Understanding User Perceptions of Trustworthiness in E-recruitment Systems

Department: Head
Editor: Name, xxxx@email

Understanding User Perceptions of Trustworthiness in E-recruitment Systems

Gideon Ogunniye
Benedicte Legastelois
Michael Rovatsos
School of Informatics,
University of Edinburgh,
Edinburgh, UK.

Liz Dowthwaite
Virginia Portillo
Elvira Perez Vallejos
Horizon Digital Economy Research Institute, University of Nottingham,
Nottingham, UK.

Jun Zhao
Marina Jirotka
Department of Computer Science,
University of Oxford, Oxford, UK.

***Abstract*—Algorithmic systems are increasingly deployed to make decisions that people used to make. Perceptions of these systems can significantly influence their adoption, yet, broadly speaking, users' understanding of the internal working of these systems is limited. To explore users' perceptions of algorithmic systems, we developed a prototype e-recruitment system called Algorithm Playground where we offer the users a look behind the scenes of such systems, and provide "how" and "why" explanations on how job applicants are ranked by their algorithms. Using an online study with 110 participants, we measured perceived fairness, transparency and trustworthiness of e-recruitment systems. Our results show that user understanding of the data**

and reasoning behind candidates' rankings and selection evoked some positive attitudes as participants rated our platform to be fairer, more reliable, transparent and trustworthy than the e-recruitment systems they have used in the past.

1. INTRODUCTION Algorithmic systems have started to penetrate almost every sector of society, disrupting markets, labour forces and institutions. The dynamics of these systems have an enormous impact on the social, political and economic life of people across the globe. Yet, broadly speaking, users' understanding of the internal workings of these systems is limited, and consequently, many of them are beginning to express concerns about their fairness, transparency and trustworthiness [4], [5], [13].

In the academic literature, factors such as fairness [13], accuracy [4], transparency [5], and trustworthiness [10] have been highlighted as key in the analysis of such algorithmic systems. For example, according to Silva and Kenney [9], social bias has long been recognised in algorithmic decision making. Barocas et al (as cited in [9]) observed that software developers are not well versed in issues such as civil rights and fairness. Herlocker et al [3] noted that many recommender systems lack transparency in terms of the recommendation process and result generation. Webb and Patel [12] observed that algorithmic processes that filter and personalise the content seen by users may lead to detrimental outcomes such as reinforcement of societal biases and gender or ethnic discrimination among others. Wang et al [14] reported that people rate an algorithm as more fair when the algorithm predicts in their favor, even compensating for the negative effects of algorithms that are biased against particular demographic groups.

In the area of e-recruitment, advances in technology and globalisation have greatly increased the global impact of the use of algorithmic systems. More and more companies rely on e-recruitment systems to recruit qualified personnel, partly because they are cheaper than classical offline recruiting and partly due to competitive pressure.

An e-recruitment system is a web-based tool that automates recruiting processes such as job publishing, resume/CV submission, job applicants' pre-screening and selection. We observe that traditional e-recruitment systems have focused more on making the front-end processes (such as posting of job adverts and filling of online job applications) transparent, while the back-end processes (such as reasoning and data behind candidates' rankings and selection) are less transparent. We argue that such decision-making processes of e-recruitment systems must be made transparent for users in order for them to trust the process. Such transparency in human-computer interactions is important to facilitate people's trust in, and attitudes toward, algorithmic decision-making [6]. Thielsch et al [11] show that the perceptions of procedural justice in algorithmic e-recruitment processes can influence applicants' satisfaction.

, .

In this paper, we investigate users' perceptions of fairness, transparency, and trustworthiness of e-recruitment systems to gain a better understanding of the expectations everyday users have towards these increasingly important systems. Fairness in the context of algorithmic systems means that algorithmic decisions should not create discriminatory or unjust consequences. Transparency means that users have sufficient understanding of the data and reasoning behind algorithmic decision making. We follow [15] to connect these key concepts to trust in algorithmic systems.

While previous work [5], [6], [10], [13], have shown the importance of these perceptions in rating algorithmic systems, they do not show how the perceptions are connected to the assessment of the inputs of the decision-making process, the processing steps, and the decisions of algorithmic systems. In our study, we introduced a new form of users' engagement with online platforms by implementing new types of user interfaces that allowed users to assess the input data and how the data is used in the decision-making process of an algorithmic system, and we showed the impact of such assessment on users' perceptions of the systems. Our study focuses on the following questions:

- What are users' perceptions of fairness, reliability, transparency, and trustworthiness in e-recruitment systems?
- What are the possible trade-offs that need to be considered when attempting to increase users' trust in e-recruitment algorithms?

