nature machine intelligence

Perspective

https://doi.org/10.1038/s42256-023-00623-7

Introducing contextual transparency for automated decision systems

Received: 19 May 2022

Accepted: 1 February 2023

Published online: 13 March 2023

Check for updates

Mona Sloane[1], Ian René Solano-Kamaiko[1], Jun Yuan[2], Aritra Dasgupta[3] & Julia Stoyanovich[4]

As automated decision systems (ADS) get more deeply embedded into business processes worldwide, there is a growing need for practical ways to establish meaningful transparency. Here we argue that universally perfect transparency is impossible to achieve. We introduce the concept of contextual transparency as an approach that integrates social science, engineering and information design to help improve ADS transparency for specific professions, business processes and stakeholder groups. We demonstrate the applicability of the contextual transparency approach by using it for a well-established ADS transparency tool: nutritional labels that display specific information about an ADS. Empirically, it focuses on the profession of recruiting. Presenting data from an ongoing study about ADS use in recruiting alongside a typology of ADS nutritional labels, we suggest a nutritional label prototype for ADS-driven rankers such as LinkedIn Recruiter before closing with directions for future work.

ADS are increasingly adopted into standard business processes, particularly in service operations and product and service development related functions across all industries[1]. Consequently, they are having a transformative impact on society. At the same time, regulators are forced to mitigate the amplifying effect that ADS can have on inequity. This issue is particularly pressing for high-stakes contexts where ADS affect decisions on, for example, employment, education and criminal justice[2–7].

One mitigation strategy has become increasingly popular among policymakers in the European Union (EU) and in North America: ADS transparency. For example, the Canadian Directive on Automated Decision-Making and the recently introduced US Algorithmic Accountability Act call for algorithmic impact assessments and other forms of documentation related to the design and effects of ADS. Similarly, the yet-to-be enacted EU Artificial Intelligence (AI) Act attempts to apply transparency obligations proportionally within their predefined risk categorization framework. This legislation aligns with a White House Executive Order issued on May 2021 by the Biden administration, which included a section directing the National Institute of Standards and Technology to initiate a pilot programme "informed by existing consumer product labeling programs to educate the public on the security capabilities of Internet-of-Things (IoT) devices and software development practices"[8]. Relatedly, the White House Office of Science and Technology Policy recently introduced the Blueprint for an AI Bill of Rights which promotes 'notice and explanation' as one of its five core principles[9].

Increasingly, these regulations acknowledge the significance of context for creating meaningful ADS transparency. For example, the EU's risk categorization framework is based on the idea that ADS deployed in the same industry, but in different contexts, can create different types of risks. Additionally, the National Institute of Standards and Technology's AI Risk Management Framework Playbook explicitly grounds risk assessment and model management in mapping the context of an AI system, in part through stakeholder engagement[10].

Despite this trend, there are currently no systematic and scalable approaches for establishing ADS context and creating ADS transparency. This is particularly problematic when ADS are used by business corporations and enterprises, given that they are often hidden from public access and scrutiny, yet can impact the public at scale.

In the first section of this Perspective, we introduce the concept of contextual transparency as an approach that integrates social science, engineering and information design to help improve ADS transparency.

[1]Tandon School of Engineering, New York University, Brooklyn, NY, USA. [2]New Jersey Institute of Technology, Department of Informatics, University Heights, Newark, NJ, USA. [3]New Jersey Institute of Technology, Department of Data Science, University Heights, Newark, NJ, USA. [4]Tandon School of Engineering and Center for Data Science, New York University, New York City, NY, USA. e-mail: mona.sloane@nyu.edu

We argue that perfect transparency is impossible to achieve. Instead, contextual transparency is an operationalizable concept for creating ADS transparency that is meaningful for distinct professions, business processes and stakeholder groups. Subsequently, we apply contextual transparency to a well-established ADS transparency tool: 'nutritional labels' that display specific information about an ADS. Empirically, we focus on the profession of recruiting. We present data from an ongoing study about ADS use in recruiting alongside a typology of ADS nutritional labels to suggest a nutritional label prototype for ADS-driven rankers such as LinkedIn Recruiter before we close the paper with directions for future work.

Contextual transparency

Across disciplines, it is widely acknowledged that meaningfulness — which we define in the pragmatist tradition as the property of conveying information that is receivable and useful to a recipient, and that has consequences in that it makes a difference to practice[11,12] — and contextuality of information are deeply entangled[13–17]. This includes scholarship and practice on systems design. The arrival of context-aware computing applications (versus theories of context-aware systems), that is, applications that gather context data and adapt systems behaviour accordingly[18], prompted scholars in computer science and systems design to seriously consider the concept of 'context'. Quickly, the problem of computationally representing context became a technology design problem.

Well-established theories of context have underlined the open-ended nature of context[19], such as Lucy Suchman's seminal framework of 'situated action'[20]. There is general agreement that context is critically important for understanding activity and information[21]. Yet, a consensual definition of context has never been achieved[22]. Technology and design scholars and practitioners continue to grapple with the issue that context is dynamic and continually evolving[19], and that meaningfulness of context only emerges through forms of practice[21]. Some have even argued that there are some human aspects of context that simply cannot be inferred technologically[23].

Therefore, the 'context problem' is a salient issue for scalable and impactful ADS transparency. To address this problem, we introduce the concept of contextual transparency. Contextual transparency stipulates that there is no perfect transparency. It centres on the idea that information about an ADS must be meaningful so that people can appropriately interpret it in the context of their (professional) practice.

Contextual transparency connects ADS transparency design with the application domain and the transparency needs of key stakeholders and their professional practices. It is an approach that addresses social science deficit of ADS[24] and allows strategic determination of what specific information and type of transparency technique is most meaningful for enhancing interpretability for professionals and the business processes where they use ADS.

Interpretability is defined here as the flows of information that allow humans to understand the cause of an (ADS-generated) decision[25] and, if necessary, contest it[26–28]. This definition pushes beyond purely technical conceptions of interpretability[29] and is aligned with more human-centric approaches[21,25,30–34].

