



+31 6xx xx xxxx
 xxxxx@xxxx.xx
<https://graus.nu>

Date of birth: x xxxx xxxx
 Nationality: Dutch

DAVID GRAUS

LEAD DATA SCIENTIST

I am **strategic**, I quickly spot relevant patterns and issues, and separate signal from noise. I like **getting things done**: I am hands-on, results-oriented, and love taking ideas from the academic world into working products. I am a **communicator**, a confident speaker and conversationalist. With a background in humanities and computer science, I like bridging the gap between these two worlds. I bring **creative energy**: I think outside the box and come up with alternative ways to proceed.

EXPERIENCE

- Present**
 March 2020

LEAD DATA SCIENTIST & CHAPTER LEAD | *Randstad Groep Nederland, Diemen, NL*
 I lead the data science chapter at Randstad Groep Nederland, the global leader in the HR services industry. Our chapter spans over a dozen data scientists across multiple teams and departments. We work on recommender systems, natural language processing and information extraction, and analytics and forecasting. As chapter lead I act as line manager for all chapter members, and handle employee development, coaching, recruiting and hiring, and other HR responsibilities.
- December 2019
 April 2018

LEAD DATA SCIENTIST | *FD Mediagroep, Amsterdam, NL*
 I led a team of data scientists on the Google DNI-funded BNR SMART Radio and FD SMART Journalism projects, which revolved around personalization. I did hands-on work, coached the team, planned, and set up the roadmap and research direction. We delivered several products, such as a content-based news recommender and automated tagger for the journalist's CMS. Our projects resulted in awards, talks at international conferences and two peer-reviewed academic publications.
- September 2018
 October 2016

DATA SCIENTIST | *Company.info (part of FD Mediagroep), Amsterdam, NL*
 As Company.info's first data scientist, I worked on machine learning with financial and company information. Projects included productizing an end-to-end custom-KB entity linking system to replace human labeling, sentiment classification of news, and business sector classification.
- 2011
 2008

SCIENCE EDITOR | *NTR, Hilversum, NL*
 After acquiring my BA in media studies I worked for several stints as science editor for public broadcaster NTR, for radio and online science portal Wetenschap24.
- September 2015
 June 2015

RESEARCH INTERN | *Microsoft Research, Redmond, WA, USA*
 Published a paper on a large-scale analysis of Cortana user logs, which received a best paper award at UMAP 2016, and a filed patent. I was mentored by Paul N. Bennett, Ryen White, and Eric Horvitz.
- February 2014
 January 2014

VISITING RESEARCHER | *University of Maryland, College Park, MD, USA*
 I visited prof. dr. Doug Oard's E-Discovery lab, where I followed a course on E-Discovery, and worked with the "Avocado" email corpus, which resulted in a publication at SIGIR 2014.

EDUCATION

- June 2017
 May 2012

PHD INFORMATION RETRIEVAL | *Informatics Institute, University of Amsterdam, NL*
 I did my PhD at the Information and Language Processing Systems (ILPS) group, a world-leading research group in Information Retrieval. My research focused on semantic search and entity retrieval, applied to digital traces. In June 2017, I defended my PhD thesis "Entities of Interest – Discovery in Digital Traces." Promotor: Prof. dr. Maarten de Rijke. Co-promotor: dr. Manos Tsagkias
- May 2012
 September 2009

MSC MEDIA TECHNOLOGY | *LIACS, Universiteit Leiden, Leiden, NL*
 Thesis: "Automatic Annotation of Cyttron Entries using the NCIthesaurus"
- September 2008
 September 2004

BA MEDIA STUDIES | *Faculty of Humanities, University of Amsterdam, NL*
 Minor: American Studies