To address these questions, we developed a prototypical e-recruitment system that allows users to look "under the hood" of the workings of such real-world tools, focusing specifically on explaining how job applicants are ranked by the algorithms used. Using this tool, we conducted an online user study to compare our prototype with traditional systems. Participants were presented with a scenario consisting of i) a description of a fictional job advert, ii) fictional profiles of three applicants, iii) three different ranking algorithms varied by the criteria they used to rank the applicants, and iv) explanations on the criteria and algorithmic outcomes. To fully understand the factors that impact users' perceptions of an e-recruitment system, participants were asked to rate i) the fairness of the presented algorithms based on the outcomes they saw and their expectations of what they think the outcomes should be as shown in Figure 3; ii) their understanding of the internal workings of the algorithms and iii) the impact of the provided explanations on their perceptions. In addition to this, participants were asked to rate our prototype system vis-à-vis the e-recruitment systems they have used in the past.

The main contributions of our article are:

- We developed a prototypical e-recruitment system that offers more explanations on the internal workings of candidate ranking algorithms when compared with traditional e-recruitment systems.
- Using the prototypical system and a mixed-methods research design, we collected a dataset of 110 participants comparing our system with the traditional e-recruitment systems they have used in the past.
- We performed statistical analysis on this dataset to understand participants' perceptions of our system in comparison to the e-recruitment systems they have used in the past.
- Our results provide insight into the interface design elements that can positively impact users' perceptions of trust in an e-recruitment system such as users' expectations regarding algorithmic outcomes, ease of use and users' control of the use of their personal data. Our results also suggest that there is a strong correlation between satisfied expectation, transparency, and trustworthiness. However, we detect a moderate correlation between fairness and trustworthiness.

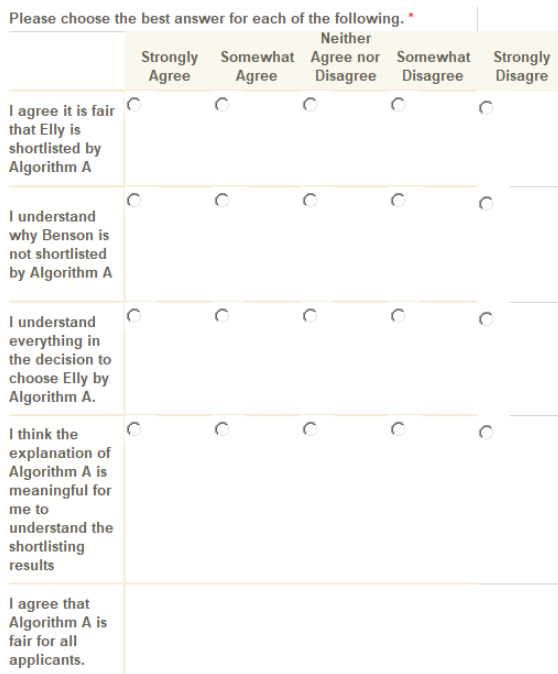

Figure 3. Exemplar statements to investigate fairness ratings.

2. METHOD

We conducted an online study in August 2020 that consisted of an online quantitative survey to understand users' perceptions of trust, fairness and transparency of e-recruitment systems, and a qualitative survey with open questions to gain insights into the impact of explanation on these perceptions.

The study is divided into three parts: pre-study questionnaire, platform engagement, and post-study questionnaire. After consenting and affirming that they were over the age of 18, participants were presented with the pre-study questionnaire. In this part, the participants assumed the roles of users of e-recruitment systems (job-seekers or recruiters). The participants completed an online survey about their perceptions of e-recruitment systems they have used in the past. After completing the pre-study questionnaire, participants were directed to our prototype e-recruitment system. The implemented system called *Algorithm Playground*[12] offers participants a look behind the scenes of a presumed e-recruitment system and provides textual explanations on how job applicants are ranked by the algorithms.

As shown **in Figure 1**, participants engage with the platform by searching for a specific job position using the provided search keywords. The search algorithm provides a list of job adverts related to the search. The participants are then expected to select a job advert, read the job description, and choose a fictional applicant of their choice that is suitable for the advertised job position. As shown in **Figure 2**, the candidate ranking algorithms return the rankings of the fictional applicants for the chosen job position together with explanations that are provided on the rankings. We presented three candidate ranking algorithms that use different parameters to rank the fictional applicants in our scenario. For instance, *Algorithm A* as shown in Figure 2 gives highest priority to candidates' job experience followed by their educational qualifications. The system is functional and derives its recommendations from a real database using algorithms.

