Ayoub Bagheri

Utrecht University
a.bagheri@uu.nl
https://nlp.sites.uu.nl

CURRENT POSITIONS

Carear-track Associate Professor in Applied Data Science (July 2022-present)

Lead of the Natural Language Processing group at the Department Methodology and Statistics (September 2022-present)

Coordinator of the Applied Data Science master profile (September 2021-present)

Chair of the M&S AI/ADS taskforce committee (January 2022-present)

Member of the Utrecht Young Academy (March 2023-present)

Board member of CUCo, Center of Unusual Collaborations (May 2023-present)

EXPERIENCE

Utrecht University 2022–present
Career-track Associate Professor

Utrecht University 2020–present
Tenured Assistant Professor

Utrecht University, University Medical Center Utrecht 2017–2020
PhD in M&S (2.5 years)

University of Kashan 2015–2017
Lecturer / Researcher

University of Twente 2012 (6 months)
Research visit Human Media Interaction Group

Isfahan University of Technology 2010–2014
Researcher in Computer Engineering

RESEARCH AND TEACHING INTERESTS

Natural Language Processing
Machine Learning
Statistical Learning
Text Mining
Applied / Human Data Science
Computational Social Science

OPEN SCIENCE

Member of the Open Science Community Utrecht
Member of ASReview team (open-source software: (*Website*))
Organizer of the data science reading group at M&S
Open and legal pre- and post-prints on arxiv, bioarxiv, and osf (*Google Scholar page*)
Author of open source Python and R packages: (*GitHub page*)
Teaching: Open materials, data, code: (*GitHub page*)

TEAM SCIENCE

Lead of the M&S Natural Language Processing group (*Website*) (2022-present)

Member of the Human Data Science team, Utrecht University (2017-present)

Member of the ASReview team (more than 15 members working on an open-source software for automatizing systematic reviewing) (2020-present)

Student co-supervision with different departments at Utrecht University, University Medical Center Utrecht, and University College London. (2020-present)

Collaboration with University College London, Text Mining projects (2019-)

Collaboration with University Medical Center Utrecht, Applied Data Science projects (2017-present)

Collaboration on research projects with different departments in Utrecht University, including Digital Humanities, Information and Computing Sciences, Social Sciences. (2017-present)

International co-authorship collaboration: Clinical Applications of Artificial Intelligence in Real-World Data, a collaboration between Utrecht University, University Medical Center Utrecht, University College London, and University of Pennsylvania (2022)

SELECTED GRANTS

Hyperparameter Optimization to accelerate active Learning Models, Netherlands eScience Center, Grant ID: NLESC.SSI.2022b.009, Principal Investigator, 2023.

NATURAL INTELLIGENCE (NI) IN HEALTH, 2M, KANSEN VOOR WEST III, Co-applicant (Principal Investigator: Vivaltes B.V., Project partners: SU Biomedicine B.V., FytagoLife B.V., Single Cell Discoveries B.V., Principelle B.V., Universiteit Utrecht, Agrobrains B.V., Alloksys B.V.), 2022.

Mining the Dutch Disposition towards Animals and Plants, PhD funding, AI Labs call to action, Utrecht University, Co-applicant (Principal Investigator: Prof. dr. Els Stronk), 2022.

Responsible use of free text notes in electronic health records in medical prediction research, 100K, NWA Route Big Data – Small Project, Co-applicant (Principal Investigator: Dr. A.M. Leeuwenberg), 2022.

Journalistic Inquiry On Public Debates, PhD funding, AI Labs call to action, Utrecht University, Co-applicant (Principal Investigator: Dr. Karin van Es, Dr. Mirko Tobias Schäfer), 2021.

Identifying Group of Relevant Antibodies in predicting Kidney Transplantation Risk using Text Mining, €5K, Focus Area Applied Data Science, Utrecht University, Co-applicant (Principal Investigator: Dr. D. M. Senejohnny), 2021.

Empowering teachers through automated text analysis of student assignments, €2K, Focus Area Applied Data Science, Utrecht University, Co-applicant (Principal Investigator: Dr. A. van Leeuwen), 2021.

Social bias detection in text using NLP and their associations with people's opinions?, €5K, Focus Area Applied Data Science, Utrecht University, Principal Investigator, 2021.

