Peng Ding

▷ **Email:** dingpeng@uchicago.edu
▷ **GitHub:** https://github.com/Oaklight
▷ **LinkedIn:** https://linkedin.com/in/peng-ding
▷ **Scholar:** 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
PhD in Computer Science (expected 2026) — *Sep 2020 – Present*
Research focus: Large Language Models, AI agent frameworks, scientific reasoning.

University of Chicago — Chicago, IL, United States
Master of Computer Science — *Oct 2018 – Dec 2019*
Graduate coursework and research in computer science (MPCS).

ShanghaiTech University — Shanghai, China
Bachelor of Science in Engineering, Computer Science — *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
▷ **Documentation:** https://toolregistry.readthedocs.io/

GitHub ★ 67 | PyPI ↓ 68.0K

Building production AI agents requires solving two entangled problems: tool protocol fragmentation and cold-start tooling. Sole-developed three complementary libraries (42K LoC, 950+ tests):

- **ToolRegistry**: Cross-protocol **tool lifecycle management**—registration, discovery, schema representation, and execution—unified under a single API across Native Python, MCP, and OpenAPI. Advanced features include **progressive disclosure**, BM25-based tool search, programmatic tool calling, and automated parallel execution (**16× throughput on network tools**). Adopted by the open-source community (68.0K PyPI downloads). Preprints on arXiv (2507.10593; 2508.02979).
- **ToolRegistry-Server**: **Auto-discovers all registered tools** from a ToolRegistry instance, computes exposure routes and schemas automatically, and serves them as **network-accessible endpoints** (OpenAPI, MCP; gRPC planned) with role-based security profiles—**zero-maintenance multi-protocol exposure** from a single tool definition. **Adopted by Argonne National Laboratory** to streamline open-source release of scientific research tools at scale.
- **ToolRegistry-Hub**: Production-grade tool collection covering **15+ tool classes** (6 search engine backends with API key failover, shell execution, multimodal file I/O, scheduling). **Replaces built-in tools of multiple coding agents** (Claude Code, Codex, Gemini CLI, etc.) in daily workflows, providing developers with greater **transparency and customization**. Ships as PyPI library, OpenAPI/MCP server, or Docker container.

zerodep - Zero-Dependency Python Library Study — Apr 2026 – Present

▷ **GitHub:** https://github.com/Oaklight/zerodep
▷ **Documentation:** https://zerodep.readthedocs.io/
▷ **Documentation:** https://arxiv.org/abs/2605.21405

GitHub ★ 12

First large-scale empirical study on **whether LLM-generated stdlib-only code can replace established third-party packages**—a key open question in AI-assisted software engineering. Built **45+ zero-dependency, single-file modules** across 12 categories (43.5K LoC), benchmarked apple-to-apple against **30+ reference libraries** (PyYAML, httpx, beautifulsoup4, pydantic, etc.). Findings **challenge the assumption that dependency-heavy development is necessary**: majority achieve correctness and performance parity; several show **5–115× speedups**. Demonstrates practical benefits in **supply-chain security** and deployment simplification—7 modules vendored into ToolRegistry for zero-dependency operation. 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 (`pip install tinyleaf`, zero runtime dependencies) providing an Overleaf-like browser workflow—CodeMirror editing, PDF preview, SyncTeX, Git integration, multi-tab editing—without MongoDB, Redis, or Node.js.

LLM-Rosetta - Universal LLM Provider Translation Layer — Mar 2025 – Present

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

GitHub ★ 94 | PyPI ↓ 57.0K

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** (9 content types, 10 stream event types), supporting **full bidirectional conversion** across 4 API formats (OpenAI Chat, OpenAI Responses, Anthropic, Google GenAI) plus **8 provider-specific shims** (OpenRouter, DeepSeek, Volcengine, etc.). Powers Argo-Proxy and ToolRegistry. Adopted by the open-source community (GitHub ★ 94, 57.0K PyPI

downloads). Preprint on arXiv (2604.09360). Companion tool **llm-comply** provides multi-format API compliance testing across all supported formats.

VeilRender - Headless Browser Rendering API Jul 2026 – Present

▷ **GitHub:** https://github.com/Oaklight/veilrender

Sole-developed a self-hostable headless browser rendering API with stealth capabilities. Accepts a URL and returns fully rendered content (HTML, Markdown, readability-extracted article, or screenshot) using a stealth Chromium browser. Features include horizontal scaling with a gateway + remote browser worker pool, mixed browser backends (Chromium CDP and Firefox/Camoufox), L1/L2 render cache (in-memory + S3-compatible), ad/tracker blocking, Prometheus metrics, and a live dashboard. Designed as a fallback for fetch tools that fail on JavaScript-rendered or bot-protected pages. Ships as PyPI package and Docker container.

Argo-Proxy Dec 2024 – Present

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

▷ **Documentation:** https://argo-proxy.readthedocs.io/en/latest/ *GitHub ★ 28 | PyPI ↓ 65.0K*

Sole-developed production-grade **universal API gateway**, **adopted across Argonne National Laboratory** (3,400+ staff). Evolved through **3 major versions** over 19 months (867 commits, 66 releases, 8 contributors incl. Argonne scientists), supporting 4 client API formats with smart upstream routing and cross-format conversion powered by LLM-Rosetta. Enables 6+ AI coding tools (Claude Code, Codex CLI, Aider, Gemini CLI, etc.) at the lab. Served as the **sole OpenAI-compatible gateway** for the lab-wide **200-participant (sold-out) Vibe Coding Hackathon**. **Impact Argonne Award – Extraordinary Impact**.

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 **5-agent pipeline** from scratch (no LangChain dependency) for design analysis, build123d code generation, multi-view 3D rendering, and **VLM-based visual feedback** with confidence scoring. Supports iterative refinement (up to 10 design $\times$ 5 code iterations) and 6 format conversions. Evaluated on 7 design tasks demonstrating text/image-to-3D generation.

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. Sole-developed a **three-layer architecture** combining parametric memory, specialized foundation models (TX-Gemma, genomic/protein models), and **5 external knowledge repositories** (PubMed, ChEMBL, PubChem, Open-Targets, BioDB) to enable **11 distinct biological reasoning modes** (phylogenetic, mechanistic, systems-level, etc.) with intelligent triage routing. **Pioneered dynamic "skill+tool" composition** 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** with **206K question pairs** testing mapping and navigation. Benchmarked 8 LLMs; revealed that even GPT-4 achieves only **58% on hard spatial tasks**, establishing spatial reasoning as a critical LLM bottleneck. **Published at COLM 2024**; **10 citations** including IJCAI 2025 and EMNLP 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 Shanghai, China

Head Teaching Assistant for Linear Algebra *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