PROFESSIONAL ACTIVITIES

- Present** ● **CO-ORGANIZER** | *Workshop RecSys in HR 2021, Amsterdam, NL*
 March 2021 Part of the organizing committee of our full-day “Workshop on Recommender Systems for Human Resources,” co-located with the 15th ACM Conference on Recommender Systems in Amsterdam.
- Present** ● **COMMITTEE MEMBER** | *DDMA AI Committee, Amsterdam, NL*
 October 2020 The Artificial Intelligence committee of the Data Driven Marketing Association (DDMA) aims to promote the use of responsible AI for marketing and other forms of interaction with customers.
- Present** ● **BOARD MEMBER** | *Stichting SETUP, Utrecht, NL*
 June 2020 SETUP is a media lab whose mission it is to create a wide group of media, algorithm and data-literate people, with a nuanced view on the significance and influence of technology on our lives.
- November 2021 ● **LOCAL OUTREACH CHAIR** | *ACM RecSys 2021, Amsterdam, NL*
 September 2020 I joined the organizing committee of the 15th ACM Conference on Recommender Systems, as local outreach chair. In this role I assist the other chairs with local matters, and link the (local) industry and companies to the conference.
- December 2019 ● **COMMITTEE MEMBER** | *“Beroepenveldcommissie” at Hogeschool Utrecht, Utrecht, NL*
 September 2018 As part of the HU University of Applied Sciences Utrecht School of Journalism’s committee that represents the (media) industry, I provided my perspective on the school’s education and curriculum.
- November 2019 ● **CO-ORGANIZER** | *DIR 2019, Amsterdam, NL*
 January 2019 I co-organized DIR 2019, the Dutch Information Retrieval Workshop. DIR is a yearly conference that showcases the best Information Retrieval work in The Netherlands and Belgium. Responsibilities included everything from venue, attracting sponsors, website, program, and local organization.
- 2015 ● **COMMUNICATIONS OFFICER** | *ILPS, University of Amsterdam, Amsterdam, NL*
 2014 To increase exposure of our research group I wrote articles of our group’s publications in collaboration with the university’s communication department. Some of our articles gained (inter)national press coverage on tech sites and in newspapers.
- 2015 ● **PHD COUNCIL CHAIR** | *Informatics Institute, University of Amsterdam, Amsterdam, NL*
 2013 The Informatics Institute PhD Council promotes interaction between members of the institute’s PhD community and communication between the PhD community and staff of the institute. I founded the council, held regular meetings with the institute management, and organized social and informational events for PhD candidates.
- 2015 ● **PUBLICITY CHAIR** | *ECIR 2014, Amsterdam, NL*
 2014 Responsible for all communication of The 36th European Conference on Information Retrieval, i.e., social media, website, calls for papers and sponsors. Supervised the design of the conference materials, e.g., logo, flyers, posters, and program booklet.
- PROGRAM COMMITTEE MEMBER**
- Various conferences ● UMAP (2021), TheWebConf (2018), DIR (2015, 2016, 2018), CIKM (2015, 2017)
- SIGIR ● On SIGIR’s PC from 2015–2020.
- Workshops ● FEAST@ECML-PKDD 2021, PodRecs@RecSys 2020, NewsIR@ECIR 2016, ESAIR@CIKM 2015



TEACHING

- February 10, 2021 ● **INVITED LECTURES**
- February 10, 2021 ● **GUEST LECTURER** | *BSc. Business Information Technology, Universiteit Twente, Online*
 Lectured on “recommender systems, algorithmic bias, and bias mitigation in hiring and recruitment.”
- November 12, 2020 ● **LECTURER** | *“Leergang Artificial Intelligence voor juristen,” Universiteit Leiden, Leiden, NL*
 At the post-academic “AI for lawyers” course at Universiteit Leiden I gave a lecture on using AI in practice for bringing news to news readers and jobs to job seekers.

September 11, 2019

LECTURER | *The ACM Summer School on Recommender Systems, Gothenburg, Sweden*

I gave a lecture and hands-on workshop at the RecSys Summerschool with Daan Odijk on “*Recommender Systems in the Media Industry – Relevance, Recency, Popularity, and Diversity.*” Our hands-on session included iPython notebooks with a custom Wikinews dataset, and an assignment of implementing re-ranking strategies for related article finding.

2019 **GUEST LECTURER** | *BSc. Information Studies, University of Amsterdam, Amsterdam, NL*

Guest lecturer at the “Collective Intelligentie” bachelor’s course on our work at FD Mediagroep around News Personalization and Ethics.

