Tuba Gokhan

Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI), Masdar City, Abu Dhabi
Email: tuba.gokhan@mbzuai.ac.ae
Homepage: https://tuba-gokhan.github.io/
LinkedIn: https://www.linkedin.com/in/tubagokhan/
Phone: +971 (0) 585420212

Research Interests

- **Areas:** Computational linguistics, artificial intelligence, natural language processing, regulatory natural language processing (RegNLP), RegTech, FinTech, machine learning, data science, data visualization

- **Keywords:** Retrieval-augmented generation (RAG), regulatory and legal question answering, multi-passage and cross-reference retrieval, text and financial narrative summarisation, dense and learning-to-rank retrieval, graph-based retrieval and knowledge graphs, synthetic data generation, evaluation metrics for LLM-based QA and summarisation, readability and linguistic quality assessment

Education

***University of Birmingham**, Birmingham, UK*
PhD in School of Computer Science, 2019-2023
- Thesis: "Graph-Based Extractive Summarisation for Long Documents"
- Supervisor: Prof. Mark Lee

***Newcastle University,** Newcastle, UK*
MSc in Advanced Computer Science, 2017-2018
- Thesis: "Stream Data Visualization Technologies"
- Supervisor: Prof. Nick Holliman

***Gazi University,** Ankara, Türkiye*
MSc in Computer Engineering, 2014-2016
- Thesis: "An Interactive Big Data Warehouse Platform Development and Implementation for Nutrition and Health Research"

***Eskisehir Osmangazi University,** Eskisehir, Türkiye*
BSc, Computer Engineering, 2008-2012

Employment

Natural Language Processing Department, MBZUAI, UAE
Senior Postdoctoral Associate, September 2023 – present

- Lead research on large-scale LLMs for regulatory compliance in partnership with Abu Dhabi Global Market (ADGM).

- Design and implement Retrieval-Augmented Generation (RAG) systems integrated with vector databases to support domain-specific conversational and QA systems.
- Develop synthetic data generation pipelines to augment regulatory QA and summarisation datasets, improving robustness and coverage.
- Develop and apply evaluation metrics tailored to LLM outputs in regulatory settings, ensuring alignment with legal and compliance requirements.
- Collaborate with interdisciplinary teams (regulators, legal experts, data scientists) to deploy AI solutions informed by real-world supervisory workflows.
- Coordinate cross-institutional RegNLP activities, including the RegNLP workshop series, shared tasks, and collaboration with regulatory partners and industry stakeholders.

(https://regnlp.github.io/).

Department of Computer Science, University of Birmingham, UK
Teaching Associate, September 2019 – May 2023

- Delivered lab-based teaching and in-class support for undergraduate and postgraduate courses, including Natural Language Processing, Storing and Managing Data, Data Structures, Algorithms and Databases, Fundamentals of Databases, and Data Structures & Algorithms.
- Provided guidance to students during lab sessions, helping with programming exercises, coursework tasks, and exam preparation.
- Held office hours to answer questions related to assignments, projects, and course content.
- Marked assignments, lab reports, and exams, and provided structured feedback to students in line with module rubrics and programme learning outcomes.

Department of Computer Engineering*,* Gazi University, Türkiye
Research Assistant, December 2013 – July 2017

- Taught undergraduate courses and labs in Machine Learning, Software Engineering I & II, Data Structures, and Object-Oriented Programming.
- Served as academic advisor for a cohort of approximately 20–25 bachelor students, supporting module selection, monitoring academic progress, and providing guidance from enrolment through to graduation.
- Served as a member of the departmental Accreditation Committee and Education Committee, contributing to curriculum development, quality assurance processes, and accreditation documentation.
- Supported departmental outreach and student recruitment activities, including representing the department at university open days and advising prospective students.
- Led a six-term introductory computer literacy course for the public in collaboration with the Ankara City Council.
- Conducted research on machine learning methods applied to health and biomedical data.

