

# Peng Ding

▷ **Email:** [dingpeng@uchicago.edu](mailto:dingpeng@uchicago.edu)  
▷ **GitHub:** <https://github.com/0aklight>  
▷ **LinkedIn:** <https://linkedin.com/in/peng-ding>  
▷ **Scholar:** [https://scholar.google.com/citations?user=R\\_cuuh4AAAAJ](https://scholar.google.com/citations?user=R_cuuh4AAAAJ)  
▷ **Website:** <https://oaklight.github.io>

## Summary

PhD candidate in Computer Science with 5+ years of experience in AI/ML research and engineering. Specialized in Large Language Models (LLMs), AI agent frameworks, and scientific AI applications. Led development of production-ready AI Infra and tools, including ToolRegistry (open-source LLM tool management library) and Argo-Proxy (enterprise AI API service). Published research in top-tier venues (COLM, ICML) with focus on

practical AI applications in biological reasoning and protein design. Proven track record of translating research into scalable solutions, with extensive experience in national lab (Argonne) environments. Strong engineering skills in Python, and modern agentic and ML frameworks. Passionate about building AI infrastructure and systems that solve real-world problems.

## Education

University of Chicago	CHICAGO, IL, UNITED STATES
<b>PhD in Computer Science (expected 2026)</b>	Sep 2020 – Present
Research focus: Large Language Models, AI agent frameworks, scientific reasoning.	
University of Chicago	CHICAGO, IL, UNITED STATES
<b>Master of Computer Science</b>	Oct 2018 – Dec 2019
Graduate coursework and research in computer science (MPCS).	
ShanghaiTech University	SHANGHAI, CHINA
<b>Bachelor of Science in Engineering, Computer Science</b>	Sep 2014 – Jul 2018
Undergraduate studies in computer science and engineering.	

## Publications

- [1] P. Ding and R. Stevens, "Stdlib or third-party? empirical performance and correctness of llm-assisted zero-dependency python libraries," *arXiv preprint arXiv:2605.21405*, 2026.
- [2] P. Ding, "Llm-rosetta: A hub-and-spoke intermediate representation for cross-provider llm api translation," *arXiv preprint arXiv:2604.09360*, 2026.
- [3] P. Ding and R. Stevens, "Toolregistry: A protocol-agnostic tool management library for function-calling llms," *arXiv preprint arXiv:2507.10593*, 2025.
- [4] P. Ding, T. Brettin, and R. Stevens, "Bior5: A three-layer architecture for biological reasoning in scientific ai," in *Proceedings of the SC'25 Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 594–601, 2025.
- [5] P. Ding, J. Fang, P. Li, K. Wang, X. Zhou, M. Yu, J. Li, H. Mei, and M. Walter, "MANGO: A benchmark for evaluating mapping and navigation abilities of large language models," in *First Conference on Language Modeling (COLM)*, 2024.
- [6] X. Liu, C.-C. Tien, P. Ding, S. Jiang, and R. Stevens, "Entropy-reinforced planning with large language models for drug discovery," in *Proceedings of the 41st International Conference on Machine Learning (ICML)*, pp. 31917–31932, 2024.
- [7] W. Lyu, P. Ding, Y. Zhang, A. Chen, M. Wu, S. Yin, and J. Yu, "Refocusable gigapixel panoramas for immersive vr experiences," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 3, pp. 2028–2040, 2019.

## Work in Progress / Under Review

- [8] P. Ding, "Cicada: Collaborative intelligent cad automation design agent." <https://cicada-agent.readthedocs.io/en/latest/>, 2025.

## Awards

Impact Argonne Award – Extraordinary Impact Category

JUL 2025

### Argonne National Laboratory

Argo-Proxy served as critical infrastructure enabling the lab-wide June 2025 Vibe Coding Hackathon—providing the only OpenAI-compatible API gateway available at the time, without which the event could not have proceeded.

Secretary of Energy Achievement Honor Award

JAN 2021

### United States Department of Energy

Acknowledging scientific contributions to the U.S. Department of Energy National Virtual Biotechnology Laboratory project "Molecular Design and Analysis to Inform Therapeutics Related to COVID-19," part of the DOE COVID-19 pandemic response in 2020.

