Aochong (Oliver) Li

(917) 945-9173 | al4143@columbia.edu | Google Scholar

EDUCATION

Cornell University   **New York, NY**
Ph.D. in Computer Science   Sep. 2024 – Present

Columbia University in the City of New York   **New York, NY**
B.A. in Computer Science; B.A. in Mathematics   *Aug. 2020 – May 2023*
- **Cumulative GPA:** 4.09/ 4.0; **CS Major GPA**: 4.20/ 4.0; **Math Major GPA**: 4.11/ 4.0
- **Award:** Russell C. Mills Award (**Top 4** undergrads awarded in CS dept. for academic excellence)
- **Latin Honors:** Summa Cum Laude; Dean's List (All Semesters)
- **Relevant Coursework:** Machine Learning, Natural Language Processing, Computer Vision, Real Analysis, Abstract Algebra, Analysis and Optimization, Probability, Statistical Inference, Stochastic Process

Washington University in St. Louis   **St. Louis, MO**
*B.A. Undeclared (***GPA:** 4.0/4.0; **Honors:** Dean's List*)*   *Aug. 2019 – May 2020 (Transferred)*

PUBLICATIONS

- **Oliver Li**, Mallika Subramanian, Arkadiy Saakyan, Sky CH-Wang, Smaranda Muresan. "NormDial: A Comparable Bilingual Synthetic Dialog Dataset For Modeling Social Norm Adherence and Violation". **Proceedings of Empirical Methods in Natural Language Processing (*EMNLP*)** 2023
- Sky CH-Wang*, Arkadiy Saakyan*, **Oliver Li**, Zhou Yu, Smaranda Muresan. "Sociocultural Norm Similarities and Differences via Situational Alignment and Explainable Textual Entailment". **Proceedings of Empirical Methods in Natural Language Processing (*EMNLP*)** 2023
- Sky CH-Wang, Evan Li, **Oliver Li**, Smaranda Muresan, Zhou Yu. "Affective Idiosyncratic Responses to Music". **Proceedings of Empirical Methods in Natural Language Processing (*EMNLP*)** 2022

RESEARCH EXPERIENCE

NormDial: A Comparable Bilingual Synthetic Dialog Dataset [paper]   **New York, NY**
Researcher | *Advisor: Smaranda Muresan*   *Jan. 2023 – Jul. 2023*
- Led a project to **generate social norm-annotated bilingual dialogues** and evaluate LLMs in **reasoning norm adherence and violation in dialogues**; drafted a paper that led to *1st author publication in EMNLP 2023*
- Proposed a novel human-in-the-loop pipeline to 1) synthesize 4231 dialogues in Chinese and English and 2) label each turn if there is a norm adherence or violation with explanation, with in-context learning (ICL) and ChatGPT
- Demonstrated that 1) the quality of NormDial **exceeded human-written dialogues** for multiple metrics 2) **lexical diversity** (N-grams) of NormDial was **higher than baselines** 3) NormDial dialogue situations covered 30+ topics
- Evaluated LLMs in reasoning norms: LLMs often failed to label norm violations and suffered from hallucination

Sociocultural Norm Similarities and Differences [paper]   **New York, NY**
Researcher | *Advisor: Zhou Yu and Smaranda Muresan*   *Sep. 2022 – Jan 2023*
- Proposed a **novel dataset of 6566 Chinese social norms**; **introduced a new task of reasoning norm similarities and differences** across US and Chinese cultures; drafted a paper that led to *publication in EMNLP 2023*
- Developed a pipeline to align Q&A pairs from Zhihu (Chinese Quora) to situations in Social-Chemistry-101 (a US social norm bank), using cross-lingual similarity between text embeddings extracted by XLM-R transformer
- Designed a **few-shot Chain-of-Thought prompt** to extract Chinese norms from Zhihu Q&A data with GPT-3
- Implemented a multivariate regression between the likelihood of cross-cultural norm contradictions and situation or norm topics derived by a LDA model; identified 3 situation and 2 norm topics correlated to norm contradictions

