



UvA-DARE (Digital Academic Repository)

Is explanation the cure?

A human-centered framework for explainable recommender systems

Wang, C.

Publication date

2026

[Link to publication](#)

Citation for published version (APA):

Wang, C. (2026). *Is explanation the cure? A human-centered framework for explainable recommender systems*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

References

References

- Abadi, A. K. M., Kalota, F., & Pour, A. J. (2025, June). Understanding Explainability in Recommender Systems-User Insights and Perspectives. In *2025 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS)* (pp. 1-4). IEEE.
- Abusitta, A., Li, M. Q., & Fung, B. C. (2024). Survey on explainable AI: Techniques, challenges and open issues. *Expert Systems with Applications*, 255, 124710.
- Ahn, Y., Lin, Y. R., Alikhani, M., & Cheon, E. (2025). Human-centered explanation does not fit all: The interplay of sociotechnical, cognitive, and individual factors in the effect AI explanations in algorithmic decision-making. *arXiv preprint arXiv:2502.12354*.
- Al-Ansari, N., Al-Thani, D., & Al-Mansoori, R. S. (2024). User-Centered Evaluation of Explainable Artificial Intelligence (XAI): A Systematic Literature Review. *Human Behavior and Emerging Technologies*, 2024(1), 4628855.
- Andrienko, N., Andrienko, G., Adilova, L., & Wrobel, S. (2022). Visual analytics for human-centered machine learning. *IEEE Computer Graphics and Applications*, 42(1), 123–133.
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, 58, 82-115.
- Ashfaq, M., Yun, J., Yu, S., & Loureiro, S. M. C. (2020). I, Chatbot: Modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents. *Telematics and Informatics*, 54, 101473. <https://doi.org/10.1016/j.tele.2020.101473>
- Atir, S., Rosenzweig, E., & Dunning, D. (2015). When knowledge knows no bounds: Self-perceived expertise predicts claims of impossible knowledge. *Psychological Science*, 26(8), 1295-1303.
- Autoriteit Persoonsgegevens. (2024). Dutch AP AI/transparency guidance (press/guidance portal). <https://autoriteitpersoonsgegevens.nl/en>
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130-1132.
- Baruh, L., Secinti, E., & Cemalcilar, Z. (2017). Online privacy concerns and privacy management: A meta-analytical review. *Journal of Communication*, 67(1), 26-53.
- Baumer, E. P. (2017). Toward human-centered algorithm design. *Big Data & Society*, 4(2), 2053951717718854., E. P. (2017). Toward human-centered algorithm design. *Big Data & Society*, 4(2), 2053951717718854.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1), 289-300.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588.
- Bhatnagar, S., & Agrawal, R. (2024). Understanding explainable artificial intelligence techniques: a comparative analysis for practical application. *Bulletin of Electrical Engineering and Informatics*, 13(6), 4451-4455.
- Bianchi, M., & Briere, M. (2021). Robo-Advising: Less AI and More XAI? *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3825110>
- Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., & Shadbolt, N. (2018, April). 'It's Reducing a Human Being to a Percentage' Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 Chi conference on human factors in computing systems* (pp. 1-14).
- Blassnig, S., Mitova, E., Strikovic, E., Urman, A., de Vreese, C., Hannák, A., & Esser, F. (2024). User Perceptions of News Recommender Systems and Trust in Media Outlets: A Five-Country Study. *Journalism Studies*, 25(10), 1182-1204.

- Boerman, S. C., Van Reijmersdal, E. A., & Neijens, P. C. (2014). Effects of sponsorship disclosure timing on the processing of sponsored content: A study on the effectiveness of European disclosure regulations. *Psychology & Marketing*, 31(3), 214-224.
- Boerman, S. C., Willemsen, L. M., & Van Der Aa, E. P. (2017). "This post is sponsored" effects of sponsorship disclosure on persuasion knowledge and electronic word of mouth in the context of Facebook. *Journal of interactive marketing*, 38(1), 82-92.
- Bojko, A. (2009). Informative or misleading? Heatmaps deconstructed. In *Human-Computer Interaction. New Trends: 13th International Conference, HCI International 2009* (pp. 30–39). Springer Berlin Heidelberg.
- Bruner, J. S. (1961). The act of discovery. *Harvard educational review*.
- Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big data & society*, 3(1), 2053951715622512.
- Buschmeier, H., Buhl, H. M., Kern, F., Grimminger, A., Beierling, H., Fisher, J., ... & Wrede, B. (2023). Forms of Understanding of XAI-Explanations. *arXiv preprint arXiv:2311.08760*.
- Carolus, A., Augustin, Y., Markus, A., & Wienrich, C. (2023). Digital interaction literacy model—Conceptualizing competencies for literate interactions with voice-based AI systems. *Computers and Education: Artificial Intelligence*, 4, 100114.
- Carter, L., Liu, D., & Cantrell, C. (2020). Exploring the intersection of the digital divide and artificial intelligence: A hermeneutic literature review. *AIS Transactions on Human-Computer Interaction*, 12(4), 253-275.
- Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13, 195–212. doi:10.1007/BF01246098
- Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: a systematic literature review. *Artificial Intelligence Review*, 57(8), 216.
- Chatti, M. A., Guesmi, M., & Muslim, A. (2024). Visualization for recommendation explainability: A survey and new perspectives. *arXiv preprint arXiv:2305.11755*.
- Chellappa, R. K., & Sin, R. G. (2005). Personalization versus privacy: An empirical examination of the online consumer's dilemma. *Information technology and management*, 6(2), 181-202.
- Chiu, T. K., Ahmad, Z., Ismailov, M., & Sanusi, I. T. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open*, 6, 100171.
- Choi, J., Lee, H. J., Sajjad, F., & Lee, H. (2014). The influence of national culture on the attitude towards mobile recommender systems. *Technological Forecasting and Social Change*, 86, 65–79.
- Chromik, M., & Butz, A. (2021). Human-XAI interaction: A review and design principles for explanation user interfaces. In *Human-Computer Interaction—INTERACT 2021* (pp. 619–640). Springer International Publishing.
- Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the national academy of sciences*, 118(9), e2023301118.
- Cotter, K. (2024). Practical knowledge of algorithms: The case of BreadTube. *New Media & Society*, 26(4), 2131-2150.
- Cotter, K., & Reisdorf, B. C. (2020). Algorithmic Knowledge Gaps: A New Dimension of (Digital) Inequality. *International Journal of Communication*, 14.
- Dang, Q., & Li, G. (2025). Unveiling trust in AI: The interplay of antecedents, consequences, and cultural dynamics. *AI & SOCIETY*, 1-24.
- DataReportal (2021), Digital 2021 Global Digital Overview. Retrieved from <https://datareportal.com/reports/-digital-2021-global-digital-overview>
- DataReportal (2022), Digital 2022: The Netherlands. Retrieved from <https://datareportal.com/reports/digital-2022-netherlands>
- De Campos, L. M., Fernández-Luna, J. M., & Huete, J. F. (2024). An explainable content-based approach for recommender systems: a case study in journal recommendation for paper submission. *User Modeling and User-Adapted Interaction*, 34(4), 1431-1465.

