

Education

2024-08 ~ 2025-12	Georgia Institute of Technology Electrical and Computer Engineering	M.S.
2020-09 ~ 2024-06	Shenzhen Technology University Internet of Things Engineering	B.Eng.

Open Source Projects

Understand Anything - Founder [81,730 GitHub stars](#)2026-03 ~ Present

Created an AI-driven codebase exploration product for understanding large repositories, exploring architecture, and accelerating developer onboarding.

- Structural analysis:** used Tree-sitter to extract functions, classes, import dependencies, and call sites into a code structure graph, with fingerprint-based incremental updates.
- Semantic graph construction:** combined code structure with LLM-generated summaries, architectural layers, and business semantics, using GraphRAG-like entity-relation modeling to support graph retrieval and codebase Q&A.
- Guided exploration:** developed architecture tours, domain views, dependency paths, graph-based search, and change-impact analysis to help users navigate unfamiliar code.
- AI coding integrations:** integrated graph generation and exploration into Claude Code, Codex, and Cursor so developers could query codebase knowledge from their existing coding environments. [Website](#)

dsh-browser - Founder [583 GitHub stars](#)2026-08 ~ Present

Built a native browser-control plugin for DeepSeek Harness, connecting its tool and session services to Chrome/Firefox MV3 extensions for operating users' existing browser sessions.

- Agent interaction protocol:** represented pages with structured snapshots and stable element IDs, returning post-action state changes alongside tool results to inform subsequent decisions.
- Human-agent collaboration:** added selection-to-prompt context and session-bound tab control, pausing actions on tab switches to confirm the intended target.
- Matched evaluation:** in six local tasks and 60 runs, achieved **20% lower mean latency** and reduced tool calls from **4.7 to 3.4** versus an in-project Playwright adapter; both achieved 30/30 successful runs.

Startup Experience

Egonex-AI - Co-Founder / CTO egonexai.com2026-01 ~ 2026-06

- January:** co-founded Egonex-AI to build emotion-aware AI recruitment products.
- March:** built Understand Anything after hours to make sense of large volumes of AI-generated code.
- June:** left following a relocation, donated Understand Anything to Egonex-AI after independently growing it to **50K+ GitHub stars**, and continued developing it independently after the donation.

Work Experience

4Paradigm - Senior Full-Stack AI Engineer	2026-06 ~ Present
Baidu - Image Search Algorithm Intern	2024-01 ~ 2024-04

- Worked on image-search algorithms, exploring ViT-based retrieval to improve semantic alignment between pretrained visual representations and downstream search tasks.
- **Representation learning:** developed semantic feature-learning strategies on ViT backbones to address the domain gap between pretraining data and image-retrieval scenarios.
 - **Hierarchical loss design:** introduced label constraints across semantic, global, and local feature spaces to jointly model category meaning, whole-image structure, and fine-grained visual details.
 - **Evaluation and iteration:** analyzed Recall@K and related retrieval metrics to compare loss functions, label granularity, and feature representations and guide subsequent model iterations.

Research Experience

Emotion-LLaMA & MER-Factory	Multimodal Emotion Understanding
Emotion-LLaMA: 618 GitHub stars MER-Factory: 119 GitHub stars	

- Contributed to **Emotion-LLaMA (NeurIPS 2024)** and developed [MER-Factory](#), the associated MERR data-generation pipeline, connecting multimodal annotation with instruction tuning for emotion recognition and reasoning.
- **Multimodal instruction data:** built a multi-agent workflow using OpenFace, FFmpeg, and multimodal models accessed through APIs, Ollama, and Hugging Face to extract emotional evidence from video, audio, and images and generate training instructions.
 - **Emotion-focused annotation:** used psychology-informed combinations of facial micro-actions to generate pseudo-labels, connecting observable facial behavior with emotional states and supporting the construction of emotion-understanding training data.
 - **Quality control:** introduced in-process and post-process checks with a gate agent using Plan-and-Solve and ReAct, alongside reference-free CLIP-Score and CLAP-Score metrics to assess cross-modal consistency.
 - **Visual representation learning:** explored CLIP- and VideoMAE-based pretraining on MER2023 videos to capture facial dynamics and strengthen visual features for emotion understanding.
 - **Model training and research output:** performed LoRA-based instruction tuning for Emotion-LLaMA, aligning audio, visual, and textual signals so the model could classify emotions and explain them using multimodal evidence; co-authored the NeurIPS 2024 paper.

Awards and Honors

- **Runner-up**, ACM Multimedia Emotion Recognition Challenge.
- **National Second Prize**, China Undergraduate Mathematical Contest in Modeling.
- **First Prize for Research Innovation** and **Craftsman Star Award**, Shenzhen Technology University.
- **Dahua Outstanding Scholarship**.
- **Outstanding Freshman Scholarship**, Shenzhen Technology University.

Selected Publications

Why We Feel: Breaking Boundaries in Emotional Reasoning with Multimodal Large Language Models

First Author | CVPR 2025, NeXD Workshop (Oral) | [Paper](#)

Introduces EIBench and coarse-to-fine self-questioning to evaluate and improve multimodal models' ability to interpret emotions and explain their causes.

Invisible Gas Detection: An RGB-Thermal Cross Attention Network and a New Benchmark

Co-First Author | Computer Vision and Image Understanding, 2024 (JCR Q2, CCF-B) | [Paper](#)

Introduces a multispectral gas-leak detection benchmark and RT-CAN, a cross-attention network that fuses RGB scene information with thermal signals.

Semi-Supervised Multimodal Emotion Recognition with Expression MAE

Second Author | ACM Multimedia 2023 (CCF-A) | [Paper](#)

Uses expression-aware masked autoencoding and unlabeled data to improve multimodal emotion recognition when high-quality annotations are scarce.

MIPS at SemEval-2024 Task 3: Multimodal Emotion-Cause Pair Extraction in Conversations with Multimodal Language Models

Third Author | SemEval 2024 | [Paper](#)

Combines dialogue context, acoustic features, and visual signals to identify emotions and their corresponding causes in conversations.

Emotion-LLaMA: Multimodal Emotion Recognition and Reasoning with Instruction Tuning

Co-Author | NeurIPS 2024 (CCF-A) | [Paper](#)

Aligns audio, video, and text through instruction tuning to recognize emotions and explain their causes using multimodal evidence.

FlexEdit: Marrying Free-Shape Masks to VLLM for Flexible Image Editing

Co-First Author | arXiv, 2024 | [Paper](#)

Combines natural-language instructions and free-shape masks for flexible image editing with fine-grained control over selected regions.