ARTUUR LEEUWENBERG, Ph.D.

Email: a.m.leeuwenberg-15@umcutrecht.nl

RESEARCH FOCUS

(Clinical) natural language processing, medical prediction methodology, machine learning

ACADEMIC POSITIONS

Since 2019 **Assistant Professor – University Medical Center Utrecht, Utrecht University**

Methodological research on clinical natural language processing and prediction modeling at the Julius Center for Health Sciences and Primary Care.

Since 2024 **Methodologist – University Medical Center Groningen**

Detachment (0,2 fte): methodological consultancy on post-radiotherapy prediction models.

Since 2024 **Methodologist – DUPROTON (Dutch Proton Therapy Centres)**

Detachment (0,2 fte): development and evaluation of post-radiotherapy prediction models for model-based selection.

2015-2019 **PhD Researcher – Katholieke Universiteit Leuven**

Conducted research on machine learning and NLP methodology for the automated extraction of events (symptoms, treatments and tests) and their timing from clinical notes, at the Language Intelligence and Information Retrieval Lab, at the Department of Computer Science.

2014-2015 **Research Assistant – Universität des Saarlandes**

Conducted (published) research on automatic extraction of synonyms for better machine translation evaluation together with Mihaela Vela, Jon Dehdari, and Josef van Genabith, at the Dept. of Language Science and Technology.

Summer 2014 **Research Intern – Inria**[1]

Conducted (published) research on automatic extraction of drug-drug interactions from biomedical articles, together with Aleksey Buzmakov, Yannick Toussaint and Amedeo Napoli, at Inria Nancy-Grand-Est.

EDUCATION

2015-2019 **Ph.D. Computer Science**

Katholieke Universiteit Leuven, Belgium
Title: From Text to Time: Machine Learning Approaches to Temporal Information Extraction from Text
Advisor: Prof. dr. Marie-Francine Moens

2013-2015 **M.Sc. Language Science and Technology**

Université de Lorraine, France (2014-2015) / Universität des Saarlandes, Germany (2015-2016)
Double degree, with 2-year Erasmus Mundus Scholarship, with distinction.

2010-2013 **B.Sc. Cognitive Artificial Intelligence**

Universiteit Utrecht, The Netherlands

SELECTED PUBLICATIONS (PEER REVIEWED)

A complete list can be found at `http://bit.ly/PubsTuur`.

[1] Merijn Rijk, Tamara Platteel, Marissa Mulder, Geert-Jan Geersing, Frans Rutten, Maarten van Smeden, Roderick Venekamp, and **Artuur Leeuwenberg**. Incomplete and possibly selective recording of signs, symptoms and measurements in free text fields of primary care electronic health records of adults with lower respiratory tract infections. *Journal of Clinical Epidemiology*, page 111240, 2023.

[2] Madhumita Sushil, Atul J Butte, Ewoud Schuit, Maarten van Smeden, and **Artuur Leeuwenberg**. Cross-institution natural language processing for reliable clinical association studies: a methodological exploration. *Journal of Clinical Epidemiology*, 167:111258, 2024.

[3] Anne de Hond*, **Artuur Leeuwenberg***, Lotty Hooft, Ilse M. J. Kant, Steven W. J. Nijman, Hendrikus J. A. van Os, Jiska J. Aardoom, Thomas P. A. Debray, Ewoud Schuit, Maarten van Smeden, Johannes B. Reitsma, Ewout W. Steyerberg, Niels H. Chavannes, and Karel G. M. Moons. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *npj Digital Medicine*, 5(1):2, 2022. (*shared first authors).

[4] **Artuur Leeuwenberg** and Marie-Francine Moens. Towards Extracting Absolute Event Timelines from English Clinical Reports. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:2710–2719, 2020.

1 French National Institute for Research in Computer Science and Automation

[5] **Artuur Leeuwenberg** and Marie-Francine Moens. A Survey on Temporal Reasoning for Temporal Information Extraction from Text. *The Journal of Artificial Intelligence Research (JAIR)*, 2019. (invited for presentation at IJCAI 2020, core A*).

[6] **Artuur Leeuwenberg** and Marie-Francine Moens. Temporal Information Extraction by Predicting Relative Time-lines. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Brussels, Belgium, 2018. ACL. (core A, oral presentation).

GRANTS & SCHOLARSHIPS

Co-applicant for "WhyMBA: Why and when to use the Model-Based Approach to evaluate clinical effects of radiotherapy techniques?" (ZonMw, €200k, 2022)

Co-applicant for RAISE: Lead a (€102k) subproject on "Responsible use of free text in medical prediction research." (NWO, NWA.1418.22.008, €300k , 2022)

Main applicant for Applied Data Science Grant: "Text mining as imputation model for clinical prediction research: does it have to be perfect to be useful?" (Utrecht University, €5k, 2021)

Erasmus Mundus Scholarship (€ 23k, European Commission, 2013-2015)

TEACHING EXPERIENCE

Co-promotor:

Lotta Mijerink: Methods for validation, updating, and implementation of **AI** prediction models, Since 2023, Utrecht Univ.

Isa Spiero: **Text mining** for systematic reviews and prediction models, Since 2022, Utrecht Univ.