- de Vries, D. A., Piotrowski, J. T., & de Vreese, C. (2025). Developing the DigIQ: A measure of digital competence. *PLOS ONE*, 20(5), e0322995.
- DeVito, M. A., Gergle, D., & Birnholtz, J. (2017, May). "Algorithms ruin everything" #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 3163-3174).
- Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital journalism*, 5(7), 809-828.
- Djamasbi, S., Siegel, M., Tullis, T., & Dai, R. (2010, January). Efficiency, trust, and visual appeal: Usability testing through eye tracking. In *2010 43rd Hawaii International Conference on System Sciences* (pp. 1–10). IEEE.
- Dogruel, L., Facciorusso, D., & Stark, B. (2020). 'I'm still the master of the machine.' Internet users' awareness of algorithmic decision-making and their perception of its effect on their autonomy. *Information, Communication & Society*, 1-22.
- Donat, E., Brandtweiner, R., & Kerschbaum, J. (2009). Attitudes and the Digital Divide: Attitude Measurement as Instrument to Predict Internet Usage. *Informing Science: The International Journal of an Emerging Transdiscipline*, 12, 037–056. <https://doi.org/10.28945/427>
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.
- Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4), 455–470.
- Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021, May). Expanding explainability: Towards social transparency in ai systems. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1-19).
- Ehsan, U., Passi, S., Liao, Q. V., Chan, L., Lee, I. H., Muller, M., & Riedl, M. O. (2024, May). The who in XAI: how AI background shapes perceptions of AI explanations. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-32).
- Eiband, M., Völkel, S. T., Buschek, D., Cook, S., & Hussmann, H. (2019, March). When people and algorithms meet: User-reported problems in intelligent everyday applications. In *Proceedings of the 24th international conference on intelligent user interfaces* (pp. 96-106).
- Elahi, E., Anwar, S., Shah, B., Halim, Z., Ullah, A., Rida, I., & Waqas, M. (2024). Knowledge graph enhanced contextualized attention-based network for responsible user-specific recommendation. *ACM Transactions on Intelligent Systems and Technology*, 15(4), 1-24.
- Elena-Bucea, A., Cruz-Jesus, F., Oliveira, T., & Coelho, P. S. (2021). Assessing the Role of Age, Education, Gender and Income on the Digital Divide: Evidence for the European Union. *Information Systems Frontiers*, 23(4), 1007–1021. <https://doi.org/10.1007/s10796-020-10012-9>
- Enoch, Y., & Soker, Z. (2006). Age, gender, ethnicity and the digital divide: university students' use of web-based instruction. *Open Learning: The Journal of Open, Distance and E-Learning*, 21(2), 99-110.
- Eslami, M., Karahalios, K., Sandvig, C., Vaccaro, K., Rickman, A., Hamilton, K., & Kirlik, A. (2016, May). First I "like" it, then I hide it: Folk Theories of Social Feeds. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 2371-2382).
- Espinoza-Rojas, J., Siles, I., & Castelain, T. (2023). How using various platforms shapes awareness of algorithms. *Behaviour & Information Technology*, 42(9), 1422-1433.
- Eubanks, B. (2022). *Artificial intelligence for HR: Use AI to support and develop a successful workforce*. Kogan Page Publishers.
- Eubanks, V. (2017). *Automating inequality: How high-tech tools profile, police, and punish the poor* (First Edition). St. Martin's Press.
- European Commission. (2016). Understanding consumer vulnerability in the EU's key markets. Factsheet, Brussels.

- European Commission. (2020). Shaping Europe's digital future. Publications Office of the European Union. https://ec.europa.eu/info/sites/default/files/communication-shaping-europes-digital-futurefeb2020_en_4.pdf
- European Commission. (2023). *Proposal for a Regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act)*. COM(2023) 270 final.
- European Parliament & Council. (2024). *Artificial Intelligence Act*. Official Journal, June 13, 2024. (consolidated OJ entry)
- European Union. (2022). *Regulation (EU) 2022/2065 of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act)*. Official Journal of the European Union, L277. <https://eur-lex.europa.eu/eli/reg/2022/2065/oj>
- Eynon, R. (2023). Algorithmic bias and discrimination through digitalisation in education: A socio-technical view. In *World Yearbook of Education 2024* (pp. 245-260). Routledge.
- Fan, M., Yang, X., Yu, T., Liao, Q. V., & Zhao, J. (2022). Human-ai collaboration for UX evaluation: effects of explanation and synchronization. *Proceedings of the ACM on human-computer interaction*, 6(CSCW1), 1-32.
- Fast, E., & Horvitz, E. (2017). Long-Term Trends in the Public Perception of Artificial Intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1), Article 1. <https://ojs.aaai.org/index.php/AAAI/article/view/10635>
- Feezell, J. T., Wagner, J. K., & Conroy, M. (2021). Exploring the effects of algorithm-driven news sources on political behavior and polarization. *Computers in Human Behavior*, 116, 106626. <https://doi.org/10.1016/j.chb.2020.106626>
- Ferrari, A. (2012). *Digital competence in practice: An analysis of frameworks* (Vol. 10, p. 82116). Luxembourg: Publications Office of the European Union.
- Festic, N. (2022). Same, same, but different! Qualitative evidence on how algorithmic selection applications govern different life domains. *Regulation & Governance*, 16(1), 85-101.
- Fletcher, J. D., & Tobias, S. (2005). The multimedia principle. In *The Cambridge Handbook of Multimedia Learning* (pp. 117–133).
- Fonseca, J. R., & Cardoso, M. G. (2007). Mixture-model cluster analysis using information theoretical criteria. *Intelligent Data Analysis*, 11(2), 155-173.
- Gillespie, T. (2014). The relevance of algorithms. *Media technologies: Essays on communication, materiality, and society*, 167(2014), 167.
- Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International journal of industrial ergonomics*, 24(6), 631-645.
- Gran, A.-B., Booth, P., & Bucher, T. (2021). To be or not to be algorithm aware: A question of a new digital divide? *Information, Communication & Society*, 24(12), 1779–1796. <https://doi.org/10.1080/1369118X.2020.1736124>
- Gruber, J., & Hargittai, E. (2023). The importance of algorithm skills for informed Internet use. *Big Data & Society*, 10(1), 20539517231168100.
- Gruber, J., Hargittai, E., Karaoglu, G., & Brombach, L. (2021). Algorithm Awareness as an Important Internet Skill: The Case of Voice Assistants. *International Journal of Communication*, 15(0), 19.
- Guesmi, M., Chatti, M. A., Joarder, S., Ain, Q. U., Siepmann, C., Ghanbarzadeh, H., & Alatrash, R. (2023). Justification vs. transparency: Why and how visual explanations in a scientific literature recommender system. *Information*, 14(7), 401.
- Guesmi, M., Chatti, M. A., Vorgerd, L., Joarder, S. A., Ain, Q. U., Ngo, T., ... & Muslim, A. (2021, September). Input or output: Effects of explanation focus on the perception of explainable recommendation with varying level of details. In *Intrs@ RecSys* (pp. 55–72).
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5), 1-42.
- Gunning, D., & Aha, D. W. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI Magazine*, 40(2), 44-58.
- Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). *Applied latent class analysis*. Cambridge University Press.

- Hallin, D. C., & Mancini, P. (2004). *Comparing media systems: Three models of media and politics*. Cambridge university press.
- Haque, A. B., Islam, A. N., & Mikalef, P. (2023). Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. *Technological Forecasting and Social Change*, 186, 122120.
- Hargittai, E. (2001). Second-level digital divide: Mapping differences in people's online skills. *arXiv preprint cs/0109068*.
- Hargittai, E., & Micheli, M. (2019). Internet Skills and Why They Matter. In *Society and the Internet* (2nd ed.). Oxford University Press.
- Helberger, N., Karppinen, K., & D'acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, communication & society*, 21(2), 191-207.
- Helberger, N., Sax, M., Strycharz, J., & Micklitz, H. W. (2021). Choice architectures in the digital economy: Towards a new understanding of digital vulnerability. *Journal of Consumer Policy*, 1-26.
- Helm, J. M., Swiergosz, A. M., Haerberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., ... & Ramkumar, P. N. (2020). Machine learning and artificial intelligence: definitions, applications, and future directions. *Current reviews in musculoskeletal medicine*, 13(1), 69-76.
- Heringer, H. J. (1983). Criteria of understanding and understandability. In M. Bock & G. Rickheit (Eds.), *Psycholinguistics Studies in Language Processing* (pp. 259–265). De Gruyter.
- Heringer, H. J. (1983). Criteria of understanding and understandability. *Psycholinguistics studies in language processing*. Berlin: de Gruyter.
- Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000, December). Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work* (pp. 241–250).
- Herm, L. V. (2023). Impact of explainable AI on cognitive load: Insights from an empirical study. *arXiv preprint arXiv:2304.08861*.
- Herm, L. V., Heinrich, K., Wanner, J., & Janiesch, C. (2023). Stop ordering machine learning algorithms by their explainability! A user-centered investigation of performance and explainability. *International Journal of Information Management*, 69, 102538.
- Hernandez-Bocanegra, D. C., & Ziegler, J. (2021, August). Effects of interactivity and presentation on review-based explanations for recommendations. In *IFIP Conference on Human-Computer Interaction* (pp. 597–618). Cham: Springer International Publishing.
- Hernandez-Bocanegra, D. C., & Ziegler, J. (2021, July). Conversational review-based explanations for recommender systems: Exploring users' query behavior. In *Proceedings of the 3rd Conference on Conversational User Interfaces* (pp. 1–11).
- Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2023). Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance. *Frontiers in Computer Science*, 5, 1096257.
- Hoffmann, A. L. (2019). Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society*, 22(7), 900-915.
- Holmqvist, K. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Huang, K. T., & Ball, C. (2024). The influence of ai literacy on user's trust in ai in practical scenarios: a digital divide pilot study. *Proceedings of the Association for Information Science and Technology*, 61(1), 937-939.
- Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The mind's eye* (pp. 573-605). North-Holland.
- Jannach, D., Pu, P., Ricci, F., & Zanker, M. (2022). Recommender systems: Trends and frontiers. *Ai Magazine*, 43(2), 145-150.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM computing surveys*, 55(12), 1-38.
- Joris, G., Vercoutere, S., De Clercq, O., Van Damme, K., Mechant, P., & De Marez, L. (2024). Nudging towards exposure diversity: Examining the effects of news recommender design

