Fiona Anting Tan

https://www.linkedin.com/in/fiona-anting-tan

Email : fiona.anting.tan@gmail.com
Mobile : +65 9022 0307

EDUCATION

- **National University of Singapore** — **Singapore**
  Ph.D. in Natural Language Processing; GPA: 4.88/5.00 — *Aug 2020 – Jun 2024*
  - Dissertation titled "To know the causes of things: Text mining for causal relations", designing **neural networks for NLP**, building and harnessing **knowledge graphs (KGs)**, and creating **datasets and evaluation** methods
  - Awards: President's Graduate Fellowship Scholarship (Full Ph.D. scholarship, annual stipend of S$43,200) [2020 – 2024], NUSGS Research Incentive Award (Additional monthly stipend of S$900 for excellent research performance) [2023 – 2024]
- **London School of Economics and Political Science** — **United Kingdom**
  B.Sc. in Econometrics and Mathematical Economics; First Class Honors (Equiv. GPA $\geq$ 3.7/4.0) — *Sep 2014 – Jul 2017*
  - Dissertation titled "Does competition affect workers' subjective wellbeing?" using trade exports and imports as **instrument variables** for **causal inference**
  - Awards: Singapore-Industry Mid-term Overseas Scholarship (Partial undergraduate scholarship, total stipend of S$60,000) [2015]

PROFESSIONAL EXPERIENCE

- **Research Scientist / Machine Learning Engineer** — **Instagram, Meta**
  Menlo Park, CA — *July 2024*
  - Ranking and **recommender systems** under Instagram Ads
- **Applied Scientist II Intern** — **Amazon**
  Seattle, WA — *Oct 2023 – Feb 2024*
  - Researched on deductive reasoning of **large language models (LLMs)** in identifying underlying cause of incidents
  - Conceptualized an internal **Root Cause Analysis** tool to assist managers in incident reporting based on **chain-of-thought (CoT)** and KG-enhanced **retrieval-augmented generation** LLMs, estimated savings of >1000 days of man-hours a year
- **Decision Science Manager Intern** — **American Express**
  Singapore — *Jul 2023 – Oct 2023*
  - Fabricated **fuzzy regex matching algorithm** to automatically identify transaction of interest from call logs
  - Introduced unsupervised classification of call transcripts by scam types. Call logs were matched to semantically similar scam definitions based on **embedding vector similarity**. Newly identified scam cases enabled (1) sizing of scam types in US market, and (2) Strategy Team to update business definitions for new scams types
- **Data Science Consultant (Part-time)** — **factorem.co (Seed-stage)**
  Singapore — *Jan 2021 – Dec 2022*
  - Designed a **CNN** model based on images of 3D CAD files and other structured data to predict supplier product price
  - Invented product similarity matching algorithms using **hashing, self organizing maps**, and **vector-based search**. Code is deployed and crucial for repeat order pricing, resulted in automation for 30% of customer enquiries on platform
- **Data Science Analyst / Management Associate** — **Charles & Keith**
  Singapore/ Shanghai, China — *Aug 2017 – Aug 2020*
  - Established an ApacheHive ETL pipeline for Tableau and R dashboards to monitor, forecast, and propose store, region, and country level stocks (>200 stores). Built a **gradient boosting tree** and **time-series** model to forecast product demand and inventory levels at SKU level (>6000 SKUs), conducted feature importance analysis using SHAP values
  - Developed an SQL ETL pipeline for a Dash App to monitor and propose international stock transfers ($\sim$10 major countries). App deployed and used weekly by inventory planners.
  - Analyzed impact of omnichannel cannibalization through **difference-in-difference (DiD) regression** for country managers to calibrate strategy for new market launches
  - Performed **hierarchical cluster analysis** of product preferences by markets ($\sim$35 countries) and presented findings to management, international merchandise planners and product designers
  - Shaped business logic to manage inventory onto dot net framework, saving >1000 man-hours
  - Pioneered **digitization** efforts in firm by: (1) Presented business insights to C-suite frequently, (2) Facilitated department heads' understanding and planning of using data science in company workflows, and (3) Created and ran a three-month company-wide Python workshop, guiding >20 participant projects on business-related automation

SKILLS

- **Programming Languages & Platforms:** Python, R, SQL, SPARQL, ApacheHive, ApacheSpark, Docker, Tableau, Stata
- **Machine Learning Libraries & Frameworks:** pytorch, tensorflow, wandb, huggingface (transformers, datasets, evaluate), CUDA, NLTK, stanza, spacy, sklearn, pandas, numpy, matplotlib

