

Saad Mahamood

Nationality: British 

Homepage: <https://www.saad.me.uk>

PROFILE

An experienced computational linguist with extensive academic and commercial experience in building, researching, and evaluating AI-based Natural Language Generation systems that convert large datasets into textual summaries. I am also capable of leading a team and I have experiences in mentoring junior colleagues or working cross-functionally as part of a team. I am looking to broaden my experiences by tackling new challenges.

EDUCATION

- 2006-2010 **University of Aberdeen, Scotland**  – PhD in Computing Science
- 2005-2006 **University of Sussex, England**  – MSc in Informatics
- 2001-2005 **Aberystwyth University, Wales**  – BSc in Computing Science

WORK EXPERIENCE

NLG Expert & Lead Data Scientist, trivago N.V., Düsseldorf, Germany  **March 2018 - Present**

Initially employed by trivago as an expert to help build NLG data-to-text solutions. I lead the creation of a project to generate accommodation descriptions automatically with a high degree of content and linguistic variation for the purposes of search engine optimisation, updatability, coverage, and to inform users. I have also worked on NLP problems such as helping to build a solution that could generate a unique selling point text from user reviews.

I have transitioned to leading a team of five data scientists and analysts. I work on leading the data science the content based solutions within trivago and working on projects to solve problems such that encompasses trivago's core content types: accommodation descriptions, reviews, images, geo-spatial information, amenities data, item matching, and pricing information. In addition, I help lead trivago's research interests by participating in research projects, writing academic research papers, applying for research projects, and supervising master students on their research theses.

I have also worked to help review research papers for academic conferences such as ACL, EMNLP, INLG, and others. In my spare time, I help to maintain the open source [SimpleNLG](#) surface realiser project and maintain an active interest in keeping up with the latest research developments in NLG and NLP.

Senior Natural Language Generation Engineer, Arria NLG, Aberdeen, Scotland  **August 2012 - February 2018**

Employed by Arria NLG to research and develop the creation of new generic tools that can enable the quicker/easier production of NLG systems for commercial clients. In the past most NLG systems contained many bespoke software components specific to a particular big data client, thus making reuse inherently difficult. Recently, I have been involved in helping to build a new NLG microplanner and multilingual (English and German) realiser that allows for the generation of texts in more than one language from a given syntactical sentence representation. This has involved dealing with aspects such as aggregation, referring expression generation, syntax, morphology, morphophonology, and orthography.

I have direct experience working and engaging with commercial clients to deliver NLG based solutions. I have also taken the lead to author an academic research paper within a commercial environment and helped to review a patent application based directly on the intellectual property that I've worked on. Additionally, in my spare time, also maintained academic interests by reviewing and publishing research papers with both internal and external collaborators.

Research Fellow, Computing Science, University of Aberdeen, Scotland  **July 2010 - June 2012**

Appointed as a research fellow to work on a follow-up project on research work done during my PhD. I was responsible for the development, deployment, refinement, and evaluation of an NLG system to convert neonatal clinical data into reports for parents of pre-term neonates. I worked closely with clinical domain experts and neonatal clinical data to help refine the quality of texts generated and bring about the

integration of our NLG system with the existing electronic health record software. I also project managed a research assistant, given the responsibility to apply for NHS ethics approval for onward evaluation, and also performed tutorial responsibilities for student demonstration classes.

PHD DISSERTATION

Thesis Title: “Generating affective Natural Language for Parents of Neonatal Infants.”

My PhD focused on the development of a Natural Language Generation (NLG) based system to convert clinical neonatal data into automatic daily summaries for parents of babies in Neonatal Intensive Care. This was done by summarising the information in a baby’s electronic patient record to produce reports focusing on matters of most interest to the parent. My thesis also focused on the informational and emotional needs of parents and tried to understand how these needs could be met by generating tailored automatic reports. During my PhD I constructed a prototype NLG system to generate such reports for parents and evaluated this system with parents that previously had a baby in neonatal care.

SKILLS

Computer Programming Languages: Java & Python.

Cloud Technologies: AWS & GCP.