Contextual transparency is inspired by Helen Nissenbaum's seminal work on privacy in context[35], which introduces the notion of contextual integrity as a normative theory of privacy, personal information and digital technologies. Contextual integrity demands that information collection and information transmission must be context appropriate[36]. Similarly, we stipulate that ADS interpretability is dependent on context-specific information. Contextual transparency serves as a practical approach that operationalizes ADS interpretability via a scalable framework for determining what information is contextually relevant and for whom to design of a meaningful nutritional label.

To develop this framework, as a guiding principle, we follow the 'imperative of interpretable machines'[28] that centres three research questions for creating mechanisms that foster ADS interpretability and trust[37]: what are we explaining? To whom are we explaining and why? Are the explanations effective?[38] We connect these questions concretely to contextual transparency by introducing three contextual transparency principles (CTPs) that connect the domains of social science, engineering and design (Fig. 1). The three CTPs are: 'stakeholder specificity', 'ADS specificity' and 'transparency and outcome specificity'.

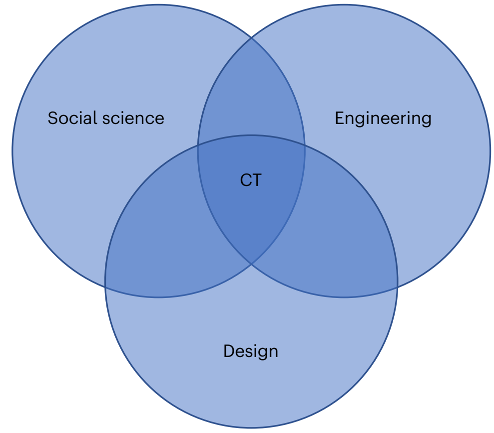

Fig. 1 | Contextual transparency Venn diagram. The convergence of social science, engineering and design in the concept of contextual transparency (CT).

To translate these CTPs into a workable method, we introduce the CTP matrix. This matrix breaks each CTP into a focus area and research question(s) that guide empirical data collection and analysis, as well as transparency design efforts (Fig. 2).

CTP 1: stakeholder specificity (social science)

CTP 1 is focused on defining a profession, relevant stakeholders within it and the professional practices into which ADS get embedded among these relevant stakeholders. Each profession and the various stakeholder groups within it integrate ADS into their professional practice in different ways. This difference must be considered when framing stakeholder specificity to inform ADS transparency interventions.

We suggest starting with existing classification systems for occupations (such as the US Occupational Information Network). These existing classification systems can help define the profession (for example, 'human resource specialist'). From there, researchers can use empirical social research methods (especially empirical qualitative work) to define the various stakeholder subgroups that are relevant for a domain of ADS use. For example, within the profession of human resource specialists, the sub-stakeholder groups of talent acquisition managers, recruiters, and sourcers may become relevant in the context of ADS deployed for screening potential job candidates.

Once general and specific stakeholders have been identified, researchers can deploy empirical research methods (quantitative, such as surveys, and qualitative, such as semi-structured interviews) to specify their professional practice vis-a-vis ADS use. For example, they can use these methods to determine how members of an occupation or community of practice use and interpret ADS, and what their transparency needs are.

CTP2: ADS specificity (engineering)

CTP 2 is focused on understanding the technical context of the ADS used by the relevant stakeholders. Different types of ADS operate with different assumptions, mechanisms and technical constraints, and successful disclosure strategies must map onto these. For example, a computer vision ADS that analyses the facial expression of job candidates to assess personality traits works based on assumptions, data

 https://doi.org/10.1038/s42256-023-00623-7

CTP 1: understanding stakeholders and professional practice (social science)			
Focus	Profession	Specific stakeholder group	Professional practice
Research questions	What is the profession and broader community of practice based on existing classification systems (e.g., the Standard Occupational Classification System [O*Net])?	What are the stakeholder sub-groups?	What is the professional practice of substakeholder groups? How do they use and interpret ADS? What are their transparency needs?

CTP 2: understanding the technical context of the ADS (engineering)				
Focus	General ADS mechanism	Specific ADS ('product')	Input features	Output description
Research questions	What is the underlying ADS mechanism?	What is the specific ADS this nutritional label design will be focused on?	What input features are used for the specific ADS?	How is the output of the ADS described and displayed?

CTP 3: creating contextual transparency (design)				
Focus	Reasonably expected change	Transparency goal	Transparency design	Evaluation notes
Research questions	What is the change in professional practice, and/or individual behaviour or belief, and/or policy that can reasonably be expected to occur after the introduction of the nutritional label?	What information should the nutritional label display so that it meaningfully connects to the professional practices of the stakeholder sub-groups within a profession and enables the desired change?	What transparency design techniques for the chosen nutritional label are most appropriate to achieve the transparency goal and the reasonably expected change?	How can the reasonably expected change be observed, assessed and/or measured?

Fig. 2 | The CTP matrix. Each CTP is broken into a focus area and research question(s) that guide empirical data collection, analysis and transparency design efforts.

and computational techniques that are very different from a ranking ADS that automatically ranks job candidate profiles. Therefore, we define ADS specificity in terms of the technical frame that underpins any given ADS (such as visual pattern analysis or algorithmic ranking).

To define ADS specificity, researchers should first work to identify the general system mechanisms that underpin relevant ADS. We treat an ADS as a system that encapsulates dynamic interactions between multiple datasets, multiple algorithmic components and people. The system view is appropriate as it accounts for the changing nature of these interactions. In the field, researchers may find that ADS are framed as 'products': complex ADS that are made up of modules of several algorithmic sub-systems that different product teams work on, but that users encounter and perceive as one system. What matters is how these modular elements inter-operate by taking specific features as input from heterogeneous sources and producing interpretable outputs for users of the ADS. Therefore, researchers should identify an ADS's input features and output descriptions. This information is key for outlining potential transparency design options and limitations, including considering if and how they can work in complex multi-step or personalized systems.

CTP 3: transparency and outcome specificity (design)

CTP 3 is focused on creating concrete ADS transparency designs. It helps define transparency goals as a function of the form (for example, text versus data visualization techniques) and intent (for example, policy inspection versus deliberative reasoning about decision biases) of information displays designed for transparency interventions. These interventions should be directly linked to measurable outcomes, such as a change in behaviour or perception, or in a policy that can subsequently be observed, documented and assessed.

