

TONGAN CAI

c-t-a.me | (734) 239-0723 | cta@psu.edu | [Google Scholar](#)

EDUCATION

Pennsylvania State University

Ph.D. Candidate in Informatics – Advisor: James Z. Wang

(Anticipated Dissertation: 11/24)

University Park, PA

08/2019 – Present

University of Michigan

B.S.E in Data Science - Magna Cum Laude

Ann Arbor, MI

09/2017 – 05/2019

Shanghai Jiao Tong University

B.S.E in Electrical & Computer Engineering - Outstanding Graduate-Class of 2019

Shanghai, China

09/2015 – 08/2019

SELECTED RESEARCH & WORK EXPERIENCE

(↔ Key Skills)

The Pennsylvania State University & Houston Methodist

08/2019 – Present | University Park, PA

Research Lead – *DeepStroke* Project

↔ [Multimodal](#) | [Computer Vision](#) | [Mobile Dev](#) | [Video/Speech Processing](#)

- Advance ER stroke triage with novel **multi-modal deep learning** frameworks that outperforms human performance in triage. Implement **real-time face tracking, and estimation** pipelines in **OpenCV/Dlib**, audio processing with **librosa**, and diagnosis models in **PyTorch/timm**. Develop an **RGB-D** video/speech audio collection tool and a processing/diagnosis tool in **iOS/Swift**.
- Leading top-tier publications: *MICCAI* 2020 Conference paper on **multimodal** deep learning with facial image and speech text, *Medical Image Analysis* 2022 Journal paper on multi-level patient video/audio feature **fusion, fairness-aware** pre-training, and **adversarial training**. Two papers in submission, with one on **Gen AI/Video retargeting** for privacy, and one on **mobile AI, state space model, and transformers**. The latest model **surpasses** clinician performance by 14% specificity and 20% sensitivity.

Nationwide Mutual Insurance Company

05/2023 – 08/2023, 05/2024 – Present | Columbus, OH/Remote

Data Scientist Intern

↔ [LLM/RAG](#) | [NLP](#) | [AWS](#) | [Object Detection/Segmentation](#) | [Foundation Model](#)

- Conduct exploratory **NLP** for Agri-Business. Implement a multi-purpose local **RAG** pipeline of **LangChain, ChromaDB** and **llama-3** with **Ollama** on policy documents. Design and optimize prompts for cause of loss querying. Perform **SpaCy NER** and **POS** tagging for risk factor analysis. Join and clean 6M~ unstructured claims and 27.7M~ raw notes on **AWS S3/SageMaker**. Embed text with **nomic-embed** and build **LDA** model with **genism**. Build **text clusters** with **scikit-learn** and **RACplusplus**.
- Prototype **PoC** framework for automated building underwriting. Handcraft street-view building image dataset from Zillow and Google with object bounding boxes using **LabelMe**. Finetuned **YOLOv8** with 91% multi-class accuracy with limited resources. Integrate Segment Anything model (SAM) **ONNX** runtime with a multithread **GUI** to provide full object segments in 3 minutes.

Tencent America - Medical AI Lab

05/2021 – 08/2021 | Palo Alto, CA/Remote

Research Scientist Intern

↔ [3D Geometry](#) | [Language-guided Image Processing](#) | [OCR](#) | [3D Rendering](#)

- Improve **3D-based** medical document **image de-warping**. Synthetic dataset rendered with real document warping **meshes** and **PDF** images, cascaded encoder networks built for shape/texture representation and refinement. DewarpNet refined with text recognition loss using **BERT** embedding distance to penalize distortions and improve OCR, validate performance on doc3D.

The Pennsylvania State University – College of Information Sciences & Technology

05/2020 – 03/2024 | University Park, PA

Research Project Lead – Applied Cognitive Science Lab

↔ [User Study](#) | [HCI](#) | [Behavior Analysis](#) | [Cognitive Modeling](#)

- Design user study on popular bibliography management applications (Mendeley, Zotero, EndNote, RefWorks) by **quantitatively** evaluating users' physical efforts using **RUI** and cognitive loads with **NASA TLX**. Pilot **prospective-retrospective** survey to reveal preference changes and usability issues. First-authored manuscript received **Best Paper** in *AsianCHI* Symposium 2021.

Research Assistant – *AI-PLAX* Project

↔ [Vision-Language Contrastive Learning](#) | [Multi-Task Learning](#) | [Digital Pathology](#)

- Develop a **multi-task vision-language contrastive learning** framework for placental diagnosis. Design a feature de-composition scheme to alleviate semantic suppression, and a distributional feature re-composition as augmentation. Improve prior state-of-the-art by 4.3% in diagnostic performance and achieve 2x computational **efficiency**. Conference paper accepted by *MICCAI* 2023.

Vision System Data Engineer

↔ [Data Curation](#) | [Crowdsourcing](#) | [GUI Programming](#)

- Curate** in-the-wild movie **video** dataset (BoLD) with 13k clips and 100k frames for **Amazon MTurker** labeling. Collaborate on large-scale Laban Movement Analysis. Build a **multi-camera Kinect/VICON** interface for **3D** body action video collection.

REPRESENTATIVE PUBLICATIONS (* INDICATES EQUAL CONTRIBUTIONS)

[Tags]

(Patent) Wang, J.Z., Yu, M., **Cai, T.**, Huang, X., Wong, K., Volpi, J., Wong, S.T.C.. *Systems and Methods for Assisting with Stroke and Other Neurological Condition Diagnosis Using Multimodal Deep Learning*. Pub. No. US 2023/0363679 A1. Patent Pending.

(Journal) **Cai, T.**, et al. (2022). *DeepStroke: Efficient Stroke Screening Framework for Emergency Rooms with Multimodal Adversarial Deep Learning*. *Medical Image Analysis*, Vol. 80. p. 102522. (IF=13.8) [Multi-Level Fusion | Fairness-Aware | Adversarial Training]

(Conference) **Cai, T.**, et al. (2021). What Makes a Good Reference Manager? A Quantitative Analysis of Bibliography Management Applications. In *Asian CHI Symposium 2021*. pp. 64-69. (Best paper) [Multi-Level Fusion | Fairness-Aware | Adversarial Training]

ADDITIONAL INFORMATION

Programming: Python (8 yoe), R, HTML/CSS, LaTeX, SQL, C++/C, Java, MATLAB, SAS, Swift C#, Shell/Bash, JavaScript, D3
Domain Expertise: Computer Vision (Medical Image, Object Detection/Recognition, Segmentation, Body/Facial Motion Recognition and Tracking, OCR, Multimodal, 3D Reconstruction, Contrastive Learning), Natural Language Processing (RAG with LLM, Speech Recognition, Sentiment Analysis, NER, Clustering), Data Science (Visualization, Management, Map-Reduce, A/B Testing, Geospatial Analysis, EDA, User Modeling, Behavior Analysis)

Updated 7/2024