Affective Idiosyncratic Responses to Music [paper]   **New York, NY**
Researcher | *Advisor: Zhou Yu and Smaranda Muresan*   *Mar. 2022 – Jul. 2022*
- Investigated **psychological and lyrical factors** that affect listeners' emotional responses to music by processing 403M comments from a music platform *NetEase Cloud Music*; drafted a paper ***published in EMNLP 2022***
- Filtered text using regular expression and a clinical lexicon, DSM-5, to identify disclosers of mental disorders
- Demonstrated that disclosers differed from the population in diurnal user activity, playlist preferences, and genre features of songs extracted by a 20-topic LDA model on lyrics, inclining towards lonely themes during nighttime

INDUSTRY EXPERIENCE

AQR Capital Management   **Greenwich, CT**
Quantitative Researcher (ML Strategies)   *Jul. 2023 – Present*
- Developed a basket of trading signals from propriety patent data owned by each company; implemented a non-linear optimization on patent signals using tree-based ML models and **improved monthly performance by 32%**
- Implemented a pipeline to use patent citation, categories, and abstract info to represent each stock as embeddings
- **Proposed a novel statistical metric** (based on Wasserstein distance) to measure stock embedding distances; prototyped 36 trading signals based on pair-wise stock embedding distances and stock price momentum
- **Trained LightGBM and XGBoost models** to use trading signals to predict monthly stock returns; fine-tuned hyperparameters using grid search, achieving 0.91 R^2 score and 310% cumulative return in 10-year backtesting

AQR Capital Management (return as a full-time)   **Greenwich, CT**
Quantitative Research Intern (NLP Strategies)   *Jun. 2022 – Aug. 2022*
- Prototyped a framework to build **NLP strategies from news sentiment** extracted from a dataset of 130M news
- **Fine-tuned a BERT model** on financial text to classify news sentiment and **achieved 91% accuracy**
- Developed a long-short portfolio by **running an OLS model** to regress stock sentiment to its portfolio weight
- Achieved a 250% cumulative return, 1.4 high Sharpe Ratio, and 10% significant alphas in 10-year backtesting

TEACHING EXPERIENCE

Artificial Intelligence (COMS 4701)   **New York, NY**
***Head Course Assistant** | Instructor: Tony Dear*   *Jan. 2023 – May 2023*
Course Assistant | Instructor: Tony Dear   *Jan. 2022 – May 2022*
- Co-organized weekly meetings among 16 CAs and Dr. Dear to adjust class agenda and address student feedback
- Held weekly office hours and review sessions to help students with class materials; initiated Numpy101 sessions with other CAs to prepare students with necessary skills; **received 4.0/5.0 average student semester reviews**
- Designed homework, midterm, and final questions with Dr. Dear and assigned CAs with grading tasks; led a team of 2 CAs to implement an **auto-grader algorithm** and improve productivity for grading coding questions

Discrete Mathematics (COMS 3203)   **New York, NY**
Course Assistant | Instructor: Tony Dear   *Sep. 2022 – Dec 2022*
- Held weekly office hours to help a class of 200 students with writing proofs, using LaTeX, and debugging code
- Graded homework, hosted review sessions to resolve students' common questions, and proctored all the exams

DIVERSITY & INCLUSION

St. Louis Queer Support Helpline (SQSH)   **St. Louis, MO**
Front-End Software Engineer; Team Mentor   *Dec. 2019 – Present*
- Led 3 students to design our internal website user interface for new member training and daily work doc sharing
- Prototyped a map to label LGBTQ friendliness of each location by classifying Google reviews with BERT models
- Served as a board member to oversee tech support operation; served as a mentor to encourage STEM education

SKILLS

Languages: Python, Java, C, SQL, R, Go, Shell, Bash, Unix/Linux
Frameworks & Tools: PyTorch, TensorFlow, Keras, Scikit-Learn, Numpy, Pandas, MySQL, MongoDB, Neo4j