Responsible AI Through the Lens of an Information Access Researcher: The Good, the Bad, and the Unknown

Damiano Spina
damiano.spina@rmit.edu.au
RMIT University
Melbourne, Australia

Udine, Italy
21–24 May 2024

ABSTRACT

The rapid advancement of Artificial Intelligence (AI) has brought immense benefits across numerous domains, but it is also well-known that it raises important challenges in terms of ethical and responsible practices in the development, application, and use of these technologies.

Given the multidisciplinary nature of such challenges, it is often not straightforward to advance knowledge in this area. In the last few years I had the opportunity to work in different multidisciplinary teams – including, in addition to computer scientists, experts with a background in media and communication, fact-checking, psychology, and law – to address problems related to responsible AI in the context of information access systems – in particular, search engines and conversational assistants. In the course, I will share my experience on these specific areas, discussing the problems faced and the solutions adopted.

I propose to start the course by facilitating a reflective/positioning activity where participants can discuss their positionality, as well as different keywords related to responsible AI and information access (e.g., "fairness", "bias", "diversity", "search engine", "recommender system", "fact-checking", "fake news", "echo chambers", "filter bubbles"). Then, through a blend of existing findings, emerging avenues, and engaging group activities, we will characterize the benefits (the "good"), challenges ("the bad"), and the opportunities (the "unknown") that we faced while advancing knowledge in specific information access tasks including presentation strategies for fact-checked content, fairness-aware rankings via search results diversification, characterizing information processing activities via physiological signals, among others.

The course learning outcomes include a better understanding of the terminology and methods used by multidisciplinary teams that aim to advance knowledge in responsible AI, as well as application of them in the context of information access and text mining research.

SHORT BIO

Dr. Damiano Spina[1] is a Senior Lecturer at RMIT University, School of Computing Technologies (Melbourne, Australia). Dr. Spina is an Associate Investigator at the ARC Centre of Excellence for Automated Decision-Making and Society (ADM+S)[2], and a research collaborator with RMIT ABC Fact Check.[3] His research expertise is in the field of Information Retrieval and Text Analytics. In particular, his research focuses on interactive information retrieval and evaluation of information access systems. Dr. Spina completed his PhD in Computer Science in 2014 (UNED, Spain). He has published more than 60 peer-reviewed scientific publications, including papers at conferences such as SIGIR, ECIR, CIKM, and CHIIR, as well as at journals such as CACM, IP&M, TOIS, and JASIST. His work attracted more than 2,500 citations. He serves as editorial board member for IP&M and TOIS, and he is an active Program Committee member of various IR conferences. He is the recipient of an ARC Discovery Early Career Researcher Award (DECRA) and the 2021 RMIT Award for Research Impact (Technology).

ACKNOWLEDGMENTS

Most of the research I presented throughout the course was conceptualized and conducted in the unceded lands of the Wurundjeri and Bunurong peoples of the eastern Kulin Nation. I pay my respects to their Ancestors and Elders, past, present, and emerging.

The research is supported by the Australian Research Council (DE200100064, CE200100005). The course was organised as a training activity for students enrolled in the PhD Course on Computer Science and Artificial Intelligence at University of Udine.

I am grateful to Kevin Roitero, Stefano Mizzaro, and Vincenzo della Mea for the opportunity to collaborate with them and the amazing PhD students who attended the course.

[1] www.damianospina.com
[2] www.admscentre.org.au
[3] www.admscentre.org.au/fact-check-research/

BIBLIOGRAPHY

Day 1: Responsible AI: A Multidisciplinary Problem – How Can We Contribute?

[1] Marwah Alaofi, Luke Gallagher, Dana Mckay, Lauren L. Saling, Mark Sanderson, Falk Scholer, Damiano Spina, and Ryen W. White. 2022. "Where Do Queries Come From?" In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR '22). Association for Computing Machinery, Madrid, Spain, 2850–2862. ISBN: 9781450387323. DOI: 10.1145/3477495.3531711

[2] Kate Crawford, Mary L Gray, and Kate Miltner. 2014. "Big Data| critiquing Big Data: Politics, ethics, epistemology| special section introduction." *International Journal of Communication*, 8, 10

[3] Treahna Hamm (Firebrace). 2023. *Information Retrieval on Country*. (2023). https://www.admscentre.org.au/information-retrieval-on-country/

[4] Binh Le, Damiano Spina, Falk Scholer, and Hui Chia. 2022. "A Crowdsourcing Methodology to Measure Algorithmic Bias in Black-box Systems: A Case Study with COVID-related Searches." In: *Proceedings of the Third Workshop on Bias and Social Aspects in Search and Recommendation* (Bias @ ECIR 2022). DOI: 10.1007/978-3-031-09316-6_5

