
SHERRY WANG / 2023-2026

PROJECT PORTFOLIO

Systems architect / AI builder

I build reliable AI systems, evidence-aware research pipelines, and products that can explain how they reached a result.

10

FEATURED CASES

18

VERIFIED ARCHIVE PROJECTS

01

PRODUCT HORIZON

BUILD SYSTEMS THAT CAN EXPLAIN THEMSELVES

blog.lifeexchange.me

Updated July 2026

PROFILE / EVIDENCE

A communication-to-systems trajectory

POSITIONING

Systems Architect / AI Builder

Cross-disciplinary work across communication research, RAG, agents, evaluation, data pipelines, full-stack products, and systems engineering.

53,954

SCOPUS RECORDS CLEANED

1,152

AGENT ACTIONS AUDITED

33%

REPORTED LEGAL RAG LIFT

1

ZERO-DAY CONTRIBUTION

TRAJECTORY

2023-24

COMMUNICATION + SECURITY

Public discourse research and LLM-assisted fuzzing.

2025

MODELS + SIMULATION

Optimization, field research, data quality, and social agents.

2026

AGENT SYSTEMS

RAG, MCP, browser evaluation, multi-agent orchestration.

NOW + NEXT

GROWTH OS

A working MVP evolving toward reflective decision support.

EVIDENCE MODEL

A

Source code, data, final report, or clearly cross-validated role.

B

Portfolio, resume, or attached project materials without an independent rerun.

SCOPE

Reported metrics keep their source qualifier; forecasts are never presented as outcomes.

FEATURED CASES / 01

Systems, decisions, and measurable evidence

FEATURED CASE / REPORTED METRIC

Legal Compliance RAG Agent

2026.01-PRESENT / R&D ENGINEER

A Singapore PDPA workflow that turns cross-document legal comparison into grounded, reusable consultation support.

SYSTEM

Parsing, chunking, retrieval, generation, multi-turn legal context, and evaluation were treated as one precision pipeline.

RESULT

Partner law-firm project records report a 33% improvement in the benchmark baseline.

EVIDENCE

RAG workflow, evaluation materials, and partner benchmark records. Metric not independently rerun for this portfolio.

Python / RAG / Prompt Engineering / Legal NLP / Evaluation

FEATURED CASE / VERIFIED ASSETS

Multi-site API Reverse Engineering + MCP

2026.01-2026.02 / ENGINEERING INTERN / INDIVIDUAL CONTRIBUTOR

A network-first workflow for discovering stable APIs and packaging fragmented site capabilities as typed, recoverable MCP tools.

SYSTEM

Endpoint and field discovery, Pydantic contracts, shared MCP core, startup checks, retries, checkpoints, and validation reports.

RESULT

14 groups of site capabilities; one consolidated change covered 231 files and roughly 20.5K added lines.

EVIDENCE

Source history, field documentation, startup tests, and validation reports. Internal sites and credentials remain private.

Python / MCP / Playwright / Pydantic / API Reverse Engineering

FEATURED CASES / 02

Systems, decisions, and measurable evidence

FEATURED CASE / VERIFIED ASSETS

Agent Evaluation Benchmark Frameworks

2026-PRESENT / INDEPENDENT RESEARCH

A confidence-aware benchmark for browser and assistant agents, designed to measure reliability beyond one successful demo.

SYSTEM

Four pillars: statistical confidence, judge agreement, task-distribution coverage, and cross-run stability.

RESULT

120 labeled tasks and 17 Chrome scripts; one 50-task batch logged 1,152 actions and 1,020 screenshots, plus 72 run audits.

EVIDENCE

Task set, automation scripts, screenshots, action traces, and run-level audit records.

Python / Wilson CI / Cohen's kappa / Bootstrap / Browser Automation

FEATURED CASE / REPORTED METRIC

Merchmind Multi-Agent Commerce System

2026.03-2026.06 / SYSTEM DESIGN + DEMO / TEAM PROJECT

Four domain agents and an orchestrator coordinate assortment, pricing, inventory, and category decisions with replayable arbitration.