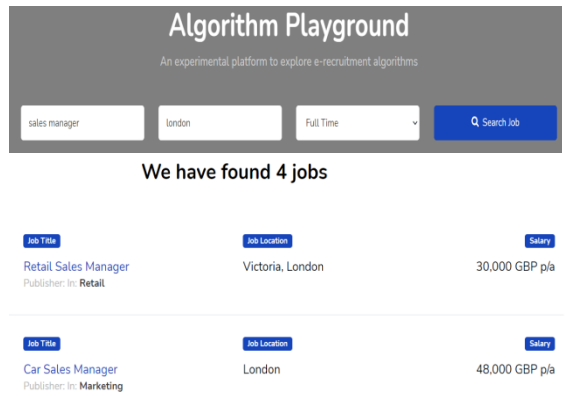

Figure 1: Our Prototype E-recruitment Platform which allows users to engage e-recruitment algorithms.

Figure 2. Results and Explanations of Results

We used a within-subject survey (pre-study/post-study questionnaire) to see if participants' perceptions of e-recruitment systems change after interacting with our explanation-enhanced prototype e-recruitment system.

After completing the tasks on our prototype platform, participants were presented with a post-study questionnaire. Some of the post-study questionnaire statements were paired with the pre-study questionnaire statements and some additional questions on the post-study questionnaire were

[1] https://psandbox.pythonanywhere.com/

[2] https://github.com/ReEnTrust/Algorithm-Playground-2

included to capture participants' perceptions of our prototype system.

Table 1. Exemplar statements from our quantitative study regarding each perception feature.

Pre-study/Post-study Questionnaire Statements	
PRE1/POST1	I don't trust e-recruitment websites that ask for personal data without explanation.
PRE2/POST2	I am more likely to trust e-recruitment websites that are easy to use.
PRE2/POST3	I am more likely to trust e-recruitment websites that I have used before.
PRE4/POST4	In general, when using e-recruitment websites, I trust that the search results will be reliable.
PRE5/POST5	In general, when using e-recruitment websites, I trust that I will get the best results for me.
PRE6/POST6	My trust in an e-recruitment website is affected by the algorithms that it uses.
PRE7/POST7	E-recruitment websites that give me control over my data are more trustworthy.
e-Rec1	E-recruitment platforms generally perform as I expect them to.
e-Rec2	E-recruitment platforms generally have access to the user information needed to generate reliable results.
e-Rec3	E-recruitment platforms are generally fair in ranking job applicants.
e-Rec4	In general, e-recruitment platforms are open and transparent in handling users' needs.
e-Rec5	Overall, most e-recruitment platforms are trustworthy.
Plat1	The platform generally performs as I expect it to.
Plat2	The platform has access to the user information needed to generate reliable results.
Plat3	The platform is fair in ranking job applicants.
Plat4	The platform is open and transparent in handling users' needs.
Plat5	Overall, the platform is trustworthy.

As shown in **Table 1**, the *PRE* statements and *e-Rec* statements were part of the pre-study questionnaire, while *POST* statements and *Plat* statements were part of the post-study questionnaire. We used the *e-Rec/Plat* statements to examine participants' perceptions of transparency, fairness and trustworthiness in traditional e-recruitment systems in comparison to our prototype e-recruitment system. Except for a few qualitative open-response questions, we asked participants to rate statements related to trust, fairness and transparency on 5-point Likert-type scales (i.e., strongly agree to strongly disagree). Other *PRE/POST* survey statements to examine participants' perceptions of trust in e-recruitment systems in comparison to our prototype e-recruitment system were as shown in **Section 4**.

To examine whether explanations enhance the understanding of an e-recruitment system, and consequently help the users to trust the system better, we asked the questions (*Plat6 to Plat11*) presented in **Table 3**. In addition to this, on our platform, we asked an open-response question: "Do you think an understanding of what a shortlisting algorithm does can increase or decrease your trust in an e-recruitment platform?".

3. PARTICIPANTS

We recruited participants through an advert sent to our various universities mailing lists and Twitter posts. The ethic for the study was approved by

the University of Nottingham, School of Computer Science Ethics Committee in November 2018. Using email, we contacted the interested participants. A total of 110 participants (53 male and 57 female) between the age of 20 and 85 years, mostly in the 26-35 age group (62.7%) took part in the study that took 45 minutes on the average to complete. The vast majority (98%) of the participants live in the UK, have a college or undergraduate degree (88%), and were in full time employment (89%). 9% worked in positions that involve hiring candidates, and a further 15.5% have done so in the past. Participants were compensated with a £10 Amazon voucher for their time, a rate above the minimum wage in the UK (£8.36 /hour).