- on audiences' news exposure behaviours and perceptions. *Digital Journalism*, 12(8), 1118-1139.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329.
- Kahneman, D. (2011). Fast and slow thinking. *Allen Lane and Penguin Books, New York*.
- Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2023). Trust in artificial intelligence: Meta-analytic findings. *Human factors*, 65(2), 337-359.
- Khan, M. L., & Idris, I. K. (2019). Recognise misinformation and verify before sharing: a reasoned action and information literacy perspective. *Behaviour & Information Technology*, 38(12), 1194-1212.
- Kim, J., Maathuis, H., & Sent, D. (2024). Human-centered evaluation of explainable AI applications: a systematic review. *Frontiers in Artificial Intelligence*, 7, 1456486.
- Kizilcec, R. F. (2016, May). How much information? Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2390–2395).
- Klein, U., Depping, J., Wohlfahrt, L., & Fassbender, P. (2024). Application of artificial intelligence: risk perception and trust in the work context with different impact levels and task types. *AI & SOCIETY*, 39(5), 2445-2456.
- Kline, R. B. (2018). Assessing statistical aspects of test fairness with structural equation modelling. In *Fairness Issues in Educational Assessment* (pp. 116–134). Routledge.
- Kline, R. B. (2023). *Principles and practice of structural equation modeling*. Guilford Publications.
- Knijnenburg, B. P., & Kobsa, A. (2013). Making decisions about privacy: information disclosure in context-aware recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 3(3), 1-23.
- Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, 22, 441–504.
- Kong, X., Liu, S., & Zhu, L. (2024). Toward human-centered XAI in practice: A survey. *Machine Intelligence Research*, 21(4), 740–770.
- Kongsted, A., & Nielsen, A. M. (2017). Latent class analysis in health research. *Journal of Physiotherapy*, 63,55–58. doi:10.1016/j.jphys.2016.05.018
- Kouki, P., Schaffer, J., Pujara, J., O'Donovan, J., & Getoor, L. (2019, March). Personalized explanations for hybrid recommender systems. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (pp. 379–390).
- Kühl, N., Meske, C., Nitsche, M., & Lobana, J. (2024). Investigating the role of explainability and ai literacy in user compliance. *arXiv preprint arXiv:2406.12660*.
- Kulesza, T., Burnett, M., Wong, W. K., & Stumpf, S. (2015, March). Principles of explanatory debugging to personalize interactive machine learning. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (pp. 126–137).
- Kulesza, T., Stumpf, S., Burnett, M., Yang, S., Kwan, I., & Wong, W. K. (2013, September). Too much, too little, or just right? Ways explanations impact end users' mental models. In *2013 IEEE Symposium on visual languages and human centric computing* (pp. 3-10). IEEE.
- Kunkel, J., Donkers, T., Michael, L., Barbu, C. M., & Ziegler, J. (2019, May). Let me explain: Impact of personal and impersonal explanations on trust in recommender systems. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-12).
- Kuran, C.H.A., Morsut, C., Kruke, B.I., Krüger, M., Segnestam, L., Orru, K., Nævestad, T.O., Airola, M., Keränen, J., Gabel, F. and Hansson, S., 2020. (2020). Vulnerability and vulnerable groups from an intersectionality perspective. *International Journal of disaster risk reduction*, 50, 101826.
- Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of communication*, 50(1), 46-70.

- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80.
- Liao, Q. V., & Varshney, K. R. (2021). Human-centered explainable ai (xai): From algorithms to user experiences. *arXiv preprint arXiv:2110.10790*.
- Liao, Q. V., & Vaughan, J. W. (2023). Ai transparency in the age of llms: A human-centered research roadmap. *arXiv preprint arXiv:2306.01941*.
- Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3), 31–57.
- Liu, J., Wu, D., & Guo, Q. (2024). Are we different? Analyzing the role of algorithmic curation and algorithmic literacy during online shopping from a gender differences perspective. *Behaviour & Information Technology*, 43(10), 1990-2006.
- Liu, X., & Cui, Y. (2025). Eye tracking technology for examining cognitive processes in education: A systematic review. *Computers & Education*, 105263.
- Liu, Y., Bi, S., Shi, Z., & Hanzo, L. (2020). When Machine Learning Meets Big Data: A Wireless Communication Perspective. *IEEE Vehicular Technology Magazine*, 15(1), 63–72. <https://doi.org/10.1109/MVT.2019.2953857>
- Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. <https://doi.org/10.1145/3313831.3376727>
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- Lutz, C. (2019). Digital inequalities in the age of artificial intelligence and big data. *Human Behavior and Emerging Technologies*, 1(2), 141-148.
- Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods in psychology: Vol. 2. Statistical analysis* (pp. 551–611). Oxford University Press.
- Mattis, N., Groot Kormelink, T., Masur, P. K., Moeller, J., & van Atteveldt, W. (2024). Nudging news readers: A mixed-methods approach to understanding when and how interface nudges affect news selection. *Digital Journalism*, 1-21.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734.
- Mayer, R. E. (2002). Multimedia learning. In *Psychology of Learning and Motivation* (Vol. 41, pp. 85–139). Academic Press.
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 31–48). Cambridge University Press.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 43–52.
- McQuail, D., & Deuze, M. (2020). *McQuail's media and mass communication theory* (7th ed.). SAGE Publications Ltd. <https://doi.org/10.4135/9781529714388>
- Mersha, M., Lam, K., Wood, J., Alshami, A. K., & Kalita, J. (2024). Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction. *Neurocomputing*, 599, 128111.
- Meske, C., Brenne, J., Uenal, E., Oelcer, S., & Doganguen, A. (2025). From Explainable to Explanatory Artificial Intelligence: Toward a New Paradigm for Human-Centered Explanations through Generative AI. *arXiv preprint arXiv:2508.06352*.
- Meza Martínez, M. A., Nadj, M., Langner, M., Toreini, P., & Maedche, A. (2023). Does this explanation help? Designing local model-agnostic explanation representations and an experimental evaluation using eye-tracking technology. *ACM Transactions on Interactive Intelligent Systems*, 13(4), 1-47.
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267, 1-38.
- Mittelstadt, B., Russell, C., & Wachter, S. (2019, January). Explaining explanations in AI. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 279-288).

- Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, 33(4), 659–684. <https://doi.org/10.1007/s13347-020-00405-8>
- Möller, J., Trilling, D., Helberger, N., & Van Es, B. (2020). Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. In *Digital media, political polarization and challenges to democracy* (pp. 45–63). Routledge.
- Moon, J. H., Kim, S., Jung, Y., Bang, J., & Sung, Y. (2025). The Effects of Explainability and User Control on Algorithmic Transparency: The Moderating Role of Algorithmic Literacy. *Cyberpsychology, Behavior, and Social Networking*.
- Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, 91(2), 358.
- Mubarak, F., Suomi, R., & Kantola, S. P. (2020). Confirming the links between socio-economic variables and digitalization worldwide: the unsettled debate on digital divide. *Journal of Information, Communication and Ethics in Society*, 18(3), 415–430.
- Müller, V. C., & Bostrom, N. (2016). Future Progress in Artificial Intelligence: A Survey of Expert Opinion. In V. C. Müller (Ed.), *Fundamental Issues of Artificial Intelligence* (pp. 555–572). Springer International Publishing. https://doi.org/10.1007/978-3-319-26485-1_33
- Naiseh, M. (2024). Social eXplainable AI (Social XAI): Towards expanding the social benefits of XAI. In *The impact of artificial intelligence on societies: Understanding attitude formation towards AI* (pp. 169–178). Cham: Springer Nature Switzerland.
- Naiseh, M., Babiker, A., Al-Shakhsi, S., Cemiloglu, D., Al-Thani, D., Montag, C., & Ali, R. (2025). Attitudes towards AI: The interplay of self-efficacy, well-being, and competency. *Journal of Technology in Behavioral Science*, 1–14.
- Naiseh, M., Babiker, A., Al-Shakhsi, S., Cemiloglu, D., Al-Thani, D., Montag, C., & Ali, R. (2025). Attitudes Towards AI: The Interplay of Self-Efficacy, Well-Being, and Competency. *Journal of Technology in Behavioral Science*, 1–14.
- Napoli, P. M. (2014). Automated media: An institutional theory perspective on algorithmic media production and consumption. *Communication theory*, 24(3), 340–360.
- Naveed, S., Stevens, G., & Robin-Kern, D. (2024). An overview of the empirical evaluation of explainable ai (xai): A comprehensive guideline for user-centered evaluation in xai. *Applied Sciences*, 14(23), 11288.
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041.
- Ng, D. T. K., Leung, J. K. L., Su, J., Ng, R. C. W., & Chu, S. K. W. (2023). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Educational technology research and development*, 71(1), 137–161.
- Norman, D. A. (2023). *Design for a better world: Meaningful, sustainable, humanity centered*. MIT Press.
- Nunes, I., & Jannach, D. (2017). A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27, 393–444.
- Oberski, D. (2016). Mixture models: Latent profile and latent class analysis. In *Modern statistical methods for HCI* (pp. 275–287). Cham: Springer International Publishing.
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, 38(1), 1–4.
- Pacheco, D. A. D. J., & Iwaszchenko, B. (2024). Unravelling human-centric tensions towards Industry 5.0: Literature review, resolution strategies and research agenda. *Digital Business*, 100090.
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin UK.
- Park, K., & Young Yoon, H. (2025). AI algorithm transparency, pipelines for trust not prisms: mitigating general negative attitudes and enhancing trust toward AI. *Humanities and Social Sciences Communications*, 12(1), 1–13.