## Recent Projects

ToolRegistry Ecosystem - Production AI Agent Infrastructure

Mar 2025 – Present

▷ **GitHub:** <https://github.com/Oaklight/ToolRegistry>

▷ **GitHub:** <https://github.com/Oaklight/toolregistry-hub>

▷ **GitHub:** <https://github.com/Oaklight/toolregistry-server>

*GitHub* ★ 55 | *PyPI* ↓ 30.0K

▷ **Documentation:** <https://toolregistry.readthedocs.io/>

Building production AI agents requires solving two entangled problems: tool protocol fragmentation and cold-start tooling. Sole-developed three complementary libraries:

- **ToolRegistry: Protocol-agnostic tool orchestration** with an **extensible modular architecture**—register, execute, and return results across MCP, OpenAPI, and LangChain (gRPC planned) with **unified schema validation**, automated parallel execution, think-augmented function calling, and **progressive BM25-based tool discovery**. Preprint on arXiv (2507.10593).
- **ToolRegistry-Server: Serve custom tools as network-accessible endpoints** (OpenAPI, MCP; gRPC planned)—capabilities stay in sync with ToolRegistry, bridging local tool definitions and deployable services for function-calling LLMs.
- **ToolRegistry-Hub: Security-hardened, ready-to-use tool collection** (shell execution, multimodal file I/O, **multi-engine web search with API key failover**, scheduled execution). Ships as PyPI library, OpenAPI/MCP server, or Docker container.

LLM-Rosetta - Universal LLM Provider Translation Layer

Mar 2025 – Present

▷ **GitHub:** <https://github.com/Oaklight/llm-rosetta>

*PyPI* ↓ 13.0K

Integrating multiple LLM providers requires  $N^2$  format-specific adapters that break with every API update. Sole-developed a **hub-and-spoke translation layer with a unified IR**, supporting **bidirectional conversion** across OpenAI Chat, OpenAI Responses, Anthropic, and Google GenAI with **multimodal, streaming, and tool-call support**. Powers Argo-Proxy and ToolRegistry; **adopted across multiple internal research projects** to reduce integration overhead.

zerodep - Zero-Dependency Python Library Study

Apr 2026 – Present

▷ **GitHub:** <https://github.com/Oaklight/zerodep>

▷ **Documentation:** <https://zerodep.readthedocs.io/>

*GitHub* ★ 7

▷ **Documentation:** <https://arxiv.org/abs/2605.21405>

Investigated whether LLM-assisted standard-library-only Python libraries can match third-party packages in correctness and performance. Built a suite of zero-dependency, single-file modules with apple-to-apple tests and benchmark dashboards, studying the tradeoffs between deployability, maintainability, correctness, and runtime efficiency. Preprint on arXiv (2605.21405).

Tinyleaf - Local-First Browser TeX Editor

May 2026 – Present

▷ **GitHub:** <https://github.com/Oaklight/tinyleaf>

▷ **Documentation:** <https://tinyleaf.readthedocs.io/>

Sole-developed a lightweight, self-hosted TeX editor that provides an Overleaf-like browser workflow without a full collaborative SaaS stack. Tinyleaf runs as a zero-runtime-dependency Python package with CodeMirror editing, PDF preview, compile logs, SyncTeX, Git tools, project search, outline, multi-tab editing, and Docker or local L<sup>A</sup>T<sub>E</sub>X compilation.

Argo-Proxy

Dec 2024 – Present

▷ **GitHub:** <https://github.com/Oaklight/argo-proxy>

*GitHub* ★ 22 | *PyPI* ↓ 41.0K

▷ **Documentation:** <https://argo-proxy.readthedocs.io/en/latest/>

Sole-developed production-grade **universal API gateway, widely adopted across Argonne National Laboratory**. Evolved from a simple proxy into a **multi-format translator** (v3.0.0), supporting 4 client API formats (OpenAI Chat, OpenAI Responses, Anthropic Messages, Google GenAI) with smart upstream routing and cross-format conversion powered by LLM-Rosetta. Features native function calling for all providers,

real-time streaming, and API key load balancing. Served as the **sole OpenAI-compatible gateway** enabling the lab-wide June 2025 Vibe Coding Hackathon (**Impact Argonne Award**).

#### Cicada - Multi-Agent CAD Automation Framework

Dec 2024 – Apr 2025

▷ **GitHub:** <https://github.com/Oaklight/cicada>

▷ **Documentation:** <https://cicada-agent.readthedocs.io>

CAD scripting requires iterating between design intent, code generation, and visual validation—a loop too complex for single-agent systems. Built a multi-agent framework from scratch (no LangChain dependency) where **agents collaborate on design analysis, DSL generation, code execution, multi-view rendering, and VLM-based visual feedback** for iterative optimization.