SYSTEM

Deterministic tools and LLM reasoning are separated; six conflict rules preserve inputs, disagreements, decisions, and actions.

RESULT

Team records report simulated conflict rates falling from about 30% to below 5%, and trial cycles from 14 to 7 days.

EVIDENCE

Runnable demo, arbitration rules, replay traces, and presentation chain. Metrics were not independently rerun.

Multi-Agent / Orchestration / ERP Concepts / Explainability

FEATURED CASES / 03

Systems, decisions, and measurable evidence

FEATURED CASE / REPORTED METRIC

BDSC2025 Multi-Agent Rumor Intervention

2025.06-2025.07 / TEAM LEAD

A social-network simulation that models rumor propagation and resistance, then compares intervention strategies at system level.

SYSTEM

Qwen2.5-14B role agents, centrality and community analysis, plus deletion, persuasion, edge-breaking, and banning strategies.

RESULT

Materials document 200 nodes and 2,000 edges; team records report a 0.3 reduction in the spread index.

EVIDENCE

Competition code, model materials, network setup, and results. The original experiment was not rerun for this portfolio.

Python / Qwen2.5-14B / NetworkX / Multi-Agent Simulation

FEATURED CASE / REPORTED METRIC

LLMForFuzzing

2023.10-2025.03 / CORE CONTRIBUTOR

A PoC-to-language-to-semantic-mutation workflow that uses LLMs to expand exploration in PDF JavaScript fuzzing.

SYSTEM

PoC reproduction, Adobe and Foxit JavaScript API mapping, semantic mutation, and automated test-PDF generation.

RESULT

Project records report 24/32 reproduced PoCs, 200+ APIs, 2,000-3,000 test PDFs, and one UI-related zero-day contribution.

EVIDENCE

Source, test samples, and closeout materials. Scale and zero-day outcome were not independently rerun.

Python / ChatGPT API / PDF JavaScript / Fuzzing / Security

FEATURED CASES / 04

Systems, decisions, and measurable evidence

FEATURED CASE / VERIFIED ASSETS

People's Daily on Bilibili

2023.12-2025.12 / INDEPENDENT RESEARCHER

A three-iteration communication study of title semantics, tags, and engagement across People's Daily videos on Bilibili.

SYSTEM

2,465 videos, 13 fields, and a 50-item random audit; jieba, Sentence-Transformers, cosine similarity, and random forest.

RESULT

Mean semantic similarity was about 0.408; random-forest test R2 was only 0.028, retained as a finding boundary.

EVIDENCE

Notebook, final dataset, audit sample, figures, and model outputs.

Python / Web Scraping / Sentence-Transformers / scikit-learn

FEATURED CASE / VERIFIED ASSETS

Turing Digital Intelligence Platform

2025.09-2026.01 / FULL-STACK DEVELOPER / TEAM PROJECT

A three-role workflow platform for applications, expert review, admission, and incubation with permissions and audit history.

SYSTEM

Applicant, Expert, and Admin flows; relational model, attachments, JWT/BCrypt, audit logs, and an operational dashboard.

RESULT

Full lifecycle implemented across roughly 103 Java files, 20 Vue files, and 15 SQL files.

EVIDENCE

Runnable system, ER model, source structure, database scripts, and course presentation materials.

Vue 3 / Spring Boot / MyBatis / MySQL / ECharts / JWT

FEATURED CASES / 05

Systems, decisions, and measurable evidence

FEATURED CASE / VERIFIED ASSETS

LexMemoryBench: Household Long-Term Memory

2026-PRESENT / EVALUATION PIPELINE ENGINEERING / TEAM

A synthetic-data pipeline for evaluating cross-member, cross-session memory in household robots and assistants.

SYSTEM

Five traceable stages connect personas, constrained QA, minimal evidence, evidence conversations, and non-leaking filler sessions.