Master Theses / internships (daily supervisor or first examiner):

Luuk van Damme: correcting evaluation measures in data sampled via active learning, 2024, UU, MSc Bioinformatics

Famke Schulting: Review on medical applications of **large language models**, 2023, UU, MSc Bioinformatics (intern)

Hamudi Saidy: Review on handling of unreported **text-based variables** in medicine, 2023, UU, MSc Medicine (intern)

Marissa Mulder: Review on causality in normal-tissue complication probability models, 2023, UU, MSc Medicine (intern)

Matthew Scheeres: **Large language models** to extract symptoms from clinical notes, 2023, UU, MSc Artificial Intelligence

Christos Chatzispyros: Validation of models to predict radiotherapeutic complications, 2023, UU, MSc Epidemiology

Xinjie Chen: The influence of overfitting on calibration of probabilistic prediction models, 2023, UU, MSc Epidemiology

Zwierd Grotenhuis: **Text mining** clinical research outcomes: how accurate should it be?, 2022, UU, MSc Artificial Intelligence

Isa Spiero: Semi-supervised learning for prediction of radiotherapeutic complications, 2022, UU, Applied Data Science

Christina Heilmaier: Predicting laryngeal edema in head- and neck cancer patients, 2021, UMC Utrecht, MSc Epidemiology

Liesbeth Allein: Correction of Dutch pronouns with neural networks, 2018-2019, KU Leuven, MSc Artificial Intelligence

Lieven Kop: Semi-supervised deep learning for **NLP**, 2018-2019, KU Leuven, MSc Computer Science

Nathan Van Laere: **Universal language Model fine-tuning** for Dutch, 2018-2019, KU Leuven, MSc Computer Science

Mariya Hendriksen: **Chatbot** dialogue breakdown Detection, 2018, KU Leuven, MSc Artificial Intelligence

Enyan Dai: Error propagation in **textual** relation extraction, 2017-2018, KU Leuven, MSc Artificial Intelligence

Ellen Vissers: Sentiment analysis **in Tweets**, 2017-2018, KU Leuven, MSc Computer Science

Lena Martens: **Clinical** Temporal relation extraction from clinical texts, 2016-2017, KU Leuven, MSc Computer Science

Vignesh Baskaran: Movie recommendation, 2015-2016, KU Leuven, MSc Artificial Intelligence

Seppe Dijkmans: Native **language identification** of English text, 2015-2016, KU Leuven, MSc Applied Informatics

Lecturer:

[MSc] Big Data: NLP and Deep Learning, 2024, Utrecht Summer School - Utrecht University.

4x [MSc] Evidence Based Medicine, 2021-2024, UMC Utrecht

4x [BSc] Clinical Scientific Research, 2021-2024 UMC Utrecht

1x [MSc] Natural Language Processing, 2018, KU Leuven

Teaching Assistantships:

3x [BSc] Clinical Scientific Research, 2019-2021, UMC Utrecht

2x [MSc] Natural Language Processing, 2017, 2018, KU Leuven

2x [BSc] Introduction to Programming, 2015, 2016, KU Leuven

1x [BSc] Natural Language Processing: Parsing-as-Deduction, 2013, Utrecht University

1x [BSc] Introduction to Logics, 2012, Utrecht University

Assessor for 8 x MSc Artificial Intelligence theses, and 7 x MSc Computer Science theses at KU Leuven (2016-2019)

PROJECTS CONTRIBUTED TO

HTx (H2020-EU.3.1.6. 825162) "Next Generation Health Technology Assessment"

ACCUMULATE (IWT-SBO-Nr. 150056) "Acquiring Crucial Medical information Using Language Technology"

MARS (KU Leuven) "Machine Reading of Patient Records"

CALCULUS (ERC Advanced Grant 788506) "Commonsense and Anticipation enriched Learning of Continuous representations sUpporting Language UnderStanding"

DAME (KU Leuven) "A platform for Deep learning Approaches for Multilingual tExt processing"

ASSESSMENT RESPONSIBILITIES

Grant reviewer

NGF AiNed XS Europa (NWO), Dutch Arthritis Society (Reuma Nederland), Netherlands Organisation for Health Research and Development (ZonMw Doelmatigheidsonderzoek), UKRI (UK Research and Innovation).

PhD examination committee

Elias Moons, Representation Learning for Automated Document Classification, KU Leuven, 2021.

Victor Milewski, Multimodal Structured Representation Learning for Language and Vision, KU Leuven, 2024

Programme committee memberships

Nature Communications, The European Heart Journal, Diagnostic & Prognostic Research, Frontiers in Epidemiology, BMJ Open, BMC Medical Informatics and Decision Making, ACL Rolling Review (ARR), Conference on Empirical Methods in Natural Language Processing (EMNLP-IJCNLP), Annual Meeting of the Association for Computational Linguistics (ACL), North American Chapter of the Association for Computational Linguistics Conference (NAACL), International Conference on Computational Linguistics (COLING) [**outstanding reviewer mention '18**], International Workshop on Semantic Evaluation (SemEval).

EXTRACURRICULAR ACTIVITIES

Part of case team for Health Innovations Netherlands (2022)

Completed the "Basic course on Regulations and Organisation for Clinical Investigators" (BROK®, 2020)

Volunteer at the Conference on Empirical Methods in Natural Language Processing (EMNLP), Brussels (2018)

Board Member, Utrecht's Study Association for Cognitive Artificial Intelligence, USCKI Incognito (2012-2013)

Board Member, Dutch Study Association for Artificial Intelligence, NSVKI (2012-2013)

Student Member of the Advisory Committee for the B.Sc. Artificial Intelligence at Utrecht University (2012-2013)

TECHNICAL SKILLS

Python, R, Torch, Keras, Java, Haskell, Prolog, git, HTML, CSS, MySQL, Javascript, PHP, LaTeX, Linux

Systemgroup Representative at LIIR; responsible for buying (deep learning) **hardware** for the lab (2015-2019)

LANGUAGES

Dutch (native), English (fluent), German (limited working proficiency), French (elementary proficiency)

Last updated: August 12, 2024