Omid **Rohanian**

Senior Research Associate · Natural Language Processing and Clinical Machine Learning · PhD

Department of Engineering Science, Old Road Campus Research Building, University of Oxford, Headington, Oxford OX3 7DQ

omidrohanian@eng.ox.ac.uk | omidrohanian | omidrohanian | @omidrohanian

Research Interests

Natural Language Processing, Biomedical and Clinical Text Analysis, Compact Transformer Models, Privacy-Aware Machine Learning in Healthcare, Model Distillation for NLP, and LLM Analysis and Evaluation.

Work Experience

University of Oxford

Oxford, UK

Senior Post-Doctoral Research Associate , Department of Engineering Science

Sep 2020 - Present

- Contributed to the CURIAL-Rapide project, focusing on developing a rapid COVID-19 testing tool using routine blood tests. Managed the machine learning pipeline, including data preprocessing and analysis.
- Led efforts in the ISARIC project to develop efficient pre-trained BioNLP transformers for biomedical and clinical texts, leveraging model distillation techniques to enhance performance.
- Played a key role in the GLOPID-R Pandemic PACT project, integrating advanced language models for automatic annotation to monitor global pathogen research activities.
- Developed and released open-source biomedical language models, which are now widely adopted within the NLP and biomedical research communities.

Education

University of Wolverhampton

West Midlands, England

PhD in Natural Language Processing

2017 - 2020

- **Dissertation project**: "Computational Treatment of Non-literal Language"
- **Research synopsis**: Designed and developed ML models for computational modelling of figurative language, leveraging advances in contextualised representation learning. The work focused on creating linguistically interpretable architectures capable of processing multimodal information.

Sharif University of Technology

Tehran, Iran

MSc in Natural Language Processing

2013 - 2016

- **Dissertation project:** "Natural Language Generation From Visual Input" (GPA: 17.93/20)
- **Research synopsis:** Developed a multimodal NLG system translating visual elements into natural language, following a classic NLG pipeline for description generation.

Soore University

Tehran, Iran

BSc in Architecture

Oct 2009 - Oct 2013

- **Dissertation project:** "Designing a cultural centre with a philosophical and hermeneutic approach," exploring the interface of metaphors, language, and design principles (GPA: 16.68/20).
- **Skill highlights:** Gained foundational knowledge in mathematics for engineering, 3D modeling, and design tools, with a focus on problem-solving and analytical thinking in spatial design contexts.

Achievements & Awards

2024	**Best Paper in NLP (Global Selection),** Yearbook of the International Medical Informatics Association (IMIA)	*Global*
2022	**Global Talent Endorsement**, UK Research and Innovation	*Oxfordshire, UK*
2016	**Full PhD studentship**, University of Wolverhampton	*West Midlands, UK*
2017	**Best Paper Award**, Recent Advances in Natural Language Processing (RANLP 2017)	*Varna, Bulgaria*
2017	**Best Paper by a Young Researcher Award**, EUROPHRAS 2017	*Las Vegas, U.S.A*
2018	**Ranked 3rd (out of 44 teams in subtask A),** SemEval 2018 shared task 3 "Irony detection in English tweets"	*New Orleans, USA*
2018	**Ranked 4th (joint),** SemEval 2018 shared task 10 "Capturing Discriminative Attributes"	*New Orleans, USA*
2018	**Ranked 1st**, Automatic Identification of Verbal Multiword Expressions (LAW-MWE-CxG 2018)	*Santa Fe, USA*

Selected Projects

ISARIC Clinical Notes Project

University of Oxford

Processing Clinical Notes

Dec 2021 - Present

- Leading an NLP project with ISARIC to process a diverse corpus of clinical notes globally. The aim is to identify disease mentions and link them to clinical knowledge bases, enhancing understanding and utility of international healthcare data.

Global Research Collaboration for Infectious Disease Preparedness (GLOPID-R)

University of Oxford

Automating Annotation in Pandemic PACT Research Tracker

Feb 2023 - Present

- Involved in the Pandemic PACT Research Tracker project, focusing on automating annotation using in-context learning via large language models. The aim is to track global infectious disease research, identify trends, and pinpoint research gaps.