Publications

- **Gokhan, T.**, & Briscoe, T. (manuscript under review). **Regulatory natural language processing: Compliance-oriented information retrieval and answer generation**. *Journal of Natural Language Processing.*

- **Gokhan, T.**, & Briscoe, T. (2025). **Grounded answers from multi-passage regulations: Learning-to-rank for regulatoryRAG**. In *Proceedings of the Natural Legal Language Processing Workshop 2025 (NLLP 2025)* (pp. 135–146), Suzhou, China. Association for Computational Linguistics. https://aclanthology.org/2025.nllp-1.10/

- **Gokhan, T.**, Wang, K., Gurevych, I., & Briscoe, T. (2025). ***Proceedings of the 1st Regulatory NLP Workshop (RegNLP 2025)***. Abu Dhabi, UAE: Association for Computational Linguistics. https://aclanthology.org/2025.regnlp-1.0/

- **Gokhan, T.**, Wang, K., Gurevych, I., & Briscoe, T. (2025). **Shared task RIRAG-2025: Regulatory information retrieval and answer generation.** In *Proceedings of the 1st Regulatory NLP Workshop (RegNLP 2025)* (pp. 1–4), Abu Dhabi, UAE. Association for Computational Linguistics. https://aclanthology.org/2025.regnlp-1.1/

- **Gokhan, T.**, Wang, K., Gurevych, I., & Briscoe, T. (2024). **RegNLP in action: facilitating compliance through automated information retrieval and answer generation**. *arXiv preprint* arXiv:2409.05677. https://arxiv.org/abs/2409.05677v1

- **Gokhan, T.** (2024). **Quantification of redundancy in text summaries.** In *Proceedings of the International Conference on Engineering Technologies (ICENTE 24)*, Konya, Türkiye. https://icente.selcuk.edu.tr/docs/icente2024_abstract_book.pdf

- **Gokhan, T.**, Price, M. J., & Lee, M. (2024). **Graphs in clusters: A hybrid approach to unsupervised extractive long document summarization using language models**. *Artificial Intelligence Review, 57*(7), 189. https://link.springer.com/article/10.1007/s10462-024-10828-w

- **Gokhan, T.** (2023). ***Graph-Based extractive summarisation for long documents*** (Doctoral thesis). University of Birmingham, Birmingham, United Kingdom. https://doi.org/10.13140/RG.2.2.16009.76647

- **Gokhan, T.**, Smith, P., & Lee, M. (2023). **Node-weighted centrality ranking for unsupervised long document summarization**. In Métais, E., Meziane, F., Sugumaran, V., Manning, W., & Reiff-Marganiec, S. (Eds.), *Natural Language Processing and Information Systems (NLDB 2023)*, *Lecture Notes in Computer Science* (Vol. 13913, pp. 299–312). Cham, Switzerland: Springer Nature. https://doi.org/10.1007/978-3-031-35320-8_21 (Best Paper)

- **Gokhan, T.**, Smith, P., & Lee, M. (2022). **GUSUM: Graph-based unsupervised summarization using sentence features scoring and sentence-bert**. In *Proceedings of TextGraphs-16: Graph-based Methods for Natural Language Processing* (pp. 44–53), Gyeongju, Republic of Korea. Association for Computational Linguistics. https://aclanthology.org/2022.textgraphs-1.5/

- **Gokhan, T.**, Smith, P., & Lee, M. (2021). **Extractive financial narrative summarisation using sentence Bert-based clustering**. In *Proceedings of the 3rd Financial Narrative Processing Workshop* (pp. 94–98), Lancaster, United Kingdom. Association for Computational Linguistics. https://aclanthology.org/2021.fnp-1.18/

- **Gokhan, T.** (2018). ***Stream data visualisation technologies*** (Master's thesis). Newcastle University, Newcastle upon Tyne, United Kingdom. https://doi.org/10.13140/RG.2.2.13485.77281

