

Ethical and Social Considerations in Automatic Expert Identification and People Recommendation in Organizational Knowledge Management Systems

Ida Larsen-Ledet
t-idal@microsoft.com
Microsoft Research
Cambridge, United Kingdom

Bhaskar Mitra
bmitra@microsoft.com
Microsoft Research
Montreal, Canada

Siân Lindley
sianl@microsoft.com
Microsoft Research
Cambridge, United Kingdom

Abstract

Organizational knowledge bases are moving from passive archives to active entities in the flow of people's work. We are seeing machine learning used to enable systems that both collect and surface information as people are working, making it possible to bring out connections between people and content that were previously much less visible in order to automatically identify and highlight experts on a given topic. When these knowledge bases begin to actively bring attention to people and the content they work on, especially as that work is still ongoing, we run into important challenges at the intersection of work and the social. While such systems have the potential to make certain parts of people's work more productive or enjoyable, they may also introduce new workloads, for instance by putting people in the role of experts for others to reach out to. And these knowledge bases can also have profound social consequences by changing what parts of work are visible and, therefore, acknowledged. We pose a number of open questions that warrant attention and engagement across industry and academia. Addressing these questions is an essential step in ensuring that the future of work becomes a good future for those doing the work. With this position paper, we wish to enter into the cross-disciplinary discussion we believe is required to tackle the challenge of developing recommender systems that respect social values.

Keywords

Enterprise knowledge modeling; Bias and fairness; Representational and allocative harms

1 Introduction

Work in large organizations generates vast amounts of information that bring challenges to knowledge management and dissemination. Recent developments have seen machine learning (ML) successfully applied in the automatic construction of knowledge bases that aim to tackle these problems [33], and products like Viva Topics [20] are bringing these capabilities to workplaces. Viva Topics identifies organizational "topics", associates them with organization members and resources, and surfaces these details via office software. As an example, the topic "Accessibility" might be associated with a company's accessibility guidelines and the contact details of an in-house expert on accessible design, and then highlighted on relevant SharePoint pages. ML-powered systems like these enable automatic creation, sharing, and maintenance of organizational knowledge

and have the potential to (re-)shape and surface that knowledge in the flow of people's work [32]. Looking ahead, knowledge bases may become an increasingly (pro-)active part of work, by recommending documents or experts that are relevant to the task a person is currently working on. In doing so, knowledge recommendations will represent people and their work, thereby influencing the visibility [4, 16] of their efforts and, potentially, their opportunities for impact [18]. This means that, in addition to known fairness challenges of recommender systems [e.g., 5, 8, 9, 21, 22], we will need to address issues particular to organizational knowledge management systems and the ways knowledge bases mediate and represent work and, hence, people.

In this position paper, we outline a set of concerns regarding issues of representation in the deployment of recommender systems that surface people and their work within their own organizations. These concerns hint at sensitivities particular to the domain of organizational knowledge bases; these are sensitivities that we and others will need to develop for these kinds of systems. We wish to start a conversation with the community, in which these social and ethical concerns can be addressed in a cross-disciplinary setting.

2 Consequences of recommending people

The design of online recommender systems raises several concerns related to fairness and transparency [5, 8, 9, 21, 22]. The stakes are high when the subject of recommendations are people, especially when these systems fail to be meaningfully transparent about the context in which these people are recommended [17] and to provide effective recourse to people in case of problematic recommendations. A person may be recommended based on their self-declared attributes, preferences, and expertise—e.g., candidate recommendations for jobs [11] and online dating applications [35]—or their attributes and expertise may be automatically inferred from contents they have authored and other traces of their relevant activities—e.g., in an organizational setting [20]. The latter approach deserves additional scrutiny, not just with respect to the correctness of inferred expertise but also the implications of potential epistemic injustices [10] and how the system may influence work practices. Extensive critical analysis is necessary to identify these biases and harms in the context of organizational knowledge and expertise modeling that uses machine learning. In particular, we should ask ourselves the question: *What are relevant forms of "social and human impact" in the context of organizational recommender systems?* Here, we enumerate some of the risks that we already anticipate along with open questions to set off a discussion to help the community approach those risks.