Instruction-based Generative LLMs for Medical NLP

University of Oxford

Lead Developer

June 2023 - Nov 2023

- Spearheaded the development of 'Llama2-MedTuned', a large-scale, instruction-tuned generative language model, in 7B and 13B versions, tailored for biomedical NLP tasks.
- One of the first LLMs specifically instruction-tuned for medical applications.
- Contributed to compiling a public medical dataset comprising 200K instruction-based samples,

Deep Learning for Tabular Data

University of Oxford

Model Design and Implementation

July 2021 - March 2022

- Investigated deep learning methods for tabular data, with a focus on privacy-preserving machine learning techniques.

Classification and Sequence Labelling with Graph Convolutional Networks (GCNs)

University of Wolverhampton

Model Design for Integration of Semantic and Syntactic Information

2018-2019

- Developed NLP architectures using GCNs to integrate structural and semantic information, enhancing the capability of syntax-aware classification and tagging models in sequence labelling tasks.

Community Service

Recent Advances in Natural Language Processing (RANLP 2021 & 2023)

Programme Committee Member

2021, 2023

Served as a Programme Committee member for RANLP, one of the leading and most competitive conferences in NLP, reviewing submissions across multiple years.

LoResLM 2025 (Low-Resource Language Models Workshop at COLING 2025)

Programme Committee Member

2025

Contributed as a Programme Committee member, reviewing papers focused on low-resource language models.

ACL, NAACL, EMNLP, EACL

Regular Reviewer for Major *CL Conferences

2018 - 2023

Cambridge Journal of Natural Language Engineering

Wolverhampton, UK

Editorial Assistant

Sep 2016 - Sep 2020

AI for Efficient Healthcare Systems, MSc in Primary Care (ADH Module 3), University of Oxford

Oxford, UK

Lecturer and Lab Instructor for 'Introduction to Biomedical NLP'

2024

Delivered a lecture and lab session on foundational concepts in Biomedical Natural Language Processing (NLP), focusing on real-world healthcare data applications.

EPSRC CDT in Health Data Science (HDS-M04), University of Oxford

Oxford, UK

Invited Lecturer for 'Introduction to Natural Language Processing'

Nov 2023 & Nov 2024

Delivered annual lectures on foundational NLP concepts as part of the 'Biomedical Time Series Analysis' module

Sharif University of Technology

Tehran, Iran

Teaching Assistant in 'Computer Programming' and 'An Introduction to NLP'

2014 - 2016

Assisted undergraduate students with programming fundamentals and introductory natural language processing concepts.

RANLP'19 Summer School on Deep Learning in Natural Language Processing (DLinNLP)

Varna, Bulgaria

Co-Organiser and Tutor

Aug 2019

Hosted a session on 'Sequence labelling and tagging' using PyTorch.

University of Wolverhampton

Wolverhampton, UK

Teaching Assistant for 'Machine Learning for NLP' and 'Computational Linguistics'

2018 - 2020

Assisted Prof Ruslan Mitkov in teaching master's students in Computational Linguistics.

Skills

Programming	Python (Numpy, Scipy, Matplotlib, Pandas, Scikit-learn), PyTorch, Transformers, PEFT, LangChain, Gensim.
NLP Expertise	Extensive experience with transformer-based models (BERT, GPT-series, T5), fine-tuning with PEFT, prompt engineering, knowledge-based systems, and building custom pipelines for NLP tasks.
Tools and Databases	Proficient with Git, Jupyter Notebooks; familiar with Bash scripting, SQL, and vector databases for NLP applications.
Languages	• **Persian** (Native speaker) • **English** (Full Professional Proficiency) • **Esperanto** (Fluent) • **Spanish** (Basic)

Selected Publications[1]

