Top 10% of qualified candidates) in India's most rigorous engineering entrance examination.