Erasure. Erasure refers to practices and collective actions that render certain individuals and groups invisible [29], resulting in their marginalization and dehumanization (representational harm [6]). An organizational knowledge management system that fails to adequately recognize different types of expertise or the expertise of certain individuals or groups contributes to the erasure of their labor. Such erasure could limit a person's opportunities for recognition and, in the long run, career advancement and work satisfaction (allocative harm [6]). There are several mechanisms by which such erasure may actualize in ML-enabled organizational knowledge management systems. For instance, if expertise is inferred from authored content, the system may over-emphasize areas and types of work that produce more textual output, and may misattribute or misrepresent expertise when the person doing the work is not the one who documents it, or when the outputs of a person's work are not as easily mined or classified [4]. Erasure can also be caused by abstractions [13, 23] and demographic data gaps [27] that skew the system towards expertise indicators historically associated with the demographics dominating the training data (typically white able-bodied cis-gendered men [27]). This prompts us to suggest that, in addition to asking "what content and data is useful," we should be asking: "What content and data is needed to adequately represent expertise?"

Bias amplification. Biases and harms are not just consequences of system design but are further compounded by the organizational practices in which these systems are situated and how users interact with them. For example, if the system presents its recommendation as a static list of people ranked by their expertise on a topic, small differences in estimated expertise may translate into disparately different levels of exposure, as presentation bias causes people to over-prioritize those at the top of the list [7, 30, 34]. This could have a self-reinforcing effect, in which "top experts" have more interactions and encounter more opportunities that help them further develop their expertise, leading to a rich-gets-richer effect and "economics of superstars" [1, 28]. Conversely, being recognized as an expert on a specific topic may put additional labor onto a person, reducing the time they have to develop or diversify their expertise. To create systems that are resilient to distortions of visibility and work practices, we need to ask: *What distortions may occur in organizational knowledge bases and what will it mean to be resilient to them?* Beyond that, we ought to explore if ML-powered knowledge bases give us the potential to strive for more than resilience, such as for knowledge bases to actively combat distortion. For instance, *what is a good balance between providing tailored recommendations and surfacing content that diverges from historical patterns and expands the individual's outlook on organizational resources?*

Creating perverse incentives. If being recognized as an expert leads to increased opportunities and rewards, people might respond by adapting their practices to ensure visibility and acknowledgment of their work. This risks producing a work culture focused on performativity rather than enduring contributions and long-term value [19], and may eventually make the applied measures of expertise less meaningful [12]. For example, people may feel incentivized to produce outputs of a form that the system picks up on even though that may not be what makes the most sense in practical terms—one might imagine designers producing high-fidelity digital

prototypes in situations where a rough paper mock-up would be more suitable. These dynamics may also encourage and normalize uncompensated labor extraction from employees, for them to ensure that their work is seen by the system and, by extension, their organization. For knowledge bases to foster healthy dynamics, we need to challenge existing designs by asking: *What gaps does the current design introduce or reproduce in terms of expertise representation?* In order to examine this question in a given organizational context, we may begin by asking: *What kinds of expertise exist in the organization? What objects are these different kinds of expertise tied to?* And, by extension: *What forms of knowledge production are, indirectly, encouraged or discouraged by existing systems and practices?* Finally, with a more generative outlook [2]: *How can knowledge bases encourage the development and nurturing of diverse expertise?*