STUDENT SUPERVISION

2021 **SARAH-JANE VAN ELS** | MSc. Information Sciences, Vrije Universiteit Amsterdam. Thesis internship at Randstad Groep Nederland. Sarah-Jane worked on her thesis “*Using synthetic data to identify bias in recommender systems for recruitment*” under supervision of prof. Emma Beauxis-Aussalet (VU), and my co-supervision.

2021 **MAURITS DE GROOT** | MSc. ICT in Business, Universiteit Leiden. Thesis internship at Randstad Groep Nederland. Maurits worked on skills knowledge graphs and wrote his thesis “*Evaluating the Skills Gap in the Labor Market.*” under my supervision. Maurits graduated with a 8/10, and we published his thesis at the RecSys in HR Workshop, ACM RecSys 2021.

2021 **MAURICE VERBRUGGE** | MSc. Data Science, Radboud Universiteit. Research internship at Randstad Groep Nederland. Maurice worked on Reinforcement Learning for algorithmic planning of personnel.

2019 **OBERON BERLAGE** | MSc. Data Science, University of Amsterdam. Thesis internship at FD Mediagroep. Oberon successfully defended his thesis “*Improving automated segmentation of radio shows with audio embeddings,*” and graduated with a 9/10. His thesis was accepted as a research paper at the IEEE ICASSP 2020 conference.

2019 **KLAUS-MICHAEL LUX** | MSc. AI & MSc. Data Science, Radboud Universiteit Nijmegen. Internship at FD Mediagroep. Klaus worked on implementing a deep learning-based audio segmentation model and co-authored our ICASSP 2020 publication.

UNIVERSITY COURSES

2015 **WEB SEARCH** | *MSc. Information Studies, University of Amsterdam*

Teaching assistant responsible for lab sessions, project supervision, and grading.

2014 **COMPLEX CRIME SCENES** | *MSc. Forensic Science, University of Amsterdam*

Co-lecturer, gave lectures on text-mining, social network analysis, and search engines. Co-responsible for lab sessions, assignments, and grading.

2013 **SOCIAL NETWORK ANALYSIS** | *BSc. Information Studies, University of Amsterdam*

Teaching assistant responsible for lab sessions and grading.



SELECTED PUBLICATIONS

GENERAL PUBLIC

2021 **Forget the Trolley Problem; Pragmatic and Fair AI in the Real World.** *TowardsDataScience.*

Opinion piece on pragmatic fair AI. Published in TowardsDataScience and selected as “editors’ pick.”

2016 **D. Graus, M. de Rijke. Wij zijn racisten, daarom Google ook.** *NRC Handelsblad, NRC.next.*

Op-ed in daily Dutch newspaper of record, on how algorithms aren’t neutral by design.

PEER-REVIEWED SCHOLARLY ARTICLES

2021 D. Lavi, V. Medentsiy, **D. Graus. conSultantBERT: Fine-tuned Siamese Sentence-BERT for Matching Jobs and Job Seekers.** *RecSys in HR 2021 Workshop, Amsterdam, The Netherlands*

2020 F. Lu, A. Dumitrache, **D. Graus. Beyond Optimizing for Clicks: Incorporating Editorial Values in News Recommendation.** *Conference on User Modeling, Adaptation, and Personalization (UMAP), Online*

2020 O. Berlage, K.-M. Lux, **D. Graus. Improving automated segmentation of radio shows with audio embeddings.** *International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Barcelona, Spain*