Consequently, data from CTP 1 and CTP 2 must serve to define the change (behaviour, perception or policy) that can reasonably be expected to occur after the introduction of the transparency intervention. The next step is to determine which existing disclosure techniques are most appropriate to achieve that transparency goal vis-à-vis the ADS at hand. Following this step, research should focus on relevant and reasonable evaluation methods.

Contextual transparency for recruiting ADS

We apply the contextual transparency approach for ADS used in professional recruiting, talent acquisition and sourcing. These professions increasingly use ADS for various business practices, ranging from outreach to screening and assessment, and onboarding[39,40]. Additionally, recruiting ADS have come under increasing scrutiny by regulators[41,42].

We use data from an ongoing qualitative study on the professional practice of (ADS-driven) recruiting (N = 33 as of October 2022). This qualitative study (approved by an institutional review board) began in summer 2021 and uses snowball sampling to identify professional recruiters, sourcers and talent acquisition managers. For the semi-structured interviews, themes were derived from literature research and the research focus. Consequently, they follow three broad themes: the professional practice of recruiting, ADS use and transparency needs. Owing to data protection regulation, research participants from the European Union and the United Arab Emirates are excluded from the research.

From these data, we learn that for professional recruiters, the foundation for any search they conduct is job specifications. These job specifications are either existing descriptions that are re-used or updated, or entirely new specifications. From these specifications, recruiters craft job advertisements that are posted on the company's internal and/or external websites and on job platforms such as LinkedIn or Indeed. Job candidates apply to these positions and recruiters manually assess submitted resumes and conduct screening interviews where they typically focus on required skills, experience and 'culture fit'. If candidates are deemed viable for the job or have a generally desirable profile, they become part of the 'slate' — a pool of attractive candidates

https://doi.org/10.1038/s42256-023-00623-7

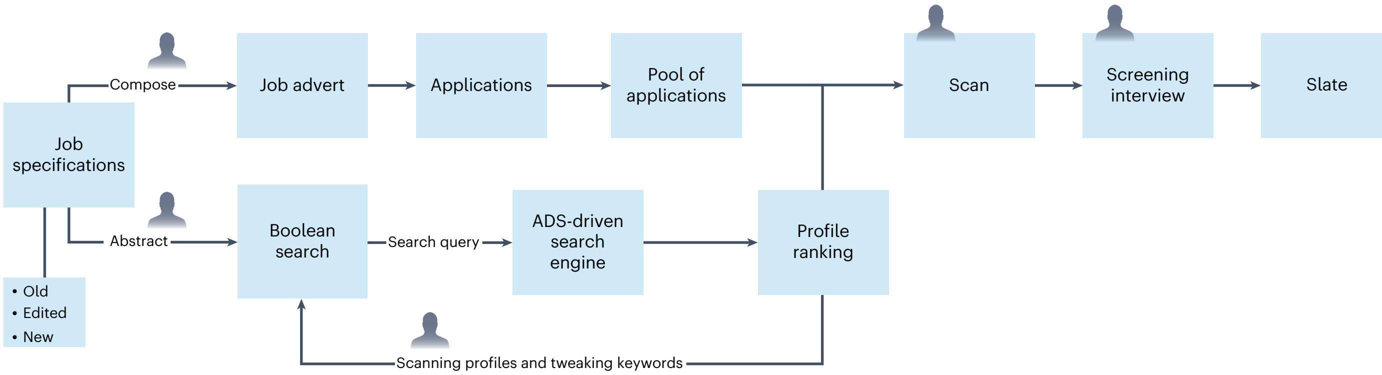

Fig. 3 | Generating the slate. An illustration of how recruiters use ADS in the process of finding job candidates.

that recruiters continually maintain. This process is illustrated in Fig. 3. It is dominant in situations where recruiters must source many candidates for the same type of role, such as internships or entry-level positions.

We also learn that for low-volume recruiting, that is, recruiting candidates with rare or sought-after skills and experience, sourcing and talent acquisition become relevant as distinct sub-professions that are focused on finding and activating relevant candidates, often those considered 'passive'. Sourcing is considered a meticulous process, as a result ADS can become useful tools for sifting through and finding relevant candidates.

> "As a recruiter, we oftentimes will need to actually source candidates, passive candidates. These are candidates who are not looking to make a move out of their current positions. And sourcing itself is kind of its own function within recruiting." – *recruiter, media and technology nonprofit*

> "Sourcing is very meticulous, like you really have to take the time." – *talent acquisition partner 1, corporate coaching company*

In the sourcing context, recruiters use job specifications to identify keywords for effective candidate searches in (ADS-driven) commercial candidate databases or platforms. The most popular product on the market is LinkedIn Recruiter, which allows recruiters to search across its vast user base. In 2021, LinkedIn reported 771.61 million users with a basic subscription worldwide, alongside 49.16 million users with a premium subscription[43].

In LinkedIn Recruiter, recruiters can use Boolean searches. Boolean searches allow combinations of words and phrases by using the terms AND, OR, NOT to refine a search input[44]. Boolean searches on ADS-driven candidate databases like LinkedIn Recruiter yield ranked profiles. These rankings are interpreted by recruiters and the desired candidate quality is assessed vis-à-vis the keywords that were used. Based on their assessments, recruiters tweak the Boolean searches until they have found the Boolean search strings that consistently yield satisfactory results. Boolean searches are seen as essential for increasing sourcing efficiency.

> "When I have a search string that's not pulling profiles for me, (...) it takes a lot longer. But when I find the right search string, it makes me get so excited about it because I'm able to find people and I'm able to find good people quickly." – *technical sourcer 1, data science company*

Our data also show that recruiters do not blindly trust ADS-derived rankings and typically double-check ranking outputs for accuracy, oftentimes going back and tweaking keywords. They are acutely aware of the difference in machine and human judgment.

> "(...) then I select all and contact [candidates] through like a message system. That's how I reach out to people on LinkedIn. I would not trust to do that without actually spending that time looking at the profiles first and making sure that they have hit the mark with what I'm looking for. (...) the computer system doesn't understand nuances that a recruiter, a human recruiter would." – *talent acquisition partner 2, corporate coaching company*

> "When you're at the beginning of your search, you just start that first page, top person, and you start going through if you're seeing that the first couple of pages aren't really producing what you're looking for. You might go back and switch up some of the filters or keywords that you've set and see if that produces better on the one or two pages." – *recruiting leader, staffing agency*

> "I don't know how [the ranking] is done. (...) I kept getting the same people. I don't know if it's like an algorithm they have or whatever, but I kept getting this like similar profiles that I was like, I don't need these profiles. And I actually went to the last page and sourced backwards so that I could find different profiles, which I did. It would be nice to figure out like why particular profiles are put up on the front in comparison to the ones that are on the last page, which are equally as good." – *technical sourcer 1, data science company*

Recruiters are concerned about bias in ADS-driven rankings and are aware that their search behaviour can exacerbate this bias.