The average online activity score of the participants (over the four weeks leading to the study) is 26 out of a range of 0 to 35 (SD = 3.9). The averages for all the individual activities (such as socialising, purchasing, finding information, entertainment, content sharing, finance and looking for jobs) range from 3.24 (finance) to 4.07 (socialising) out of 0 to 5 (SD ≈ 1). Most of the participants (95.5%) had used e-recruitment websites before; with nearly half (42.7%) saying they often used them, 30% using them sometimes, and 15.5% rarely. Two thirds (69%) indicated they were somewhat or moderately familiar with such sites and 12% are extremely familiar with them; 18% are slightly or not familiar with e-recruitment websites.

The participants have fair digital confidence scores. The average digital confidence score (alpha = 0.899) which is a measure of the self-reported digital literacy of the participants is 3.7 on a scale from 0 to 5 (SD = 0.6). The average trust measure score of the participants is 3.54 out of 1 to 5 (SD = 0.66). We used this metric to measure the importance of trust to the participants when they are online. It showed that the majority of the participants think about trust when online. The average trusting belief score (alpha = 0.883/0.858 *pre/post*) which measured how trusting the user is when they are online is 3.86 (SD = 0.617) for pre-study and 3.88 (SD=0.577) for post-study on a scale from 1 to 5. There is a strong degree of correlation between the online activity scores of the participants and their digital confidence (r_s = 0.640, p = 0.000) and between the online activity scores of the participants and their trusting beliefs (r_s = 0.672, p = 0.000). There is a moderate degree of correlation between the online activity scores of the participants and their average trust measure score (r_s = 0.278, p = 0.003). These participants' statistics show that they are fairly expert users of online digital systems including e-recruitment systems, have reasonable digital confidence and expectations of trust in e-recruitment systems.

4. FINDINGS

In this section, we summarised relevant findings based on the following questions.

What are users' perceptions of fairness, transparency and trustworthiness in e-recruitment systems?

Comparing participants' responses before and after they engaged with our platform, our results (see **Table 2**) suggest that compared to what they generally believed about the existing e-recruitment platforms, they were more likely to agree that our prototype platform performed as expected (t(109)=-2.965, p=0.004) and had all the information necessary to generate reliable results (t(109)=-2.067, p=0.041). They also believed that the platform was significantly more fair (t(109)=-6.067, p=0.000) and more trustworthy (t(109)=-3.805), p=0.000) than e-recruitment platforms in general, but not more transparent (t(109)=-3.805, NS). There are a number of interesting findings in these results. The perception of fairness (e-*Rec3*) scores is the lowest for participants' preconception of e-recruitment platforms in general. Participants are generally indifferent and somewhat disagree that e-recruitment platforms are fair. This value substantially increased after the participants engaged with our platform.

As shown in this section, (see Table 3 and **Table 5**), user understanding of the data and reasoning behind candidates' rankings and selection evoked some positive attitudes as our platform was considered fairer and more trustworthy by the study participants. The answers to the post-questionnaire

questions (c.f., Table 3) on the impact of explanations and the open-response answers analysed in the section (c.f., Table 5) allow us to understand why participants have such positive attitudes towards our prototype system more than the traditional e-recruitment systems that they have used in the past. For example, 47.3% of the participants claimed they somewhat understood the algorithms that our platform uses, 35.5% of the participants claimed they strongly understood the algorithms that our platform uses, a total of 82.8% of the participants (See Table 3 for full details).

Table 2. Descriptive statistics for variables measured in the study to compare our prototype platform to participants' general preconceptions of e-recruitment platforms, showing mean and (in brackets) standard deviation values.

	Satisf-ied expec-tation	relia-bility	fairn-ess	trans-pare-ncy	Trust-wort-hine-ss
e-recruit-ment platfor-ms (PRE)	3.71 (0.734)	3.94 (0.610)	3.63 (0.752)	3.82 (0.693)	3.81 (0.710)
our prototy-pe platfom (POST)	3.97 (0.723)	4.11 (0.805)	4.15 (0.866)	3.98 (0.888)	4.13 (0.791)

Further, we analysed several correlations between variables to see how perceptions related to each other. Some items (*Plat* statements) were significantly skewed and/or kurtotic, so Spearman's rho correlations were carried out where relevant, otherwise Pearson's r was used. In particular, we carried out correlations to relate participants' perceptions of expectation satisfaction, reliability, fairness and transparency to trust in our prototype platform. Our results show that there is a strong degree of correlation between expectation and trustworthiness (r_s = 0 .622, p = 0 .000). There is a strong degree of correlation between reliability and trustworthiness (r_s = 0 .592, p = 0 .000). There is a moderate degree of correlation between fairness and trustworthiness (r_s = 0 .322, p = 0 .001). There is a strong degree of correlation between transparency and trustworthiness (r_s = 0 .511, p = 0 .001).

