Explainability and the intention to use AI-based conversational agents.

An empirical investigation for the case of recruiting*

Fleiß, Jürgen[1], Bäck, Elisabeth[2], and Thalmann, Stefan[1]

[1] University of Graz, Attemsgasse 11, 8010 Graz, Austria
[2] Silicon Austria Labs, Inffeldgasse 25F, 8010 Graz, Austria

Abstract. The use of conversational agents (CA) based on artificial intelligence (AI) is increasing in the field of recruiting. Recruiting is considered a particular sensitive domain, especially if CAs also make (pre)selection decisions. The black box character of AI decisions may hinder the acceptance and use of CAs as they are not considered to be fair, accountable and transparent (FAT). Explainable AI (XAI) has the goal to make AI decisions more transparent and thus to increase its FAT. But little is known about the perception of XAI by potential job candidates and their intention to use CAs. To investigate this research gap, we conducted a vignette-style questionnaire survey filled out by 490 persons from a quota-representative population sample for Germany and Austria. Scenarios are varied by (a) the type of XAI approach and (b) by whether the explanations refer to measurable qualification or soft skills. The results indicate that XAI increases the intention to use CA in recruiting, compared to CA relying on black box AI.

Keywords: Conversational Agent · Explainable AI · User Study

1 Motivation and Background

Conversational agents (CA) and Artificial Intelligence (AI) fundamentally change the way information systems (IS) interact with humans. AI enables interactions between IS and humans that are similar to the way that humans interact with each other [11]. However, AI is usually based on black box models and the behavior of conversational agents is thus opaque [9]. This is an unpleasant situation for users as they might perceive the CAs as unfair, in-transparent or less trustworthy, and this in turn influences the acceptance of the IS, especially in high-stake situations [7].

One recent example of such a sensitive application of CAs is in the field of recruiting: CAs now conduct job interviews online and even preselect candidates based on their resumes and responses [6]. This application is considered especially sensitive due to the black box character of AI, as the stakes for applicants are

high and thus it is reasonable to assume that applicants will expect explanations [4]. Furthermore, such explanations are also seen as required by the European General Data Protection Regulation [10].

Research on AI has recently proposed approaches to make AI explainable and, through those explanations, more transparent [1]. First results indicate that XAI can reduce the negative perceptions towards AI in general [8]. However, there is little research on the influence in critical decision situations and in particular on the influence of certain explainability features on the acceptance and intention to use CAs by (potential) job applicants. To tackle this research gap, we conducted a vignette-style questionnaire survey with a total of 490 persons from a quota-representative population sample for Germany and Austria. In the next section, we will develop scenarios to study the effect of explainability and the type of skills that those explanations refer to, to investigate their effect on the willingness of potential applicants to use such CAs.

2 Research Model Development

We investigate the use and overall acceptance of CA (pre)selection decisions using a vignette-style method, suitable to be combined with an experimental design in surveys [2]. In this approach we present subjects with scenarios that are varied by (a) the type of XAI approach and (b) by whether the explanations refer to measurable qualification or soft skills. For each scenario subjects evaluate their intention to use such an CA and their overall acceptance of the CA decision.

The survey starts with a general introduction for the scenarios, namely that they apply for a job and on the company website a chatbot appears and informs them that it will make the preselection of candidates instead of a human recruiter. This CA will communicate by chat and ask all the necessary questions to assess their fit for the open position. After this introduction, seven scenarios will subsequently be presented to subjects in random order referring to the outcome of this preselection process.[3] In all scenarios, subjects will be informed that they were rejected by the CA in the preselection process. In a baseline scenario, BASE, subjects will simply be informed that the CA decided to reject their application. This mimics the result-focused decision of a typical CA based on a black box AI. We vary BASE with regard to two factors derived from the literature: explainability (EXPLAIN) and type of skill that is used in the explanation (SKILLTYPE).

In the three EXPLAIN variations, we distinguished between the explanation of black box models and interpretable models [3]. Two of the variations of the EXPLAIN factor offer explanations of black box model decisions. In EXPLAIN_LIST, subjects are provided with a list of three criteria that the rejection is based on. In EXPLAIN_COMPARE, subjects see a visualization of the score that the conversational agent assigned to them and the average

[3] Decisions in later scenarios can be affected by previous scenarios. This can be tested by comparing results for the scenarios when presented first to those for all scenarios.

