Human Reliance on Machine Learning-based Systems for Decision-Making: An Experiment in Recruiting

1. Research Design

Research Aim and Research Questions:

This research aims to analyze the human-machine collaboration in the form of an experimental set-up by identifying the differences in decision-making between humans with and without support of ML-based recommendations. Thus, we investigate the following research questions:

RQ1: To what extent do humans rely on ML-based systems in their decision-making process?

RQ2: Which implications does the collaboration between humans and ML-based systems in the decision-making process have on the business and its stakeholders?

Experiment Set-up

The study was performed through a mixed-method approach consisting of a lab experiment followed by expert interviews with the instruction to create a shortlist of the top-10 candidates

Fictional job postings: *Junior Fullstack Developer* and one for a *Junior Online Marketing Manager*

Resumes: Diverse set of 30 resumes for each job posting from computer science, information systems, business, and marketing students of higher education institutions

Quantitative Study - Experiment Groups:

1) Personalized resumes / non-ML support
2) Anonymized resumes / non-ML support
3) Personalized resumes / ML support
4) Anonymized resumes / non-ML support

Qualitative Study - Expert Interviews: We interviewed 24 recruiting professionals from 14 companies who had previously participated in the quantitative study - 66.66% female, and 75% between 25 and 44 years, 15 partcipants with ML-support and 9 with non-ML-support

ML support: ML-based recruiting system for resume matching purposes based on word vectors and keyword matching

Participants: 74 recruiters of 22 large multinational companies - 74% were female, 83% were between 25 and 44 years old and 78% had at least three years of experience in recruiting

2. Results of Mixed-Methods Study

RQ1 – Reliance

In order to determine to what extent humans rely on ML-based systems in their decision-making process, we analyzed the **correlation** between the **ranking** provided by professionals and the **given order of resumes** (RQ1). Based on the quantitative results, we were able to prove a higher "Ranking Correlation" score of the ML-supported setting compared to the non-ML-supported setting for the developer and the marketing job posting (see Figure 1). Regarding the development job posting, the unweighted marginal means of "Ranking Correlation" scores for ML-supported and non-ML-supported rankings were .589 +/- .038 and .066 +/- .042, respectively. ML-supported rankings were associated with a mean "Ranking Correlation" score .523 points higher than non-ML-supported rankings, a statistically significant difference, $p < .0005$. We also analyzed the main effect for resume anonymity which was slightly significant, $F(1, 70) = 2.795$, $p = .099$, partial $\eta^2 = .038$. Here, the unweighted marginal means of "Ranking Correlation" for personalized and anonymous settings were .280 +/- .038 and .375 +/- .042, respectively. Anonymized rankings exhibited a mean "Ranking Correlation" score .095 points higher than personalized rankings, a slightly statistically significant difference, p = .099. For the Marketing job posting, the ML-supported rankings were associated with a mean "Ranking Correlation" score .523 points higher than non-ML-supported rankings, a statistically significant difference, $p < .0005$. The main effect for resume anonymity was not significant for this task, $F(1, 70) = 2.579$, $p = .113$, partial $\eta^2 = .036$.

Through these quantitative results, we were able to disprove the perception of recruiters in the ML-supported setting that they did not rely on the ML-based matching score which was stated during the qualitative expert interviews