Table 3. The impact of the presented explanations on participants' perceptions of our prototype platform. The percent column shows the percentage of the participants that somewhat and strongly agree with each of the metrics used.

	Ave rage (out of 0 to 5)	SD	Per-cent
Plat6: I understand what the algorithms the platform uses are doing.	4.16	0.761	82.8 %
Plat7: The explanations provided increase my understanding of the data and reasoning behind the presented algorithms.	4.13	0.768	80.9 %
Plat8: The explanations provided increase my understanding of the data and reasoning behind e-recruitment algorithms in general.	4.16	0.807	80.9 %
Plat9: The explanations provided increase my trust in the presented algorithms.	4.23	0.820	81.8 %
Plat10: The explanations provided decrease my trust in the presented algorithms.	2.38	1.39	22.7 %
Plat11: I trust the results of the presented algorithms to be transparent, reliable and fair to all applicants.	3.89	0.902	72.7 %

What are the possible trade-offs required to increase users' trust in e-recruitment systems?

Our results (see **Table 4**) show that on average, participants somewhat agreed to all the statements about trust in e-recruitment websites. They felt that they did not really trust e-recruitment websites that asked for their personal data without explanation. They were more likely to trust e-recruitment websites that are easy to use, but they also did trust that they would get the best results from them. Participants also indicated that their trust in the sites is affected by the algorithms used, and that having more control over their own data increases their trust. These opinions did not significantly change after interacting with our prototype platform (that is, p-value is larger than 0.05 in all variables).

Table 4. Descriptive statistics about each variable measured in the study as the other factors that may impact trust in e-recruitment algorithms, with their means and standard deviation values.

	data with-out expla-nation	ease of use	best result satisf-action	algori-thms	cont-rol over data
PRE	3.85 (0.768)	3.75 (0.818)	3.78 (0.722)	3.82 (0.666)	3.82 (0.744)
POST	3.91 (0.761)	3.77 (0.797)	3.93 (0.713)	3.92 (0.692)	3.93 (0.568)

.

What is the impact of understanding on participants' perceptions in e-recruitment?

Our results show that the vast majority of the participants agree that an understanding of what e-recruitment algorithms does increases their trust in the e-recruitment systems. **Table 5** shows the exemplar responses of the vast majority of the participants. This result is consistent with the result presented in Tables 2 and 3 where participants rated our explanation-enhanced prototypical platform to be more trustworthy than the traditional e-recruitment systems.

Table 5. An exemplar question and responses from our qualitative survey regarding understanding and perception of trust. The unique ID of the participant that gave each response is enclosed in brackets.

Exemplar Question:
Do you think an understanding of what a shortlisting algorithm does can increase or decrease your trust in an e-recruitment platform?
Exemplar responses:
It certainly increases trust, because you know the internal algorithms, you know why the rankings are the way they are (Y2MDNVZ).
Knowing the algorithm, I know more about the platform and trust it (SJ4Y32Z).
Increased trust because it felt very public (YN0HR99).
It increases my trust. I understand that everything is recommended according to the requirements of the recruiter (T21ZYKC).
After knowing the algorithm, I like and trust this website more (X3JRCVZ).
More trust in websites because they are transparent (F9WJVPH).
Yes, it allows me to understand the criteria used for the shortlisting (YQ3TTXK).
Knowing the algorithm that the site is very fair, more trust (2VKL28V).
Understanding increases my confidence (GHHFAF9).

5. DISCUSSION

This study is an effort to broaden the current discussion of the issues surrounding fairness, reliability, transparency and trustworthiness in e-recruitment systems. Our findings show the following: i) Users rated our explanation-enhanced e-

recruitment system to be fairer and more trustworthy than traditional e-recruitment systems. ii) There is a strong correlation between trust in e-recruitment systems and the expectation of users, reliability and transparency. However, our findings show moderate correlation between user perception of fairness and trustworthiness. This result implies that while most of the participants in general are indifferent and somewhat disagree that e-recruitment systems are fair, they have a reasonable degree of trust in the systems (see Table 2). iii) Other factors such as explanation, perceived ease of use, best result satisfaction, algorithms used and perceived control over data may impact users' trust in e-recruitment systems. These opinions of the participants remain consistent across e-recruitment systems in general and our prototype system specifically. iv) An understanding of what e-recruitment algorithms do increases user trust in the e-recruitment system.