- 2018 ● **D. Graus, D. Odijk, M. de Rijke.** **The Birth of Collective Memories: Analyzing Emerging Entities in Text Streams.** *Journal of the Association for Information Science and Technology (JASIST)*
- 2016 ● **D. Graus, P. N. Bennett, R. White, E. Horvitz.** **Analyzing and Predicting Task Reminders.** *Conference on User Modeling, Adaptation, and Personalization (UMAP)*, Halifax, Canada
- 2016 ● **D. Graus, E. Tsagkias, W. Weerkamp, E. Meij, M. de Rijke.** **Dynamic Collective Entity Representations for Entity Ranking.** *International Conference on Web Search and Data Mining (WSDM)*, San Francisco, US
- IN THE PRESS**
- June 4, 2019 ● **EUROPEAN SCIENCE-MEDIA HUB,** *A scientist's opinion: Interview with David Graus about AI-tools for journalists.* Interview in the context of a three-day Summer School AI and journalism on how we apply AI in Journalism at FD Mediagroep.
- May 10, 2019 ● **DE NATIONALE AI CURSUS,** *Geef niet het algoritme de schuld.* Interview in response to my op-ed on the alleged promoting of anti-vaxx books by Bol.com's algorithms in Het Parool.
- November 26, 2018 ● **BNR,** *The case of: AI in Media.* BNR Podcast where I talk about our goals in the SMART Journalism project in the context of "applied AI" in the media industry.
- October 12, 2018 ● **DE VOLKSKRANT,** *Met deze eigen zoekmachine spit de politie schatten aan digitaal bewijs door.* Interview about the search technology developed by the Netherlands Forensic Institute, and its use in the high-profile "Mocromaffia" court cases, with my view on the role of such technology in the collection of digital evidence.
- October 2, 2018 ● **NRC,** *Waar blijft de Nederlandse robotjournalist?* Article on automated journalism, where I mention our SMART Journalism project.
- July 10, 2018 ● **COMMUNICATIONS OF THE ACM,** *Dutch Police Fight Crime by Cracking PGP Phones.* Article about my work as an expert in digital forensics and search engines for Inez Weski on the "Ennetcom" case.
- November 12, 2017 ● **DENKTANK,** *Baas over brein.* Television show for youngsters on futuristic topics. In this episode: how companies influence us online through persuasion/behavior analysis.
- June 17, 2017 ● **NOS,** *David onderzocht de sporen die wij onbewust online achterlaten.* Article on my research and PhD thesis: *Entities of Interest – Discovery in Digital Traces.*

SELECTION OF INVITED TALKS

I enjoy speaking about my work and field of expertise. I also feel a responsibility as a computer scientist to participate in the public debate and share my expertise. Below a selection of talks aimed at the general public, for an exhaustive overview see <https://graus.nu/talks>

- September 17, 2021 ● **KEYNOTE ON "DATA SCIENCE IN HR: RECOMMENDER SYSTEMS, KNOWLEDGE GRAPHS, VACANCIES AND RESUMES" | FEAST Workshop, Online**
I was invited as the keynote speaker at the International Workshop on Fair, Effective And Sustainable Talent management using data science (FEAST) at ECML-PKDD. 2021.
- September 1, 2021 ● **WEBINAR "CONVERSATIONS ON GOVERNING AI WITH RANDSTAD & DATAIKU" | Dataiku Webinar, online**
Together with Helen Hulsker (Senior International Legal Counsel Randstad) I joined of a webinar on governing AI by Dataiku where we discussed implementing AI in a high-risk domain from a technical, practical, and jurisdictional perspective.
- February 11, 2021 ● **"RESPONSIBLE DATA SCIENCE IN PRACTICE" PANEL | ICT.Open 2021, Online**
I was invited as a panelist on "Challenges of responsible use of data and algorithms in practice," the session was part of the Responsible Data Science track in the Digital Society research program of de Vereniging van Universiteiten (VSNU).
- November 20, 2020 ● **AI FOR BURSTING FILTER BUBBLES AND ADDRESSING BIAS IN HIRING | BNAIC/BENELEARN 2020, Leiden/Online**
I was invited to give a talk at the Industry & Society day of The annual Benelux Conference on AI for bursting filter bubbles and addressing bias in hiring.
- October 31, 2020 ● **ALGORITHMIC MATCHING, BIAS, AND BIAS MITIGATION IN HR | Anti-Discrimination Hackathon, Online**
I was invited to give a talk on algorithmic matching, algorithmic bias, and bias mitigation in the domain of automated recruitment.