RESULT

The core typed pipeline runs and produces staged artifacts; the complete benchmark haystack and scaled runs remain in progress.

EVIDENCE

Pipeline documentation and 15 attributable commits. Current submission bundle is not a final benchmark result.

Python / Typed Schemas / Prompt Engineering / Synthetic Data QA

FEATURED CASE / VERIFIED ASSETS

Tang Poetry Imagery Universe

2026.06 / DATA ANALYSIS + INTERACTIVE VISUALIZATION

An interactive digital-humanities site that turns imagery co-occurrence across 57,607 Tang poems into an explorable network.

SYSTEM

Cleaning, Jieba segmentation, imagery matching, windowed co-occurrence, NetworkX centrality, and multi-view comparison.

RESULT

29,760 active windows produced 28 nodes and 342 edges; the moon-wind edge was strongest at weight 764.

EVIDENCE

Processing code, CSV/JSON, summary metrics, methodology, interaction design, and completed site.

Python / Jieba / NetworkX / PyEcharts / ECharts

PROJECT ARCHIVE / 01

Breadth with traceable delivery

PROJECT ARCHIVE / 2025.09-2026.03 / DATA RESEARCH

DEA Evolution + Bibliometric Pipeline

Cleaned 53,954 Scopus records into about 33,690 application papers, then mapped domains, scholars, collaboration, similarity, journals, and SPC main paths.

DELIVERY / RESULT

Reusable Python pipeline, structured datasets, network outputs, and research figures.

Raw and processed data / scripts / charts / Python / Scopus / DeepSeek / Bibliometrics / Network Analysis

PROJECT ARCHIVE / 2025.01 / MATHEMATICAL MODELING / TEAM PROJECT

2025 MCM Sustainable Tourism

Modeled Juneau overtourism through revenue, glacier, and social-impact submodels, entropy weighting, NSGA-II Pareto search, ARIMA, and sensitivity analysis.

DELIVERY / RESULT

22-page English paper and a tourism-committee decision memo.

Code / figures / paper / policy memo / Python / NSGA-II / ARIMA / LaTeX

PROJECT ARCHIVE / 2025.02-2025.07 / EVALUATION DESIGN

LLM Review Data-Value Assessment

Compared DeepSeek and G-EVAL with human ratings across authenticity, information value, and relevance using MSE, MAE, and error review.

DELIVERY / RESULT

2,559 raw and 1,589 processed samples are reported as different pipeline stages, not one cohort.

Datasets / scoring scripts / report versions / Python / DeepSeek / G-EVAL / Human Evaluation

PROJECT ARCHIVE / 02

Breadth with traceable delivery

PROJECT ARCHIVE / 2025.11 / INDEPENDENT DATA ANALYSIS

AI + Employment Opportunity Restructuring

Automated cleaning, analysis, visualization, and reporting for 30,000 job postings across education, skills, industries, regions, salary, fairness, and automation risk.

DELIVERY / RESULT

Reproducible analysis pipeline and structured Markdown report.

[Data-processing code](#) / [figures](#) / [generated report](#) / [Python](#) / [pandas](#) / [Visualization](#) / [Reporting](#)

PROJECT ARCHIVE / 2025.12-2026.01 / PARALLEL COMPUTING

CUDA SGEMM Optimization

Built a naive CUDA baseline and shared-memory tiled SGEMM, then benchmarked sizes from 256 to 2048 with bank-conflict and anomaly analysis.

DELIVERY / RESULT

Report records 999.51 GFLOPS at 512x512, 85.5% above the project baseline under its test conditions.

[CUDA C++](#) / [benchmark data](#) / [performance report](#) / [CUDA C++](#) / [SGEMM](#) / [GPU Optimization](#) / [Benchmarking](#)

PROJECT ARCHIVE / 2026.04-2026.06 / SYSTEMS + TEST ENGINEERING

Compiler + Regression + Practice Platform

Packaged a SysY compiler, LLVM/RISC-V outputs, compile-only and multi-backend regression, and an offline React/TypeScript practice site.

