

## CURRICULUM VITAE

# JOHANNES B. GRUBER

GESIS – Leibniz Institute for the Social Sciences

✉ [Johannes.Gruber@gesis.org](mailto:Johannes.Gruber@gesis.org)

🦋 [@JohannesBGruber](https://twitter.com/JohannesBGruber) | [📺 JBGruber](https://www.youtube.com/channel/UCJBGGruber) | [🌐 JBGruber](https://www.linkedin.com/company/jbgruber)

🌐 [JohannesBGruber.eu](https://www.JohannesBGruber.eu)

---

## EMPLOYMENT

---

- Apr. 2025—**current**: **Team Leader** at GESIS Leibniz Institute for the Social Sciences, [Data Services for the Social Sciences](#). Team: Metadata Standards and Interoperability.
- Jan. 2025—**current**: **Data Editor** at [Political Communication](#).
- Apr. 2024—Mar. 2025: **Post-Doc Researcher** at the Department of Language, Literature and Communication at Vrije Universiteit Amsterdam. Project: [NEWSFLOWS. Modeling news flows: How feedback loops influence citizens beliefs and shape democracy](#).
- Apr. 2023—Mar. 2024: **Post-Doc Researcher** at the Department of Communication Science at the University of Amsterdam. Project: [NEWSFLOWS. Modeling news flows: How feedback loops influence citizens beliefs and shape democracy](#).
- Oct. 2022—Sep. 2023: **Post-Doc Researcher** at the Department of Communication Science at Vrije Universiteit Amsterdam. Project: [OPTED – Observatory for Political Texts in European Democracies](#).
- Jul. 2021—Sep. 2022: **Post-Doc Researcher** at the Chair of Political Theory and Digital Democracy, European New School of Digital Studies (ENS), European University Viadrina Foundation Frankfurt (Oder).
- Nov. 2015—Jun. 2021: Public Relations Officer at Heinrich-Böll-Stiftung Hessen e.V. (Niddastr. 64; 60329 Frankfurt/Main). Roles: Student Assistant, Research Assistant (Wissenschaftliche Hilfskraft; since 2020). Work included: managing website, creating and managing social media channels (Facebook, Twitter, Instagram, YouTube), planning and realisation of digital political education events.
- Sep. 2020—Nov. 2020: Technical director Symposium “Politische Dimensionen kultureller Bildung” (Political Dimensions of Cultural Education) at Jewish Museum Frankfurt. Work included: planning and implementation of video conferences between panellists as well as broadcasting the online symposium to participants.
- Dec. 2017—May. 2018: Research Assistant for Dr. Philip Leifeld at the University of Glasgow. Development and documentation of Discourse Network Analyzer (DNA) and its companion R package rDNA.
- Sep. 2016—Nov. 2016: Freelance Researcher for Prof. Dr. Daniela Grunow, Faculty of Social Sciences Goethe-University. Project: “THE APPARENT PROJECT. International and national studies of norms and gender division of work at the life course transition to parenthood” Role: creating a newspaper article dataset on parenthood depictions in *The Guardian*.

---

Updated: February 23, 2026

- Mar. 2015—Dec. 2016: Research Assistant for Dr. Iñaki Sagarzazu and Dr. Ana Inés Langer at the University of Glasgow.  
Project: “Personalisation of UK politics in the media”.  
Work included: downloading, parsing and cleaning newspaper articles about UK cabinet members in Westminster parliament from 1992 to 2015.
- Nov. 2015—Sep. 2016: Freelance Researcher for Dr. Patrick Sachweh at Goethe-University Frankfurt, Department of Sociology.  
Project: “Comparative media analysis (US/GER) The One Percent during and after the global economic and financial crisis”.  
Work included: designing and conducting a comparative study on German and US media coverage of extreme wealth.