We contribute to the related work in the following ways:

First, participants consider our explanation-enhanced e-recruitment system to be fairer and more trustworthy. The vast majority of the participants agree that an understanding of what e-recruitment algorithms do increases their trust in the e-recruitment system. By relying on the studies that have shown strong connections between explanation and perceptions of fairness [1], reliability [7] and trustworthiness [8], in algorithmic systems, we can say that the textual explanations provided in our prototype system may have impacted participants' perceptions. Also, related work [5] has shown that explanation significantly increased the perception of transparency in algorithmic systems. Our results suggest that explanations of the data and reasoning behind e-recruitment systems as well as their results significantly influenced how people perceive the systems. Participants felt that explanations give them more understanding of the criteria used for shortlisting candidates, make the systems more transparent, increase their confidence in the systems and thereby increase their trust.

Second, the general opinion of participants was also reflected in the lowest ratings of fairness among other perceptions in e-recruitment systems (See Table 2). Based on this result, we can assume that participants still have more concerns about fairness of e-recruitment systems than the other perceptions investigated. Maybe, as there is a lot of diversity in the population of users of e-recruitment systems, participants did not think existing systems are significantly fair to all users. This result is corroborated by the findings in [11] where participants gave mediocre ratings to fairness among other perceptions. Although fairness ratings increased significantly in our explanation-enhanced e-recruitment system, Binns et al [1] have suggested that explanations may or may not help individuals to evaluate the fairness of algorithmic systems. In this regard, our future research will investigate issues surrounding perception of fairness and examine testable factors that can positively impact this perception.

Third, our findings show that in addition to fairness, reliability and transparency of e-recruitment systems, participants still thought that factors such as ease of use, data privacy and having more control over their data, reliability of information, more details about algorithmic processes, platform reputation and quality feedback from other users are equally important (c.f., Section 4). In this regard, we suggest that organisations need to optimize their e-recruitment systems to be easy to use and to find a trade-off between the minimum information about users that is required for decision-making and the amount of information the users are willing to provide. For instance, ease of use can be implemented by providing some flexibility in the filling of online applications, having clear contact information for enquiries and some flexibility in username registration and log in amongst others. Likewise, control over data can be implemented by finding the right balance between an applicant's desire not to supply certain personal information and whether that information is actually required for decision-making. We suggest that optimised systems that adequately address these concerns may be considered more trustworthy than those that do not.

Our results regarding the correlation between trustworthiness and other perceptions such

as transparency, reliability and the satisfaction of users' expectation suggest that to engender trust in e-recruitment systems, increased understanding of these perceptions are important.

6. LIMITATIONS

There are some limitations to be considered when interpreting our results. First, although our study was based on a large sample and the vast majority (98%) are UK residents, we do not consider their cultural backgrounds and test whether diversified cultural backgrounds will impact our results differently. Second, we still need to examine the impact of different explanation styles on the perceptions we investigated in this study. This is also an avenue for future work that we are considering. Lastly, the vast majority of our study participants are expert users of e-recruitment systems. We assumed they understood the explanations provided on our prototype system. The results may be different for non-expert users.

7. CONCLUSIONS

We presented a study to explore users' perceptions in e-recruitment systems. Our results suggest that users have more positive perceptions in our prototype system that provides explanation on the data and reasoning behind e-recruitment algorithmic results than existing systems. By implication, we suggest that to improve users' perceptions of fairness, reliability and transparency of e-recruitment systems, and to engender trust in them, the data and reasoning behind algorithmic results must be explained. We also discussed other fundamental factors that may impact trust in e-recruitment systems. In particular, we suggest that e-recruitment systems must be optimised to be easy to use and to provide users with more control over their data.

REFERENCES

1 R. Binns, M. V. Kleek, M. Veale, U. Lyngs, J. Zhao, and N. Shadbolt, "'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, 2018.

2. A. L. García-Izquierdo, P. J. Ramos-Villagrasa, and A. M. Castano, "e-Recruitment, gender discrimination, and organizational results of listed companies on the Spanish Stock Exchange" Revista de Psicología del Trabajo y de las Organizaciones, vol. 31, no. 3, pp. 155–164, 2015.

3. J. L. Herlocker, J. A. Konstan, and J. T. Riedl, "Explaining collaborative filtering recommendations", In Proceedings of the 2000 ACM conference on Computer supported cooperative work, pp. 241–250, 2000.

4. J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems" ACM Transactions on Information Systems (TOIS), vol. 22, no. 1, pp. 5–53, 2004.

5. R. F. Kizilcec, "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, 2016.

6. M. K. Lee, "Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management", Big Data & Society, vol. 5, no. 1, 2018.