> "I think that AI still has a way to go to avoid the bias. For example, if I'm immediately knocking out anybody on the search that comes up from AI with a name that looks like it's foreign (...) I am now teaching the system to not show me anybody with a foreign sounding name, so that inherently will create a bias in the system going forward." – *founder and CEO, recruiting firm*

Furthermore, recruiters have a specific transparency need with regards to ADS-driven rankings of candidates: they are keen to understand the functionality of the system, especially in terms of input features.

> "If I were to do the same LinkedIn search three times, I might get different results all three times. (...) It feels like it depends on how the algorithm is feeling based on that day. (...) I have no idea how or why people are dished to me and the way that they're served to me and what the relevancy is. That's a huge, huge gap in their product. (...) if I'm thinking about AI-driven selection, right, and thinking about what, was it about these, I gave you a hundred profiles, you returned these six. What was it about them that

Table 1 | CTP matrix filled in with empirical data on ADS-driven recruiting

	Focus	Findings/research questions
CTP 1: understanding stakeholders and professional practice (social science)	Profession	Human Resources
	Specific stakeholder group	Recruiters, talent acquisition managers, sourcers
	Professional practice	Finding and engaging passive and active candidates using ADS
CTP 2: understanding the technical context of the ADS (engineering)	General ADS mechanism	Learned rankings
	Specific ADS ('product')	LinkedIn Recruiter
	Input features	Candidate profiles, platform activity
	Output description	Ranked list of candidates
CPT 3: creating contextual transparency (design)	Reasonably expected change	ADS literacy among recruiters; improved diversity of candidate pools
	Transparency goal	Surface absence in the ranking and uncover 'hidden' candidates to create a more diverse pool
	Suitable disclosure techniques	Nutritional label for ADS functionality that shows general influential factors as well as keyword influence per Boolean search string and intra-ranking explanation of candidate position within the ranking; potentially add general diversity data per job title to facilitate ranking interpretation vis-à-vis demographic information
	Evaluation notes	Stakeholder interviews about potential change in use and perception of ADS; participant diaries documenting professional practice; A/B testing

made you select them? And as you narrowed, what were those criteria?" – *director of talent acquisition, data science company*

Recruiters see the lack of ADS transparency as challenging efforts to recruit for diversity. This can lead to various practices of manually inferring 'diverse' identity markers such as race and gender, for example from surnames or profile pictures.

"At my former company, [we] actually had requirements as far as the number of women and the number of ethnic diverse candidates that needed to be included in the interview slate. And if you just were not feeling like you were getting that just from the candidate pool you were given [by the software], you just really had to go out and identify additional candidates on your own through different methods of just diversity sourcing and reaching out to appropriate organizations to help you get those types of candidates." – *talent acquisition and HR process leader, data science company*

"I would say probably I still think that the majority of the diverse folks (...) I found through LinkedIn (...) just from literally sifting through hundreds of profiles and just quickly looking (...) at the name or looking at the profile. It can get tricky because sometimes they don't always have profile pictures available to the public." – *technical sourcer 2, data science company*

As Table 1 illustrates, we use these data to fill out the contextual transparency matrix and narrow down a recruiting-specific ADS disclosure technique.

Specifically, we follow domain-specific developments in ADS disclosure techniques (in this case, employment, hiring, recruiting), specifically[45], which outlines an approach for designing ADS nutritional labels for job seekers who are engaging with hiring ADS as a way for catering towards the transparency needs of a specific stakeholder group. Instead of job seekers, we are targeting recruiters who use ADS to make high-stakes decisions about job seekers.

Nutritional labels have historically been used in the food industry to relay ingredients and nutritional values to consumers[46,47]. First introduced in the US in the 1940s with disclosure of calories or sodium content, nutritional labels for food share the trajectory of labels used for tobacco products: they have gradually become more specific as scientific knowledge about the links between the consumption of a product, such as a cigarette, and its health impacts have expanded[46]. Generally, nutritional labels display specific, rather than exhaustive information in a standardized way. These labels are designed to be useful and interpretable for specific stakeholder groups, such as consumers or professionals.

Nutritional labels for ADS have recently emerged as a pragmatic design approach to address transparency needs in ADS. They are part of a wider push in the ADS research community towards transparency-by-design, an approach that sets out to include 'normative, relational and social factors' in the 'meaningful realization of transparency'[48] across all stages of the software development process[49,50]. Transparency-by-design seeks to address information asymmetries[48] and was inspired by privacy-by-design, which integrates and establishes privacy protections in (technology) design processes[51–53].

Here, privacy labels – akin to ADS nutritional labels – emerged as a standard design approach[54–57] for displaying privacy policies in an understandable way[58] or for giving users more agency over data collection and data brokerage, such as in the case of Apple's privacy labels[59,60]. Visualization techniques are key for transparency- and privacy-by-design[61,62]. In the context of ADS nutritional labels, they can provide insights into the training data, demographic overviews, model performance measures, feature relevance and impact, fairness measures, model stability highlights and ranking comparisons, amongst many others[63,64].

We develop a nutritional labels typology and identify three types of nutritional labels that currently exist for ADS: nutritional labels for datasets, nutritional labels for functionality, and nutritional labels for outcomes. To identify these types, we conducted a literature review on existing scholarship on ADS nutritional labels using a snowballing approach. This approach uses a selection of known key papers on the topic and then deploys forward and backward snowballing to map the topical field and construct the literature review[65,66]. It is suitable for nascent fields and fields that are interdisciplinary and where, consequently, the selection of relevant databases and search strings is challenging and can produce a lot of noise. The nutritional labels typology follows five key questions: What is the underlying assumption of the label? What is its overarching goal? What is the guiding question? What are the core elements? What are representative case examples? Table 2 illustrates the typology and provides an example for each type that is drawn from the current literature.

