

TAO AN 安涛

AI Researcher · M.S. in Artificial Intelligence, Hawaii Pacific University

Beijing | tan1@my.hpu.edu | tony@fim.ai

tao-hpu.github.io | [Google Scholar](#) | [Semantic Scholar](#) | [ORCID](#) | [GitHub](#) | [LinkedIn](#)

♦ RESEARCH PROFILE

M.S. in Artificial Intelligence from Hawaii Pacific University (2026), combining 13+ years of production software engineering with five years delivering enterprise-grade AI systems. My research formalizes a decade of applied practice into reproducible work on **retrieval-augmented generation**, **LLM long-term memory**, and **knowledge graphs** — one position paper accepted at HHAI 2026, two papers under review (TMLR and ACL Rolling Review), and a further arXiv preprint — alongside **FIM One**, an open-source enterprise agent platform with 1,100+ GitHub stars.

♦ EDUCATION

M.S. in Artificial Intelligence, Hawaii Pacific University — Honolulu, Hawaii, USA

JAN 2025 – MAY 2026

Graduated May 2026. Research interests: retrieval-augmented generation, LLM memory architectures, and knowledge graphs, with a focus on enterprise-grade AI for the legal and healthcare domains.

♦ PUBLICATIONS, PREPRINTS & PATENTS

- 01 **AI as Equalizer or Amplifier? Task Complexity as the Moderating Factor for Human Expertise in Hybrid Intelligence Systems**
An, T. (2026). Position paper accepted at the 5th International Conference on Hybrid Human-Artificial Intelligence (HHAI 2026), Brussels, July 6–10, 2026. Task complexity determines whether AI equalizes or amplifies expertise gaps.
- 02 **The Preference Centroid: Consensus Density Governs Output Dispersion in Aligned LLMs**
An, T., & Feng, S. (2026). Under review at TMLR. Output dispersion in aligned LLMs is governed by the **consensus density** of the task (Spearman $\rho = 0.85$; predicted by prompt-only judges at $\rho = -0.91$); alignment amplifies a gradient the base model already carries, grounding the "AI as equalizer vs. amplifier" boundary at the distribution level.
- 03 **Verbatim Chunks Beat Extracted Artifacts: A Controlled Ablation of Memory Representations for Long LLM Conversations**
An, T. (2026). arXiv preprint; under review via ACL Rolling Review (EMNLP/AAACL 2026). A controlled ablation of the stored memory representation: **verbatim chunks beat LLM-extracted artifacts by 15.9 pp on LoCoMo and 22.0 pp on LongMemEval-S** under an identical pipeline — lossy distillation, so structured memory should augment, not replace, verbatim text. CODE: github.com/tao-hpu/cog-canvas
- 04 **Cognitive Workspace: Active Memory Management for Large Language Models**
An, T. (2025). arXiv preprint. An active memory-management paradigm enabling functional infinite context, achieving a **58.6% memory-reuse rate**. CODE: github.com/tao-hpu/cognitive-workspace · 99★
- 05 **A Graph-Neural-Network Method for Data-Information Recommendation**
An, T. (2026). Chinese invention patent — under revision, to be re-filed in 2026. A graph-neural-network method for recommendation over heterogeneous data-information graphs.

♦ FIM ONE — FLAGSHIP OPEN-SOURCE PROJECT

"Wire every system you already run — global SaaS to the China stack — through one agent core."

Founder & lead architect of **FIM One** ★ 1,100+ STARS — a self-hosted, model-agnostic enterprise agent platform, deployable as a standalone assistant, an embedded Copilot, or a central orchestration Hub.

- RUNTIME** Dynamic DAG planning that decomposes goals into dependency graphs · ReAct reasoning loop · approval & audit hooks outside the LLM loop.
- RETRIEVAL** Full RAG pipeline — Jina embeddings, LanceDB vector store, hybrid retrieval · visual workflow editor with 25 node types.
- INTEGRATION** Connectors for 8 enterprise & China-stack databases (PostgreSQL, Oracle, DM, KingbaseES...) · 6 IM channels (Slack, Teams, Feishu, WeCom, DingTalk...).
- STACK** Python 3.11 · FastAPI · Next.js 14 · React · multi-tenant with JWT auth & org isolation · six-language i18n.