7. F. Nothdurft, T. Heinroth, and W. Minker, "The impact of explanation dialogues on human-computer trust", In International Conference on Human-Computer Interaction, Springer, pp. 59–67, 2013.

8. A. Preece, "Asking 'Why' In AI: Explainability of intelligent systems–perspectives and challenges" Intelligent Systems in Accounting, Finance and Management, vol. 25, no. 2, pp. 63–72, 2018.

9. S. Silva and M. Kenney, "Algorithms, platforms, and ethnic bias" Commun. ACM, vol. 62, no. 11, pp. 37–39, 2019.

10. A. Skulmowski, Y. Augustin, S. Pradel, S. Nebel, S. Schneider, and G. D. Rey, "The negative impact of saturation on website trustworthiness and appeal: A temporal model of aesthetic website perception", Computers in Human Behavior, vol 61, pp. 386–393, 2016.

11. M. T. Thielsch, L. Träumer, and L. Pytlik, "E-recruiting and fairness: the applicant's point of view", Information Technology and Management, vol. 13, no. 2, pp. 59–67, 2012.

12. H. Webb and M. Patel, "Can video-based qualitative analysis help us understand user-algorithm interaction?" In Proceedings of the 32nd International BCS Human Computer Interaction Conference, pp. 1–5, 2018.

13. A. Woodruff, S. E Fox, S. Rousso-Schindler, and J Warshaw, "A qualitative exploration of perceptions of algorithmic fairness", In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–14, 2018.

14. R. Wang, F. M Harper, and H Zhu, "Factors influencing perceived fairness in algorithmic decision-making: algorithm outcomes, development procedures, and individual differences", In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1-14, 2020.

15. D. Shin, and Y. J. Park, "Role of fairness, accountability, and transparency in algorithmic affordance" Computers in Human Behavior, vol. 98, pp. 277-284, 2019.

SHORT AUTHORS' BIOS

1. **Gideon Ogunniye**

Dr Gideon Ogunniye is a Research Associate on CONTEXT project at the Artificial Intelligence and its Applications Institute (AIAI), School of Informatics, University of Edinburgh. The project is part of the REPHRAIN project. The project focuses on understanding the role of contexts in managing privacy online. Prior to working on the CONTEXT project, he worked as a Research Associate on ReEnTrust project. Gideon completed his PhD at University of Aberdeen, UK, which was about modelling trust and trust dialogues using augmentation theories. He has worked on a variety of issues relating to the following themes: i) knowledge representation and reasoning, ii) argumentation theories, iii) argumentation-based agent dialogues, iv) trust computing and reputation mechanisms, v) explainable and responsible AI, vi) agent communication and language semantics, vii) AI ethics, viii) Machine reasoning, and ix) human- computer interaction.
Contact him at: g.ogunniye@ed.ac.uk

2. **Benedicte Legastelois**

Dr Bénédicte Legastelois is a Research Associate who used to work at the University of Edinburgh. She is now part of the TAS Hub at King's College London. Her research focuses on a large spectrum of domains, from formal logic for AI to explainable AI. Since she has interest in ethical issues and, especially, trustworthiness of AI systems, she works on using formal methods to provide safer, fairer and more reliable algorithms. Nowadays, she works in the domain of medical diagnosis and trustworthiness of prediction algorithms.
Contact her at: benedicte.legastelois@kcl.ac.uk

3. **Michael Rovatsos**

Prof. Michael Rovatsos is a Professor of Artificial Intelligence at the School of Informatics at the University of Edinburgh. His research interests are in Artificial Intelligence with a specific focus on multiagent systems and human-friendly and ethical algorithm design. Within this wider field, he has made contributions to a range of topics from multiagent communication, planning, and learning to the design of argumentation, trust and reputation, and normative systems. His most recent work has focused on designing smart orchestration platforms for human collaboration, developing diversity-aware coordination algorithms, and developing methods to elicit fairness criteria from human users to translate these to ethical resource allocation mechanisms. Michael has been involved in research projects that have received over £17 million of external funding, has given over 50 invited academic and industry talks and tutorials, and regularly serves on the senior committees of key AI conferences such as IJCAI, AAAI, and AAMAS. Michael obtained his PhD from the Technical University of Munich in 2004, and his first degree from the University of Saarbruecken in 1999.
Contact him at: michael.rovatsos@ed.ac.uk

4. **Liz Dowthwaite**

Dr Liz Dowthwaite is a Research Fellow in the Horizon Digital Economy Research Institute at the University of Nottingham. She received an undergraduate degree (BSc) in Psychology and a Masters in the Body and Representation (MA), both from the University of Reading. Her PhD at Nottingham was grounded in social psychology from an HCI and Human Factors perspective, looking at motivations for providing financial support to online artists who create free content. She is currently looking at how psychological factors relate to participation in online citizen science; examining whether and how autonomous systems in the home can be used to aid value-based decision-making related to wellbeing; and how to embed responsible research and innovation into the digital economy. She is especially interested in how motivation, human values, and attitudes relate to behaviour and online participation.