We note that the delineations between these three types are porous in both research and practice. However, the types are sufficiently

 https://doi.org/10.1038/s42256-023-00623-7

Table 2 | An overview of the nutritional label typology including a summary of the three nutritional label types

	Assumption	Goal	Guiding question	Elements	Examples
Nutritional labels for datasets	Datasets may not be fit to be used for a task	Assessing the fitness of a dataset for use in a particular task or family of tasks	Is this data fit for use in an ADS?	Standardized qualitative and quantitative measures focused on data provenance and content	Data Nutrition Project[87] Datasheets for Datasets[68] Responsible Data Management[88]
Nutritional labels for functionality	Interpretability is conditioned on transparency about the ADS design components	Achieving interpretability by describing the components of an ADS and how they are intended to function	What is the ADS designed to consider and how?	Documentation of an ADS' purpose, data, data transformation, and metadata used, model performance characteristics, safety and security aspects, ADS evaluation criteria and other potentially relevant factors	Model Cards for Model Reporting[64] Factsheets[89] Nutritional Labels for Data and Models[45,67,90]
Nutritional labels for outcomes	Interpretability is key for establishing trust in ADS and enhancing equity	Providing opportunities for equity-focused decision-making by describing how a system performed for a certain group, subgroup, and/or individual	What information on a certain group did the ADS consider and how?	Model performance measures, diversity of outcomes, fairness and model stability	Ranking Facts[63] Counterfactuals[91,92]

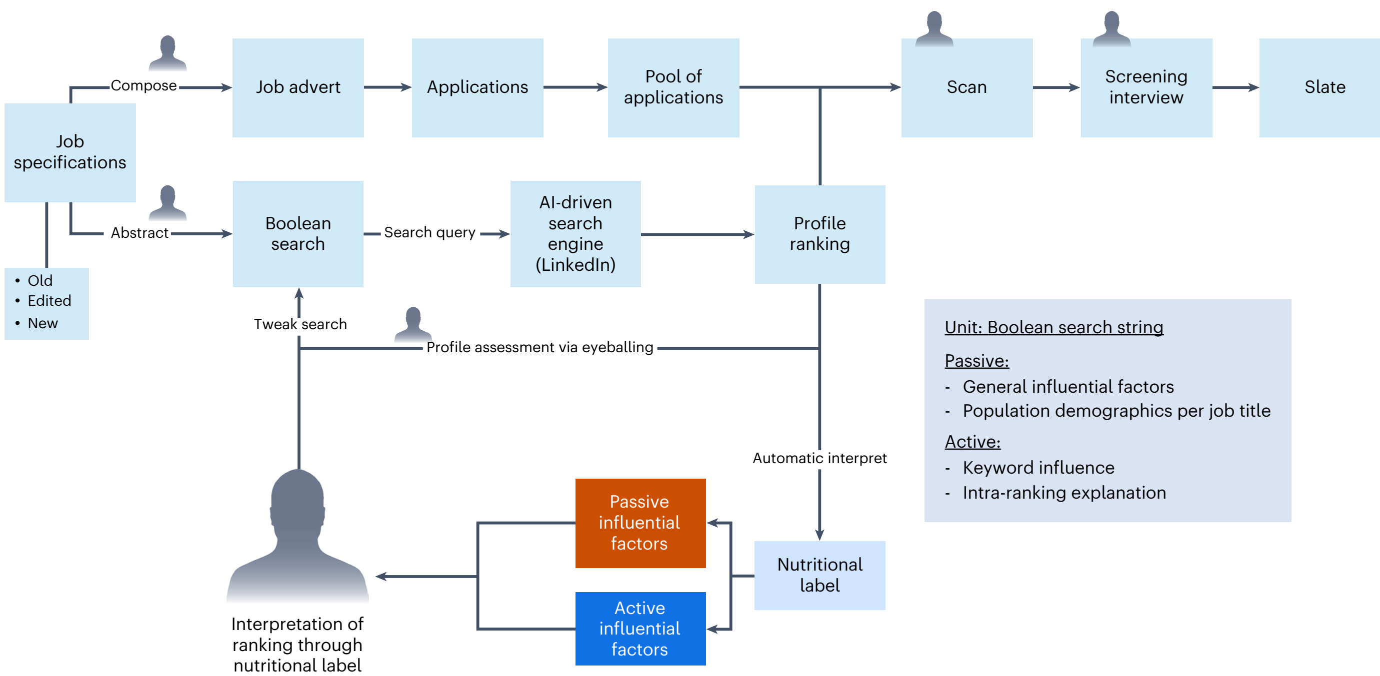

Fig. 4 | Nutritional label in slate generation. The location and design of the nutritional label in the ADS-driven recruiting process.

distinct for us to use this typology to assess which ones are most suited to our needs based on the CTP matrix. To meet the needs of the recruiters outlined in the qualitative study and the corresponding data from the CTP matrix, we suggest a combination of nutritional labels for functionality and outcomes. To ensure scalability of transparency while accounting for the flow of data as well as shifting contexts[67] we suggest that this label be automatically generated rather than hand-drawn[68], building on respective approaches that have been proposed for privacy labels[69,70].

Specifically, we build on our data to suggest that the nutritional label for functionality for ADS-driven rankings of candidates should focus on the unit of any given Boolean search string. To address the transparency needs of recruiters, we suggest that the nutritional label include passive and active factors. Passive factors comprise information that is relevant to the general functioning of the ADS and the professional practice of recruiting in general, while active factors comprise information that is specific to the Boolean search string and therefore changes. Based on the information displayed in the CTP matrix in Fig. 4, we suggest that the passive factors of the nutritional label comprise the general influential factors of ADS-driven rankers such as LinkedIn Recruiter, as well as population and diversity demographics pertaining to industry or profession to allow for a contextual interpretation of the ranking output vis-à-vis diversity goals. Furthermore, we suggest that active factors comprise information on the influence of the specific keywords in the Boolean search string, as well as information on the difference of position within the ranking. We illustrate this concept in Fig. 4, outlining where in the ADS-driven recruiting process the nutritional label could best be placed.

To evaluate whether this ADS transparency intervention did achieve the change that can reasonably be expected (Fig. 4), we suggest using stakeholder interviews about potential change in use and perception of ADS alongside participant diaries documenting professional practice and A/B testing (if possible).