♦ PROFESSIONAL EXPERIENCE

Founder & Chief Executive Officer, FIM Labs Pte. Ltd.

BEIJING · MAR 2026 – PRESENT

Lead R&D on enterprise document intelligence, contract automation, and RAG-based knowledge systems with private / on-premise deployment, reaching 80%+ document-processing accuracy in production. Productized FIM One into a commercial offering bridging global SaaS with the China enterprise stack.

Founder & Chief Executive Officer, Beijing Feimu Network Technology Co., Ltd.

BEIJING · 2013 – PRESENT

Built and scaled a full-stack software and AI delivery firm for government and enterprise clients; led a five-year pivot to AI-native products (LLMs, RAG, agentic workflows) and founded the SLAN Research Institute. Delivered production AI knowledge systems with leading medical and academic institutions — a gynecology knowledge graph and a surgical-recommendation system (Peking Union Medical College Hospital), an HIV drug-interaction reference system (Beijing Ditan Hospital / WHO Collaborating Centre), and a quantitative-paleontology platform (Peking University).

Full-Stack Engineer, JD.com

BEIJING · 2016

Developed full-stack components of JD's e-commerce platform under high-traffic, high-availability constraints.

♦ SKILLS & LANGUAGES

TECHNICAL Retrieval-augmented generation · GraphRAG · Knowledge graphs · LLM memory architectures · Agentic workflows & ReAct · Supervised fine-tuning · Model Context Protocol (MCP) · Document intelligence · Private / on-premise LLM deployment.

LANGUAGES Chinese — native or bilingual proficiency · English — professional working proficiency.

♦ ACADEMIC SERVICE

Ethics Reviewer, NeurIPS 2026 — Ethics Review Committee

2026

Review flagged submissions against the conference Code of Ethics — data provenance & consent, dual-use risk, human-subjects, and societal impact.

安涛 TAO AN

人工智能研究者 · 夏威夷太平洋大学 人工智能理学硕士

北京 | tan1@my.hpu.edu | tony@fim.ai

tao-hpu.github.io | [Google Scholar](#) | [Semantic Scholar](#) | [ORCID](#) | [GitHub](#) | [LinkedIn](#)

◆ 研究简介

夏威夷太平洋大学人工智能理学硕士（2026 年），拥有 13 年以上的生产级软件工程经验，并在过去五年专注于交付企业级 AI 系统。研究工作将十余年的应用实践沉淀为可复现的成果，聚焦于检索增强生成（RAG）、大语言模型长期记忆与知识图谱——一篇立场论文获 HHAI 2026 录用，两篇论文审稿中（TMLR 与 ACL Rolling Review），另有一篇 arXiv 预印本——同时主导开源企业级智能体平台 FIM One（GitHub 星标 1,100+）。