---

## EDUCATION

---

- Sep. 2016—Jul. 2021 PhD in Politics (Funded by College of Social Sciences Scholarship).  
University of Glasgow, School of Social and Political Sciences.  
Thesis title: “Troublemakers in the streets? A framing analysis of newspaper coverage of protests in the UK 1992-2017”.  
Supervisors: Dr. Ana Inés Langer, Dr. Sergi Pardos-Prado & Dr. Brian Fogarty.
- Sep. 2014—Sep. 2015: MSc Political Communication, University of Glasgow.  
Dissertation: “Media coverage and public opinion towards protests. A case study of media frames in the coverage of the protests against Stuttgart 21, testing supervised machine learning for content analysis of frames”.  
Classification: With Distinction
- Oct. 2012—Jun. 2017: MA Political Theory, Goethe University of Frankfurt/Main & TU Darmstadt.  
Thesis: “The 1% in the news. Critique and justifications of income and wealth concentration in German and US news media during and after the global financial crisis”.  
Classification: Very Good (1,3)
- Oct. 2009—Sep. 2012: BA Political Science; Economics and Economic Studies in History, RWTH Aachen-University.  
Bachelor Thesis: “Social acceptance of (major) projects in the German public sphere”.  
Classification: Very Good (1,5)

---

## PUBLICATIONS

---

- Balluff, P., Ho, J. C.-t., **Gruber, Johannes B.**, Palicki, S., Palmer, A., Rossi, L., Shklovski, I., and Chan, C.-h. (2026). Newer, larger, better? A critique of the unreflective LLM adoption in Communication Research. *Political Communication*. Accepted for publication. <https://doi.org/10.1080/10584609.2026.2618486>.
- Rouquette, O. Y., Peters, Y., **Gruber, Johannes B.**, Stier, S., and Weller, K. (2026). Enabling Access to Online Platform Data: Research Data Infrastructures in the DSA Era. [https://osf.io/emrqa\\_v1](https://osf.io/emrqa_v1).

---

\*All authors contributed to this project equally. The order of the names was determined by a random draw.

- Baumann, J., Röttger, P., Urman, A., Wendsjö, A., Plaza-del Arco, F. M., **Gruber, Johannes B.**, and Hovy, D. (2025). Large Language Model Hacking: Quantifying the Hidden Risks of Using LLMs for Text Annotation. arXiv:2509.08825 [cs]. <http://arxiv.org/abs/2509.08825>.
- Gruber, Johannes B.** (2025a). paperboy — A Collection of News Media Scrapers in R. [https://osf.io/hu6qw\\_v1](https://osf.io/hu6qw_v1).
- Gruber, Johannes B.** (2025b). traktok — Making TikTok Data Accesible for Research. [https://osf.io/xrgc6\\_v1](https://osf.io/xrgc6_v1).
- Gruber, Johannes B.**, Ortega, A. L., and Van Der Velden, M. A. C. G. (2024). Weaponizing rights: Political Uses of LGBTQ+ Rhetoric in Media Outlets. <https://osf.io/dyuwa>.
- Gruber, Johannes B.** and Votta, F. (2025). Large language models. In Nai, A., Grömping, M., and Wirz, D., editors, *Elgar Encyclopedia of Political Communication*. Edward Elgar Publishing. Accepted version. <https://doi.org/10.31219/osf.io/s7qx2>.
- Chan\*, C.-h., Schatto-Eckrodt, T., and **Gruber, Johannes B.** (2024). What makes computational communication science (ir)reproducible? *Computational Communication Research*. <https://doi.org/10.5117/CCR2024.1.5.CHAN>.
- Gruber, Johannes B.** and van der Velden, M. A. C. G. (2024). A research note on rethinking software documentation: Creating inclusive computational tools for social sciences. <https://doi.org/10.31219/osf.io/a3kwf>.
- Gruber, Johannes B.** and Weber, M. (2024). rollama: An R package for using generative large language models through ollama. <https://doi.org/10.48550/arXiv.2404.07654>.
- Gruber, Johannes B.** and van Atteveldt, W. (2025). Sharing is caring (about research): Three avenues for sharing (copyrighted) text collections and the need for non-consumptive research. [https://doi.org/10.31235/osf.io/gz3xf\\_v1](https://doi.org/10.31235/osf.io/gz3xf_v1).
- Vaughan, M., **Gruber, Johannes B.**, and Langer, A. I. (2023). The tension between connective action and platformisation: Disconnected action in the GameStop short squeeze. *New Media & Society*. <https://doi.org/10.1177/14614448231182617>.
- Gruber, Johannes B.** (2023). Troublemakers in the Streets? A Framing Analysis of Newspaper Coverage of Protests in the UK 1992-2017. *The International Journal of Press/Politics*, 28(2):414–433. <https://doi.org/10.1177/19401612221102058>.
- Weigand, M., Weber, M., and **Gruber, Johannes B.** (2022). Conspiracy narratives in the protest movement against COVID-19 restrictions in Germany. a long-term content analysis of telegram chat groups. In *Proceedings of the Fifth Workshop on Natural Language Processing and Computational Social Science (NLP+CSS)*, pages 52–58, Abu Dhabi, UAE. Association for Computational Linguistics. <https://aclanthology.org/2022.nlpcss-1.8>.
- Gruber, Johannes B.** (2021). *Troublemakers in the Streets? A Framing Analysis of Newspaper Coverage of Protests in the UK 1992-2017*. <https://doi.org/10.5525/gla.thesis.82346>.
- Langer, A. I. and **Gruber, Johannes B.** (2021). Political Agenda Setting in the Hybrid Media System. Why Legacy Media Still Matter a Great Deal. *The International Journal of Press/Politics*, 26(2):313–340. <https://journals.sagepub.com/doi/10.1177/1940161220925023>.
- Leifeld, P., **Gruber, Johannes B.**, and Bossner, F. R. (2018). *Discourse Network Analyzer Manual*. <https://github.com/leifeld/dna/releases/download/v2.0-beta.24/dna-manual.pdf>.