These evaluation mechanisms are important steps for sounding out the limitations of contextual transparency. Although designed to empower professionals to make more equitable decisions in the context of ADS use, contextual transparency is no silver bullet for achieving systemic change. It is an approach that only works in tandem with wider changes in policy, professional practice and technology design. Within the contextual transparency framework, professionals are not 'human infrastructure'[71] that merely upholds the accountability of an ADS. To the contrary, contextual transparency acknowledges that it is impossible to satisfy all the tradeoffs that are necessary within any given design and instead allows professionals to understand these tradeoffs so that they can intervene with their expert judgment applied in context.

Conclusion

Rapidly creating ADS transparency that is responsive to context while remaining effective is important for putting guard rails around ADS use in professional practice. Contextual transparency can serve as a unified approach for providing meaningful ADS information for distinct professions, business processes, stakeholder groups and decisions. As we show through the case of recruiting ADS, the contextual transparency matrix and the ADS typology can serve as practical tools for creating context-specific ADS transparency interventions.

Future work involves operationalizing contextual transparency for an ADS that is already on the market and focus on measuring its effectiveness. New research should combine contextual transparency with work on explainability and transparency in ranking and recommender systems[72–74] as well as the existing body of work on privacy labels and standardizing privacy notices[54–60,62,70]. Relatedly, future work should compare the potential of the contextual transparency approach with the effectiveness of standardization efforts inside and outside of ADS regulation. This work should also engage with the dialogue on new regulations for ranking transparency, such as the EU guidelines on ranking transparency[75].

We also see a new line of research emerging that connects the concept of contextual transparency to the idea of distributed accountability within socio-technical systems[76,77] potentially providing a pathway for discussing issues around risk and liability in ADS.

Lastly, work on contextual transparency must be integrated with existing scholarships on the professions[78–81] as well as user studies on the efficiency of transparency and explainability interventions for algorithms and ADS[59,60,82] and their relationship to trust[31,83–86].

References

1. AI capabilities deployed in standard business processes 2020. *Statista* http://www.statista.com/statistics/1112998/ai-capabilities-standard-business-processes/ (2022).
2. Bailey, D. E. Emerging technologies at work: policy ideas to address negative consequences for work, workers, and society. *ILR Rev.* **75**, 527–551 (2022).
3. Ajunwa, I., Crawford, K. & Schultz, J. Limitless worker surveillance. Preprint at *SSRN* https://ssrn.com/abstract=2746211 (2016).
4. Kizilcec, R. F. & Lee, H. Algorithmic fairness in education. Preprint at https://doi.org/10.48550/arXiv.2007.05443 (2021).
5. Baker, R. S. & Hawn, A. Algorithmic bias in education. *Int. J. Artif. Intell. Educ.* **32**, 1052–1092 (2022).
6. Gipson Rankin, S. Technological tethereds: potential impact of untrustworthy artificial intelligence in criminal justice risk assessment instruments. Preprint at *SSRN* https://doi.org/10.2139/ssrn.3662761 (2020).
7. Angwin, J., Larson, J., Mattu, S. & Kirchner, L. Machine Bias. *ProPublica* (23 May 2016).
8. Executive Order on Improving the Nation's Cybersecurity. *The White House* https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/ (12 May 2021).
9. *Blueprint for an AI Bill of Rights*. *The White House* https://www.whitehouse.gov/ostp/ai-bill-of-rights/ (2022).
10. *AI RMF Playbook* https://pages.nist.gov/AIRMF/ (NIST, 2022).
11. Peirce, C. S. *Philosophical Writings of Peirce* (Dover Publications, 1955).
12. James, W. *Pragmatism, and Other Essays* (Washington Square Press, 1963).
13. de Lange, F. P., Heilbron, M. & Kok, P. How do expectations shape perception? *Trends Cognit. Sci.* **22**, 764–779 (2018).
14. Dretske, F. *Knowledge and the Flow of Information* (Basil Blackwell, 1981).
15. Glynn, D. In *Corpus Methods for Semantics* 7–38 (John Benjamins Publishing Company, 2014).
16. Otten, M., Seth, A. K. & Pinto, Y. A social Bayesian brain: How social knowledge can shape visual perception. *Brain Cognit.* **112**, 69–77 (2017).
17. Snyder, J. S., Schwiedrzik, C. M., Vitela, A. D. & Melloni, L. How previous experience shapes perception in different sensory modalities. *Front. Hum. Neurosci.* **9**, 594 (2015).
18. Baldauf, M., Dustdar, S. & Rosenberg, F. A survey on context-aware systems. *Int. J. Ad Hoc Ubiquitous Comput.* **2**, 263–277 (2007).
19. Greenberg, S. Context as a dynamic construct. *Hum. Comput. Interact.* **16**, 257–268 (2001).
20. Suchman, L. *Human-Machine Reconfigurations: Plans and Situated Actions* (Cambridge Univ. Press, 1987).
21. Dourish, P. What we talk about when we talk about context. *Pers. Ubiquit. Comput.* **8**, 19–30 (2004).
22. Bazire, M. & Brézillon, P. In *Modeling and Using Context* (eds. Dey, A., Kokinov, B., Leake, D. & Turner, R.) Vol. 3554, 29–40 (Springer, 2005).
23. Bellotti, V. & Edwards, K. Intelligibility and accountability: human considerations in context-aware systems. *Hum. Comput. Interact.* **16**, 193–212 (2001).
24. Sloane, M. & Moss, E. AI's social sciences deficit. *Nat. Mach. Intell.* **1**, 330–331 (2019).
25. Miller, T. Explanation in artificial intelligence: insights from the social sciences. Preprint at https://doi.org/10.48550/arXiv.1706.07269 (2018).
26. Hirsch, T., Merced, K., Narayanan, S., Imel, Z. E. & Atkins, D. C. Designing contestability: interaction design, machine learning, and mental health. in *Proceedings of the 2017 Conference on Designing Interactive Systems* 95–99 (Association for Computing Machinery, 2017); https://doi.org/10.1145/3064663.3064703
27. Lyons, H., Velloso, E. & Miller, T. Conceptualising contestability: perspectives on contesting algorithmic decisions. *Proc. ACM Hum. Comput. Interact.* **5**, 106:1–106:25 (2021).
28. Stoyanovich, J., Van Bavel, J. J. & West, T. V. The imperative of interpretable machines. *Nat. Mach. Intell.* **2**, 197–199 (2020).
29. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1**, 206–215 (2019).
30. Marcinkevičs, R. & Vogt, J. E. Interpretability and explainability: a machine learning zoo mini-tour. Preprint at https://doi.org/10.48550/arXiv.2012.01805 (2020).
31. Ribeiro, M. T., Singh, S. & Guestrin, C. 'Why should i trust you?': Explaining the predictions of any classifier. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1135–1144 (Association for Computing Machinery, 2016); https://doi.org/10.1145/2939672.2939778
32. Datta, A., Sen, S. & Zick, Y. Algorithmic transparency via quantitative input influence: theory and experiments with learning systems. In *2016 IEEE Symposium on Security and Privacy* 598–617 (2016); https://doi.org/10.1109/SP.2016.42
33. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. in *Advances in Neural Information Processing Systems* Vol. 30 (Curran Associates, 2017).