- Park, S., & Humphry, J. (2019). Exclusion by design: intersections of social, digital and data exclusion. *Information, Communication & Society*, 22(7), 934-953.
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Patel, K., & Patel, H. B. (2020). A state-of-the-art survey on recommendation system and prospective extensions. *Computers and Electronics in Agriculture*, 178, 105779.
- Patil, A., & Patil, M. (2023, February). A comprehensive review on explainable AI techniques, challenges, and future scope. In *International Conference on Intelligent Computing and Networking* (pp. 517-529). Singapore: Springer Nature Singapore.
- Petty, R. E., & Cacioppo, J. T. (1986). Message elaboration versus peripheral cues. In *Communication and persuasion: Central and peripheral routes to attitude change* (pp. 141-172). New York, NY: Springer New York.
- Petty, R. E., & Cacioppo, J. T. (2012). *Communication and persuasion: Central and peripheral routes to attitude change*. Springer Science & Business Media.
- Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, 32(1), 3-25.
- Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Wortman Vaughan, J. W., & Wallach, H. (2021, May). Manipulating and measuring model interpretability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-52).
- Powers, E. (2017). My News Feed is Filtered?: Awareness of news personalization among college students. *Digital Journalism*, 5(10), 1315-1335.
<https://doi.org/10.1080/21670811.2017.1286943>
- Pu, P., & Chen, L. (2007). Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems*, 20(6), 542-556.
- Pu, P., Chen, L., & Hu, R. (2011, October). A user-centric evaluation framework for recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems* (pp. 157-164).
- Radensky, M., Downey, D., Lo, K., Popovic, Z., & Weld, D. S. (2022, April). Exploring the role of local and global explanations in recommender systems. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (pp. 1-7).
- Rader, E., Cotter, K., & Cho, J. (2018, April). Explanations as mechanisms for supporting algorithmic transparency. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1-13).
- Ragnedda, M. (2020). *Enhancing digital equity: Connecting the digital underclass*. Springer Nature.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.
- Rayner, K. (2009). Eye movements in reading: Models and data. *Journal of Eye Movement Research*, 2(5), 1.
- Redmiles, E. M., Kross, S., & Mazurek, M. L. (2017). Where is the Digital Divide?: A Survey of Security, Privacy, and Socioeconomics. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 931-936.
<https://doi.org/10.1145/3025453.3025673>
- Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A., & Meira Jr, W. (2020, January). Auditing radicalization pathways on YouTube. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 131-141).
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:1606.05386*.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135-1144).
- Ricci, F., Rokach, L., & Shapira, B. (2021). Recommender systems: Techniques, applications, and challenges. *Recommender systems handbook*, 1-35.

- Riedl, R. (2022). Is trust in artificial intelligence systems related to user personality? Review of empirical evidence and future research directions. *Electronic Markets*, 32(4), 2021-2051.
- Riefle, L., Hemmer, P., Benz, C., Vössing, M., & Pries, J. (2022). On the influence of cognitive styles on users' understanding of explanations. *arXiv preprint arXiv:2210.02123*.
- Rong, Y., Leemann, T., Nguyen, T. T., Fiedler, L., Qian, P., Unhelkar, V., ... & Kasneci, E. (2023). Towards human-centered explainable AI: A survey of user studies for model explanations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(4), 2104-2122.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1-36.
- Rothrock, L., & Narayanan, S. (2011). *Human-in-the-loop simulations* (pp. 26-29). London, UK: Springer.
- Russell, Norvig, P., & Chang, M.-W. (2022). *Artificial intelligence : a modern approach* (Fourth edition.). Pearson.
- Rutjes, H., Willemsen, M., & IJsselsteijn, W. (2019, May). Considerations on explainable AI and users' mental models. In *CHI 2019 Workshop: Where is the Human? Bridging the Gap between AI and HCI*. Association for Computing Machinery, Inc.
- Sadeghi, Z., Alizadehsani, R., Cifci, M. A., Kausar, S., Rehman, R., Mahanta, P., ... & Pardalos, P. M. (2024). A review of explainable artificial intelligence in healthcare. *Computers and Electrical Engineering*, 118, 109370.
- Scheerder, A., Van Deursen, A., & Van Dijk, J. (2017). Determinants of Internet skills, uses and outcomes. A systematic review of the second-and third-level digital divide. *Telematics and informatics*, 34(8), 1607-1624.
- Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards artificial intelligence scale. *Computers in Human Behavior Reports*, 1, 100014.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.
- Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, 146, 102551. <https://doi.org/10.1016/j.ijhcs.2020.102551>
- Shin, D. (2022). How do people judge the credibility of algorithmic sources? *AI & SOCIETY*, 37(1), 81-96. <https://doi.org/10.1007/s00146-021-01158-4>
- Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction*, 36(6), 495-504.
- Siles, I., Segura-Castillo, A., Solís, R., & Sancho, M. (2020). Folk theories of algorithmic recommendations on Spotify: Enacting data assemblages in the global South. *Big Data & Society*, 7(1), 2053951720923377.
- Siles, I., Valerio-Alfaro, L., & Meléndez-Moran, A. (2024). Learning to like TikTok... and not: Algorithm awareness as process. *New Media & Society*, 26(10), 5702-5718.
- Silva, D. E., Chen, C., & Zhu, Y. (2024). Facets of algorithmic literacy: Information, experience, and individual factors predict attitudes toward algorithmic systems. *New Media & Society*, 26(5), 2992-3017.
- Skulmowski, A., Nebel, S., Remmele, M., & Rey, G. D. (2022). Is a preference for realism really naive after all? A cognitive model of learning with realistic visualizations. *Educational Psychology Review*, 34(2), 649-675.
- Smit, E. G., Van Noort, G., & Voorveld, H. A. (2014). Understanding online behavioural advertising: User knowledge, privacy concerns and online coping behaviour in Europe. *Computers in human behavior*, 32, 15-22.
- Smith, E. R., & Collins, E. C. (2009). Dual-process models: A social psychological perspective. In J. S. B. T. Evans & K. Frankish (Eds.), *Two Minds: Dual Processes and Beyond* (pp. 197-216). Oxford University Press.
- Smith, M., Szongott, C., Henne, B., & von Voigt, G. (2012). Big data privacy issues in public social media. *2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*, 1-6. <https://doi.org/10.1109/DEST.2012.6227909>
- Stein, J. P., Messingschlagler, T., Gnamb, T., Huttmacher, F., & Appel, M. (2024). Attitudes towards AI: measurement and associations with personality. *Scientific Reports*, 14(1), 2909.