- Çetin, A., & **Gokhan, T.** (2018). **Differential diagnosis of erythematous squamous diseases with feature selection and classification algorithms**. In Kose, U., et al. (Eds.), *Nature-inspired intelligent techniques for solving biomedical engineering problems* (pp. 103–129). IGI Global. https://doi.org/10.4018/978-1-5225-4769-3.ch005

- **Gokhan, T.**, & Çetin, A. (2017). **Estimation of obesity with machine learning approaches based on socio-demographic data.** In *Proceedings of the IIII International Conference on Engineering and Natural Sciences (ICENS 2017)*, Budapest, Hungary. https://www.icens.eu/sites/default/files/2017_icens_absract_v9_1.pdf

- **Gokhan, T.**, & Çetin, A. (2017). **Evaluation of classification methods for early diagnosis of chronic kidney diseases.** In *Proceedings of the IIII International Conference on Engineering and Natural Sciences (ICENS 2017)*, Budapest, Hungary. https://www.icens.eu/sites/default/files/2017_icens_absract_v9_1.pdf

- Sahin, B., **Gokhan, T.**, & Çetin, A. (2017). **Identification of factors causing diabetes using data mining methods**. In *Proceedings of Akademik Bilişim 2017 (AB2017)*, Kırıkkale, Türkiye. https://ab.org.tr/ab17/bildiri/146.pdf

- **Gokhan, T.**, & Çetin, A. (2015). **A comparative analysis of current approaches in high-performance distributed computing systems**. In *Proceedings of Akademik Bilişim 2015 (AB2015)*, Eskişehir, Türkiye. https://ab.org.tr/ab15/bildiri/345.pdf

- **Gokhan, T.**, & Çetin, A. (2015). **Creating database for collaborative study on health with use of social media and mobile technologies**. In *Proceedings of the Conference of the International Journal of Arts and Sciences (IJAS 2015)* (pp. 277–280), Vienna, Austria. https://www.universitypublications.net/proceedings/0804/pdf/V5NA430.pdf

Teaching Experience

- Lecturing (Academic)

  - **Teaching Associate,** *Natural Language Processing (Lab Sessions)*, MSc Course, Department of Computer Science, University of Birmingham, Spring 2022-2023

  - **Teaching Associate,** *Storing and Managing Data (Lab Sessions)*, MSc Course, Department of Computer Science, University of Birmingham, Fall 2022-2023

  - **Teaching Associate,** *Storing and Managing Data (Lab Sessions)*, MSc Course, Department of Computer Science, University of Birmingham, Spring 2021-2022

  - **Teaching Associate,** *Systems Programming in C/C++ (Lab Sessions)*, BSc Course, Department of Computer Science, University of Birmingham, Fall 2021-2022

  - **Teaching Associate,** *Storing and Managing Data (Lab Sessions)*, MSc Course, Department of Computer Science, University of Birmingham, Spring 2020-2021

  - **Teaching Associate,** *Data Structures, Algorithms, and Databases (Lab Sessions)*, BSc Course, Department of Computer Science, University of Birmingham, Fall 2020-2021

  - **Teaching Associate,** *Fundamentals of Databases (Lab Sessions)*, BSc Course, Department of Computer Science, University of Birmingham, Spring 2019-2020

- **Teaching Associate,** *Data Structures & Algorithms (Lab Sessions)*, BSc Course, Department of Computer Science, University of Birmingham, Fall 2019-2020

- **Lecturer,** *Software Engineering II,* BSc Course, Department of Computer Engineering, Gazi University, Spring 2016-2017 *(Developing course material and lecturing)*

- **Lecturer,** *Machine Learning*, BSc Course, Department of Computer Engineering, Gazi University, Fall 2016-2017 *(Developing course material and lecturing)*

- **Lecturer,** *Software Engineering I,* BSc Course, Department of Computer Engineering, Gazi University, Fall 2016-2017 *(Developing course material and lecturing)*