- Gruber, Johannes B.** (2023). *askgpt. A chat package connecting to API endpoints by 'OpenAI' to answer questions (about R)*. R package version 0.1.3. <https://github.com/JBGruber/askgpt>. Cranlogs download count: 13,638.
- Gruber, Johannes B.**, Guinaudeau, B., and Votta, F. (2024). *atrrr: Wrapper for the AT Protocol Behind 'Bluesky'*. R package version 0.0.2, <https://github.com/JBGruber/atrrr>. <https://jbgruber.github.io/atrrr/>. Cranlogs download count: 9,751.
- Gruber, Johannes B.** (2022). *rwhatsapp. An R package for working with WhatsApp data*. R package version 0.2.4. <https://github.com/JBGruber/rwhatsapp>. Cranlogs download count: 40,932.
- Gruber, Johannes B.** and Weber, M. (2024). *rollama: Communicate with 'Ollama'*. R package version 0.0.2. <https://CRAN.R-project.org/package=rollama>. Cranlogs download count: 14,411.
- Gruber, Johannes B.** (2023a). *paperboy. A comprehensive collection of news media scrapers*. R package version 0.0.5.9000. <https://github.com/JBGruber/paperboy>.
- Gruber, Johannes B.** (2023b). *traktok. Getting TikTok data through the official and unofficial APIs*. R package version 0.1.1. <https://github.com/JBGruber/traktok>. Cranlogs download count: 536.
- Gruber, Johannes B.** (2023c). *cookiemonster: Your Friendly Solution to Managing Browser Cookies in R*. R package version 0.0.3. <https://github.com/JBGruber/cookiemonster>. Cranlogs download count: 6,744.
- Gruber, Johannes B.** (2024). *LexisNexisTools. An R package for working with newspaper data from 'LexisNexis'*. R package version 1.0.0. <https://github.com/JBGruber/LexisNexisTools>. Cranlogs download count: 51,379.
- Benoit, K., Matsuo, A., and **Gruber, Johannes B.** (2023a). *spacyr: Wrapper to the 'spaCy' 'NLP' Library*. R package version 1.3.0. <https://CRAN.R-project.org/package=spacyr>. Cranlogs download count: 405,246.
- Benoit, K., Watanabe, K., Wang, H., Perry, P. O., Lauderdale, B., **Gruber, Johannes B.**, and Lowe, W. (2023b). *quanteda.textmodels: Scaling Models and Classifiers for Textual Data*. R package version 0.9.6. <https://CRAN.R-project.org/package=quanteda.textmodels>. Cranlogs download count: 201,914.
- Leifeld, P., **Gruber, Johannes B.**, and Henrichsen, T. (2019). *rDNA. A Package to Control Discourse Network Analyzer from R*. University of Glasgow, School of Social and Political Sciences, Glasgow. R package version 2.1.18. <http://www.philipleifeld.com>.