- Stray, J., Halevy, A., Assar, P., Hadfield-Menell, D., Boutilier, C., Ashar, A., ... & Vasan, N. (2024). Building human values into recommender systems: An interdisciplinary synthesis. *ACM Transactions on Recommender Systems*, 2(3), 1-57.
- Strömbäck, J. (2005). In search of a standard: Four models of democracy and their normative implications for journalism. *Journalism studies*, 6(3), 331-345.
- Susser, D. (2019). Invisible influence: Artificial intelligence and the ethics of adaptive choice architectures. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. pp. 403-408.
- Swart, J. (2021). Experiencing Algorithms: How Young People Understand, Feel About, and Engage With Algorithmic News Selection on Social Media. *Social Media & Society*, 7(2), 20563051211008828. <https://doi.org/10.1177/20563051211008828>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285.
- Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Academic Press.
- Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and Instruction*, 8(4), 351-362.
- Tambe, P. (2018). The Growing Importance of Algorithmic Literacy. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3776492>
- Thorson, E. (2008). Changing patterns of news consumption and participation: News recommendation engines. *Information, communication & society*, 11(4), 473-489.
- Thurman, N., Moeller, J., Helberger, N., & Trilling, D. (2019). My friends, editors, algorithms, and I: Examining audience attitudes to news selection. *Digital journalism*, 7(4), 447-469.
- Tichenor, P. J., Donohue, G. A., & Olien, C. N. (1970). Mass media flow and differential growth in knowledge. *Public opinion quarterly*, 34(2), 159-170.
- Tintarev, N., & Masthoff, J. (2007). A survey of explanations in recommender systems. In *IEEE 23rd International Conference on Data Engineering Workshop* (pp. 801-810).
- Tintarev, N., & Masthoff, J. (2011). Designing and evaluating explanations for recommender systems. *Foundations and Trends in Information Retrieval*, 5(1), 1-97.
- Tintarev, N., & Masthoff, J. (2015). Explaining recommendations: Design and evaluation. In *Recommender Systems Handbook* (pp. 353-382). Boston, MA: Springer US.
- Tso, W. W., Reichert, F., Law, N., Fu, K. W., de la Torre, J., Rao, N., ... & Ip, P. (2022). Digital competence as a protective factor against gaming addiction in children and adolescents: A cross-sectional study in Hong Kong. *The Lancet Regional Health-Western Pacific*, 20, 100382.
- Turpin, M., Michael, J., Perez, E., & Bowman, S. (2023). Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting. *Advances in Neural Information Processing Systems*, 36, 74952-74965.
- Valtonen, T., Tedre, M., Mäkitalo, K., & Vartiainen, H. (2019). Media Literacy Education in the Age of Machine Learning. *Journal of Media Literacy Education*, 11(2), 20-36.
- Van Deursen, A. J., & Van Dijk, J. A. (2014). The digital divide shifts to differences in usage. *New media & society*, 16(3), 507-526.
- Van Deursen, A. J., Helsper, E. J., & Eynon, R. (2016). Development and validation of the Internet Skills Scale (ISS). *Information, communication & society*, 19(6), 804-823.
- Van Deursen, A. J., Helsper, E., Eynon, R., & Van Dijk, J. A. (2017). The compoundness and sequentiality of digital inequality. *International Journal of Communication*, 11, 452-473.
- Van Dijk, J. (2020). *The digital divide*. John Wiley & Sons.
- Van Dijk, J. A. (2005). *The deepening divide: Inequality in the information society*. Sage publications.
- Verhagen, T., Hooff, B. V. D., & Meents, S. (2015). Toward a better use of the semantic differential in IS research: An integrative framework of suggested action. *Journal of the Association for Information Systems*, 16(2), 1.
- Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision sciences*, 22(2), 219-240.

- Vilone, G., & Longo, L. (2021). Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion*, 76, 89–106.
- Voorveld, H. A., Meppelink, C. S., & Boerman, S. C. (2024). Consumers' persuasion knowledge of algorithms in social media advertising: identifying consumer groups based on awareness, appropriateness, and coping ability. *International Journal of Advertising*, 43(6), 960-986.
- Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harv. JL & Tech.*, 31, 841.
- Wang, C., Boerman, S. C., Kroon, A. C., Möller, J., & de Vreese, C. H. (2024). The artificial intelligence divide: Who is the most vulnerable? *New Media & Society*, 14614448241232345.
- Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019, May). Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–15).
- Wang, X., & Yin, M. (2022). Effects of explanations in ai-assisted decision making: Principles and comparisons. *ACM Transactions on Interactive Intelligent Systems*, 12(4), 1-36.
- Wardatzky, K., Inel, O., Rossetto, L., & Bernstein, A. (2025). Whom do Explanations Serve? A Systematic Literature Survey of User Characteristics in Explainable Recommender Systems Evaluation. *ACM Transactions on Recommender Systems*, 3(4), 1-35.
- Wei, K. K., Teo, H. H., Chan, H. C., & Tan, B. C. (2011). Conceptualizing and testing a social cognitive model of the digital divide. *Information Systems Research*, 22(1), 170-187.
- Whittaker, J., Looney, S., Reed, A., & Votta, F. (2021). Recommender systems and the amplification of extremist content. *Internet Policy Review*, 10(2).
- Wiard, V., Lits, B., & Dufresne, M. (2022). “The Spy Who Loved Me”: A Qualitative Exploratory Analysis of the Relationship Between Youth and Algorithms. *Frontiers in Communication*, 7, 778273.
- Wienrich, C., Carolus, A., Roth-Isigkeit, D., & Hotho, A. (2022). Inhibitors and enablers to explainable AI success: a systematic examination of explanation complexity and individual characteristics. *Multimodal Technologies and Interaction*, 6(12), 106.
- Wojdyski, B. W., & Evans, N. J. (2016). Going native: Effects of disclosure position and language on the recognition and evaluation of online native advertising. *Journal of Advertising*, 45(2), 157-168.
- Wolfe, J. M. (2000). Visual attention. In *Seeing* (pp. 335–386).
- Xu, L., Jiang, C., Qian, Y., Li, J., Zhao, Y., & Ren, Y. (2017). Privacy-accuracy trade-off in differentially-private distributed classification: A game theoretical approach. *IEEE Transactions on Big Data*, 7(4), 770-783.
- Ystgaard, K. F., Atzori, L., Palma, D., Heegaard, P. E., Bertheussen, L. E., Jensen, M. R., & De Moor, K. (2023). Review of the theory, principles, and design requirements of human-centric Internet of Things (IoT). *Journal of Ambient Intelligence and Humanized Computing*, 14(3), 2827–2859.
- Yu, W. (2025). A Conceptual Framework for AI Literacy with a Focus on Competency. *International Journal of Artificial Intelligence in Education*, 1-18.
- Yu, X., Haroon, M., Menchen-Trevino, E., & Wojcieszak, M. (2024). Nudging recommendation algorithms increases news consumption and diversity on YouTube. *PNAS nexus*, 3(12), pga518.
- Zafari, S., de Pagter, J., Papagni, G., Rosenstein, A., Filzmoser, M., & Koeszegi, S. T. (2024). Trust Development and Explainability: A Longitudinal Study with a Personalized Assistive System. *Multimodal Technologies and Interaction*, 8(3), 20.
- Zarouali, B., Boerman, S. C., & de Vreese, C. H. (2021). Is this recommended by an algorithm? The development and validation of the algorithmic media content awareness scale (AMCA-scale). *Telematics and Informatics*, 62, 101607.
- Zarouali, B., Boerman, S. C., Voorveld, H. A., & van Noort, G. (2022). The algorithmic persuasion framework in online communication: conceptualization and a future research agenda. *Internet Research*, 32(4), 1076-1096.

- Zarouali, B., Helberger, N., & De Vreese, C. H. (2021). Investigating algorithmic misconceptions in a media context: Source of a new digital divide?. *Media and Communication*, 9(4), 134-144.
- Zerilli, J., Bhatt, U., & Weller, A. (2022). How transparency modulates trust in artificial intelligence. *Patterns*, 3(4).
- Liao, Q. V., & Sundar, S. S. (2022, June). Designing for responsible trust in AI systems: A communication perspective. In *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency* (pp. 1257-1268).
- Zhang, B., & Sundar, S. S. (2019). Proactive vs. reactive personalization: Can customization of privacy enhance user experience?. *International journal of human-computer studies*, 128, 86-99.
- Zhang, B., Wang, N., & Jin, H. (2014). Privacy concerns in online recommender systems: influences of control and user data input. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)* (pp. 159-173)
- Zhang, Q., Lu, J., & Jin, Y. (2021). Artificial intelligence in recommender systems. *Complex & intelligent systems*, 7(1), 439-457.
- Zhang, Y., & Chen, X. (2020). Explainable recommendation: A survey and new perspectives. *Foundations and Trends® in Information Retrieval*, 14(1), 1–101.
- Zuiderveen Borgesius, F. J., Trilling, D., Möller, J., Bodó, B., De Vreese, C. H., & Helberger, N. (2016). Should we worry about filter bubbles?. *Internet policy review*, 5(1).

Summary

English Summary

As artificial intelligence becomes deeply embedded in everyday life, algorithms increasingly decide what we see: from the next series to watch to the news stories that shape public opinion. Yet the reasoning behind these automated choices often remains opaque, raising concerns about fairness, bias, and accountability in algorithmic decision-making. Explainable AI (XAI) has emerged as a response to this opacity, promising greater transparency and user empowerment. But this promise raises a fundamental question at the heart of this dissertation: Is explanation really the cure for the challenges of algorithmic transparency?