Knowledge production vs. knowledge extraction. As we have hinted at above, there is more at stake than fairness: Active knowledge bases not only *represent* extracted knowledge but also *shift* the way knowledge production happens, by mediating interactions between people and influencing work through suggestions and surfaced content. This underscores the need to understand local practices of knowing [25], so as to be deliberate in how the system will slot into and alter those practices. In most organizations, the identification of expertise is inherently a social and situated process [24], whereas classification systems tend to codify expertise into rigid structures that do not always exhibit the same contextual flexibility [4]. This leads us to ask how these systems can leave room for the situated sense-making and consensus building that helps people trust knowledge and expertise [24], through questions such as: *What would "sociality" mean in these kinds of systems? What will consensus building and trust look like if automated knowledge bases and recommender systems become key infrastructures and mediators (or what do we want them to look like)?* Furthermore, to ensure that the *extractive* process of constructing organizational knowledge bases does not adversely affect knowledge *production* within the organization, we also need to address how to make systems that respect people's autonomy [4]. This hints at questions about workplace monitoring and control of one's data (see below), but more directly related to knowledge production, we may ask: *What additional work does increased or altered visibility impose on people? [16] Who is work shifted from and onto, and do those who are doing the work also benefit (sufficiently) from it? [14]*

Organizational knowledge vs. control and organizational power. Our final critique on the application of machine learning for extracting and surfacing organizational knowledge is that, unless deliberately countered, these systems will reflect existing power structures, with the risk that existing power imbalances manifest in abuse such as workplace surveillance [31]. This prompts questions like the following: *What kinds of accountability does the visibility brought about by these systems impose on the people being represented through the system? In what ways might this accountability result in abuse?* One topic to address here is the relationship between control and the temporality of visibility: Organizational knowledge objects are not static. Documents and other content produced by employees have a life cycle, moving through stages from work in progress, to finished, to approved, to outdated, etc. The content picked up by an

automated knowledge base will often be work in progress and, as such, knowledge objects may be made available to others, or even actively distributed, before they are “ready,” thereby taking away the authors’ control over how they appear to colleagues through the objects of their work [15]. From a design perspective, it is worth asking: *How can automated knowledge bases be attentive to context such as timing?*

3 Beginning the conversation

These kinds of systems are still in their early stages, but as researchers and practitioners we have a responsibility to anticipate and mitigate risks like the ones outlined here, as early as possible [26]. Any meaningful mitigation strategy necessitates causal attribution of the sources of said risks—e.g., (i) the framing of the problem that the machine learning system should solve may itself encode assumptions, such as that expertise can be inferred from authored content, raising questions of construct validity [13], (ii) datasets collected under unjust social biases and power inequities may contribute to erasure and unfair outcomes, (iii) the choice of optimization objectives, such as mean performance, may overlook poor recommendation quality for minority groups of users, (iv) the models themselves, for example when they do not consider causality, may result in stereotyping, and (v) presentation bias in how the system displays the recommendation results, such as a static ranked list, may further amplify inequities in exposure. Here, we have only begun to speculate about potential issues, but even this demonstrates the urgency of more thoroughly identifying pitfalls. In other words, while predictive machine learning may, indeed, have much to offer organizational knowledge management, we must remain cognizant that no vision for the “future of work” (a seemingly pervasive phrase across industry, business, and academia,¹ to name a few areas) is complete without adequate emphasis on *the future of workers*, which in turn may emphasize the need for participatory design [3] and protection for worker rights and the labor movement. Some of the questions we have posed in the section above pertain primarily to one of the sources mentioned here, while others can be asked from multiple angles. For instance, being attentive to context (the final question posed) may be a question of what signals are included in the data or how they are used in the model; but it may also be a question of what we assume different contextual signals to mean or how we choose to design the user interface based on them. At a more political level we may pose the questions: *Who should influence and control the construction of organizational recommender systems at various stages, from design to adoption and use? Who owns the knowledge extracted from activities at work? What does it do to a person’s “ownership” of a process if that process is modelled and made replicable? What would be pros and cons, from an end-user perspective, of being given control over how one’s work data is interpreted and when it is used?* The questions are many, and we are only just beginning the conversation.

References

[1] Nina Baranchuk, Glenn MacDonald, and Jun Yang. 2011. The economics of super managers. *The Review of Financial Studies* 24, 10 (2011), 3321–3368.

¹See, e.g., Microsoft (www.microsoft.com/research/theme/future-of-work), Forbes (www.forbes.com/feature/future-of-work), OECD (www.oecd.org/future-of-work), and Harvard University (www.harvard.edu/the-future-of-work).