Contact her at: liz.dowthwaite@nottingham.ac.uk

5. **Virginia Portillo**

Dr Virgina Portillo is a Researcher in the Horizon Digital Economy Research Institute at the University of Nottingham. Her current research focuses on investigating citizen's perceptions, experiences and ethical concerns regarding autonomous decision-making systems. She is particularly interested in embedding and promoting Responsible Innovation practice into research and to put forward citizens' recommendations to guide the development of responsible design, co-creation, regulation and governance of digital technologies. Other areas of interest include: trustworthy autonomous systems, transparency and data privacy by design, participatory research, policy impact and STEM public engagement. Contact her at virginia.portillo@nottingham.ac.uk

6. **Elvira Perez Vallejos**

Dr. Elvira Perez Vallejos is Associate Professor of Mental Health and Technology at the University of Nottingham (UoN), NIHR Nottingham Biomedical Research Centre (BRC) and NIHR MindTech. Elvira's research portfolio is highly interdisciplinary and she specialises in assessing the impact that technology has on the mental wellbeing of groups with protected characteristics (children, young people, older adults) applying co-design and participatory methods. She is driving world-leading research on digital mental health including issues of data ethics and privacy, user and stakeholder engagement, co-production, and responsible research and innovation (RRI). Over the last six years, she has been working with groups of young people and brought to the forefront of policy innovation their concerns and recommendations on digital mental health. She is currently collaborating with Aardman Productions on the co-development of animated movies to increase their digital literacy. She is director of RRI at UKRI Trustworthy Autonomous Systems Hub.

Contact her at: elvira.perez@nottingham.ac.uk

7. **Jun Zhao**

Dr Jun Zhao is a Senior Researcher from University of Oxford. Her research focuses on investigating the impact of algorithm-based decision makings upon our everyday life, especially for families and young children. She takes a human-centric approach, focusing on understanding real users' needs, to design technologies that can make a real impact. For this, she works closely with schools, children, families as well as technologists for children, to understand the technological, societal and regulatory challenges that we are facing, to inform national and international policymakers, technology designers and families.
Contact her at: jun.zhao@cs.ox.ac.uk

8. **Marina Jirotka**

Prof. Marina Jirotka is a renowned RI scholar in the UK. She is Professor of Human Centred Computing in the Department of Computer Science at the University of Oxford and Governing Body Fellow of St Cross College. She leads an interdisciplinary research group investigating the responsible development of technologies that are more responsive to societal acceptability and desirability. She is an EPSRC Established Career Fellow - a five year investigation into Developing Responsible Robotics for the Digital Economy. She is Director of the newly established Responsible Technology Institute at Oxford and she is co-director of the Observatory for Responsible Research and Innovation in ICT (ORBIT) which provides RI services and training to ICT researchers and practitioners. She is also currently PI on the EPSRC Digital Economy TIPS project Rebuilding and Enhancing Trust in Algorithms (ReEnTrust). Her recently concluded projects involve a range of topics in RI: she led the Responsible Innovation strand of work for the Networked Quantum Information Technologies Hub (NQIT); she was PI on ERSCs Digital Wildfires project; and she was Co-PI on EPSRC Digital

Economy TIPS project, Emancipating Users Against Algorithmic Biases for a Trusted Digital Economy (UnBias). From her work on analysing the spread of hate speech and misinformation on social media (Digital Wildfires), she was appointed specialist advisor to the House of Lords Select Committee on Communications for their inquiry into Children and the Internet. Marina is a Chartered IT Professional of the British Computer Society where she is a member of the ICT Ethics Specialist Group committee. She sits on EPSRCs Programme Advisory Board for the Digital Economy, the Advisory Board of the Society for Computers and Law and is a member of the Steering Committee for the All Party Parliamentary Group on Data Analytics. She is also a member of the UK Committee for IIASA (International Institute for Applied Systems Analysis). In 2018 was elected to the UK Computing Research Community (UKCRC) a highly selective grouping of the UK's leading computing academics. She has published widely in international journals and conferences on, Human Computer Interaction, Computer Supported Cooperative Work and Requirements Engineering. Contact her at: marina.jirotka@cs.ox.ac.uk