---

## AWARDS AND GRANTS

---

- *AI4SOCIALPLUS: AI Readiness for Social Impact* Jan 2027 – Dec 2029 [HORIZON-INFRA-2025-01-EOSC-03](#) — Advancing AI-readiness and Machine-Actionability in the EOSC Ecosystem (Stream B: AI-driven tools for the machine-actionability of data repositories) **7,486,799.08 (GESIS: 327,500 €)**
- Top Paper Award, ICA Computational Methods Division, 74th Annual International Communication Association Conference (ICA 2024)
- Ninth Conference on New Directions in Analyzing Text as Data (TADA 2018) travel grant and registration fee waiver
- Text Learning Workshop at the LSE travel funding 2017
- University of Glasgow College of Social Sciences PGR experience fund 2017
- University of Glasgow College of Social Sciences PhD Scholarship 2016/17 (funding period: 2016 – 2020)
- Best Performance Prize, 2014 MSc/MRes in Political Communication

- Deans list Award, 2011 (best 3% of cohort Political Science)
- Deans list Award, 2010 (best 3% of cohort Political Science)

---

## CONFERENCES AND TALKS

---

- 2025 KODAS Workshop “Intrinsic/Extrinsic – Transferring and Developing Quality Indicators for Online Platform Data”, Cologne, 27-28/08/2025, presentation: *“Gaslight, Gather, Gatekeep? Comparing Access Strategies for Research on TikTok Data”*.
- 75th Annual International Communication Association Conference (ICA 2025), Denver, presentations: *“Sharing Is Caring (About Research): Three Avenues for Sharing (Protected) Text Collections and the Need for Non- Consumptive Research”*; *“[Tool Demo] Traktok: Making TikTok Data Accessible for Research”*; *“[Tool Demo] Paperboy: A Collection of News Media Scrapers”*.
- Annual Conference of the International Association for Social Science Information Service and Technology (IASSIST 2025), Bristol, presentation: *“Sharing is Caring (about Research): Addressing Challenges in Sharing Protected Text Data Collections through Non-Consumptive Research”*.
- Etmaal van de Communicatiewetenschap conference (Etmaal 2025), Bruges, presentation: *“Textual Feelings: A Review of Computational Tools for Measuring Opinions in Text”*.
- Invited Talk *“TikTok Research with traktok & Short-Form, Broad Reach”* at Institut für Publizistik - Johannes Gutenberg-Universität Mainz, 07/11/2024.
- Invited Talk *“We should use local gLLMs (and its surprisingly easy to do!)”* at the European Policy Data Science Network Meetup, 17/10/2024.
- 2024 ECREA 2024 10th European Communication Conference, Ljubljana, presentation: *“How media storms are framed: Relations between legacy, social and web media in the hybrid media system”*, roundtable: *“Advancing the study of opinion expressions in digital media: The OPINION Network approach ”*.
- The 2nd Post-API Conference, *“Social media data acquisition in a Post-API world”*, Queensland University of Technology, Brisbane, presentations: *“Sharing is Caring (about Research): Four Avenues for Sharing (Copyrighted) Text Collections and the Need for Non-Consumptive Research”*.
- 74th Annual International Communication Association Conference (ICA 2024), Gold Coast, presentations: *“What Makes Computational Communication Science (Ir)reproducible?”*; *“Whose Voices Are Heard? Making Sense of Framing Patterns in Protest Coverage”*.
- OPINION | COST Action CA21129 Working Group Meeting 2024, University of Salamanca, Masterclass: *“The Good, The Bad, and The Generated: Opinion Mining with (local) AI”*.
- COMPTXT Conference 2024, Amsterdam, Keynote: *“We should use local GPTs (and it's very easy to do!)”*.
- COMPTXT Conference 2024, Amsterdam, presentations: *“Short-Form, Broad Reach: Multimodal Political Communication on TikTok”*; *“The TikTok EuroParty database: Analyzing Political Party Activity on TikTok across 30 European Countries”*.
- Etmaal van de Communicatiewetenschap conference (Etmaal 2024), Rotterdam, presentation: *“Dancing to the partisan beat. A comparative analysis of political parties TikTok use across Europe”*.
- Invited Talk *“Using open source generative large language models with Ollama”* at the Vrije Universiteit Amsterdam Societal Analytics Lab New Year's meetup and LLM-workshop, 30/01/2024.