34. Weld, D. S. & Bansal, G. The challenge of crafting intelligible intelligence. *Commun. ACM* **62**, 70–79 (2019).
35. Nissenbaum, H. *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford Univ. Press, 2009); https://doi.org/10.1515/9780804772891
36. Zimmer, M. Addressing conceptual gaps in big data research ethics: an application of contextual integrity. *Social Media Soc.* **4**, 2056305118768300 (2018).
37. Jacovi, A., Marasović, A., Miller, T. & Goldberg, Y. Formalizing trust in artificial intelligence: prerequisites, causes and goals of human trust in AI. in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* 624–635 (Association for Computing Machinery, 2021); https://doi.org/10.1145/3442188.3445923
38. Celino, I. Who is this explanation for? Human intelligence and knowledge graphs for explainable AI. Preprint at https://doi.org/10.48550/arXiv.2005.13275 (2020)
39. Ajunwa, I. An auditing imperative for automated hiring. Preprint at *SSRN* https://doi.org/10.2139/ssrn.3437631 (2019).
40. Black, J. S. & van Esch, P. AI-enabled recruiting: what is it and how should a manager use it? *Bus. Horiz.* **63**, 215–226 (2020).
41. Cumbo, L. A. *NYC Local Law 144 of 2021* (The New York City Council, 2021).
42. *Artificial Intelligence Video Interview Act* (Illinois General Assembly, 2020).
43. Statista. Number of LinkedIn users worldwide from 2019 to 2021, by subscription type. *Statista* http://www.statista.com/statistics/1335947/linkedin-global-users-by-subscription/ (2022).
44. Burns, S. What is Boolean search? *The New York Public Library* https://www.nypl.org/blog/2011/02/22/what-boolean-search (2011).
45. Stoyanovich, J. Hiring and AI: let job candidates know why they were rejected. *WSJ* https://www.wsj.com/articles/hiring-job-candidates-ai-11632244313 (2021).
46. Wartella, E. A., Lichtenstein, A. H. & Boon, C. S. *History of Nutrition Labeling. Front-of-Package Nutrition Rating Systems and Symbols: Phase I Report* (National Academies Press, 2010).
47. Guidance for Industry: Food Labeling Guide. *US Food and Drug Administration* https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-industry-food-labeling-guide (2013).
48. Felzmann, H., Fosch-Villaronga, E., Lutz, C. & Tamò-Larrieux, A. Towards transparency by design for artificial intelligence. *Sci. Eng. Ethics* **26**, 3333–3361 (2020).
49. Kay, J., Kuflik, T. & Rovatsos, M. Transparency by design (Dagstuhl seminar 21231). *Dagstuhl Rep.* **11**, 1–22 (2021).
50. Zieglmeier, V. & Pretschner, A. Trustworthy transparency by design. Preprint at https://doi.org/10.48550/arXiv.2103.10769 (2021).
51. Cavoukian, A. *The 7 Foundational Principles* (Information and Privacy Commissioner of Ontario, 2009).
52. Pattakou, A., Mavroeidi, A.-G., Diamantopoulou, V., Kalloniatis, C. & Gritzalis, S. Towards the design of usable privacy by design methodologies. In *2018 IEEE 5th International Workshop on Evolving Security & Privacy Requirements Engineering (ESPRE)* (2018); https://doi.org/10.1109/ESPRE.2018.00007
53. Romanou, A. The necessity of the implementation of privacy by design in sectors where data protection concerns arise. *Comput. Law Secur. Rev.* **34**, 99–110 (2018).
54. Emami-Naeini, P., Agarwal, Y., Faith Cranor, L. & Hibshi, H. Ask the experts: what should be on an IoT privacy and security label? In *2020 IEEE Symposium on Security and Privacy* 447–464 (2020); https://doi.org/10.1109/SP40000.2020.00043
55. Johansen, J. et al. A multidisciplinary definition of privacy labels. *Inf. Comput. Secur.* **30**, 452–469 (2022).
56. Kelley, P. G., Cesca, L., Bresee, J. & Cranor, L. F. Standardizing privacy notices: an online study of the nutrition label approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 1573–1582 (Association for Computing Machinery, 2010); https://doi.org/10.1145/1753326.1753561
57. Shen, Y. & Vervier, P.-A. in *Privacy Technologies and Policy* (eds. Naldi, M. et al.) 136–147 (Springer, 2019); https://doi.org/10.1007/978-3-030-21752-5_9
58. Kelley, P. G., Bresee, J., Cranor, L. F. & Reeder, R. W. A 'nutrition label' for privacy. in *Proceedings of the 5th Symposium on Usable Privacy and Security* (Association for Computing Machinery, 2009); https://doi.org/10.1145/1572532.1572538
59. Kollnig, K., Shuba, A., Van Kleek, M., Binns, R. & Shadbolt, N. Goodbye tracking? Impact of iOS app tracking transparency and privacy labels. In *2022 ACM Conference on Fairness, Accountability, and Transparency* 508–520 (Association for Computing Machinery, 2022); https://doi.org/10.1145/3531146.3533116
60. Scoccia, G. L., Autili, M., Stilo, G. & Inverardi, P. An empirical study of privacy labels on the Apple iOS mobile app store. In *Proceedings of the 9th IEEE/ACM International Conference on Mobile Software Engineering and Systems* 114–124 (Association for Computing Machinery, 2022); https://doi.org/10.1145/3524613.3527813
61. Mascharka, D., Tran, P., Soklaski, R. & Majumdar, A. Transparency by design: closing the gap between performance and interpretability in visual reasoning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* 4942–4950 (2018). https://doi.org/10.1109/CVPR.2018.00519
62. Bargh, M., van de Mosselaar, M., Rutten, P. & Choenni, S. On using privacy labels for visualizing the privacy practice of SMEs: challenges and research directions. In *DG.O 2022: The 23rd Annual International Conference on Digital Government Research* 166–175 (Association for Computing Machinery, 2022).
63. Yang, K. et al. A nutritional label for rankings. In *Proceedings of the 2018 International Conference on Management of Data* 1773–1776 (Association for Computing Machinery, 2018); https://doi.org/10.1145/3183713.3193568
64. Mitchell, M. et al. Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* 220–229 (Association for Computing Machinery, 2019); https://doi.org/10.1145/3287560.3287596
65. Wohlin, C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering* (Association for Computing Machinery, 2014); https://doi.org/10.1145/2601248.2601268
66. Wnuk, K. & Garrepalli, T. Knowledge management in software testing: a systematic snowball literature review. *E-Informatica Softw. Eng. J.* **12**, 51–78 (2018).
67. Stoyanovich, J. & Howe, B. Nutritional labels for data and models. In *A Quarterly bulletin of the Computer Society of the IEEE Technical Committee on Data Engineering* (2019).
68. Gebru, T. et al. Datasheets for datasets. *Commun. ACM* **64**, 86–92 (2021).
69. Harkous, H. et al. Polisis: automated analysis and presentation of privacy policies using deep learning. Preprint at https://arxiv.org/abs/1802.02561 (2018).
70. Fang, L. & LeFevre, K. Privacy wizards for social networking sites. In *Proceedings of the 19th International Conference on World Wide Web* 351–360 (Association for Computing Machinery, 2010); https://doi.org/10.1145/1772690.1772727
71. Mateescu, A. & Elish, M. C. AI in context. *Data & Society* https://datasociety.net/library/ai-in-context/ (2019).

