Boya Zhang

boya.zhang3@gmail.com tinaboya.github.io
linkedin.com/in/tinaboya 0000-0002-4439-8212



Research Interests

Trustworthy machine learning for high-stakes decision making:

- **Transparent human–AI interaction**: reliable evidence-based decision support systems.
- **Robustness**: failure-mode analysis of models deployed in real-world settings.
- **Fairness**: how model behaviour differs across languages and cultures.

Research Experience

Nov 2025 – Apr 2026
Cambridge, MA, USA

Visiting Researcher, MIT
Laboratory for Computational Physiology. Supervisor: Prof. Leo A. Celi

- Led **PAINT**, a study of disparities in how pain is described in medical text across four languages, with 17 co-authors from 16 institutions. PAINT is built to support fairness audits of both healthcare systems and clinical language models. Paper at MRL, EMNLP 2026.
- Mentored **datathon** teams at MIT Critical Data events with Mayo Clinic, Harvard Medical School, Brown University and Bowdoin College, helping clinicians and students work with AI on real clinical data and real-life prompts.
- Facilitated workshops and served as a panelist at *AI as a Catalyst* with the American Medical Association, Stanford and Threshold Ventures, raising awareness of cross-institutional collaboration in healthcare AI and medical education.
- Founded and hosted the **MIT AI Comedy Night**, a live benchmark of humans against AI on humor, supported by the De Florez Fund for Humor. The evening brought joy and curiosity back into scientific exploration.

Nov 2021 – Sep 2026
Geneva, Switzerland

Doctoral Researcher, Université de Genève
Data Science for Digital Health group. Supervisor: Prof. Douglas Teodoro
Thesis: *Your Health Questions Answered by Reliable Language Models.*

- **Retrieval**: built a ranking framework based on discriminative models that automatically detects usefulness, supportiveness and credibility signals across a billion web pages. The work improves the quality of health search results for lay users. Paper at *JMIR AI* 2024.
- **Verification**: built the **CliniFact** benchmark for clinical claim verification and reasoning. We found that small discriminative models can outperform large generative models at clinical natural language inference. The work supports workflow optimisation in clinical trial design. Paper at *Scientific Data* 2025.
- **Generation**: built the **HealthContradict** benchmark for biomedical retrieval-augmented generation. We found that in-domain fine-tuning helps models identify the correct context. This work supports safer deployment of healthcare digitalization. Paper at *npj Digital Medicine* 2026.

Teaching: gave lectures, prepared slides and exercises, and supervised student projects for *Applied Machine Learning in Healthcare: ChatGPT and beyond*, helping students understand how to apply language models to clinical workflows.

Industry: collaborated with TriNetX (formerly Clinerion), RisKlick and public health partners on NLU4EHR, CTxAI and GESICA projects, bridging research outcomes into real-world products and services.

Jun – Oct 2021
Zürich, Switzerland

Research Intern, ETH Zürich
Center for Law & Economics. Supervisor: Prof. Elliott Ash

- Trained claim-verification models and built an evidence dataset with the Google Search API to test how the choice of knowledge base affects automated claim checking. Paper at *ACM JDIQ* 2023.

Education

2021–2026 **PhD, Life Sciences, AI for Healthcare**, Université de Genève, Faculty of Medicine

2018–2021 **MSc (ir.), Electrical Engineering, with Distinction**, Vrije Universiteit Brussel / ULB

2015–2019 **BEng, Information Engineering**, Northwestern Polytechnical University, Xi'an

Technical Skills & Languages

- **Programming:** Coding agents, Python, PyTorch, Hugging Face, vLLM, CUDA, SQL, Git, Docker, SLURM/HPC, C/C++.
- **Languages:** English (C2), Chinese (native), French (A2–B1).

List of Publications

First-author publications

1. **Zhang B.**, Naderi N., Mishra R., Teodoro D. *Online health search via multidimensional information quality assessment based on deep language models: algorithm development and validation.* JMIR AI, 3:e42630, 2024.