- January 23, 2020 ● **ALGORITHMS AND AI-DRIVEN TECHNOLOGIES IN THE INFORMATION SOCIETY** | *CPDP2020 Conference on Data Protection and Artificial Intelligence, Brussels, Belgium*
 I was invited to join a panel discussion with privacy and legal scholars on the role of AI-driven innovation such as news recommender systems, chatbots, deepfakes in society.
- June 20, 2019 ● **EXPLAINABLE AI: SHOULD COMPANIES JUSTIFY ALGORITHMIC DECISIONS?** | *Academisch-Cultureel Centrum SPUI25, Amsterdam, NL*
 Talk on our "Reading News with a Purpose" work on explainable recommender systems, and participation in a panel session with prof. Marcel Worring (UvA) and Hinda Haned (Ahold).
- March 28, 2018 ● **ZOEKEN, VINDEN, EN AANBEVELEN: PERSONALISATIE VS. PRIVACY** | *vogin-ip-lezing 2018: Zoeken & Vinden, Amsterdam, NL*
 Lecture on recommender systems and (the need for) personalization.
- June 18, 2017 ● **DE MACHT VAN DATA—HOE ALGORITMEN ONS LEVEN VORMGEVEN** | *De Balie, Amsterdam, NL*
 I gave the opening address on AI and machine learning, and participated in a panel on the impact of algorithms on our daily lives, at an evening at De Balie on the topic of "the power of data."

PATENTS

- July 21, 2020 ● **US10719757B2** | *System for interpreting and managing imprecise temporal expressions*
 Assigned to Microsoft Technology Licensing LLC. Authors: Adam Fourney, Paul N. Bennett, Ryan White, Eric Horvitz, Xin Rong, David Graus. This patent is based on Rong's internship work, which in turn was a follow-up from my work at Microsoft Research.

AWARDS

- 2019 ● **DUTCH INTERACTIVE AWARDS** For BNR SMART Radio we won two Dutch Interactive Awards: Gold in the category "Content," and silver in category "Disruptor."
- 2019 ● **ANNUAL MASTERS OF MEDIA (AMMA) AWARDS** Our news personalization efforts at FD Mediagroep won the AMMA 2019 Award in the category "Best Media Innovation."
- 2019 ● **MARCONI ONLINE AWARD** Our BNR SMART Radio project won the first Marconi Online Award for innovative radio at "het RadioGala van het Jaar."
- 2016 ● **BEST STUDENT PAPER AWARD** James Chen Best Student Paper Award at UMAP 2016 for "Analyzing and Predicting Task Reminders," the publication from my internship at Microsoft Research.
- 2014 ● **ILPS MVP** Most Valuable Player Award given (by peer votes) to an ILPS member that has made significant contributions to the group.

SOFTWARE SKILLS

- Coding & scripting ● Python, bash, Clojure, C#, \LaTeX , JavaScript, HTML, CSS
- Tools/libraries ● scikit-learn, xgboost, numpy, pandas, spaCy, NLTK, D3, git
- AWS ● Lambda, ECS, ECR, SSM, Redshift, ElastiCache, SageMaker

OPEN SOURCE CONTRIBUTIONS

- 2018 ● **TEXTPIPE** | <https://github.com/textpipe/textpipe>
 Python package for converting raw text into clean, readable text, and metadata extraction. At FD Mediagroep we use textpipe in production for our content-based recommender system.
- 2016 ● **SEMANTICIZER** | <https://github.com/semanticize/semanticizer>
 Machine learning-based entity linking (to Wikipedia) framework in Python. I contributed a novel global linking algorithm to the Semanticizer codebase.

LANGUAGES

- Dutch ● Native language
- English ● Fluent in speaking, reading and writing, both technical and conversational
- French ● Conversationally fluent and reading and writing with high proficiency
- German ● Basic level of speaking, reading and writing

HOBBIES

● I enjoy visiting **museums** and **art galleries** (contemporary art and photography), **reading** (science fiction), **craft beer**, and **photography**. I like to dabble in **graphic design**, to **travel** the world, and **running** and **yoga**.