Text mining as imputation model for clinical observational studies: does it have to be perfect to be useful?, €10K, Focus Area Applied Data Science, Utrecht University, Co-applicant (Principal Investigator: Dr. A. Leeuwenberg), 2021.

Prediction of prognosis of patients with heart failure using deep learning on electrocardiograms and text mining, €5K, Focus Area Applied Data Science, Utrecht University, Co-applicant (Principal Investigator: Prof. F.W. Asselbergs), 2020.

DEEP-ENIGMA: A DEEP nEural Network for Image seGMentation to classify and quantify Atherosclerotic disease based on high-resolution scanned histological slides, €5K, Focus Area Applied Data Science, Utrecht University, Co-applicant (Principal Investigator: Dr. S. van der Laan), 2020.

CONVOCALS: A CONVOlutional neural network to predict symptoms and major secondary CArdiovascuLar events based on high-resolution scanned histological Slides, €5K, Focus Area Applied Data Science, Utrecht University, Co-applicant (Principal Investigator: Dr. S. van der Laan), 2019.

Deep neural networks for ICD-10 classification of diagnosis registration in cardiovascular notes to allow data mining in electronic health records, €5K, Focus Area Applied Data Science, Utrecht University, Principal Investigator, 2018.

Automated information extraction from patients' clinical data using text analysis techniques, €5K, Focus Area Applied Data Science, Utrecht University, Principal Investigator, 2018.

Integrated decision support systems for municipal spatial data, €10K, Isfahan Municipality, Isfahan University, Co-applicant (Principal Investigator: Dr. S. Nadi), 2013.

ACADEMIC SERVICE

General chair of the BNAIC2024 conference at Utrecht University

Committee member in data science and NLP conferences such as CLIN2020, EMNLP2022, BNAIC2023

Staff member of the Interuniversity Graduate School of Psychometrics and Sociometrics (IOPS), 2022–present

Coordinator of the Applied Data Science Master Profile, Utrecht University, 2021–present

Co-Organizer of the SIG Active Learning, Focus Area Applied Data Science & Human-centered Artificial Intelligence, Utrecht University, 2021–present

Member of Association for Computational Linguistics (ACL), 2021–present

Organizer of Data Science Reading Group, Methodology and Statistics, Utrecht University, 2021–present

Co-Organizer of the SIG Text Mining, Focus Area Applied Data Science, Utrecht University, 2020–present

Member of the ASReview team, Utrecht University, 2020–present

Member of the Association for Computing Machinery (ACM), 2020–present

Member of the Netherlands Society for Statistics and Operations Research (VVSOR), 2020–present

Member of the Institute for Systems and Technologies of Information, Control and Communication (INSTICC), 2020–present

Member of the PhD Council in the Faculty of Social & Behavioural Sciences, Utrecht University, 2019–2020

Organizer of Deep Learning Reading Group, Methodology and Statistics, Utrecht University, 2018–2021

Member of the Open Science Community Utrecht, 2018–present

Member of the Human Data Science Lab (previous MSDSLab), Methodology and Statistics, Utrecht University, 2017–present

Reviewer, Guest editor:
Knowledge-Based Systems (Elsevier), *Journal of Information Science* (SAGE), *Artificial Intelligence Review* (Springer), *Journal of Experimental & Theoretical Artificial Intelligence* (Taylor and Francis), *Soft Computing Journal* (Springer), *Transactions on Multimedia Computing Communications and Applications* (ACM), *International Journal of Data Science and Analytics* (Springer), *Association for Computational Linguistics* (ACL), *IEEE Transactions on Computational Social Systems* (IEEE), *European Heart Journal*

SELECTED TEACHING AND COORDINATING EXPERIENCES

UMC Utrecht (Summer 2018, 2019, 2022), *Big Data in Health Research*, Developer (20%), Teacher.

Utrecht University (2021-2022), *Battling the Curse of Dimensionality*, `https://uudav.nl/`, Developer (22%), Teacher.

Utrecht University (Summer 2021, 2022), *Introduction to Text Mining with R*, `https://ayoubbagheri.nl/r_tm/`, New course started in 2021, Coordinator, Developer (75%), Teacher.

Utrecht University (Summer 2021, 2022), *Applied Text Mining*, `https://ayoubbagheri.nl/applied_tm/`, New course started in 2021, Coordinator, Developer (70%), Teacher.

Utrecht University (2020-2021, 2021-2022), *Data Analysis and Data Wrangling*, Coordinator, Developer (22%), Teacher.

