Xiaohang Yu

Email: xhyumiracle@gmail.com
linkedin.com/in/xhyumiracle scholar.google.com/citations?user=WURsNDYAAAAJ github.com/xhyumiracle

PROFILE

Research at the intersection of agentic AI, cryptography, and system security. Recent work spans cryptographic protocols for agent identity and credential delegation, reinforcement-learning post-training for LLM reasoning, and agentic frameworks for blockchain forensics. PhD candidate at Imperial College London.

EDUCATION

Ph.D. Computer Science **Imperial College London, UK** 2021.10 – 2027

- Research: agentic systems, RL post-training, cryptographic protocols for AI, and blockchain forensics.

M.S. Security Informatics **Johns Hopkins University, US** 2016.08 – 2018.01

- Focus: practical cryptography, applied machine learning, network security, and security & privacy.

B.S. Computer Science **Beijing Normal University, China** 2012.09 – 2016.06

RESEARCH

- **Agent Security**
  Formalized the Agent Secret Use problem and designed a credential-delegation protocol that provably satisfies seven required security properties; the cryptographic foundation of SafeClaw.
  Xiaohang Yu (first author), et al. *SUDP: Secret-Use Delegation Protocol for Agentic Systems.* arXiv preprint (under submission).
- **Agentic Blockchain Forensics**
  An agentic framework for blockchain forensics; PhD work.
  Xiaohang Yu (first author), William Knottenbelt. *LOCARD: An Agentic Framework for Blockchain Forensics.* **IEEE ICBC 2026**.
- **LLM Reinforcement Learning**
  Empirical study of scaling behaviors in RL post-training for mathematical reasoning; a survey of agentic RL for LLMs.
  Zelin Tan, Hejia Geng, **Xiaohang Yu**, et al. *Scaling Behaviors of LLM Reinforcement Learning Post-Training.* **ACL 2026** (main).
  Guibin Zhang[1], Hejia Geng[1], **Xiaohang Yu**[1], et al. *The Landscape of Agentic Reinforcement Learning for LLMs: A Survey.* **TMLR 2025** ([1]co-first).
- **Verifiable AI**
  Verifiable machine-learning inference on-chain; core technical contributor to open-source standards including opML (Optimistic Machine Learning), ERC-7007, and ERC-7641.

WORK EXPERIENCE

Co-Founder **SafeClaw (safeclaw.pro)** 2026.03 – present

- Built SafeClaw, an Agent Identity Solution: security infrastructure that puts agents' secret use under human control, giving agents scoped, per-use, user-approved delegation instead of reusable secrets that carry the user's full authority.
- Led the design of SUDP, the cryptographic protocol underneath (see Research): human approval becomes an unforgeable, single-use authorization bound to the exact operation, so reusable authority never crosses the agent boundary.
- Shipped the entire production stack solo in about three months: a Rust egress proxy with inline secret injection, an end-to-end encrypted vault with multi-device sync sealed under the user's passkey, a web console and cloud control plane; launched at safeclaw.pro in July 2026, first seed users and positive feedback within a week; open source at github.com/SafeClaw-OSS/safeclaw.
- Also built *NodPay.ai*, an open-source wallet stack for AI-agent payments; prototyped in one week.

Technical Co-Founder **Blockchain Infrastructure Startup** 2023 – 2025

- Focused on the research and implementation of core protocols for verifiable AI systems and zero-knowledge proof (ZKP) infrastructure.
- Core technical contributor to open-source standards including *opML* (Optimistic Machine Learning), *ERC-7007*, and *ERC-7641*.

Head of Blockchain Security **Chaitin Tech** 2018.03 – 2023.03

- Founded and led the blockchain security practice from zero, one of the earliest teams of its kind; tech lead and PM.
- Led protocol and cryptographic security audits across major L1s (BTC, ETH, Cosmos, IOST, Nervos CKB, EveriToken, Quarkchain, and others); discovered $\sim$100 0-days across consensus (PBFT, DPoS, PoC, VRF), P2P, and verification layers; 5 CVEs in smart-contract audits.
- Technical lead for China's first national blockchain vulnerability scoring standard; contributed to three sector-level blockchain-security standards.
- Co-organized and designed blockchain security challenges for *Real World CTF*.
- Independent research: adapted JOP (Jump-Oriented Programming) from classical binary exploitation to the EVM, first demonstrated as a *Real World CTF 2018* challenge and later established as a canonical blockchain-CTF category; invited talk at *DEFCON 27 Village*.

SELECTED TALKS

- IEEE ICBC 2026 Talk "LOCARD: An Agentic Framework for Blockchain Forensics" at Brisbane, Australia (*upcoming*) 2026
- DEFCON 27 Village Talk "EVM Opcode JOP" at Las Vegas, US 2019
- DEFCON China 1.0 Village Talk "Practical Security of Blockchain in Industry" at Beijing, China 2019

QUALIFICATIONS & AWARDS

- 3rd Place, DEF CON CTF World Finals, team *Tea Deliverers* 2021
- Project Management Professional (PMP) 2021
- Offensive Security Certified Professional (OSCP) 2017
- Fully Funded Master Scholarship from China Scholarship Council (CSC) 2016
- Silver prize in ACM/ICPC Asia Regionals, Guangdong (92/798) and Mudanjiang (19/154) 2014

SKILLS

Cryptography • Agent Security • Agentic AI • Blockchain Security • LLM RL Post-Training