ACADEMIC ACTIVITIES

- **Shared Task Organizer**
  Held under CASE Workshop @ EMNLP 2022 and RANLP 2023 *2021 – 2023*
  - Led a team of annotators and curators to create an Event Causality Identification dataset from scratch. Worked with linguists to establish annotation guidelines. Designed evaluation metrics and baseline models.
  - Implemented an annotation tool as a static website hosted on DigitalOcean cloud service
  - Hosted shared task on Codalab with customized evaluation Docker environment
  - Executed two iterations of shared task, attracting ∼50 teams and model submissions that beat baseline scores
- **Reviewer**
  Multiple Venues: COLING 2022, IJCAI 2023, JBCB 2023, IPM 2023, LREC-COLING 2024, NAACL 2024 *2022 – Now*
- **Teaching Assistant**
  BT5126: Post-graduate business analytics module *2021*

PUBLICATIONS

- Fiona Anting Tan. To Know the Causes of Things: Text Mining for Causal Relations. **AAAI-DC 2024**.
- Fiona Anting Tan, H. Hettiarachchi, A. Hürriyetoglu, N. Oostdijk, T. Caselli, T. Nomoto, O. Uca, F.F. Liza, and S.K. Ng. RECESS: Resource for extracting cause, effect, and signal spans. **IJCNLP-AACL 2023**.
- Fiona Anting Tan and See-Kiong Ng. 2023. Economics assistant for robustness checks (EconARC): Identifying confounders from causal knowledge graphs. **ISWC 2023**.
- Fiona Anting Tan, H. Hettiarachchi, A. Hürriyetoglu, N. Oostdijk, O. Uca, S. Thapa, and F.F. Liza. Event Causality Identification with Causal News Corpus - Shared Task 3. **CASE @ RANLP 2023**.
- Fiona Anting Tan, Debdeep Paul, Sahim Yamaura, Miura Koji, and See-Kiong Ng. Constructing and Interpreting Causal Knowledge Graphs from News. **AAAI-SS 2023**.
- Fiona Anting Tan, Xinyu Zuo, and See-Kiong Ng. UniCausal: Unified benchmark and repository for causal text mining. **DAWAK 2023**.
- K. Dhole, V. Gangal, S. Gehrmann, A. Gupta, . . . , Fiona Anting Tan, . . . (many authors). NL-Augmenter: A Framework for Task-Sensitive Natural Language Augmentation. **NELTJ 2023**.
- Fiona Anting Tan, H. Hettiarachchi, A. Hürriyetoglu, T. Caselli, O. Uca, F.F. Liza, and N. Oostdijk. Event Causality Identification with Causal News Corpus - Shared Task 3. **CASE @ EMNLP 2022**.
- Fiona Anting Tan, A. Hürriyetoğlu, T. Caselli, N. Oostdijk, T. Nomoto, H. Hettiarachchi, I. Ameer, O. Uca, F.F. Liza, and T. Hu. The Causal News Corpus: Annotating Causal Relations in Event Sentences from News. **LREC 2022**.
- Fiona Anting Tan, D. Hazarika, S.K. Ng, S. Poria, and R. Zimmermann. Causal Augmentation for Causal Sentence Classification. **CI+NLP @ EMNLP 2021**.
- Fiona Anting Tan and See-Kiong Ng. NUS-IDS at FinCausal 2021: Dependency Tree in Graph Neural Network for Better Cause-Effect Span Detection. **FNP 2021**.

AWARDS

- AAAI Travel Award (Granted complimentary conference registration and $2600 for travel and accommodation) [2024]
- 1st place in FinCausal Shared Task on cause-and-effect span detection (Prize money of $650) [2021]
- Finalist (Top 7 out of 60 teans) in AWS x WomenWhoCode ML Hackathon (Awarded $500 AWS credits) [2020]

TALKS & PRESENTATIONS

- Panelist, “Women in Science at Amazon - A Conversation with our Amazonians”, Amazon Programs, Seattle, WA, Nov 2023
- Poster Presentation, “Mining for Causal Relations in Text”, Hyundai Vision Conference, Seoul, Korea, Aug 2023
- Talk, “Extraction of Causal Relations in Text”, IDS PhD-Teach-PhD Workshops 2022, Singapore, Sep 2022
- Talk, “Fairness in NLP”, WING Reading Group, Singapore, Oct 2020