- 13th annual conference of the European Political Science Association (EPSA 2023), Glasgow, presentations: *“How Political Agendas Form: Relations Between Legacy, Social and Web Media in the Hybrid Media System”*, *“Whose voices are heard? Making sense of framing patterns in protest coverage”*.
- 73rd Annual International Communication Association Conference (ICA 2023), Toronto, presentations: *“Why WallStreetBets Isn't Occupy 2.0: How the Rise of Platforms Can Disrupt Connective Action”*; *“[Tool Demo] AmCAT(s): The Amsterdam Content Analysis Toolkits Suite”*.
- COMPTTEXT Conference 2023, Glasgow, presentation: *“How Political Agendas Form: Relations Between Legacy, Social and Web Media in the Hybrid Media System”*.
- 2022 (In)Visibility in the Digital Age, DGPK-Fachgruppe Digitale Kommunikation, Düsseldorf, presentation: *“Fragmentierung im Facebook-Feed? Die (Un-)Sichtbarkeit politischer Inhalte im Bundestagswahlkampf 2021”*.
- 2022 ECREA 2022 9th European Communication Conference, Aarhus, presentation: *“Fragmentation in Your Feeds? What voters saw on Facebook in the 2021 German election campaigns”*.
- 2022 APSA Annual Meeting & Exhibition, Montreal, presentation: *“Why WallStreetBets isn't Occupy 2.0: How the Rise of Platforms can Disrupt Connective Action”*.
- Invited Talk “Semi-Automated Content Analysis of Media Frames” at Text as Data PoliSci Reading Group, 05/07/2022.
- 72nd Annual International Communication Association (ICA 2022) Conference, Paris, presentation: *“Troublemakers in the Streets? A Framing Analysis of Newspaper Coverage of Protests in the UK 1992-2017”*, <https://youtu.be/uA4rs2at-bc>.
- Third Annual COMPTTEXT Conference, online event (COMPTTEXT 2020), attendee only.
- European Consortium for Political Research 2019 General Conference (ECPR 2019), Wrocław, presentation: *“Troublemakers in the streets? 26 years of UK newspaper coverage of protest”*.
- Fifth International Conference on Computational Social Science (IC<sup>2</sup>S<sup>2</sup> 2019), Amsterdam, presentation: *“Troublemakers in the streets? 26 years of UK newspaper coverage of protest”*.
- Ninth annual conference of the European Political Science Association (EPSA 2019), Belfast, presentation: *“Semi-automated Content Analysis of Media Frames”*.
- Quantitative Text Analysis Dublin (QTA-DUB 2019), Dublin, presentation: *“How to Analyse Media Reports of Protest? An Automated Content Analysis Approach”*.
- Invited talk “How to Analyse Media Reports (Fast)” at Mittelbaukolloquium, Goethe University of Frankfurt/Main, 13/06/2019.
- Ninth Conference on New Directions in Analyzing Text as Data (TADA 2018), Seattle, poster-presentation: *“Semi-automated Content Analysis of Media Frames”*.