Utrecht University (2020-2021, 2021-2022), *Applied Data Analysis and Visualisation*, Developer (10%), Teacher.

Utrecht University (Summer 2019, 2021), *Data Science: Statistical Programming with R*, Teacher.

Utrecht University (2018-2019, 2019-2020, 2020-2021), *Data Analysis and Visualisation*, Teacher.

University of Kashan (2016-2017), *Web Mining and Search Engines*, Coordinator, Developer (100%), Teacher.

University of Kashan (2016-2017), *Advanced Data Mining*, Coordinator, Developer (100%), Teacher.

University of Kashan (2015-2016), *Algorithm Design*, Coordinator, Developer (60%), Teacher.

University of Kashan (2015-2016), *Advanced C++ Programming*, Coordinator, Developer (50%), Teacher.

Teacher of the Year Award (2017), Computer Science, University of Kashan.

SUPERVISION

Postdoc:

Dr. Huyen Nguyen (2022-), Text & acoustic data mining tools to understand gender differences in debate tournament speeches and negotiation exchanges, Co-supervision with Frank van Tubergen, Utrecht University.

PhD:

Arjan van Dalfsen (2023-), Mining the Dutch disposition towards animals and plants, in a co-supervision with Folgert Karsdorp and Els Stronks, Utrecht University.

Hadi Mohammadi (2023-), Explainable natural language processing for social sciences, in a co-supervision with Anastasia Giachanou and Daniel Oberski, Utrecht University.

Paul Keuren (2023-), Knowledge-based natural language processing for real world applications, in a co-supervision with Marc Ponsen (Statistics Netherlands, CBS) and Daniel Oberski, Utrecht University.

Niek van de Pas (2022-), Discovering Dutchness through text mining, in a co-supervision with Willy Sier, Wil Pansters and Daniel Oberski, Utrecht University.

Jelle Teijema (2022-), FAIR metadata of current and next generation systematic reviews, in a co-supervision with Rens van de Schoot, Lars Tummers, and Matthieu Brinkhuis, Utrecht University.

Mahsa Afsharizade (2016-2021), Improving text summarization using deep learning and noun phrase coreference resolution, Co-supervision with Hossein Ebrahimpour, University of Kashan.

Master students:

Christine Hedde-von Westernhagen (M&S UU, 2023), Daniel Anadria (M&S UU, 2023), Kevin Patyk (M&S UU, 2022), Daan de Jong (M&S UU, 2021), Zhenwei Jang (M&S UU, 2021), Martijn Strum (INFO UU, 2021), Jelle Teijema (ADS UU, 2021), Pablo Vizan (ADS UU, 2021), Govert Verberg (ADS UU, 2021), Bart-Jan Boverhof (M&S UU, 2021), and more before 2020. You can contact me for the full list.

KEY PUBLICATIONS

Fang, Q., Giachanou, A., **Bagheri, A.**, Boeschoten, L., van Kesteren, E.J., Shafiee Kamalabad, M., Oberski, D.L. (2023). On Text-based Personality Computing: Challenges and Future Directions. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, ACL 2023.*

van Tubergen, F., Nadi Njafabadi, S., **Bagheri, A.** (Submitted, 2023). Media portrayals of immigrants, refugees, and Muslims: Group hierarchies in online news sources in the Netherlands. *Journal of Ethnic and Migration Studies.*

Soppe, K., **Bagheri, A.**, Nadi Njafabadi, S., Klugkist, I., Wubbels, T., Wijngaards-de Meij, L. (Submitted, 2023). Pre-university motivation and first year study success: Text mining for pre-enrollment questionnaires. *Higher Education.*

Keuren, P., Ponsen, M., **Bagheri, A.** (2022). WordGraph2Vec: Combining domain knowledge with text embeddings. In Joint International Scientific Conferences on AI and Machine Learning (BNAIC/BeNeLearn 2022).

de Jong, D., **Bagheri, A.** (2022). The Case of Imperfect Negation Cues: A Two-Step Approach for Automatic Negation Scope Resolution. In International Conference on Applications of Natural Language to Information Systems (pp. 413-424). Springer, Cham.

Afsharizadeh, M., Ebrahimpour-Komleh, H., **Bagheri, A.**, Chrupała, G. (2022). A Survey on Multi-document Summarization and Domain-Oriented Approaches. Journal of Information Systems and Telecommunication (JIST), 1(37), 68.