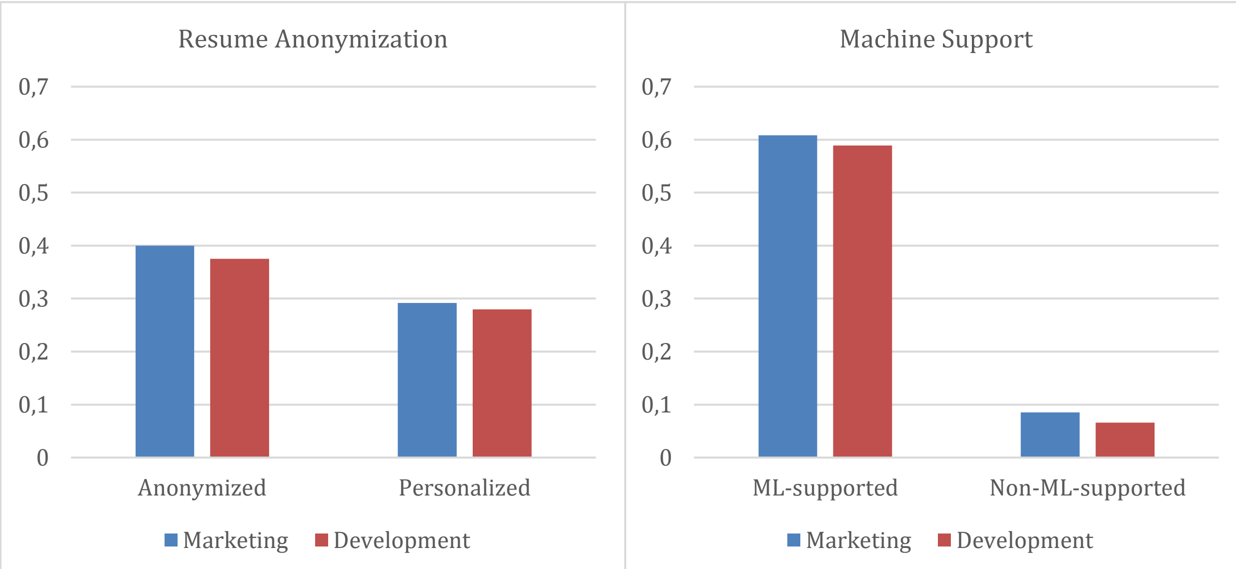

Figure 1: Comparison of Rank Correlation Scores

Correlations are measured using Kendall's Tau and averaged over all participants in the respective group

RQ2

With regards to RQ2, we aim to identify specific effects on businesses, the recruiters and the candidates when implementing a ML-based system in the candidate screening decision-making process. According to *H2*, we anticipated that **hard skills** have a positive impact on the ranking among the top-10. This assumption has correctly been supported by the quantitative results for the developer and marketing job postings in regard to the study duration and the relevant working experiences.

Figure 2: Hard skills in ranked and unranked candidates

Study duration and working experience measured in years, stated values represent median values over grouped candidates

These findings were corroborated through the qualitative expert interviews as hard skills are considered as must-have criteria for the job postings:

> *'I primarily focused on work experience and secondarily on education, and for candidates whose past experience was not clear, I have considered aspects such as international experience.'*
> *– HR Specialist Recruiting –*

The most frequently mentioned examples for using a ML-based recruiting system were **high-volume and entry-level roles**. These roles are characterized by standardized job profiles and by hard skills which require a less-nuanced assessment and exhibit high candidate volumes. Therefore, these roles would be better suited for an initial ranking by a ML-based system:

> *'I could see this for entry-level roles and/or a first screening for our trainee program where requirements are clearly defined. Or in the area of our production workers […] and we can clearly define which degrees are required and if they are part of a candidate's profile or not […]. My gut feeling tells me, the more complex a position gets, the harder it gets for an algorithm to distinguish who are we looking for.'*
> *– Talent Attraction Manager –*

However, building trust in the ML-based system and its results is a concern. Additionally, participants mentioned the risk of dismissing candidates with an unconventional skillset due to inflexible or incorrect criteria within the ML-based system.

When asked about the impact of ML-based systems on their own job duties, recruiters also see **operational efficiency** as a key benefit, primarily for high-volume roles and to reduce the number of interviews they need to conduct:

> *'If recruiters no longer need to review resumes 15 to 30, that would be very helpful. […] Every recruiter is happy when they don't need to review resumes which are not relevant for the search.'*
> *– Senior Recruiting Expert –*

Additionally, professionals are aware of potential pitfalls due to the use of a ML-based system such as their reliance on the ML-based matching score for their own decision-making and losing awareness for the type of applicants that they receive. While the participants perceive a ML-based system as a suitable decision support, they are concerned about its limitations compared to the perceived strengths of a human expert. Therefore, they propose a **symbiosis of humans and ML-based systems**:

> *'I think that it can very well be of help to recruiters, but I don't think it can replace a recruiter completely. […] I could imagine that ML-based systems cannot filter out what a human sees in the documents.'*
> *– Talent Attraction Manager –*