- [2] Michel Beaudouin-Lafon, Susanne Bødker, and Wendy E. Mackay. 2021. Generative Theories of Interaction. *ACM Trans. Comput.-Hum. Interact.* 28, 6, Article 45 (nov 2021), 54 pages. <https://doi.org/10.1145/3468505>
- [3] Susanne Bødker and Morten Kyng. 2018. Participatory Design That Matters—Facing the Big Issues. *ACM Trans. Comput.-Hum. Interact.* 25, 1, Article 4 (feb 2018), 31 pages. <https://doi.org/10.1145/3152421>
- [4] Geoffrey C. Bowker and Susan Leigh Star. 2000. *Sorting Things Out: Classification and Its Consequences*. The MIT Press.
- [5] Robin Burke, Michael D Ekstrand, Nava Tintarev, and Julita Vassileva. 2021. Preface to the special issue on fair, accountable, and transparent recommender systems. *User Modeling and User-Adapted Interaction* 31, 3 (2021), 371–375.
- [6] Kate Crawford. 2017. The trouble with bias. In *Conference on Neural Information Processing Systems, invited speaker*.
- [7] Fernando Diaz, Bhaskar Mitra, Michael D Ekstrand, Asia J Biega, and Ben Carterette. 2020. Evaluating stochastic rankings with expected exposure. In *Proceedings of the 29th ACM international conference on information & knowledge management*. 275–284.
- [8] Michael D Ekstrand, Anubrata Das, Robin Burke, Fernando Diaz, et al. 2022. Fairness in Information Access Systems. *Foundations and Trends® in Information Retrieval* 16, 1–2 (2022), 1–177.
- [9] Michael D Ekstrand, Mucun Tian, Ion Madrazo Azpiazua, Jennifer D Ekstrand, Oghenemaro Anuyah, David McNeill, and Maria Soledad Pera. 2018. All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness. In *Conference on fairness, accountability and transparency*. PMLR, 172–186.
- [10] Miranda Fricker. 2007. *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
- [11] Sahin Cem Geyik, Qi Guo, Bo Hu, Cagri Ozceglar, Ketan Thakkar, Xianren Wu, and Krishnaram Kenthapadi. 2018. Talent search and recommendation systems at LinkedIn: Practical challenges and lessons learned. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 1353–1354.
- [12] Charles AE Goodhart. 1984. Problems of monetary management: the UK experience. In *Monetary theory and practice*. Springer, 91–121.
- [13] Abigail Z Jacobs and Hanna Wallach. 2021. Measurement and fairness. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 375–385.
- [14] Airi Lampinen, Naja Holten Møller, Riyaz Sheikh, Asbjørn Ammitzbøll Flügge, Kristin Kaltenhäuser, and Baki Cakici. 2022. CSCW and Algorithmic Systems. In *Proceedings of 20th European Conference on Computer-Supported Cooperative Work*. European Society for Socially Embedded Technologies (EUSSET). https://doi.org/10.48340/ecscw2022_ws03
- [15] Ida Larsen-Ledet and Henrik Korsgaard. 2019. Territorial Functioning in Collaborative Writing: Fragmented Exchanges and Common Outcomes. *Computer Supported Cooperative Work (CSCW)* 28 (08 May 2019), 391–433. <https://doi.org/10.1007/s10606-019-09359-8>
- [16] Ida Larsen-Ledet and Siân Lindley. 2022. Ways of Seeing and Being Seen: People in the Algorithmic Knowledge Base. (June 2022). <https://www.microsoft.com/en-us/research/publication/ways-of-seeing-and-being-seen-people-in-the-algorithmic-knowledge-base/> Workshop on CSCW and Algorithmic Systems at the 20th European Conference on Computer-Supported Cooperative Work (ECSW 2022)..
- [17] Ruohan Li, Jianxiang Li, Bhaskar Mitra, Fernando Diaz, and Asia J Biega. 2022. Exposing Query Identification for Search Transparency. In *Proceedings of the ACM Web Conference 2022*. 3662–3672.
- [18] Siân Lindley, Denise Wilkins, and Britta Burlin. 2021. Actions and their Consequences? Implicit Interactions with Workplace Knowledge Bases.
- [19] Carol Linehan. 2011. *New Ways of Organizing Work: Developments, Perspectives and Experiences* (1st ed.). Routledge, Chapter 5: Flexible Work, Flexible Selves? The Impact of Changing Work Practices on Identity, 62–77.
- [20] Microsoft. 2022. Microsoft Viva Topics. <https://www.microsoft.com/en-gb/microsoft-viva/topics> Accessed: 01 August 2022.
- [21] Nicola Neophytou, Bhaskar Mitra, and Catherine Stinson. 2021. Revisiting Popularity and Demographic Biases in Recommender Evaluation and Effectiveness. *CoRR* abs/2110.08353 (2021). arXiv:2110.08353 <https://arxiv.org/abs/2110.08353>
- [22] Alexandra Olteanu, Jean Garcia-Gathright, Maarten de Rijke, Michael D Ekstrand, Adam Roegiest, Aldo Lipani, Alex Beutel, Ana Lucic, Ana-Andreea Stoica, et al. 2021. FACTS-IR: fairness, accountability, confidentiality, transparency, and safety in information retrieval. In *ACM SIGIR Forum*, Vol. 53. ACM New York, NY, USA, 20–43.
- [23] Cathy O’neil. 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway books.
- [24] Naomi Oreskes. 2021. *Why Trust Science?* Princeton University Press.
- [25] Wanda J. Orlikowski. 2002. Knowing in Practice: Enacting a Collective Capability in Distributed Organizing. *Organization Science* 13, 3 (2002), 249–273. <http://www.jstor.org/stable/3086020>
- [26] Nassim Parvin and Anne Pollock. 2020. Unintended by Design: On the Political Uses of “Unintended Consequences”. *Engaging Science, Technology, and Society* 6 (Aug. 2020), 320–327. <https://doi.org/10.17351/ests2020.497>