2. **Zhang B.**, Bornet A., Yazdani A., Khlebnikov P., Milutinovic M., Rouhizadeh H., Amini P., Teodoro D. *A dataset for evaluating clinical research claims in large language models.* Scientific Data, 12:86, 2025.

3. **Zhang B.**, Bornet A., Yang R., Liu N., Teodoro D. *HealthContradict: evaluating biomedical knowledge conflicts in language models.* npj Digital Medicine, 2026.

4. **Zhang B.**, Naderi N., Jaume-Santero F., Teodoro D. *DS4DH at TREC Health Misinformation 2021: multi-dimensional ranking models with transfer learning and rank fusion.* In The Thirtieth Text REtrieval Conference (TREC 2021) Proceedings, 2022.

5. **Zhang B.**, Mishra R., Teodoro D. *DS4DH at MEDIQA-Chat 2023: leveraging SVM and GPT-3 prompt engineering for medical dialogue classification and summarization.* In Proceedings of the 5th Clinical Natural Language Processing Workshop, pages 536–545, Toronto, Canada, 2023. Association for Computational Linguistics.

6. **Zhang B.**, Ma Y., Wu J., Chaisutyakorn K., Gameiro R., Han A., *et al.*, Celi L.A., Teodoro D. *PAINT: a framework for multilingual pain representation from medical licensing exams to real-world clinical texts.* In Proceedings of the 6th Workshop on Multilingual Representation Learning (MRL) at EMNLP 2026. aclanthology.org/2026.mrl-1.51

Co-authored publications

7. Jaume-Santero F., **Zhang B.**, Proios D., Yazdani A., Gouareb R., Bjelogrlic M., *et al. Cluster analysis of low-dimensional medical concept representations from electronic health records.* Int. Conf. on Health Information Science (HIS), 313–324, 2022.

8. Stammbach D., **Zhang B.**, Ash E. *The choice of textual knowledge base in automated claim checking.* ACM Journal of Data and Information Quality (JDIQ), 15(1):1–22, 2023.

9. Rouhizadeh H., Nikishina I., Yazdani A., Bornet A., **Zhang B.**, Ehrsam J., *et al. A dataset for evaluating contextualized representation of biomedical concepts in language models.* Scientific Data, 11:455, 2024.

10. Bornet A., Khlebnikov P., Meer F., Haas Q., Yazdani A., **Zhang B.**, Amini P., *et al. Analysis of eligibility criteria clusters based on large language models for clinical trial design.* Journal of the American Medical Informatics Association (JAMIA), 32(3):447–458, 2025.

11. Yazdani A., Bornet A., Khlebnikov P., **Zhang B.**, Rouhizadeh H., Amini P., *et al. An evaluation benchmark for adverse drug event prediction from clinical trial results.* Scientific Data, 12:424, 2025.

12. Rouhizadeh H., Yazdani A., **Zhang B.**, Teodoro D. *Exploring zero-shot cross-lingual biomedical concept normalization via large language models.* In Intelligent Health Systems — From Technology to Data and Knowledge, pages 788–792. IOS Press, 2025.

13. Teodoro D., Naderi N., Yazdani A., **Zhang B.**, Bornet A. *A scoping review of artificial intelligence applications in clinical trial risk assessment.* npj Digital Medicine, 8:486, 2025.

14. Reichenpfader D., Zaghir J., Cécilia-Joseph E., Delange B., Shaitarova A., Sylvestre E., Dennstädt F., **Zhang B.**, Teodoro D., Meystre S., Müller H., Denecke K. *The consensus-based CINEX guideline for reporting clinical information extraction studies.* Journal of the American Medical Informatics Association (JAMIA), 33(10):2040–2050, 2026.