Teijema, J., Hofstee, L., Brouwer, M., de Bruin, J., Ferdinands, G., de Boer, J., Vizan, P., van den Brand, S., Bockting, C., van de Schoot, R., **Bagheri, A.** (Submitted, 2022). Systematic reviewing using Active Learning-based neural networks: the case of the onset, maintenance, and relapse of depressive disorders. *Expert Systems with Applications.*

Felix, S. EA., **Bagheri, A.**, Ramjankhan, F. R., Spruit, M. R., Oberski, D. L., de Jonge, N., van Laake, L. W., Suyker, W. JL., Asselbergs, F. W., (2021). A data mining-based cross-industry process for predicting major bleeding in mechanical circulatory support. *European Heart Journal-Digital Health*, 2(4), 635-642.

Bagheri, A., Groenhof, T. K. J., Asselbergs, F. W., Haitjema, S., Bots, M. L., Veldhuis, W. B., de Jong, P. A., Oberski, D. L. (2021). Automatic prediction of recurrence of major cardiovascular events: A text mining study using chest X-ray reports. *Journal of Healthcare Engineering*, 1-11.

Boeschoten, L., van Kesteren, E. J., **Bagheri, A.**, Oberski, D. L. (2021). Achieving fair inference using error-prone outcomes. *International Journal of Interactive Multimedia & Artificial Intelligence*, 6(5).

Sammani, A., **Bagheri, A.**, van der Heijden, P. G. M., Te Riele, A., Baas, A., Oberski, D. L., Asselbergs, F. W. (2021). Automatic multilabel detection of ICD10 codes in Dutch cardiology discharge letters using neural networks. *NPJ digital medicine*, 4(1), 1-10.

Yang, Z., **Bagheri, A.**, van der Heijden, P. G. (2021). Neural Networks for Latent Budget Analysis of Compositional Data. *arXiv preprint* arXiv:2109.04875.

Bagheri, A., Sammani, A., van der Heijden, P. G. M., Asselbergs, F. W., Oberski, D. L. (2020). ETM: Enrichment by topic modeling for automated clinical sentence classification to detect patients' disease history. *Journal of Intelligent Information Systems*, 55(2), 329-349.

Afsharizadeh, M., Ebrahimpour-Komleh, H., **Bagheri, A.** (2020). Automatic Text Summarization of COVID-19 Research Articles Using Recurrent Neural Networks and Coreference Resolution. *Frontiers in Biomedical Technologies*, 7(4), 236-248.

Bagheri, A., Groenhof, T. K. J., Veldhuis, W. B., de Jong, P. A., Asselbergs, F. W., Oberski, D. L. (2020). Multimodal learning for cardiovascular risk prediction using EHR data. *In Proceedings of the 11th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB 2020).*

Bagheri, A., Sammani, A., van der Heijden, P. G. M., Asselbergs, F. W., Oberski, D. L. (2020). Automatic ICD-10 classification of diseases from Dutch discharge letters. *In Proceeding of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies*, 281-289.

Boeschoten, L., van Kesteren, E. J., **Bagheri, A.**, Oberski, D. L. (2020). Fair inference on error-prone outcomes. *ECAI conference 2020 workshop: Artificial Intelligence for a Fair, Just and Equitable World.*

van de Leur, R. R., Boonstra, M. J., **Bagheri, A.**, Roudijk, R. W., Sammani, A., Taha, K., Doevendans, P. A. F. M., van der Harst, P., van Dam, P. M., Hassink, R. J., van Es, R., Asselbergs F. W. (2020). Big data and artificial intelligence: Opportunities and threats in electrophysiology, *Arrhythmia & Electrophysiology Review*, 9(3), 146-154.

Ferdinands, G., Schram, R., de Bruin, J., **Bagheri, A.**, Oberski, D., Tummers, L., van de Schoot, R. (2020). Interactive screening prioritization in systematic reviews by employing active learning.

Sammani, A., Jansen, M., Linschoten, M., **Bagheri, A.**, de Jonge, N., Kirkels, H., van Laake, L., Vink, A., van Tintelen, J. P., Dooijes, D., Te Riele, A., Harakalova, M., Baas, A. F., Asselbergs, F. W. (2019). UNRAVEL: Big data analytics research data platform to improve care of patients with cardiomyopathies using routine electronic health records and standardized biobanking. *Netherlands Heart Journal*, 27(9), 426-434.