- [27] Caroline Criado Perez. 2019. *Invisible women: Data bias in a world designed for men*. Abrams.
- [28] Sherwin Rosen. 1981. The economics of superstars. *The American economic review* 71, 5 (1981), 845–858.
- [29] Parul Sehgal. 2016. Fighting 'Erasure.'. *New York Times Magazine* (2016).
- [30] Ashudeep Singh and Thorsten Joachims. 2018. Fairness of exposure in rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2219–2228.
- [31] Zephyr Teachout. 2022. The Boss Will See You Now. *The New York Review of Books* 69, 13 (2022).
- [32] Denise Wilkins, Siân Lindley, Max Meijer, Richard Banks, and Britta Burlin. 2020. Designing AI Systems That Make Organizational Knowledge Actionable. *Interactions* 27, 6 (nov 2020), 72–75. <https://doi.org/10.1145/3424690>
- [33] John Winn, Matteo Venanzi, Tom Minka, Ivan Korostelev, John Guiver, Elena Pochernina, Pavel Mishkov, Alex Spengler, Denise Wilkins, Siân Lindley, Richard Banks, Sam Webster, and Yordan Zaykov. 2021. Enterprise Alexandria: Online High-Precision Enterprise Knowledge Base Construction with Typed Entities. In *3rd Conference on Automated Knowledge Base Construction*. <https://openreview.net/forum?id=0xDHnrK6S5B>
- [34] Haolun Wu, Bhaskar Mitra, Chen Ma, Fernando Diaz, and Xue Liu. 2022. Joint Multisided Exposure Fairness for Recommendation (*SIGIR '22*). Association for Computing Machinery, New York, NY, USA, 703–714. <https://doi.org/10.1145/3477495.3532007>
- [35] Yong Zheng, Tanaya Dave, Neha Mishra, and Harshit Kumar. 2018. Fairness in reciprocal recommendations: a speed-dating study. In *Adjunct publication of the 26th conference on user modeling, adaptation and personalization*. 29–34.