[5] Jason Edward Lewis, Angie Abdilla, Noelani Arista, Kaipulaumakaniolono Baker, Scott Benesiinaabandan, Michelle Brown, Melanie Cheung, Meredith Coleman, Ashley Cordes, Joel Davison, Kūpono Duncan, Sergio Garzon, D. Fox Harrell, Peter-Lucas Jones, Kekuhi Kealiikanakaoleohaililani, Megan Kelleher, Suzanne Kite, Olin Lagon, Jason Leigh, Maroussia Levesque, Keoni Mahelona, Caleb Moses, Isaac ('Ika'aka) Nahuewai, Kari Noe, Danielle Olson, 'Ōiwi Parker Jones, Caroline Running Wolf, Michael Running Wolf, Marlee Silva, Skawennati Fragnito, and Hēmi Whaanga. 2020. "Indigenous Protocol and Artificial Intelligence Position Paper." DOI: 10.11573/SPECTRUM.LIBRARY.CONCORDIA.CA.00986506

[6] Dang Nguyen, James Meese, Jean Burgess, Julian Thomas, Louisa Bartolo, Nicholas Carah, Dominique Carlon, Jeffrey Chan, Sam Kininmonth, Amanda Lawrence, Ramon Lobato, Ariadna Matamoros-Fernández, Silvia Ximena Montaña-Niño, Lucinda Nelson, Jing Qian, Aaron Snoswell, Damiano Spina, Arjun Srinivas, Nicholas Suzor, Avantik Tamta, Patrik Wikstrom, and Joanna Williams. 2023. *AI and Automated Decision-Making in News and Media: Key Technologies and Emerging Challenges*. Tech. rep. ARC Centre of Excellence for Automated Decision-Making and Society, RMIT University. https://www.admscentre.org.au/news-and-media-report/

[7] Alexandra Olteanu, Michael Ekstrand, Carlos Castillo, and Jina Suh. 2023. *Responsible AI Research Needs Impact Statements Too*. (2023). arXiv: 2311.11776 [cs.AI]

[8] Raymond Williams. 2014. *Keywords: A vocabulary of culture and society*. Oxford University Press

[9] Tyson Yunkaporta. 2023. *Right Story, Wrong Story: Adventures in Indigenous Thinking*. Text Publishing Company. ISBN: 9781922790439

Day 2: Mixed Methods for Designing and Evaluating Presentation Strategies for Fact-checked Content

[10] Assunta Cerone, Elham Naghizade, Falk Scholer, Devi Mallal, Russell Skelton, and **D. Spina**. 2020. "Watch 'n' Check: Towards a Social Media Monitoring Tool to Assist Fact-Checking Experts." In: *Proc. DSAA*, 607–613

[11] Danula Hettiachchi, Kaixin Ji, Jenny Kennedy, Anthony McCosker, Flora D. Salim, Mark Sanderson, Falk Scholer, and Damiano Spina. 2023. "Designing and Evaluating Presentation Strategies for Fact-Checked Content." In: *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management* (CIKM '23). Association for Computing Machinery, Birmingham, United Kingdom. DOI: 10.1145/3583780.3614841

[12] Lauren L. Saling, Devi Mallal, Falk Scholer, Russell Skelton, and **D. Spina**. 2021. "No One is Immune to Misinformation: An Investigation of Misinformation Sharing by Subscribers to a Fact-Checking Newsletter." *PLOS ONE*, 16, 8, 1–13

[13] **D. Spina**, Danula Hettiachchi, Anne Kruger, Devi Mallal, and Michelle Riedlinger. 2023. *Can I Get a Fact-Check? News, Misinformation and Fact-Checking*. https://youtu.be/FnKgyZjZatM

[14] Damiano Spina, Mark Sanderson, Daniel Angus, Gianluca Demartini, Dana Mckay, Lauren L. Saling, and Ryen W. White. June 2023. "Human-AI Cooperation to Tackle Misinformation and Polarization." *Commun. ACM*, 66, 7, (June 2023), 40–45. DOI: 10.1145/3588431

Day 3: Fairness and Diversity: Two Sides of the Same Coin?

[15] Enrique Amigó, Hui Fang, Stefano Mizzaro, and ChengXiang Zhai. 2018. "Are we on the Right Track? An Examination of Information Retrieval Methodologies." In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (SIGIR '18). Association for Computing Machinery, Ann Arbor, MI, USA, 997–1000. ISBN: 9781450356572. DOI: 10.1145/3209978.3210131

[16] Enrique Amigó, Damiano Spina, and Jorge Carrillo-de-Albornoz. 2018. "An Axiomatic Analysis of Diversity Evaluation Metrics: Introducing the Rank-Biased Utility Metric." In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (SIGIR '18). Association for Computing Machinery, Ann Arbor, MI, USA, 625–634. ISBN: 9781450356572. DOI: 10.1145/3209978.3210024