DELIVERY / RESULT

About 133KB of core implementation with repeatable tests and troubleshooting documentation.

[C source](#) / [backend outputs](#) / [PowerShell](#) / [web app](#) / [C](#) / [LLVM](#) / [RISC-V](#) / [PowerShell](#) / [React](#) / [TypeScript](#)

PROJECT ARCHIVE / 03

Breadth with traceable delivery

PROJECT ARCHIVE / 2025.04-2025.06 / RESEARCH + CAMPAIGN PLANNING / TEAM

Women's Health Public-Interest Campaign

Researched and segmented students aged 18-25, then designed campus touchpoints, an ad script, campaign site, media schedule, and budget.

DELIVERY / RESULT

Complete research-to-execution plan; forecast metrics are not presented as realized outcomes.

Research / strategy / script / site / media plan / User Research / Segmentation / Campaign Planning

PROJECT ARCHIVE / 2025.06 / THREE-PERSON DEVELOPMENT TEAM

DreamOfStickman 2D Roguelike Shooter

Built a playable Java/libGDX/LWJGL3 game with ECS, Box2D, procedural maps, enemy state machines, and custom weapon and projectile systems.

DELIVERY / RESULT

Runnable build and core gameplay demo, explicitly attributed to the team.

Team source / runnable build / gameplay demo / Java / libGDX / LWJGL3 / ECS / Box2D

PROJECT ARCHIVE / 2024.12-2025.01 / FIELD RESEARCH

Mingde Underground Public-Space Study

Combined 22 days of observation, 160 valid surveys, and seven interviews to study noise, learning, reservation, and governance conflicts.

DELIVERY / RESULT

Proposals for soundproofing, time slots, spatial redesign, and online negotiation.

Observation logs / surveys / interviews / report / Field Research / Survey / Interview / Public Communication

PROJECT ARCHIVE / 04

Breadth with traceable delivery

PROJECT ARCHIVE / 2026-PRESENT / FOUNDER-BUILDER

Life Exchange Working MVP

A local-first growth-partner MVP for daily settlement, time allocation, identity voting, periodic review, and multi-device continuity.

DELIVERY / RESULT

Web PWA, Expo mobile, shared domain package, and sync server run; public sync remains hidden pending release and policy checks.

Public repo / tests / build and export records / React / TypeScript / Expo / Node.js / PWA

PROJECT ARCHIVE / 2026 / ENGINEERING CONTRIBUTOR / TEAM

Ascend 910C Training + Inference Diagnostics

Contributed training diagnosis, benchmark analysis, and a dynamic-routing fix for three operator input formats in an LLM infrastructure project.

DELIVERY / RESULT

Team records report MFU moving from about 2% to 25%-30%; a distributed HCCL deadlock remained open.

Internal fix + run records / reported team metric / Python / Ascend 910C / Operator Compatibility / HCCL

PROJECT ARCHIVE / 2026-PRESENT / INDEPENDENT FULL-STACK PROTOTYPE

PhiloRevol Chinese Philosophy Platform

A digital-humanities system combining timelines, text mining, knowledge graphs, and East-West comparison.

DELIVERY / RESULT

11 API route groups and 12 frontend pages; core prototype works while corpus integration and de-hardcoding remain incomplete.

Public repo / README / changelog / routes and pages / FastAPI / React / PostgreSQL / Neo4j / Elasticsearch

PROJECT ARCHIVE / 05

Breadth with traceable delivery

PROJECT ARCHIVE / 2025-2026 / INDEPENDENT COURSEWORK ENGINEERING

Advanced Distributed Data Systems Lab

A reproducible lab portfolio across Hadoop, Spark, HBase, streaming, inverted indexes, prediction, and performance tuning.

DELIVERY / RESULT

Includes 12.6M-order aggregation, a 120K-record stream run, and driver time reduced from 431ms to 202ms.