In pursuit of this question, this dissertation develops a human-centered framework that integrates four key dimensions: explanation design (i.e., scope and modality), underlying mechanisms (i.e., visual attention), user outcomes (i.e., cognitive and affective responses such as understanding of, trust in, and attitude toward the system), and user competence (i.e., AI-related knowledge, skills, and attitudes). It examines *what* should be explained and *in what form, how* explanations work through visual attention, and *for whom* they are effective given differences in users' AI competence. Bridging communication science, cognitive psychology, and human-computer interaction, the framework reconceptualizes explainability not as a fixed property of an algorithm but as a communicative and contextual process that re-centers human interpretation and agency within algorithmic interaction.

The framework is examined through three independent yet conceptually connected empirical studies. A population-representative survey in the Netherlands maps individuals' AI competence in everyday AI use, identifying distinct competence profiles and confirming the presence of an emerging AI divide. A controlled laboratory experiment using eye-tracking investigates how explanation scope and modality influence visual attention, and, in turn, users perceived and objective understanding, trust, and attitudes toward an entertainment recommender system. Finally, an online experiment in a news context tests whether explanations

improve these outcomes across different levels of AI competence in a broader population sample.

Across studies, four overarching insights emerge. First, explanation design is decisive. Joint explanations that integrate both local and global reasoning, providing specific justifications alongside system-level logic, consistently enhance user understanding of how the system operates. However, added visual illustrations aid comprehension only when they are well aligned with textual information. Second, visual attention functions as the proximal mechanism through which design features translate into cognitive gains. Explanations that capture and sustain users' visual attention foster higher perceived and objective understanding of the system's reasoning, although attention alone does not account for affective responses such as trust or attitudes toward the system. Third, these affective responses are conditional and often unfold indirectly. Improved understanding of the system enables more trust and more positive evaluations toward the system rather than producing them directly. Finally, explanation benefits are not evenly distributed. Users with greater AI knowledge, skills, and positive attitudes derive the most cognitive and affective gains, whereas less competent users often show limited improvement. This pattern underscores that explainability is inherently stratified and that one-size-fits-all transparency designs risk deepening, rather than reducing, existing competence divides.

These results advance theory in several ways. The framework replaces a binary view of explanations with a design-alignment perspective that balances informational completeness with cognitive accessibility. It specifies attention as a bounded mechanism linking design to cognition, while analytically distinguishing cognitive outcomes from affective evaluations and indicating a cognitive-before-affective sequence. It also embeds inclusiveness into theory by treating competence differences as a core moderator rather than statistical noise. Together, these insights offer a structured and testable model of explainability as a process of learning and judgment, bounded by cognition, mediated by attention, and conditioned by competence.

The work also yields practical and policy guidance. For design, explanations should be treated as coordinated choices about scope and modality, with visuals used to clarify with formats that accommodate user diversity. For

evaluation, outcome measures should be complemented with process measures of attention and engagement to inform iterative improvement. For inclusiveness, layered and adaptive strategies can better serve diverse competence profiles, but explanations alone will not close skill gaps; literacy initiatives remain essential. For governance, transparency obligations should be context-sensitive and competence-aware and paired with inclusion policies so that accountability does not inadvertently reinforce inequality. A roadmap for future research follows from these insights. Priorities include testing the framework across cultures and over time, extending it to additional high- and low-stakes domains, broadening the design space to interactive, personalized, and contrastive forms, strengthening measurement with behavioral and process-tracing methods at scale, and adapting explanation concepts to generative AI.

In conclusion, the dissertation shows that explanations can help users make sense of algorithmic choices and support trustworthy human-AI interaction, but only when aligned with what is explained, how it is presented, who the user is, and the context of use. Explainability is not a static interface feature; it is a practice of communication. Ultimately, in the age of AI, the goal is not only interpretability but empowerment. By advancing a load-sensitive, attention-mediated, and competence-aware framework, this work contributes to building AI that enhances human understanding and serves societal well-being.

Nederlandse Samenvatting

Naarmate kunstmatige intelligentie steeds dieper verweekeld raakt in het dagelijks leven, bepalen algoritmen in toenemende mate wat wij te zien krijgen: van de volgende tv-serie die we bekijken tot de nieuwsverhalen die de publieke opinie vormen. De redenering achter deze geautomatiseerde keuzes blijft echter vaak ondoorzichtig, wat zorgen oproept over eerlijkheid, bias en verantwoordingsplicht in algoritmische besluitvorming. Explainable AI (XAI) is ontstaan als een antwoord op deze ondoorzichtigheid en belooft meer transparantie en gebruikersvriendelijkheid. Deze belofte roept echter een fundamentele vraag op die centraal staat in dit proefschrift: is uitleg (XAI) werkelijk het middel tegen de uitdagingen van algoritmische transparantie?

Om deze vraag te verkennen ontwikkelt dit proefschrift een mensgerichte benadering die drie kern-dimensies integreert: explanation design (zoals de reikwijdte en vorm), aandachtmechanismen (zoals visuele aandacht), de effecten bij gebruikers (bijvoorbeeld cognitieve en emotionele reacties zoals begrip, vertrouwen en houding ten opzichte van het systeem) en gebruikerscompetentie (hun kennis, vaardigheden en houding ten aanzien van AI). Het onderzoekt wat moet worden uitgelegd en in welke vorm; hoe deze uitleg de visuele aandacht trekt; en voor wie uitleg over AI en recommender systems effectief is, rekening houdend met verschillen in AI-gerelateerde kennis, vaardigheden en houdingen. Door communicatiewetenschap, cognitieve psychologie en human-computer interaction te verbinden, wordt een framework voorgesteld dat uitlegbaarheid niet als een vaste eigenschap van een algoritme, maar als een communicatief, dynamisch en contextueel proces dat menselijke interpretatie en handelingsvermogen centraal stelt in de interactie met algoritmische systemen.

Het framework is onderzocht in drie onafhankelijke maar conceptueel samenhangende empirische studies. Een representatieve bevolkingsenquête in Nederland brengt de AI-competentie van individuen in kaart binnen alledaags AI-gebruik, identificeert verschillende competentieprofielen en bevestigt het bestaan van een opkomende AI-kloof. Een gecontroleerd laboratoriumexperiment met eye-tracking onderzoekt hoe de reikwijdte (scope) en vorm van uitleg (tekstuele en/of visuele informatie) visuele aandacht beïnvloeden en vervolgens de waargenomen en objectieve begrip, het vertrouwen en de houding van gebruikers ten aanzien van

een entertainmentaanbevelingssysteem. Ten slotte test een online experiment in een nieuwscontext of de effecten van uitleg verschilt tussen gebruikers met verschillende niveaus van AI-competentie.

Uit de studies komen vier overkoepelende inzichten naar voren. Ten eerste is explanation design doorslaggevend. Verklaringen die zowel lokale als globale redeneringen combineren, waarbij specifieke toelichtingen worden gekoppeld aan de onderliggende logica van het systeem, vergroten consequent het begrip van gebruikers over hoe het systeem werkt.. Toegevoegde visuele elementen bevorderen het begrip echter alleen wanneer ze goed zijn afgestemd op de tekstuele informatie; anders vergroten ze eerder dan dat ze de cognitieve belasting verminderen. Ten tweede fungeert visuele aandacht als het directe mechanisme waardoor ontwerpelementen zich vertalen in cognitieve winst. Een uitleg die de aandacht van gebruikers weet te trekken en vast te houden, bevordert een hoger waargenomen en objectief begrip van de redenering van het AI systeem. Aandacht biedt echter geen verklaring voor affectieve reacties zoals vertrouwen of houding ten aanzien van het systeem. Ten derde zijn deze affectieve reacties voorwaardelijk en treden ze vaak indirect op: beter begrip leidt indirect tot meer vertrouwen en positievere evaluaties. Ten slotte zijn de voordelen van uitleg ongelijk verdeeld. Gebruikers met meer AI-kennis, vaardigheden en positieve houdingen ten opzichte van AI behalen de grootste cognitieve en affectieve winst, terwijl minder competente gebruikers vaak weinig vooruitgang boeken. Dit patroon onderstreept dat uitlegbaarheid inherent gelaagd is en dat uniforme transparantieontwerpen het risico lopen bestaande competentieverschillen te verdiepen in plaats van te verkleinen.

Deze resultaten leveren verschillende theoretische bijdragen. Het voorgestelde framework vervangt een binaire benadering van uitleg door een design-alignment-perspectief dat informatieve volledigheid afweegt tegen cognitieve toegankelijkheid. Het specificceert aandacht als een begrensd mechanisme dat ontwerp koppelt aan cognitie, terwijl het cognitieve uitkomsten analytisch onderscheidt van affectieve evaluaties en een cognitieve-vóór-affectieve volgorde suggereert.