Bagheri, A., Oberski, D. L., Sammani, A., van der Heijden, P. G. M., Asselbergs, F. W. (2019). SALTClass: Classifying clinical short notes using background knowledge from unlabeled data. *bioRxiv 801944.*

Bagheri, A. (2019). Integrating word status for joint detection of sentiment and aspect in reviews. *Journal of Information Science*, 45(6), 736-755.

Sedighi, Z., Ebrahimpour-Komleh, H., **Bagheri, A.**, Kosseim, L. (2019). Learning to identify fake opinion reviews using a neural network with attention. *In Proceeding of the 32nd International Artificial Intelligence Research Society Conference*, 245-248.

Asgarnezhad, R., Monadjemi, S. A., Soltanaghaei, M., **Bagheri, A.** (2018). SFT: A model for sentiment classification using supervised methods on Twitter. *Journal of Theoretical & Applied Information Technology*, 96(8), 2242-2251.

Afsharizadeh, M., Ebrahimpour-Komleh, H., **Bagheri, A.** (2018). Query-oriented text summarization using sentence extraction technique. *In Proceeding of the 4th IEEE International Conference on Web Research*, 128-132.

Sedighi, Z., Ebrahimpour-Komleh, H., **Bagheri, A.** (2017). RLOSD: representation learning based opinion spam detection. *In Proceeding of the 3rd IEEE Conference on Intelligent Systems and Signal Processing*, 74-80.

Zaghian, A., **Bagheri, A.** (2016). A combined model of clustering and classification methods for preserving privacy in social networks against inference and neighborhood attacks. *International Journal of Security and Its Applications*, 10(1), 95-102.

Bagheri, A., Saraee, M., Nadi Njafabadi, S. (2015). PSA: A hybrid feature selection approach for Persian text classification. *Journal of Computing and Security, Special Issue in Data Mining and Knowledge Discovery*, 1(4), 261-272.

Bagheri, A., Saraee, M., de Jong, F. (2014). ADM-LDA: An aspect detection model based on topic modeling by using structures of reviews. *Journal of Information Science*, 40(5), 621-636.

Bagheri, A., Saraee, M., de Jong, F. (2014). Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews. *Knowledge-Based Systems*, 52, 201–213.

Bagheri, A., Saraee, M. (2014). Persian sentiment analyzer: A framework based on a novel feature selection method. *International Journal of Artificial Intelligence*, 12(2), 115-129.

Bagheri, A., Saraee, M., de Jong, F. (2013). An unsupervised aspect detection model for sentiment analysis of reviews. *Natural Language Processing and Information Systems NLDB2013*, 140-151.

Saraee, M., **Bagheri, A.** (2013). Feature selection methods in Persian sentiment analysis. *Natural Language Processing and Information Systems NLDB2013*, 303-308.

Bagheri, A., Saraee, M., de Jong, F. (2013). Latent Dirichlet Markov allocation for sentiment analysis. *In Proceeding of the Fifth European Conference on Intelligent Management Systems in Operations*, 90-96.

Bagheri, A., Saraee, M., de Jong, F. (2013). Sentiment classification in Persian introducing a mutual information-based method for feature selection. *In Proceedings of the 21st Iranian Conference on Electrical Engineering*, 1-6.

Nadi Njafabadi, S., Saraee, M., **Bagheri, A.** (2011). A hybrid recommender system for dynamic web users. *International Journal of Multimedia and Image Processing*, 1(1), 3-8.

Moghimi, M., Saraee, M., **Bagheri, A.** (2011). Modeling batch annealing process using data mining techniques for cold rolled steel sheets. *In Proceedings of the First International Workshop on Data Mining for Service and Maintenance*, 18-22.

Nadi Njafabadi, S., Saraee, M., **Bagheri, A.** (2010). FARS: Fuzzy ant based recommender system for web users. *International Journal of Computer Science Issues*, 7(6), 203-209.

Norouzzadeh, M., **Bagheri, A.**, Saraee, M. (2009). Web search personalization: A fuzzy adaptive approach. *In Proceeding of the 2nd IEEE International Conference on Computer Science and Information Technology*, 143-148.

Bagheri, A., Akbarzadeh, M., Saraee, M. (2008). Finding shortest path with learning algorithms. *International Journal of Artificial Intelligence*, 1(8), 86-95.