

Jeng-Yu Chou

📍 Amherst, MA, USA 📩 jchou@umass.edu 🌐 <https://jelchou.github.io/> 🌐 jelchou

Education

University of Massachusetts Amherst PhD in Computer Science	Expected: May 2027
University of Massachusetts Amherst MS in Computer Science	Sep 2022 - May 2025
○ GPA: 3.925/4.0 ○ Advised by Brian Levine ○ Main Coursework: Secure Distributed Systems, Adv. Algorithms, NLP, Neural Networks, Machine Learning	

University of Massachusetts Amherst BS in Computer Science	Sep 2018 – Dec 2021
○ GPA: 3.869/4.0 ○ Main Coursework: Computer & Network Security, Applied Cryptography, Algorithms for Data Science, AI	

Publications

A Quantitative Analysis of Inappropriate Content, Age Rating Compliance, and Risks to Youth on the Whisper Platform Jeng-Yu Chou, Brian Levine	ARES (COSH) Jul 2024
Enabling Cross-Platform Comparison of Online Communities Using Content and Opinion Similarity Prasanna Lakkur Subramanyam, Jeng-Yu Chou, Kevin K. Nam, Brian Levine	EMNLP (Findings) Nov 2024

Research Experience

Graduate Research Assistant University of Massachusetts Amherst, UMass Rescue Lab	Amherst, MA Sep 2022 – Present
○ Research into AI-based content moderation of multimodal data, potential and dangers of generative AI, characterization of large online communities, and the application of ML to child rescue and AI safety topics ○ Designed and implemented an automated data collection and classification pipeline using Android Studio, Appium, and LLMs for detecting sexual, illicit, and inappropriate content in social media posts, providing quantitative insights into platform safety, community behavior, and age rating compliance challenges ○ Implemented knowledge distillation using VLM synthetic supervision to improve CLIP's performance on harmful content detection across multiple moderation tasks while maintaining computational efficiency	

Undergraduate Research Assistant University of Massachusetts Amherst, UMass Rescue Lab	Amherst, MA Jun 2021 – Dec 2021
○ Advisor: Brian Levine ○ Evaluated the danger social applications pose to minors ○ Utilized Python libraries and Twitter and Discord APIs to scrape and analyze toxicity patterns in text data	

Industry Experience

Software Development Engineer Automated Controversy Detection, Inc.	Amherst, MA Mar 2022 – Jul 2022
○ Engineered full-stack dashboard platform with React frontend and FastAPI backend, implementing JWT authentication, automated data refresh intervals, and real-time visualization of Twitter analytics including toxicity scoring, controversy detection, and sentiment analysis across multi-service architecture ○ Developed NLP service integrations and data pipeline infrastructure, including text summarization and	

stance detection services, implementing PDF keyword search with context extraction and highlighting capabilities, and integrating third-party APIs for geographic interest mapping and news article aggregation

Software Developer

Automated Controversy Detection, Inc.

Amherst, MA

May 2020 – May 2021

- Developed Chrome extension for content filtering and constructed data mining web crawlers for text analysis (Elasticsearch, Kibana)

Other Experience

Graduate Student Representative

University of Massachusetts Amherst

Amherst, MA

Oct 2025 - Present

- Represent the student body, participate in faculty meetings, interview and evaluate faculty candidates

Teaching Assistant

University of Massachusetts Amherst

Amherst, MA

Jan 2024 - May 2025

- CMPSCI345(Data Mgmt), CMPSCI 563(Internet Law/Policy), CMPSCI 220(Programming Methodology)

CSWomen Social Events Coordinator

University of Massachusetts Amherst

Amherst, MA

Dec 2023 - Present

- Organized networking, professional development, and professor/PhD research and career talk events

Director of Outreach

TechTogether Boston Hackathon

Boston, MA

May 2020 - Apr 2021

- Oversaw outreach initiatives at the high school, collegiate, and post-grad levels

Co-Founder

Microbial Identifier: iSPY Startup

Amherst, MA

Feb 2019 - May 2020

- Utilized Google AutoML Vision to identify morphology of bacteria

Projects

A Survey on Privacy and Safety in Diffusion Models

Dec 2024

- Conducted survey of attack vectors and defense mechanisms, categorizing privacy threats to generative models (e.g. membership inference, backdoor insertion, data leakage, and adversarial perturbations)
- Evaluated mitigation strategies for ensuring safe deployment of diffusion models, examining differential privacy techniques, content filtering, bias prevention, and backdoor detection while identifying trade-offs in computational cost, output fidelity, and scalability for real-world sensitive applications

Stochastic Meta-Learning for Augmentation Policy (SMAP):

Dec 2023

Enhancing Fine-Grained Image Classification

github: SMAP ↗

- Developed a novel optimized augmentation policy, Stochastic Meta-Learning for Augmentation Policy (SMAP), that leverages meta-learning to optimize augmentation strategies for enhanced classifier performance
- Utilized a ResNet-50 model as the backbone classifier and compared the impact of SMAP against traditional usage of augmentation techniques

Examining Medical Narratives of Eating Disorder Recovery

May 2023

on Reddit

github: narrative-analysis ↗

- Fine-tuned BERT models to identify narratives in text and generated domain-specific sentiment lexicons
- Utilized ChatGPT instruction prompting for trigger and factor extraction and experimented with custom and NLTK stop words for topic modeling

Technical Skills

Python, Java, Javascript, Typescript, SQL, PyTorch, Multimodal Learning (Vision), Pandas, NumPy, Scikit-learn, Transformers, Git/GitHub, Slurm/HPC Environments