#### BioR5 - Three-Layer Tool-use Biological Reasoning Framework

Jun 2025 – Sep 2025

▷ **GitHub:** <https://github.com/brettin/biological-reasoning>

General-purpose LLMs lack the structured reasoning patterns needed for biological research (e.g., phylogenetic, mechanistic, systems-level reasoning). Sole-developed a **three-layer architecture** combining parametric memory, specialized foundation models, and external knowledge repositories to enable **reasoning-mode-aware biological inquiry**. **Pioneered dynamic “skill+tool” composition management** before similar approaches like Claude Skills. **Accepted at SC Workshops '25**.

#### MANGO - Benchmark for Evaluating Mapping and Navigation Abilities of LLMs

2023 – 2024

▷ **GitHub:** <https://github.com/Oaklight/mango>

▷ **Website:** <https://mango.ttic.edu>

No standard benchmark existed to evaluate whether LLMs can build and use spatial mental models from text. Introduced MANGO, a benchmark of 53 text-game mazes that tests mapping and navigation through **complex multi-hop spatial questions**. Revealed that **spatial reasoning is a critical bottleneck** for downstream tasks such as game-playing, providing actionable insights for LLM evaluation. **Published at COLM 2024**.

## Experience

### University of Chicago

CHICAGO, IL, UNITED STATES

#### Research Assistant

Jan 2021 – Present

Conducted independent research on LLM reasoning, spatial cognition, AI agent infrastructure, and scientific AI under Prof. Rick Stevens (committee: Ian Foster, Kyle Chard). Published as first author at COLM and contributing author at ICML. Built and open-sourced multiple AI infrastructure tools adopted by external organizations.

#### Teaching Assistant for Deep Learning System

Fall 2020, Fall 2021, Fall 2022

Assisted instructors in course management, Q&A, grading, and student access coordination.

### Argonne National Laboratory

LEMONT, IL, UNITED STATES

#### Research Aide (recurring summer appointments)

2019, 2020–2021, 2022, 2024, 2025

- Collaborated with senior scientists across 5 summer appointments spanning biological reasoning, drug discovery, LLM evaluation, scientific NLP, and neuroimaging.
- Authored and co-authored papers published at COLM 2024, ICML 2024, and SC Workshops '25.
- Built Argo-Proxy, now widely adopted across Argonne for AI API integration.

#### Predoctoral Appointee

Mar 2020 – Sep 2020

- Researched image-based molecular docking score prediction using TensorFlow 2.
- Devised large-scale data preprocessor to facilitate chemical compound docking score prediction
- Contributed to work recognized with Secretary of Energy Achievement Honor Award.

### iStark AI Ltd (Startup)

SHENZHEN, GUANGDONG, CHINA (REMOTE)

#### R&D Tech Lead

Apr 2024 – Dec 2024

Served as consultant and tech lead for iStark AI Mechanic, overseeing design and prototyping of an agentic AI assistant.

- Led architecture and development of a production-ready assistant powered by LLM agents
- Built a RAG system with hybrid retrieval and reranking for domain-specific AI queries

### DGene Digital Technology (Shanghai) Co., Ltd

SHANGHAI, CHINA

#### Software Engineering Intern

Oct 2017 – Apr 2018

- Developed and maintained software for light-field camera array in C++ and Python.
- Collaborated with hardware engineering team for frame design testing.

### Shanghai Roche Pharmaceutical Ltd.

SHANGHAI, CHINA

#### Market Data Analyst Intern, Herceptin Central Market Team

Oct 2017 – Mar 2018

- Processed monthly patient flow and sales data via Python and VBA

- Analyzed sales variance for supply allocation.

ShanghaiTech University

**Head Teaching Assistant for Linear Algebra**

SHANGHAI, CHINA

*Fall 2017, Fall 2018*

- Led tutorial sessions, managed TA team and online forum, designed MATLAB PCA project for course enrichment

## Skills

**Programming Languages:** Python, Golang, MATLAB, Shell/Bash

**AI/ML Related:** Agentic AI, Large Language Models (LLMs), Prompt Engineering, Retrieval-Augmented Generation (RAG) Systems, Vector Databases, Model Deployment

**Frameworks & Libraries:** PyTorch, FastAPI, Pydantic, FastMCP, MCP (Model Context Protocol), Aiohttp

**DevOps & Tools:** Docker, RESTful API, Git, CI/CD (GitHub Actions), Linux administration, Private Cloud Management

**Other Skills:** Data analysis, Scientific computing, Software architecture, Team leadership, Technical consulting

## Interests

**Professional Interests:** Large language models, AI agent frameworks, scientific reasoning, protein design, private cloud infrastructure

**Personal Interests:** Server management (self-hosted multi-region network with GeoDNS), open-source software, music, travel