[17] Enrique Amigó, Stefano Mizzaro, and Damiano Spina. 2022. "Ranking Interruptus: When Truncated Rankings Are Better and How to Measure That." In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR '22). Association for Computing Machinery, Madrid, Spain, 588–598. ISBN: 9781450387323. DOI: 10.1145/3477495.3532051

[18] Michael D. Ekstrand, Anubrata Das, Robin Burke, and Fernando Diaz. 2022. "Fairness in Information Access Systems." *Foundations and Trends® in Information Retrieval*, 16, 1-2, 1–177. DOI: 10.1561/1500000079

[19] Djoerd Hiemstra. Jan. 2023. "Was Fairness in IR Discussed by Cooper and Robertson in the 1970's?" *SIGIR Forum*, 56, 2, Article 19, (Jan. 2023), 5 pages. DOI: 10.1145/3582900.3582924

[20] Johannes Kiesel, Damiano Spina, Henning Wachsmuth, and Benno Stein. 2021. "The Meant, the Said, and the Understood: Conversational Argument Search and Cognitive Biases." In: *CUI 2021 - 3rd Conference on Conversational User Interfaces* (CUI '21) Article 20. Association for Computing Machinery, Bilbao (online), Spain, 5 pages. ISBN: 9781450389983. DOI: 10.1145/3469595.3469615

[21] Sachin Pathiyan Cherumanal, Damiano Spina, Falk Scholer, and W. Bruce Croft. 2021. "Evaluating Fairness in Argument Retrieval." In: *Proceedings of CIKM'21*. DOI: https://doi.org/10.1145/3459637.3482099

[22] Sachin Pathiyan Cherumanal, Damiano Spina, Falk Scholer, and W. Bruce Croft. 2022. "RMIT at TREC 2021 Fair Ranking Track." In: *Proceedings of TREC 2021*. https://trec.nist.gov/pubs/trec30/papers/RMIT-IR-F.pdf

[23] Sachin Pathiyan Cherumnal, Marwah Alaofi, Reham Abdullah Altalhi, Elham Naghizade, Falk Scholer, and Damiano Spina. 2023. "RMIT CIDDA IR at the TREC 2022 Fair Ranking Track." In: *Proceedings of TREC 2022*

[24] Sachin Pathiyan Cherumanal, Kaixin Ji, Danula Hettiachchi, Johanne R. Trippas, Falk Scholer, and Damiano Spina. 2023. "RMIT_IR at the NTCIR-17 FairWeb-1 Task." In: *Proceedings of 17th Conference on Evaluation of Information Access Technologies* (NTCIR-17), 5 pages. DOI: 10.20736/0002001315

[25] Stefano Mizzaro. 1997. "Relevance: The Whole History." *Journal of the American Society for Information Science*, 48, 9, 810–832. DOI: 10.1002/(SICI)1097-4571(199709)48:9<810::AID-ASI6>3.0.CO;2-U

[26] Mark Sanderson. 2010. "Test Collection Based Evaluation of Information Retrieval Systems." *Foundations and Trends® in Information Retrieval*, 4, 4, 247–375. DOI: 10.1561/1500000009

Day 4: Characterizing Information Processing Activities with Physiological Signals from Multiple Wearable Devices

[27] Kaixin Ji. 2023. "Quantifying and Measuring Confirmation Bias in Information Retrieval Using Sensors." In: *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing* (UbiComp/ISWC '23 Adjunct). Association for Computing Machinery, <conf-loc>, <city>Cancun, Quintana Roo</city>, <country>Mexico</country>, </conf-loc>, 236–240. DOI: 10.1145/3594739.3610765

[28] Kaixin Ji, Damiano Spina, Danula Hettiachchi, Flora D. Salim, and Falk Scholer. 2023. "Examining the Impact of Uncontrolled Variables on Physiological Signals in User Studies for Information Processing Activities." In: *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR '23). ACM, Taipei, Taiwan. DOI: 10.1145/3539618.3591981

[29] Kaixin Ji, Damiano Spina, Danula Hettiachchi, Flora Dylis Salim, and Falk Scholer. 2023. "Towards Detecting Tonic Information Processing Activities with Physiological Data." In: *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing* (UbiComp/ISWC '23 Adjunct). DOI: 10.1145/3594739.3610679

[30] Kaixin Ji, Danula Hettiachchi, Flora D. Salim, Falk Scholer, and Damiano Spina. 2024. "Characterizing Information Seeking Processes with Multiple Physiological Signals." In: *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR '24). ACM, Washington, DC, USA. DOI: 10.1145/3626772.3657793

[31] Kaixin Ji, Sachin Pathiyan Cherumanal, Johanne R. Trippas, Danula Hettiachchi, Flora D. Salim, Falk Scholer, and Damiano Spina. 2024. *Towards Detecting and Mitigating Cognitive Bias in Spoken Conversational Search*. (2024). arXiv: 2405.12480 [cs.HC]