Tevens verankert het framework inclusiviteit in de theorie door verschillen in gebruikerscompetentie te erkennen als een kernfactor die de werking en effectiviteit van verklaringen mede bepaalt, in plaats van als willekeurige ruis.

Gezamenlijk bieden deze inzichten een gestructureerd en toetsbaar model van uitlegbaarheid als een proces van leren en oordelen, begrensd door cognitie, gemedieerd door aandacht en beïnvloed door competentie.

Het onderzoek levert daarnaast praktische en beleidsmatige richtlijnen op. Bij het ontwerp van een uitleg over AI, moet men rekening houden met zowel de reikwijdte als de modaliteit, waarbij visuele elementen worden gebruikt om te verduidelijken en de presentatie is afgestemd op de diversiteit van gebruikers. Voor evaluatie dienen uitkomstmaten te worden aangevuld met procesmaten die inzicht geven in aandacht en betrokkenheid van gebruikers, om zo iteratieve verbetering te ondersteunen. Voor inclusiviteit kunnen adaptieve strategieën beter aansluiten bij uiteenlopende competentieprofielen, maar uitleg alleen zal vaardigheidskloof niet dichten; geletterdheidsinitiatieven blijven essentieel. Voor beleid en toezicht moeten transparantieplichtingen contextgevoelig en competentiebewust zijn en gepaard gaan met inclusiebeleid, zodat verantwoordingsmechanismen niet onbedoeld ongelijkheid versterken. Een onderzoeksagenda voor de toekomst vloeit voort uit deze inzichten. Prioriteiten zijn het toetsen van het framework over culturen en tijd heen, het uitbreiden ervan naar aanvullende domeinen met hoge en lage inzet, het verbreden van de ontwerpruimte naar interactieve, gepersonaliseerde en contrasterende vormen, het versterken van metingen via gedrags- en proces-tracingmethoden op grotere schaal, en het aanpassen van uitlegconcepten aan generatieve AI.

Het proefschrift laat zien dat uitleg over AI en recommender systems gebruikers kan helpen om algoritmische keuzes te begrijpen en een betrouwbare interactie tussen mens en AI te ondersteunen, maar alleen wanneer de uitleg is afgestemd op wat wordt uitgelegd, hoe het wordt gepresenteerd, wie de gebruiker is en in welke context deze plaatsvindt. Uitlegbaarheid is geen statische interface-functie; het is een communicatie-strategie. Uiteindelijk gaat uitlegbaarheid in het tijdperk van AI niet alleen over interpreteerbaarheid, maar ook over de versterking van het handelingsvermogen van gebruikers. Door een framework te ontwikkelen dat gevoelig is voor cognitieve belasting, gemedieerd wordt door aandacht en

rekening houdt met gebruikerscompetentie, draagt dit werk bij aan de ontwikkeling van AI die menselijk begrip bevordert en het maatschappelijk welzijn dient.

中文摘要

随着人工智能日益深入人类生活，算法在很大程度上决定了人们所看到的内容——从下一部观看的剧集，到塑造公众舆论的新闻报道。然而，这些算法选择背后的逻辑常常不透明，引发了关于公平性、偏见与问责机制的广泛关注。可解释人工智能（XAI）由此兴起，旨在回应这种不透明性，提升算法系统的透明度与用户自主性。这也引出了本论文的核心问题：可解释性算法说明是否真能成为应对算法透明性挑战的有效途径？

为探讨这一问题，本论文构建了一个以人为中心的分析框架，整合解释设计、视觉关注机制与用户能力三个核心维度。研究围绕三项核心议题展开：应当向用户呈现哪些信息及其合适的表达形式；解释如何通过视觉关注影响理解与评价；以及在用户 AI 相关知识、技能与态度存在差异的情况下，解释对不同群体的作用效果。该框架连接传播学、认知心理学与人机交互学，重新界定了“可解释性”——它并非算法的固有属性，而是一种以人为核心、具有交流性、动态性与情境性的实践。

本研究通过三项相互衔接的实证研究检验该框架。第一项基于荷兰全国代表性调查，描绘了公众在日常 AI 使用中的能力分布，识别出不同的 AI 能力群体，并揭示了“AI 鸿沟”的初步证据。第二项控制实验利用眼动追踪技术，考察解释范围与呈现方式如何影响视觉关注，并进一步影响用户对娱乐推荐系统的理解、信任与态度。第三项在线实验在新闻推荐情境中检验解释在更广泛样本及不同 AI 能力水平下的作用差异。

综合研究结果，呈现出四项主要发现。第一，解释设计具有决定性意义：融合局部与整体逻辑的联合型解释能显著提升用户对系统运作机制的理解，而视觉元素只有在与文字信息契合、减少认知负荷时才有助于理解。第二，视觉关注是解释设计转化为认知收益的关键机制；吸引并维持用户注视的设计能提升对系统逻辑的主观与客观理解，但注意力本身并不直接影响信任或态度。第三，情感反应具有条件性，往往通过理解提升间接实现更恰当的信任与更积极的评价。第四，解释效益呈现不均：具备更高 AI 知识、技能与积极态度的用户受益更多，而能力较低者提升有限。这表明可解释性本身具有分层性，统一的透明化设计可能加深而非缩小现有差距。

本研究在理论上以“设计契合”的视角取代对解释的二元理解，强调信息完整性与认知可及性之间的平衡；明确了视觉关注在设计与认知之间的中介作用，并区分了认知与情感的不同路径；同时将包容性纳入理论核心，将能力差异视为关键调节机制。实践上，研究提出在设计、评估、包容与治理层面的具体启示——从解释设计的多维协调，到以注意与参与为导向的评估方法，从自适应解释策略到 AI 素养建设与情境化监管。未来研究应跨文化与时间检验该框架，拓展高低风险应用场景，发展交互式与个性化解释形式，并在生成式 AI 时代探索溯源机制、不确定性呈现及能力敏感型设计原则。

综上所述，算法说明可以帮助用户理解算法决策并促进可信赖的人机互动，但唯有在内容、形式、用户与情境的契合下才能发挥真正作用。可解释性不是静态的界面属性，而是一种交流性实践。面对 AI 时代的变革，其目标不仅是系统的可理解性，更是用户的赋权。本研究通过提出一个关注认知负荷、以视觉关注为中介、兼顾能力差异的框架，为构建促进人类理解并服务社会福祉的人工智能提供了理论与实践贡献。

Author contributions

This section details the specific author contributions to the individual chapters, listed in order of their level of contribution. All studies were conducted under the project Understanding Explainable AI as Digital Competence.

Authors' initials

Chenyue Wang (CW)

Dr. Sophie C. Boerman (SB)

Dr. Anne C. Kroon (AK)

Prof. Dr. Judith Möller (JM)

Prof. Dr. Claes H. de Vreese (CV)

Chapter 2. The artificial intelligence divide: Who is the most vulnerable?

Authors: CW, SB, JM, CV

Conceptualization: CW, SB, JM, CV

Methodology: CW, SB, CV, JM

Data Collection: SB, CV

Analysis: CW

Writing (original draft preparation): CW

Writing (review and editing): CW, SB, AK, JM, CV

Visualization: CW

Chapter 3. When recommendations are explainable: An eye-tracking study comparing how and what to explain

Authors: CW, SB, AK, JM, CV

Conceptualization: CW, SB, AK, JM, CV

Methodology: CW, SB, AK, JM, CV

Data Collection: CW

Analysis: CW

Writing (original draft preparation): CW

Writing (review and editing): CW, SB, AK, JM, CV

Visualization: CW

Chapter 4. The role of AI competencies: How do AI knowledge, skills, and attitudes shape users' understanding and evaluation of explainable news recommendations?

Authors: CW, SB, AK, JM, CV

Conceptualization: CW, SB, AK, JM, CV

Methodology: CW, SB, AK, JM, CV

Data Collection: CW

Analysis: CW

Writing (original draft preparation): CW

Writing (review and editing): CW, SB, AK, JM, CV

Visualization: CW

Acknowledgements

As this PhD journey comes to an end, I find myself looking back with gratitude. It has been a journey that challenged me in more ways than I expected, not only intellectually, but also personally, emotionally, and physically. For a long time, I did not know whether what I was doing truly amounted to anything meaningful. Research can be strange in that way: the more you learn, the more aware you become of all that you do not know. There was also loneliness: working from home and spending long days navigating uncertainty in my own mind became a familiar routine. On top of that, I went through a period of illness, surgery, and recovery that lasted for more than a year. Looking back, I think what makes this dissertation meaningful to me is not only that I finished it, but that I somehow found a way to continue through all of that.