Signed reports / code / run captures / explicit synthetic-data note / Hadoop / Spark / HBase / Python / Java

PROJECT ARCHIVE / 2026-PRESENT / INDEPENDENT PRODUCT + ENGINEERING

LifeExchange Multilingual Knowledge Archive

An evidence-aware Chinese, English, and German archive for essays, projects, Zhihu QA, tags, search, RSS, and a downloadable portfolio.

DELIVERY / RESULT

Production site with schema checks, static generation, Pagefind indexing, and repeatable Vercel deployment.

Public repo / content checks / production builds / Astro / TypeScript / Pagefind / Supabase / Vercel

PROJECT ARCHIVE / 2026 / INDEPENDENT PRODUCT PROTOTYPE

Career Decision Evidence Engine

An explainable interface for comparing graduate study, employment, international paths, entrepreneurship, and parallel strategies.

DELIVERY / RESULT

Interactive weighting, linked evidence, O*NET inputs, local persistence, and export; all personal decision data stays private.

Working local HTML/JavaScript prototype / JavaScript / LocalStorage / Decision Modeling / O*NET

PROJECT ARCHIVE / 06

Breadth with traceable delivery

PROJECT ARCHIVE / 2026 / EXTERNAL EVALUATOR

Headless Browser Benchmark Audit

A methodology audit of reproducibility, baselines, variance, macOS resource metrics, concurrency, and release readiness.

DELIVERY / RESULT

Found high variance and major coverage gaps; 8 of 9 formal publish-readiness gates were not yet satisfied.

Debug/release builds / smoke outputs / evaluator report / Rust / Browser Benchmarking / Statistical QA / macOS

PROJECT ARCHIVE / 2026-PRESENT / OPEN-SOURCE EXTENSION + SYSTEM DESIGN

Quant Research + Paper-Trading Workbench

A personal extension of open-source InStock connecting evidence snapshots, decision state, deterministic risk, paper intents, and review.

DELIVERY / RESULT

25 tracked upstream files modified plus new services, pages, and tests; current workflow intentionally stops at PAPER_PENDING.

Local diff / design docs / dedicated tests / Python / Tornado / MySQL / MLflow / State Machine

PROJECT ARCHIVE / 2026 / INDEPENDENT FRONTEND TOOLS

CFA + 408 Adaptive Learning Tools

Two local-first study prototypes for timed practice, mistake tracking, mastery, review dates, and subjective self-assessment.

DELIVERY / RESULT

Working CFA study hub and 180-minute 408 mock exam; longitudinal learning impact has not been tested.

Local HTML/JavaScript/CSS / README / working interactions / HTML / JavaScript / CSS / LocalStorage

PRODUCT DIRECTION

Life Exchange / Growth OS

WORKING MVP / LONGER HORIZON

A local-first growth partner for reflective decisions

The current MVP connects daily settlement, time allocation, identity voting, weekly and monthly review, custom assets, and optional multi-device continuity without pretending that a model can choose for the user.

CURRENT STATE

Web PWA, Expo mobile, shared domain logic, tests, and a lightweight sync server exist. Public sync remains hidden until release, privacy, support, and accessibility checks are complete.

SYSTEMS I BUILD

RELIABLE AI

RAG, MCP, agent evaluation, typed tools, recovery, and evidence-aware metrics.

RESEARCH PIPELINES

Scraping, cleaning, modeling, network analysis, field evidence, and honest uncertainty.

PRODUCT SYSTEMS

Workflow modeling, multi-agent orchestration, full-stack delivery, and humane interfaces.

SELECTED STACK

Python / JavaScript / TypeScript / C / CUDA / Java / SQL / Git / Linux / RAG / LLM / MCP / Agent Evaluation / NLP / HCI / Data + Network Analysis

BUILD SYSTEMS THAT CAN EXPLAIN THEMSELVES

github.com/SherryWang2005 / lifeexchange.me / blog.lifeexchange.me

hi@lifeexchange.me