---

## TEACHING EXPERIENCE

---

- GESIS Fall Seminar in Computational Social Science : [Introduction to Computational Social Science with R](#), 01/09 - 05/09/2025
- Essex Summer School in Social Science Data Analysis: [Introduction to Web Scraping and Data Management for Social Scientists](#), 07/07-18/07/2025
- GESIS Fall Seminar in Computational Social Science : [Introduction to Computational Social Science with R](#), 30/08 - 05/09/2024

- Essex Summer School in Social Science Data Analysis: [Introduction to Web Scraping and Data Management for Social Scientists](#), 22/07-02/08/2024
- Comptext24 Workshop: [Intro to web scraping](#), 02/05/2024
- Essex Summer School in Social Science Data Analysis: [Introduction to Web Scraping and Data Management for Social Scientists](#), 24-28/07/2023
- 2-day workshop: Text Analysis in R at Vienna Doctoral School of Social Sciences, 06-07/07/2023
- NCRM/Exeter Computational Communication Methods Spring School: [Collaborative annotation projects for supervised machine learning with AmCAT](#), 28/04/2023
- Text Analysis in R: An Introduction at Goethe-University Frankfurt, 27-28/03/2023.
- [Python for R users](#) (#WorkshopsForUkraine), 16/02/2023.
- Introduction to Data Science with R for Social Scientists. European New School of Digital Studies (ENS). Summer Semester 2022
- Introduction to Political Communication. European New School of Digital Studies (ENS). Winter Semester 2021/22
- Introduction to Data Science with R for Social Scientists. European New School of Digital Studies (ENS). Winter Semester 2021/22
- Preparatory Colloquium for Master Students. European New School of Digital Studies (ENS). Winter Semester 2021/22
- Teaching Assistant, Scottish Graduate School of Social Science, course: “Quantitative Approaches to Linguistic and Textual Analysis and Data Visualisation Using R” by Dr. Stefan Müller (University College Dublin) and Dr. Theresa Gessler (University of Zürich). 4-18/11/2020
- Invited guest lecture in course “Automatisierte Inhaltsanalyse für Textdaten mit R” (Automated content analysis for text data with R), Goethe-University Frankfurt, topic: “Introduction to *LexisNexisTools*“. 29/05/2019.
- Teaching Assistant, ECPR Winter School, course: “Introduction to Discourse Network Analysis (DNA)” by Philip Leifeld (University of Glasgow). 22-23/02/2019.
- Tutor for Stata and MAXQDA at Goethe-University Frankfurt, Department of Social Sciences. 2013-2014.

---

## PROFESSIONAL MEMBERSHIPS

---

- International Communication Association.
- COST action network [OPINION](#) – Chair Working Group 2 (Tools).

---

## SERVICE TO PROFESSION

---

### Referee for Journals

- [Computational Communication Research](#)
- [EPJ Data Science](#)
- [Political Analysis](#)
- [Mass Communication and Society](#)

- [Media and Communication](#)
- [New Media & Society](#)
- [Medien & Kommunikationswissenschaft \(M&K\)](#)
- [Political Psychology](#)
- [Politics and Governance](#)
- [Social Science Computer Review](#)
- [The International Journal of Press/Politics](#)

2023—2024

Co-organizer of COMPTExT Conference, Amsterdam May 2024

---

## SKILLS

---

### Languages

- German (native), English (fluent)

### Programming/Statistical Software

- R, Python, Stata, Git

### Markup languages

- Quarto/R Markdown, Markdown,  $\text{\LaTeX}$ , HTML/CSS

### System administration

- Linux, Docker, virtual machines, cloud computing