And I did not do so alone.

This book exists because many people, in different ways, carried me when I could not fully carry myself. Some offered guidance, some patience, some care, some trust, and some simply stayed. I am deeply grateful to all of them.

First and foremost, I want to thank my supervisors: **Sophie Boerman, Anne Kroon, Judith Möller,** and **Claes de Vreese.** I feel incredibly fortunate to have had not just supervisors, but mentors who supported me with both academic guidance and genuine care. We began this journey during the COVID period, when most of our meetings took place online. Yet even through a screen, your guidance was always constructive, insightful, and encouraging. What I appreciated most was that you never asked me to follow suggestions blindly. Instead, you encouraged me to think, to question, and to make my own decisions. That trust helped me grow not only as a researcher, but also as a person. In moments when I wanted to quit, or when I felt that what I was doing had no value, thinking of your faith in me often gave me just enough strength to continue.

Sophie, thank you for being my first daily supervisor and for guiding me so closely throughout this process. I could not have wished for a better person in that role. Your feedback was always remarkably timely, detailed, direct, and reliable. You never softened your academic feedback unnecessarily, but it always came from a place of honesty and kindness. I always knew I could rely on you, whether it was for a manuscript, a difficult decision, or simply for help when things felt too heavy to carry alone. I also want to thank you for the many small but memorable acts of care: the lunches and coffees you treated me to, the walks and conversations, and the way you continued to make time and space for me even after you joined Wageningen University & Research. When I became too ill to continue the final round of revisions for our first paper, you and Anne carried it forward and brought it to publication. I have never forgotten that. Nor will I forget your phone call to congratulate me when my dissertation assessment was positive. You have been steadfast, kind, and dependable at every stage of this process, and that has meant more to me than I can fully express.

Anne, thank you for joining the team and becoming such an important part of my PhD journey. Your kindness, sincerity, and willingness to help have been constant sources of comfort and strength. You were always there, always ready to think along with me, whether about my manuscripts, my scripts and codes, or the methodological puzzles that seemed impossible until you helped untangle them. At the same time, what I will remember just as much is your warmth. The coffees we had on campus, the conversations that made difficult periods feel lighter, and the

little group chat among the three of us that began during our lab experiment but somehow never disappeared are all things I remember with great fondness. During my illness, your compassion touched me deeply. You made it clear, again and again, that my health mattered more than my work, and that I should take the time I needed to recover. That support meant more to me than I can say.

Judith, thank you for the generosity, gentleness, and calm strength with which you supported me throughout this PhD. You were always smiling, whether through a screen or in person, and you always had a way of making me feel calmer, more capable, and more understood. I still remember the moment before my very first class as a tutorial teacher, when I was standing outside the classroom feeling nervous and you encouraged me with such warmth that I immediately felt steadier. That is the kind of presence you have had throughout my PhD. When I was down, stressed, or unsure of myself, you were there. You always left more positive comments on my manuscripts than critical ones, and somehow I always came away from them with more confidence than before. For someone who often struggles with self-doubt, that mattered immensely. Even after you moved to Germany, the distance never really created distance. Although we did not meet as often after that, your support remained just as present.

Claes, thank you for your energy, sharpness, and remarkable ability to see the bigger picture whenever I felt stuck in my research. You always seemed to know how to help me make decisions when it mattered most, whether about direction, framing, or the larger meaning of the work. In that sense, you were the captain who kept the ship moving forward. I have often wondered how one person can be in so many places, doing so many things at once, often with an automatic reply switched on, and yet still somehow answer emails almost immediately: usually briefly, often unconventionally, but always helpfully. There is something very alive in the way you work and communicate, and I have always admired that. I still remember my first presentation at an international conference, scheduled in the earliest possible slot at ICA in Paris. You advertised it on social media, and you, Sophie, and Judith were all there to support me. That meant a great deal. And somehow, every time you wrote a simple “nice” next to a sentence in my manuscript, it made me disproportionately happy. Your comments were often the shortest, but they were always at the core of the issue. Thank you for your trust,

your clarity, and your ability to guide both the team and me without ever making it feel heavy.

Together, the four of you formed a supervisory team that complemented one another beautifully. Despite being a relatively large team, you almost never placed me in the difficult position of having to navigate conflicting expectations or competing directions. I know how rare that is, and I never took it for granted. Working with you has been one of the greatest gifts of my PhD journey. Long after the dissertation is finished, that will remain one of my greatest gains from these years.

I would also like to thank my committee members for taking the time to read this dissertation and to serve on my defense committee: **Jochen Peter**, **Theo Araujo**, **Joanna Strycharz**, **Tibor Bosse**, and **Marjolijn Antheunis**. I am truly grateful for your time, attention, and engagement with my work.

This dissertation also depended on the support of many people and teams who contributed directly to the research behind it. I am grateful to the **DIGCOM project** team, whose work and data laid the foundation for my first study. In particular, I want to thank Dian, who kindly provided the dataset and generously shared her scripts and relevant files for reference. I also want to thank the lab staff and my research assistant, **Eli**, for their help with data collection during the lab experiment for my second study. Last but certainly not least, I am deeply grateful to all the participants who took part in my research. In a dissertation that is concerned with designing more human-centered systems, your role was especially meaningful. You were not only participants, but the human perspective at the center of the questions I was trying to answer. Thank you for giving your time and attention to this work. Without your participation, this dissertation would not exist.

Beyond the research, I was also fortunate to be surrounded by an academic community that made this journey less solitary. To my fellow PhD researchers at the University of Amsterdam, thank you for the feedback, the PhD club meetings, the conference trips, the conversations, the laughter, and yes, the parties too. We were, in many ways, comrades in the same battle, and it means a great deal to have shared this journey with people who truly understood what it was like. My thanks also go to the people I met at conferences and summer schools, whose conversations broadened my thinking and reminded me that research is also a

community. I want to thank my students as well, especially those who supported me when I taught for the first time. Teaching was one of the experiences that made me feel more grounded in academia, and your engagement gave me confidence.

I am also deeply grateful to the people who supported me through illness, recovery, and reintegration. In particular, I want to thank **Linda Bos** for the regular reintegration meetings, for your practical advice, and for helping me find realistic ways to return step by step. I also want to thank **Monique**, my career coach, for supporting me during that period of reintegration and transition. I am equally grateful to my therapist, **Yulan**, for helping me through the darkest time and for helping me rebuild confidence in myself. To the doctors and medical staff who treated me and, in a very real sense, saved my life, thank you. Your work allowed me not only to recover, but to be here at all, writing these words.

To my friends, thank you for bringing warmth back into my life in ways both big and small. **Lan**, you always somehow know when I need socialization the most and invite me out at exactly the right time. Sometimes a cup of bubble tea under the sunset can be unexpectedly healing. **Zijun**, thank you for listening to all my complaints and even my weird dreams without ever judging me. **Shanhong**, thank you for the sense of ease you bring; I always feel more relaxed after talking to you. I also want to thank the friends I met through yoga, meditation clubs, workshops, and other corners of life. Some people arrive quietly, but still leave light behind. I am grateful for all of you.

To my parents and my grandma, thank you for always reminding me to rest, to take care of myself, and not to stay up too late. That care may sound simple, but during a long and difficult journey, it means more than I can say.

And finally, to my husband, **Vincent**, thank you for being beside me in every stage of this journey, from the first paper I submitted, to the first class I taught, to the final steps of preparing for my defense. You have listened to my endless complaints more times than anyone should have to. You have been the first audience for my presentations, the calm voice when I could not stay calm myself, and the person who celebrated every small win with me when those small wins were all I had. This dissertation could not have been completed without your emotional support, your patience, and your steady presence. You have seen me at

my most uncertain, most exhausted, and most overwhelmed, and you were still there. I could not have completed this journey without you.

This dissertation is the product of research, but also of recovery, encouragement, and care. Finishing it means more than completing a piece of academic work. It also means that, with the help of many people, I kept going. For that, I am profoundly grateful.

Why am I seeing this recommendation and not something else? In a world where algorithms quietly shape what we see online, this dissertation asks whether explanations can help people better understand the recommender systems influencing their choices. Through a human-centered framework and three empirical studies, it examines for whom, in what ways, and under what conditions such explanations can make recommender systems more comprehensible, effective, and inclusive. The findings show that explanations can improve users' understanding, but their effectiveness depends on how they are designed, whether users notice and engage with them, and on differences in users' AI competence. Rather than treating explainability as a simple cure-all, this dissertation presents it as a human-centered challenge: designing explainable recommender systems that support meaningful engagement in AI-shaped communication environments.