



+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** ● **PRINCIPAL INVESTIGATOR** | *FINDHR project*
November 2022
I am Randstad's PI in the "Fairness and Intersectional Non-Discrimination in Human Recommendation" (FINDHR) project, a Horizon Europe-funded research project with a budget of over EUR 3.3M, which runs from 2022 to 2025. I participated in writing the proposal, and lead a work package on exploitation and demonstration.
- Present** ● **COMMITTEE MEMBER** | *Veld Adviesraad Master Program Applied AI*
September 2022
I am a committee member at the advisory board of the Applied AI Master program of the University of Applied Sciences Amsterdam (Hogeschool van Amsterdam) to provide input to the curriculum, and represent the industry's perspective.
- Present** ● **CO-ORGANIZER** | *RecSys in HR Workshop*
March 2021
Part of the organizing committee of the full-day "Workshop on Recommender Systems for Human Resources," co-located in 2021 with the 15th ACM Conference on Recommender Systems in Amsterdam, in Seattle in 2022, and Singapore in 2023.
- 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.
- 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.
- PROGRAM COMMITTEE MEMBER**
- Various conferences ● UMAP (2021-2023), SIGIR (2015-2022), CIKM (2015, 2017)
- Workshops ● FEAST@ECML-PKDD 2021, PodRecs@RecSys 2020, NewsIR@ECIR 2016, ESAIR@CIKM 2015



TEACHING

- INVITED LECTURES**
- June 16, 2023 ● **LECTURER** | *The ACM Summer School on Recommender Systems, Copenhagen, Denmark*
I gave a lecture at the RecSys Summerschool with Mesut Kaya on "Job recommendations."
- September 23, 2022 ● **LECTURER** | *UvA in DeLaMar Business Seminar Series, DeLaMar Theatre, Amsterdam*
I was invited to give a seminar to around 800 master students of the Amsterdam Business School (ABS) of the University of Amsterdam on the future of the labor market.

- 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

- 2023 ● **HILKE SCHUURMANS** | MSc. International Technology Law, Vrije Universiteit Amsterdam. Thesis internship at Randstad Groep Nederland. Hilke wrote her thesis *“Bridging liability gaps for discriminatory AI systems—A study on the effectiveness of the reforms of the liability framework by the European Union in addressing algorithmic discrimination”*. Hilke’s thesis was awarded an 8.5/10.
- 2022 ● **ADAM ARAFAN** | MSc. Data Science, Universiteit van Amsterdam. Thesis internship at Randstad Groep Nederland. Adam wrote his thesis *“End-to-end bias mitigation in candidate recommender systems with fairness gates”* under supervision of myself, Fernando Santos (Civic AI Lab) and Emma Beaxis-Aussalet (Civic AI Lab). Adam’s thesis was awarded a 9/10, and he published his thesis at the RecSys in HR ’22 workshop at ACM RecSys 2022.
- 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 Beaxis-Aussalet (Civic AI Lab), and myself. Her thesis was published at the CompJobs workshop at WSDM 2022.
- 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 published his thesis at the RecSys in HR Workshop, at ACM RecSys 2021.
- 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.

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

- 2022 ● A.M. Arafan, **D. Graus**, F.P. Santos, E. Beauxis-Aussalet. **End-to-End Bias Mitigation in Candidate Recommender Systems with Fairness Gates**. *RecSys in HR'22, Seattle, WA, USA*
- 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*

IN THE PRESS

- June 9, 2023 ● **THE NETHERLANDS INSTITUTE FOR HUMAN RIGHTS** Interview: 'Collega's zullen nooit volledig afgaan op wat onze algoritmes zeggen'. Interview on ethical AI in the context of HR and recruitment.
- November 16, 2022 ● **RECSPEERS** *Recommender systems in Human Resources with David Graus*. Podcast with Marcel Kurovski where we touch on many topics, from vacancy over talent recommendation to mitigating biases effectively.
- February 22, 2022 ● **REFLEX** (Quarterly magazine of ABH), *Dwarsdenkers: 'AI is onbevooroordeeld en leidt tot meer inclusieve arbeidsmarkt.'* Interview on the promises of AI in the labor market.
- 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.
- October 12, 2018 ● **DE VOLKSKRANT**, *Met deze eigen zoekmachine spit de politie schatten aan digitaal bewijs door*. Interview about the search engine at the Netherlands Forensic Institute, and its use in the high-profile "Mocromaffia" court cases, with my perspective on the role of such technology in collecting 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