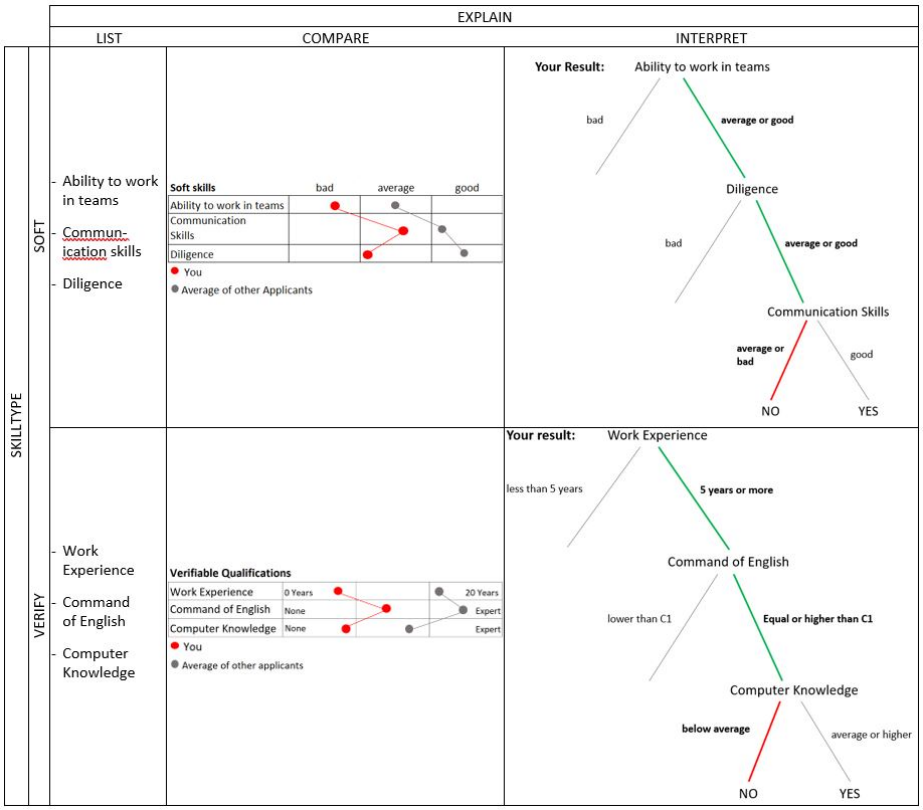

Fig. 1. Scenario Overview including Stimuli

score of other applicants. In the third variation of the factor EXPLAIN, EXPLAIN_INTERPRET, participants are shown a simple decision tree using the same criteria as in the first two variations of EXPLAIN. The path to the decision "reject" is highlighted in the decision tree and paths to the decision "accept" are visible. Such a simple decision tree is a typical example of a simple rule based model, which can be intuitively interpreted by humans [9].

We again vary all three EXPLAIN variations, with two variations of the factor SKILLTYPE, resulting in a three by two design. The factor SKILLTYPE is a natural consequence of explaining a hiring decision, as such decisions must be based on the match between the skills of the candidate and the position to be filled. The two variations of the factor SKILLTYPE we choose capture the distinction between "emotional" and "cognitive" judgements, also used in a previous scenario study on human perceptions of AI decisions. This study distinguishes between "mechanical" and "human" skills, the latter of which are meant to capture emotional capabilities or subjective judgements [5]. Mechanical skills refer to objective measures. For the recruiting application, we incorporate

human skills as soft skills in SKILLTYPE_SOFT and mechanical skills as more objectively verifiable qualifications in SKILLTYPE_VERIFY. For soft skills, we use the ability to work in teams, communication skills and diligence, for verifiable qualifications work experience, command of English and computer knowledge.

Combining each of the EXPLAIN variations with each of the SKILLTYPE variations results in six scenarios in addition to BASE. These six scenarios and the corresponding key elements of the explanations as they will be shown to subjects are displayed in Figure 1. The full questionnaire is available upon request.

3 Outlook

We conducted the survey described before with 490 persons from a quota-representative population sample for Germany and Austria. A preliminary and raw analysis of the results indicates that XAI increases the intention to use CAs in recruiting, compared to CAs relying on black box AI. The next step is to rigorously analyze the collected data. We believe that the developed scenarios capture important aspects of CAs in the field of recruiting, but also of AI in general. XAI, by overcoming the black box nature of many algorithms, is seen as an important step to create fair, accountable and transparent (FAT) AI solutions. This in turn should also increase the trust of those affected by the decisions.

References

1. Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al.: Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion **58**, 82–115 (2020)
2. Aviram, H.: What would you do? Conducting web-based factorial vignette surveys. In: Gideon, L. (ed.) Handbook of survey methodology for the social sciences, pp. 463–473. Springer, New York, NY. (2012)
3. Biran, O., Cotton, C.: Explanation and justification in machine learning: A survey. In: IJCAI-17 workshop on explainable AI (XAI). vol. 8, pp. 8–13 (2017)
4. Kim, B., Park, J., Suh, J.: Transparency and accountability in AI decision support: Explaining and visualizing convolutional neural networks for text information. Decision Support Systems **134**, 1–11 (2020)
5. Lee, M.K.: Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. Big Data & Society **5**, 1–16 (2018)
6. Leong, C.: Technology & recruiting 101: How it works and where it's going. Strategic HR Review **17**, 50–52 (2018)
7. Rai, A.: Explainable AI: From black box to glass box. Journal of the Academy of Marketing Science **48**, 137–141 (2020)
8. Ribeiro, M.T., Singh, S., Guestrin, C.: "Why should I trust you?" Explaining the predictions of any classifier. In: Proceedings of NAACL-HLT 2016 (Demonstrations). pp. 97–101 (2016)
9. Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence **1**, 206–215 (2019)

10. Selbst, A., Powles, J.: Meaningful information and the right to explanation. International Data Privacy Law **7**, 233–242 (2017)
11. Wang, W., Benbasat, I.: Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. Journal of Management Information Systems **23**, 217–246 (2007)