◆ 教育背景

人工智能理学硕士，夏威夷太平洋大学 — 美国夏威夷檀香山

2025 年 1 月 — 2026 年 5 月

2026 年 5 月毕业。研究方向：检索增强生成、大语言模型记忆架构与知识图谱，专注于面向法律与医疗领域的企业级 AI。

◆ 论文、预印本与专利

- 01** *AI as Equalizer or Amplifier? Task Complexity as the Moderating Factor for Human Expertise in Hybrid Intelligence Systems*
An, T. (2026). 立场论文，已被第五届国际混合人类—人工智能大会（HHAI 2026）录用，比利时布鲁塞尔，2026 年 7 月 6–10 日。以任务复杂度为调节变量，调和 AI“均衡器”与“放大器”两类对立证据，并提出认知放大框架：人类贡献的三个层次与人机协作的三个介入层级。
- 02** *The Preference Centroid: Consensus Density Governs Output Dispersion in Aligned LLMs*
An, T., & Feng, S. (2026). TMLR 审稿中。对齐大模型的输出离散度由任务的共识密度决定（Spearman $\rho = 0.85$ ；仅凭提示词、由保留判别模型即可预测， $\rho = -0.91$ ）；对齐放大了基座模型本已携带的梯度，将“AI 究竟是均衡器还是放大器”的边界落到分布层面，而非假设的任务复杂度。
- 03** *Verbatim Chunks Beat Extracted Artifacts: A Controlled Ablation of Memory Representations for Long LLM Conversations*
An, T. (2026). arXiv 预印本，经 ACL Rolling Review 审稿中（目标 EMNLP/AAACL 2026）。一项受控消融实验，只替换存储的记忆表示：在完全相同的检索流水线，逐字 chunk 比大模型提取式 artifact 在 LoCoMo 上高 15.9 个百分点、在 LongMemEval-S 上高 22.0 个百分点。根因是有损蒸馏，因此结构化记忆应当是逐字文本的补充而非替代。CODE: github.com/tao-hpu/cog-canvas
- 04** *Cognitive Workspace: Active Memory Management for Large Language Models*
An, T. (2025). arXiv 预印本。一种主动式记忆管理范式，实现功能性的无限上下文，达到 58.6% 的记忆复用率。CODE: github.com/tao-hpu/cognitive-workspace · 99★
- 05** *A Graph-Neural-Network Method for Data-Information Recommendation*
An, T. (2026). 中国发明专利，完善中、年内重新提交。面向异构“数据—信息”图的图神经网络推荐方法。

◆ FIM One — 旗舰开源项目

“将你已在运行的所有系统——从全球 SaaS 到中国国产技术栈——统一接入同一个智能体内核。”

FIM One 创始人兼首席架构师 **★ 1,100+ STARS** —— 一个自托管、模型无关的企业级智能体平台，可作为独立助手、嵌入式 Copilot 或中枢编排 Hub 部署。

运行时	动态 DAG 规划，将目标分解为依赖图 · ReAct 推理循环 · 位于 LLM 循环之外的审批与审计钩子。
检索	完整 RAG 流水线——Jina 嵌入、LanceDB 向量库、混合检索 · 含 25 种节点类型的可视化工作流编辑器。
集成	8 种企业级与国产数据库连接器（PostgreSQL、Oracle、达梦 DM、人大金仓 KingbaseES...） · 6 个 IM 渠道（Slack、Teams、飞书、企业微信、钉钉...）。
技术栈	Python 3.11 · FastAPI · Next.js 14 · React · 多租户，JWT 鉴权与组织隔离 · 六语言国际化。

◆ 工作经历

创始人兼首席执行官，FIM Labs Pte. Ltd.（新加坡）

北京 · 2026 年 3 月 — 至今

主导企业文档智能、合同自动化与基于 RAG 的知识系统研发，支持私有化 / 本地部署，生产环境文档处理准确率达 80%+。将 FIM One 产品化为商业产品，打通全球 SaaS 与中国企业技术栈。

创始人兼首席执行官，北京飞沐网络科技有限公司

北京 · 2013 — 至今

创建并发展了一家面向政府与企业客户的全栈软件与 AI 交付公司；主导历时五年向 AI 原生产品（大模型、RAG、智能体工作流）的转型，并创立 SLAN 研究院。与顶尖医疗及学术机构共同交付生产级 AI 知识系统——妇科知识图谱与手术推荐系统（北京协和医院）、HIV 药物相互作用参考系统（北京地坛医院 / WHO 合作中心）、定量古生物学平台（北京大学）。

全栈工程师，京东（JD.com）

北京 · 2016

在高流量、高可用约束下开发京东电商平台的全栈组件。

◆ 技能与语言

技术能力 检索增强生成 · GraphRAG · 知识图谱 · 大语言模型记忆架构 · 智能体工作流与 ReAct · 监督微调（SFT） · 模型上下文协议（MCP） · 文档智能 · 私有化 / 本地大模型部署。

语言 中文——母语或双语水平 · 英文——专业工作能力。

◆ 学术服务

伦理审稿人，NeurIPS 2026 — 伦理审稿委员会

2026 年

依据大会伦理准则评审被标记的投稿——数据来源与知情同意、双重用途风险、人类受试者及社会影响。