15. Rouhizadeh H., Yazdani A., **Zhang B.**, Vicente Alvarez D., Hüser M., Vanobberghen A., *et al.*, Teodoro D. *Knowledge-enhanced LLMs for multilingual biomedical concept normalization: a multilingual benchmarking and behavioral analysis.* npj Digital Medicine, 2026. doi:10.1038/s41746-026-03224-x

16. He J., Feng W., Ferdowsi S., Yazdani A., **Zhang B.**, Koraş O.A., Khattab S., Corbeil J.-P., Müller H., Teodoro D. *SeMLI-RAG: semantics-guided multimodal late interaction for retrieval-augmented radiology report generation.* In Proceedings of EMNLP 2026 (main conference). aclanthology.org/2026.emnlp-main.2379

17. He J., **Zhang B.**, Rouhizadeh H., Chen Y., Yang R., Lu J., Chen X., Liu N., Teodoro D. *Retrieval-augmented generation in biomedicine: a survey of technologies, datasets, and clinical applications.* Artificial Intelligence Review, 2026. doi:10.1007/s10462-026-11743-y

Academic Activities

Scientific communications

1. Flash talk, Health Misinformation Track, Thirtieth Text REtrieval Conference (TREC 2021), National Institute of Standards and Technology, Gaithersburg, Maryland, USA (remote), 2021.
2. Flash talk, CUSO Winter School *Data Science in the Digital Era*, Hôtel Suisse, Champéry, Switzerland, 2022.
3. Flash talk, *Fostering Quality and Affording Development on Doctoral Researchers*, 16th General Conference of the International Association of Universities, University College Dublin, Ireland, 2022.
4. Flash talk, SciMed PhD Conference, PhD School of Life Sciences, University of Geneva, Maison des congrès, Les Diablerets, Switzerland, 2023.
5. Flash talk, CUSO Winter School *Human–AI Collaboration*, Champéry, Switzerland, 2025.
6. Poster presentation, Women in Data Science (WiDS) Conference, Zürich, Switzerland, 2025. *Best Poster Award.*
7. Panellist, *AI as a Catalyst: Reimagining Health Innovation*, convened by MIT Critical Data, Stanford University, and Threshold Ventures, Palo Alto and Stanford, California, USA, 2026.

Outreach

1. PhD Booster Programme (7th edition), Faculty of Medicine, University of Geneva, 2025. Project presented: *A health question? Ask a reliable AI.*
2. Organiser and host, MIT AI Comedy Night: *Can AI Learn to Be Funny? A Stand-up Experiment on April Fools' Day*, 2026. *De Florez Fund for Humor Award.*
3. i-Teams commercial-translation project, Faculty of Medicine, University of Geneva, taken as *Better Buddy* to the first GATE Challenge, Shanghai, 2025. *Second place.*

Teaching and mentoring

1. Teaching assistant, *Applied Machine Learning in Healthcare: ChatGPT and beyond*, University of Geneva.
2. Coordinator of the Data Science for Digital Health (DS4DH) Journal Club on artificial intelligence for healthcare, University of Geneva, 2022–present.
3. Datathon mentor for Team Signal2Scan, Mayo Clinic Datathon & Data Summit, Jacksonville, Florida, USA, 2026. *Recognition Award.*
4. Datathon mentor for events organised by MIT Critical Data with Bowdoin College, Brown University, and Harvard Medical School, USA, 2026.
5. Workshop facilitator, *AI as a Catalyst: Reimagining Medical Education and Workforce Development*, MIT, with the American Medical Association, Cambridge, Massachusetts, USA, 15 January 2026.

Peer-review service

- **Conferences:** EMNLP 2026; ClinicalNLP @ ACL 2023 and 2025.
- **Journals:** *npj Digital Medicine*, *Communications Medicine*, *Scientific Reports*, *JAMIA Open*, *BMC Bioinformatics*, *BMC Medical Informatics and Decision Making*, *BMC Medical Education*, *Journal of Healthcare Informatics Research*, *Diagnostic and Prognostic Research*, *International Journal of Machine Learning and Cybernetics*.