Bowen Wang

Mobile: +44 07354560949 Email: wwwency2003@outlook.com

Education

University of Bristol **Bristol, UK**
BSc Mathematics 2:1 Class Honours **2023.09 - 2026.06**
Core Courses: Statistical Machine Learning, Stochastic Optimisation, Mathematical Programming, Multivariable Calculus and Complex Functions, Ordinary Differential Equations 2, Optimisation, Metric Spaces, Probability 2, Statistics 2, Linear Algebra 2, etc.
MSc Engineering with Artificial Intelligent **2026.09 -**

Publication

SURE: Standardized Multimodal Model Onboarding and Reproducible Evaluation Framework
https://github.com/PigeonDan1/sure
Co-first Author | Accepted by NCMMSC 2026.03 – present
Supervisor: Prof. Kai Yu, IEEE Fellow

- Led standardized onboarding of **18 public speech/audio models across 6 task families**, integrating ASR, TTS, VC, SD, SA-ASR, and multimodal audio models into unified CLI-callable components.
- Built the root CLI, ModelRegistry, and **4-stage validation pipeline**, improving one-shot model onboarding from **12/18 to 18/18 (100%)**.
- Developed reproducible runtime and inference validation with standardized manifests, dependency checks, failure classification, and model-local handoff bundles.
- Led the reproducible evaluation workflow across **7 ASR conditions and 2 TTS subsets**, revealing **0.02–0.52-point** protocol-dependent deviations from reported results.

SAER: Decomposed Evaluation Metric for Speaker-Attributed ASR
First Author | Submitted by ICASSP 2026.03 – present
Supervisor: Prof. Kai Yu, IEEE Fellow

- Led the end-to-end design of **SAER**, formulating a **time-constrained partial-order dynamic programming** alignment that decouples recognition error (E_W) from speaker-attribution error (E_S) across heterogeneous SA-ASR outputs.
- Derived an exact decomposition of **cpCER−CER** into a **speaker-consistency penalty** and a **cross-speaker order-relaxation gain**, proving that Δcp is not monotonic in speaker-attribution error and can be positive, zero, or negative.
- Evaluated **7 public SA-ASR pipelines** on AISHELL-4 and AliMeeting; controlled overlap tests increased (E_S) from **7.19%→29.58%** while Δcp changed only **13.36→14.46 pp**, demonstrating substantially cleaner error attribution. - Designed overlap and timestamp-sensitivity analyses over **4,539 AISHELL-4** and **11,062 AliMeeting** speaker turns; found **97.6%/98.3%** fall within the standard 5 s window and tcpCER can shift by **>20 pp** solely from collar settings, motivating a collar-free formulation.

Research Experience

LLM Agents • AI Evaluation • Speech & Multimodal AI • Mathematical Modellin
Shanghai Jiao Tong University 2026.03 – 2026.09
Supervisor: Prof. Kai Yu, IEEE Fellow

- Researched **LLM-agent-based and mathematically grounded evaluation methods for speech and multimodal AI**, combining agentic orchestration, reproducible execution, and principled metric design.

- Co-developed an **agentic evaluation framework** for heterogeneous AI models, standardizing **18 public speech/audio models across six task families** through structured execution contracts, runtime verification, and deterministic scoring, improving one-shot onboarding from **12/18 to 18/18**.
- Developed a **time-constrained partial-order dynamic programming metric** for speaker-attributed ASR that decomposes lexical and speaker-attribution errors; theoretical and controlled experiments exposed fundamental limitations of cpCER/tcpCER, including severe overlap and timestamp-induced evaluation distortions.

LLM-assisted AI • Trustworthy AI • Privacy & Security Analytics

Institute of Information Engineering, Chinese Academy of Sciences
2025.06 - 2026.03
Supervisor: Prof. Yan Zhang

- Researched LLM-assisted methods for scalable and trustworthy data annotation and risk analysis, using heterogeneous language models, statistical modelling, and structured evidence to replace labor-intensive and heuristic decision processes.
- Designed a multi-agent-style collaborative LLM workflow in which heterogeneous LLMs independently annotate privacy documents, critique conflicting predictions, and escalate unresolved cases for human adjudication; produced 28,000+ bilingual paragraphs with 13-way multi-label annotations and character-level evidence spans, reducing human intervention to 12.4%.
- Developed the associated data, quality-control, and evaluation infrastructure, including application-level 8:1:1 splits and reproducible BERT-family baselines; evidence-span localisation reached Micro-F1 = 0.91, demonstrating that LLM-generated supervision could support reliable downstream lightweight models.
- Applied statistical distribution modelling and graph-based structural mining to security-risk analysis, combining Zipf/log-normal modelling, path n-grams, domain–path–parameter associations, and heterogeneous security indicators into interpretable risk scores evaluated with ROC/AUC, KS, Precision@K, and Recall.

Competition

Aegis Professor Undergraduate Competition: Image Classification Comparison, Bristol, UK

Project Leader | 2025.05 – 2026.03

- Two end-to-end pipelines: Completed both a keypoint-based route (MLP/random forest) and a full-image CNN route, covering the full workflow of training, evaluation, and visualization.
- Four core figures: Created a pipeline overview, accuracy comparison, occlusion ablation experiment, and Grad-CAM heatmap for poster and presentation use.
- Interpretable findings for three facial regions: Tested >=6 occlusion combinations (ears/eyes/nose-mouth: occlude-only/display-hide), concluding that the nose-mouth region contributed the most, followed by the eyes, while the ears had the weakest impact.
- Evidence-based argument construction: Aligned key poster claims with literature (e.g., 48-keypoint methods, the importance of the nose-mouth region, and random forest robustness on heterogeneous samples), improving the persuasiveness of the presentation.

Skills & Interests

- **Technical:** Python, R, MATLAB, SQL, Git, Linux, Docker, LaTeX.
- **AI & Math:** LLM Agents, AI Evaluation, Machine Learning, Optimization, Dynamic Programming.
- **Languages:** Mandarin (native), English (IELTS 7.0), Korean (proficient).