72. Volokhin, S., Collins, M., Rokhlenko, O. & Agichtein, E. Generating and validating contextually relevant justifications for conversational recommendation. In *ACM SIGIR Conference on Human Information Interaction and Retrieval* 284–289 (ACM, 2022); https://doi.org/10.1145/3498366.3505789
73. Balog, K. & Radlinski, F. Measuring recommendation explanation quality: the conflicting goals of explanations. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* 329–338 (Association for Computing Machinery, 2020); https://doi.org/10.1145/3397271.3401032
74. Sinha, R. & Swearingen, K. The role of transparency in recommender systems. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems* 830–831 (Association for Computing Machinery, 2002); https://doi.org/10.1145/506443.506619
75. *2020/C 424/01* (European Union, 2020).
76. Bathaee, Y. The artificial intelligence black box and the failure of intent and causation. *Harvard J. Law Technol.* **31**, 889–938 (2018).
77. Wagner, B. Liable, but not in control? Ensuring meaningful human agency in automated decision-making systems. *Policy Internet* **11**, 104–122 (2019).
78. Susskind, R. E. & Susskind, D. *The Future of the Professions: How Technology Will Transform the Work of Human Experts* (Oxford Univ. Press, 2015).
79. Klegon, D. The sociology of professions: an emerging perspective. *Sociol. Work Occup.* **5**, 259–283 (1978).
80. Abbott, A. *The System of Professions: An Essay on The Division of Expert Labor* (Univ. Chicago Press, 1988).
81. Wenger, E. *Communities of Practice: Learning, Meaning, and Identity* (Cambridge Univ. Press, 1999).
82. Kartikeya, A. in *Intelligent Computing* (ed. Arai, K.) 353–358 (Springer, 2022); https://doi.org/10.1007/978-3-031-10464-0_23
83. Schmidt, P., Biessmann, F. & Teubner, T. Transparency and trust in artificial intelligence systems. *J. Decis. Syst.* **29**, 260–278 (2020).
84. Elia, J. Transparency rights, technology, and trust. *Ethics Inf. Technol.* **11**, 145–153 (2009).
85. Ashwin, Agnew, W., Pajaro, U., Jethwani, H. & Subramonian, A. Rebuilding trust: queer in AI approach to artificial intelligence risk management. Preprint at https://doi.org/10.48550/arXiv.2110.09271 (2022).
86. LaRosa, E. & Danks, D. Impacts on trust of healthcare AI. in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* 210–215 (Association for Computing Machinery, 2018); https://doi.org/10.1145/3278721.3278771
87. Chmielinski, K. S. et al. The Dataset Nutrition Label (2nd Gen): Leveraging Context to Mitigate Harms in Artificial Intelligence. Preprint at https://arxiv.org/abs/2201.03954 (2022).
88. Stoyanovich, J., Abiteboul, S., Howe, B., Jagadish, H. V. & Schelter, S. Responsible data management. *Commun. ACM* **65**, 64–74 (2022).
89. Arnold, M. et al. FactSheets: increasing trust in AI services through supplier's declarations of conformity. Preprint at https://doi.org/10.48550/arXiv.1808.07261 (2019).
90. Sun, C., Asudeh, A., Jagadish, H. V., Howe, B. & Stoyanovich, J. MithraLabel: flexible dataset nutritional labels for responsible data science. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* 2893–2896 (Association for Computing Machinery, 2019); https://doi.org/10.1145/3357384.3357853
91. Wachter, S., Mittelstadt, B. & Russell, C. Counterfactual explanations without opening the black box: automated decisions and the GDPR. Preprint at https://doi.org/10.48550/arXiv.1711.00399 (2018)
92. Byrne, R. M. J. Counterfactuals in explainable artificial intelligence (XAI): evidence from human reasoning. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence* 6276–6282 (2019).

Acknowledgements
This research was supported in part by National Science Foundation awards 1916505, 1922658 and 1928627.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence should be addressed to Mona Sloane.

Peer review information *Nature Machine Intelligence* thanks Aurelia Tamo-Larrieux and Silvia Milano for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.