Languages: English 🇬🇧 (native), German 🇩🇪 (B1.2)

SELECTED PUBLICATIONS

- Lining Zhang, João Sedoc, Simon Mille, Yufang Hou, Sebastian Gehrmann, Daniel Deutsch, Elizabeth Clark, Yixin Liu, Miruna Clinciu, **Saad Mahamood**, Khyathi Chandu (2023). Needle in a Haystack: An Analysis of Finding Qualified Workers on MTurk for Summarization. ACL 2023. Toronto, Canada 🇨🇦.
- Kaustubh Dhole, Varun Gangal, Sebastian Gehrmann, Aadesh Gupta, Zhenhao Li, **Saad Mahamood**, et al. (2023). NL-Augmenter: A Framework for Task-Sensitive Natural Language Augmentation. NEJLT. Vol. 9 No. 1.
- Emiel van Miltenburg, Miruna Clinciu, Ondřej Dušek, Dimitra Gkatzia, Stephanie Inglis, Leo Leppänen, **Saad Mahamood**, Stephanie Schoch, Craig Thomson, Luou Wen (2023). Barriers and enabling factors for error analysis in NLG research. NEJLT. Vol. 9 No. 1.
- Simon Millie, Kaustubh D. Dhole, **Saad Mahamood**, Laura Perez-Beltrachini, Varun Gangal, Mihir Kale, Emiel van Miltenburg (2021). Automatic Construction of Evaluation Suites for Natural Language Generation Datasets. NeurIPS 2021.
- Emiel van Miltenburg, Miruna Clinciu, Ondrej Dusek, Dimitra Gkatzia, Stephanie Inglis, Leo Lappänen, **Saad Mahamood**, Emma Manning, Stephanie Schoch, Craig Thompson, and Lulu Wen (2021). Underreporting of errors in NLG output, and what to do about it. INLG 2021. Aberdeen, Scotland 🇬🇧.
- Miruna Clinciu, Dimitra Gkatzia, and **Saad Mahamood** (2021). It’s common sense, isn’t it? Demystifying Human Evaluations in Commonsense-enhanced NLG systems. Workshop on Human Evaluation of NLP systems (HumEval) at EAACL, 2021. Kyiv, Ukraine 🇺🇦.
- David M. Howcroft, Anya Belz, Miruna-Adriana Clinciu, Dimitra Gkatzia, Sadid A. Hasan, **Saad Mahamood**, Simon Mille, Emiel van Miltenburg, Sashank Santhanam, Verena Rieser (2020). Twenty Years of Confusion in Human Evaluation: NLG Needs Evaluation Sheets and Standardised Definitions. INLG 2020. Dublin, Ireland 🇮🇪.
- **Saad Mahamood** and Maciej Zembrzusi (2019). Hotel Scribe: Generating High Variation Hotel Descriptions. INLG 2019. Tokyo, Japan 🇯🇵.
- Dimitra Gkatzia and **Saad Mahamood** (2015). A Snapshot of NLG Evaluation Practices 2005-2014. ENLG 2015 Brighton, United Kingdom 🇬🇧.
- **Saad Mahamood**, William Bradshaw, and Ehud Reiter (2014). Generating Annotated Graphs using the NLG Pipeline Architecture. INLG 2014. Philadelphia, PA, USA 🇺🇸.
- **Saad Mahamood** and Ehud Reiter (2012). Working with Clinicians to Improve a Patient-Information NLG system. INLG 2012. Starved Rock, IL, USA 🇺🇸.
- **Saad Mahamood** and Ehud Reiter (2011). Generating Affective Natural Language for Parents of Neonatal Infants. ENLG 2011. Nancy, France 🇫🇷.
- Albert Gatt, François Portet, Ehud Reiter, Jim Hunter, **Saad Mahamood**, Wendy Moncur, Somayajulu Sripada (2009). From Data to Text in the Neonatal Intensive Care Unit: Using NLG Technology for Decision Support and Information Management. AI Communications 22:153-186.
- **Saad Mahamood**, Ehud Reiter, Chris Mellish (2008). Neonatal Intensive Care Information for Parents — An Affective Approach. Computer-Based Medical Systems (CBMS) 2008. Jyväskylä, Finland 🇫🇮.
- **Saad Mahamood**, Ehud Reiter, Chris Mellish (2007). A Comparison of Hedged and Non-hedged NLG Texts. ENLG 2007. Schloss Dagstuhl, Germany 🇩🇪.

PATENTS

- Method and apparatus for document planning — *Inventor:* Saad Mahamood (2014), *Patent Number:* [US20160232152A1](#).