

Wenyuan “Rafael” Wu

Binzmuehlestrasse 11, 8050 Zurich

Signal  rafael.wu@pm.me

 Portfolio

 UZH Page

 LinkedIn

 GitHub



Expertise

AI engineer and researcher specializing in LLM agent architecture for privacy-sensitive, regulated domains. Co-author of *PROMISE*, an open-source framework for stateful prompt orchestration, and of five peer-reviewed papers on conversational agent design and evaluation. I take systems from architecture through self-hosted deployment to measured outcomes with real users, in an Innosuisse-funded project delivered with clinical and industry partners.

Skills

AI Engineering	LLM applications on hosted APIs (Anthropic, OpenAI, Google) and self-hosted models; agent architecture and stateful prompt orchestration; prompt engineering, RAG, fine-tuning, and evaluation design; AI-assisted development with Claude Code.
ML & NLP	Natural language processing, machine learning, data science; PyTorch, Hugging Face.
Infrastructure	Linux administration, Ansible, Nginx, Docker; Azure OpenAI, GPU-backed model serving, on-premises deployment.
Programming	Python, Java, Go, R, SQL, Bash, JavaScript; Git.
Methods	Design science research, mixed methods (surveys, interviews, user data), experiment design, statistics.
Languages	Mandarin Chinese (native), English (C2), German (C1, TestDaF 5-4-4-5).

Employment History

- 2023 – present  **Research Assistant & PhD Candidate.** Information Management Research Group, Department of Informatics, University of Zurich, Switzerland.
- Architect, build, and evaluate LLM agents for privacy-sensitive domains: prompt design and dialogue orchestration, deployment on managed cloud (Azure OpenAI) and self-hosted Linux servers, and user studies with real patients. Delivered within a four-partner Innosuisse consortium (ZHAW, UZH, COBEDIAS, Helmedica AG).
- Teaching assistant for *Computer-Supported Cooperative Work* (2024 – 2027); supervision of master’s and bachelor’s theses.
- **Digital Companion** (Innosuisse, with Helmedica AG): conversational agent maintaining continuity between consultations through personalized dietary and activity messaging. GPT-3.5 via Azure OpenAI; field-tested with 29 users in clinics across Canton Zurich.
 - **PROMISE** (open source, with ZHAW): model-driven framework that drives multi-turn LLM dialogue through hierarchically nested state machines, giving deterministic control where free-form prompting is too brittle. Peer-reviewed [4]. github.com/zhaw-iwi/promise
 - **AI-Blended Psychotherapy**: CBT companion supporting patients between therapy sessions with reminders, guided exercises, and progress tracking, plus a therapist-facing view into patient effort. Self-hosted LLM stack on Linux (Ansible, Nginx) keeps patient data on premises. Expert-reviewed; clinical evaluation pending ethics approval.

Employment History (continued)

- 2025 – present  **Technical Advisor** (side projects).
• **AI in Public Administration:** advise on LLM use cases, system design, and privacy-compliant deployment in the public sector.
• **AI in University Career Services:** advise on LLM-assisted guidance tools and their evaluation.
- 2021 – 2022  **Assistant to Doping Control Officer.** Clearidium A/S, Denmark.
Sample collection and documentation under regulated anti-doping testing protocols (EU/CH), with strict procedural compliance and athlete data confidentiality.
- 2018  **Business Consultant.** Cromine GmbH, Germany.
Advised on German–Chinese market entry in both directions, including real-estate ventures. Owned end-to-end delivery for a German client exhibiting at a Shanghai trade fair for overseas property, immigration, education, and investment.

Education

- 2023 – 2027  **PhD in Information Systems (expected mid-2027), University of Zurich.**
Thesis title: *Personalized Adherence Companion: Realizing Human-Centric Persuasion Through Generative Artificial Intelligence.*
- 2018 – 2022  **M.A., University of Zurich.** Computational Linguistics.
- 2012 – 2017  **B.A., Lanzhou University.** German Language and Literature.

Extracurricular Activities

- 2027  **Program Committee,** PERSUASIVE conference.
- 2026  **Volunteer,** Davos Tech Summit.

Selected Publications

- 1 **W. Wu**, S. Egger, A. Bucher, I. Vashkite, M. Dolata, and G. Schwabe, “Designing persuasive artificial intelligence for mental health: A prioritization framework to enhance trust and engagement,” in *Persuasive Technology*, K. Sumi, R. Ali, and R. Legaspi, Eds., Cham: Springer Nature Switzerland, 2026, pp. 420–434.  DOI: 10.1007/978-3-032-19687-3_31.
- 2 A. Bucher, S. Egger, I. Vashkite, **W. Wu**, and G. Schwabe, “It’s Not Only Attention We Need’: Systematic review of large language models in mental health care,” *JMIR Mental Health*, vol. 12, no. 1, e78410, 2025.  DOI: 10.2196/78410.
- 3 **W. Wu**, M. Dolata, A. de Spindler, and G. Schwabe, “Persuasive prompting: The case of digital health,” in *ECIS 2025 Proceedings*, vol. 10, 2025, ECIS2025–1051.  URL: https://aisel.aisnet.org/ecis2025/health_it/health_it/10.
- 4 **W. Wu**, J. Heierli, M. Meisterhans, A. Moser, A. Färber, M. Dolata, E. Gavagnin, A. de Spindler, and G. Schwabe, “PROMISE: A framework for model-driven stateful prompt orchestration,” in *Intelligent Information Systems*, S. Islam and A. Sturm, Eds., Cham: Springer Nature Switzerland, 2024, pp. 157–165.  DOI: 10.1007/978-3-031-61000-4_18.

Full list — five peer-reviewed publications, two manuscripts under review — at wenyuanwu.com.