- June 22, 2023 ● **PANELIST** | *CPDP 2023, Brussels, Belgium*
Panelist on "Beyond Ethics Washing: Impact Assessments, Audits, and Oversight for AI" moderated by Mirko Tobias Schäfer (Utrecht University) with Iris Muis (Utrecht Data School), Paul Nemitz (European Commission), and Maria Koomen (Open Governance Network for Europe).
- October 31, 2022 ● **GUEST SPEAKER** | *Felix Meritis, Amsterdam*
I was invited as a guest at the "AI for Good" series, organized at Felix Meritis to discuss the opportunities of AI, alongside Dasha Simons (IBM) and Oumaima Hajri (Hogeschool Rotterdam).
- October 14, 2022 ● **PANELIST** | *Reshaping Work 2022 Conference. De Rode Hoed, Amsterdam*
I was invited as an "AI and Ethics" panelist, moderated by Ting Li (Erasmus University Rotterdam), alongside Robert Seamans (New York University) and Anastasia Sergeeva (Vrije University Amsterdam).

- May 18, 2022 ● **PANELIST** | *"Hey Siri: Vind een geschikte kandidaat," DeFabrique, Utrecht*
 I was invited as a panelist to talk about AI in HR and recruitment for a more inclusive labor market, alongside Siri Beerends (SETUP), Rina Joosten-Rabou (Seedlink) and Hans de Zwart (HvA).
- September 17, 2021 ● **KEYNOTE SPEAKER** | *FEAST Workshop, Online*
 I was invited as the keynote speaker on "Data Science in HR: recommender systems, knowledge graphs, vacancies and resumes" at the International Workshop on Fair, Effective And Sustainable Talent management using data science (FEAST) at ECML-PKDD 2021.
- September 1, 2021 ● **PANELIST** | *Dataiku Webinar, online*
 Together with Helen Hulsker (Senior International Legal Counsel Randstad) I joined a webinar "Conversations on Governing AI with Randstad & Dataiku" where we discussed implementing AI in our high-risk domain from technical, practical, and legal perspectives.
- February 11, 2021 ● **PANELIST** | *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 ● **INDUSTRY SPEAKER** | *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 ● **KEYNOTE SPEAKER** | *Anti-Discrimination Hackathon, Online*
 I was invited to give an online keynote on "Algorithmic matching, bias, and bias mitigation" in automated recruitment at the Anti-Discrimination Hackathon co-organized by the Netherlands Ministry of Education, Culture and Science and Ministry of the Interior and Kingdom Relations of Netherlands.
- January 23, 2020 ● **PANELIST** | *CPDP 2020, Brussels, Belgium*
 I was invited to join a panel on "Algorithms and AI-driven technologies in the information society" with privacy and legal scholars on the role of AI-driven innovation such as news recommender systems, chatbots, deepfakes in society.
- June 20, 2019 ● **SPEAKER AND PANELIST** | *Academisch-Cultureel Centrum SPUI25, Amsterdam, NL*
 At the "Explainable AI: Should companies justify algorithmic decisions?" event I gave a talk on our "Reading News with a Purpose" work on explainable recommender systems, and participated in a panel with prof. Marcel Worring (UvA) and Hinda Haned (Ahold).
- March 28, 2018 ● **SPEAKER** | *VOGIN-IP-lezing 2018: Zoeken & Vinden, Amsterdam, NL*
 At the VOGIN-IP Lezing I gave a lecture on recommender systems and (the need for) personalization.
- June 18, 2017 ● **OPENING ADDRESS AND PANELIST** | *De Balie, Amsterdam, NL*
 I gave the opening address at "De Macht van Data—Hoe Algoritmen ons Leven Vormgeven" 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
- Databases ● MongoDB, elasticsearch, SQL, Redis, redshift
- Software ● Adobe Creative Suite, Microsoft Office

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**.