- **Lecturing,** *Bioinformatic (Lab Sessions),* BSc Course, Department of Computer Engineering, Gazi University, Fall 2016-2017

- **Lecturer,** *Software Engineering II,* BSc Course, Department of Computer Engineering, Gazi University, Spring 2015-2016 *(Developing course material and lecturing)*

- **Lecturer,** *Data Structures (Lab Sessions)*, BSc Course, Department of Computer Engineering, Gazi University, Fall 2015-2016

- **Lecturer,** *Software Engineering I,* BSc Course, Department of Computer Engineering, Gazi University, Fall 2015-2016 *(Developing course material and lecturing)*

- **Lecturer,** *Software Engineering II,* BSc Course, Department of Computer Engineering, Gazi University, Spring 2014-2015 *(Developing course material and lecturing)*

- **Lecturer,** *Data Structures (Lab Sessions)*, BSc Course, Department of Computer Engineering, Gazi University, Fall 2014-2015

- **Lecturer,** *Software Engineering I,* BSc Course, Department of Computer Engineering, Gazi University, Fall 2014-2015 *(Developing course material and lecturing)*

- **Lecturer,** *Object-Oriented Programming (Lab Sessions),* BSc Course, Department of Computer Engineering, Gazi University, Spring 2013-2014

- **Lecturer,** *Data Structures (Lab Sessions)*, BSc Course, Department of Computer Engineering, Gazi University, Fall 2013-2014

- Lecturing (Outreach)

  - **Lecturer,** *Introductory Computer Literacy Course*, Gazi University-Ankara City Council, Spring 2015-2016

  - **Lecturer,** *Introductory Computer Literacy Course*, Gazi University-Ankara City Council, Fall 2015-2016

  - **Lecturer,** *Introductory Computer Literacy Course*, Gazi University-Ankara City Council, Spring 2014-2015

  - **Lecturer,** *Introductory Computer Literacy Course*, Gazi University-Ankara City Council, Fall 2014-2015

  - **Lecturer,** *Introductory Computer Literacy Course*, Gazi University-Ankara City Council, Spring 2013-2014

  - **Lecturer,** *Introductory Computer Literacy Course*, Gazi University-Ankara City Council, Fall 2013-2014

Other Professional Activities

- **Co-organizer and Committee member,** Workshop on Regulatory Natural Language Processing (RegNLP)
- **Lead organiser**, Regulatory Information Retrieval and Answer Generation (RIRAG)-2025 shared task
- **Scientific committee member and Reviewer,** International Conference on Engineering Technologies (ICENTE) 2024 and 2025
- **Reviewer,** *Information and Software Technology* (special issue: Regulatory Compliance in Software Engineering)
- **Editor,** *Emerging Trends and Applications in Computer Engineering 2026* (book)
- **External Research Collaborator** on Regulatory NLP and RegTech, Abu Dhabi Global Market (ADGM)
- **Member,** RegNLP Research Group (regulatory NLP research collaboration)

Awards and Scholarships

- YLSY Scholarship, Republic of Türkiye Ministry of National Education – full scholarship for MSc studies abroad
- YLSY Scholarship, Republic of Türkiye Ministry of National Education – full scholarship for PhD studies abroad
- Best Paper Award, *Natural Language Processing and Information Systems (NLDB 2023)*, for the paper "Node-weighted centrality ranking for unsupervised long document summarization"
- First prize, Savronik Project Competition 2012 (SPY-2012), for the project "*Secerem*: a 3D social media platform for exploring family trees"

Trainings

HEFI (Higher Education Futures Institute), *University of Birmingham, UK*

- Introduction to Learning and Teaching in Higher Education for Postgraduates
- Large Group Teaching (Lectures)
- Principles of Assessment and Feedback
- Small Group Teaching (Seminars)
- Small Group Teaching (Labs)

Updated: 29 Nov. 25