1. **Rohanian, O.**, Nouriborji, M., Seminog, O., Furst, R., Mendy, T., Levanita, S., Kadri-Alabi, Z., Jabin, N., Toale, D., Humphreys, G., Antonio, E., Bucher, A., Norton, A., Clifton, D. A. *Rapid Biomedical Research Classification: The Pandemic PACT Advanced Categorisation Engine*. (2024 - link to preprint, arXiv, 2407.10086, cs.CL).
2. **Rohanian, O.**, Nouriborji, M., Kouchaki, S., Nooralahzadeh, F., Clifton, L., Clifton, D. A. *Exploring the Effectiveness of Instruction Tuning in Biomedical Language Processing*. (2024 - link to publication, Artificial Intelligence in Medicine, 158, 103007).
3. **Rohanian, O.**, Nouriborji, M., Jauncey, H., Kouchaki, S., Nooralahzadeh, F., Clifton, L., Merson, L., Clifton, D. A., ISARIC Clinical Characterisation Group. *Lightweight Transformers for Clinical Natural Language Processing*. (2023 - link to paper, Natural Language Engineering, pp. 1-28).
4. **Rohanian, O.**, Nouriborji, M., Kouchaki, S., Clifton, D. A. *On the Effectiveness of Compact Biomedical Transformers*. (2023 - Bioinformatics, Vol. 39, No. 3, Article btad103, DOI: 10.1093/bioinformatics/btad103).
5. Taylor, N., Ghose, U., **Rohanian, O.**, Nouriborji, M., Kormilitzin, A., Clifton, D. A., Nevado-Holgado, A. *Efficiency at Scale: Investigating the Performance of Diminutive Language Models in Clinical Tasks*. (2024 - link to publication, Artificial Intelligence in Medicine, 157, 103002).
6. Seminog, O., Furst, R., Mendy, T., **Rohanian, O.**, Levanita, S., Kadri-Alabi, Z., Jabin, N., Humphreys, G., Antonio, E., Bucher, A., Norton, A. *A Protocol for a Living Mapping Review of Global Research Funding for Infectious Diseases with a Pandemic Potential–PANDEMIC PACT*. (2024 - link to paper, Wellcome Open Research, Vol. 9, Issue 156, pp. 156).
7. Rohanian, M., Nooralahzadeh, F., **Rohanian, O.**, Clifton, D., Krauthammer, M. *Disfluent Cues for Enhanced Speech Understanding in Large Language Models*. In Findings of the Association for Computational Linguistics: EMNLP 2023, pages 3676–3684, December 6-10, 2023. (link to paper)
8. **Rohanian, O.**, Kouchaki, S., Soltan, A. A. S., Yang, J., Rohanian, M., Yang, Y., Clifton, D. A. *Privacy-Aware Early Detection of COVID-19 Through Adversarial Training*. (2023 - Journal of Biomedical and Health Informatics, Vol. 27, No. 3, pp. 1249-1258, DOI: 10.1109/JBHI.2022.3230663) (link to paper).
9. Nouriborji, M., **Rohanian, O.**, Kouchaki, S., Clifton, D. A. *MiniALBERT: Model Distillation via Parameter-Efficient Recursive Transformers*. (2023 - link to paper, Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, pp. 1161-1173, Dubrovnik, Croatia).

   Rohanian, O., Jauncey, H. et al *Using Bottleneck Adapters to Identify Cancer in Clinical Notes under Low-Resource Constraints*. In Proceedings of the 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks (BioNLP 2023) (link to paper)
10. Nouriborji, M., **Rohanian, O.**, Clifton, D. A. (2022) *Nowruz at SemEval-2022 Task 7: Tackling Cloze Tests with Transformers and Ordinal Regression*. In Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022) (link to paper).
11. Soltan, Andrew. A. S., Yang, J., Pattanshetty, R., Novak, A., Yang, Y., **Rohanian, O.**, et al. *Real-world evaluation of AI-driven COVID-19 triage for emergency admissions: External validation & operational assessment of lab-free and high-throughput screening solutions* (2022 - *Lancet Digital Health*)(link to paper).

[1] For the full list of publications, please visit my Google Scholar page.

12. **Rohanian, O.**, Rei, M., Taslimipoor, S., Ha L. A. (2020) *Verbal Multiword Expressions for Identification of Metaphor*. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)(link to paper).

13. **Rohanian, O.**, Taslimipoor, S., Kouchaki, S., Ha, L.A and Mitkov, R. *Bridging the Gap: Attending to Discontinuity in Identification of Multiword Expressions*. In Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2019)(link to paper).

14. Chauhan, V. K., Thakur, A., O'Donoghue, O., **Rohanian, O.**, Molaei, S., Clifton, D. A. *Continuous Patient State Attention Model for Addressing Irregularity in Electronic Health Records*. (2024 - link to paper, BMC Medical Informatics and Decision Making, Vol. 24, Issue 1, pp. 117).

15. Liu, F., Li, Z., Zhou, H., Yin, Q., Yang, J., Tang, X., Luo, C., Zeng, M., Jiang, H., Gao, Y., Nigam, P., Nag, S., Yin, B., Hua, Y., Zhou, X., **Rohanian, O.**, Thakur, A., Clifton, L., Clifton, D. A. *Large Language Models Are Poor Clinical Decision-Makers: A Comprehensive Benchmark*. (2024 - link to publication, Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, pp. 13696–13710).