Deep Learning in HR: A Technical Report on the Ethics in AI Video Analysis in Interviews

Student ID 2491092

November 2023

Contents

1 Abstract **2**

2 Introduction **2**

3 AI Project Background **3**

4 Responsible AI Assessment **4**
4.1 **Problem Formulation** 4
4.2 **Data Extraction and Acquisition** 6
4.3 **Data Processing** 7
4.4 **Modeling** 7
4.5 **Testing and Validation** 9
4.6 **Deploy, Monitor and Release** 9

5 Conclusion **10**

1 Abstract

This report explores the increasing use of AI in human resources, focusing on AI Video Analysis for assessing applicants. Acknowledging the potential benefits of efficiency, it critically examines such technology's ethical and privacy implications. The study proposes a Process-Based Governance (PBG) Framework, integrating FAST Track Principles and SUM Values, to ensure responsible AI implementation in HR. The framework prioritises fairness, transparency, accountability, and respect, addressing concerns about bias, privacy, and user experience. It emphasises continuous monitoring, external validation, and adherence to legal standards. The report evaluates the challenges and benefits of AI Video Analysis, recognising its potential to reduce human bias but highlighting technical and cultural limitations. The PBG Framework serves as a comprehensive guide for organisations to navigate the ethical complexities of AI in HR, fostering trust, inclusivity, and ethical practices in the recruitment process.

2 Introduction

Human resources (HR) is part of a business responsible for sourcing, recruiting, retaining, screening, training and firing people, managing rewards and staffing needs Cardon and Stevens 2004. Since the early 2000s, there has been an increase in the use of Artificial Intelligence (AI) within HR. The increase in demand has further pushed the need for automation to increase speed and efficiency when recruiting the best talent while reducing costs with leaner teams Jatobá et al. 2019. Some AI applications include chatbots for screening candidates and keeping them informed during the process, bots with social media scraping tools that collect applicants' data to predict behaviours and AI-based gamification that tests memory, focus and the ability to read social cues. For this report, the focus will be on the more advanced AI solution using AI-assisted interviews, also known as AI Video Analysis. HireVue developed a programme that scores an individual based on their tone of voice, body language, keywords and emotional state. Affectiva has a programme that measures a person's emotional intelligence and honesty Ahmed 2018. Nevertheless, surely, such technology infringes a person's right to privacy, especially when their emotional state is being measured? One bad day can result in an AI giving a candidate a low score. AI accelerates the hiring process and reduces cost through automation and reducing administrative duties Chen 2023 and it gives the ability to have an excess of candidates to apply for jobs no matter where they are Suen, Hung, and Lin 2019. However, AI can also passively absorb human biases and reflect unconscious bias Fernández and Fernández, 2019; Ong, 2019, as cited by Chen 2023. This was the case with Amazon's recruitment tool, favouring men rather than women and ignoring features such as relevant skill sets due to the biased model Pena et al. 2020.

There are three building blocks to delivering a responsible AI project: en-

suring it is fair, trustworthy, ethical and justifiable. Using the SUM (Support, Underwrite and Motivate) Values Connect, Respect, Care and Protect to assess the AI technology and the FAST Track Principles to examine developmental compliance of Fairness, Accountability, Sustainability and Transparency, this report will explain a new design Process-Based Governance (PBG) Framework to set up a transparent process for AI Video Analysis to be aligned with responsible AI Implementation. This report will outline the background of AI Video Analysis and provide a description; an assessment will be made on how it can be responsible, followed by a conclusion.

3 AI Project Background

As mentioned, AI Video Analysis scores a candidate on specific behaviours, be it tone of voice, keywords, body language and emotional state, for example, whether they are happy, angry or confident. The system would record everything a person says and does, paying close attention to the key indicators it is trained on Kammerer 2021. During the interview, the candidates are given a series of questions to answer within a certain time frame, which is recorded using their devices. Face recognition using computer vision algorithms is used to know where certain features of the face are, such as the eyes, nose and mouth, to understand the expressions. Further algorithms are used to analyse the emotional state and vocal cues to aid in understanding a person's emotional intelligence and personality Escalante et al. 2018. The audio recorded analyses tone of voice and use of language and speech patterns, which are further analysed using Natural Language Processing (NLP), turning speech into text to identify keywords Kaufmann 2014. Based on the measurable features, a score is given along with objective feedback. Unfortunately, the applicants are never given feedback. The other drawback is not correctly turning speech into text for those with an accent or misinterpreting emotions due to cultural differences, exposing the applicant to algorithmic bias Köchling et al. 2021. On the other hand, it can reduce human bias as it does not care about how a person looks Kammerer 2021. Then, there are the technical problems that candidates may face, as the recording has to be done on their own devices. Applicants may be instructed to record in a quiet and well-lit room; however, there are instances when the unexpected happens and the interview is recorded in poor lighting. There could be online issues or background noise that affects the quality of the video. It is highly unlikely that the AI will be able to give a fair score as it is assumed that the applicants will record the interview according to the instructions. The other unfortunate problem is disregarding a person's personality and how they could be a good fit based on their personality, as only measurable input data is considered. The over-reliance on this technology could potentially overlook the ideal candidates as some may find it intimidating or impersonal, hindering them from performing their best Vivek 2023. However, AI video analysis saves time and resources for both the company and the applicant. The AI can process a large number of videos quicker and can generate vast data to assist in future decisions, enabling

the recruiter to spend more time on other tasks. Although the data generated can help improve the selection process over time, there is a privacy issue when considering UK data protection laws. Is the information gathered just for assessing whether the candidate is a good fit, improving the process and if the candidate has the right to claim the extracted data Kammerer 2021.

4 Responsible AI Assessment

The image above illustrates the PBG Framework and the five stages used to design and redevelop AI Video Analysis to be aligned with responsible AI. This framework will form the basis of this section.

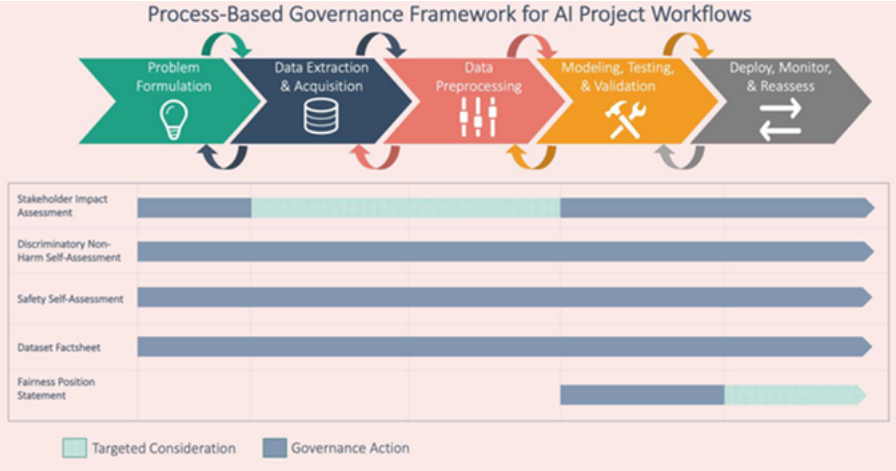

Figure 1: An image of the PBG Framework Leslie 2019.

Organisations can ensure responsible AI implementation by integrating the FAST track principles and SUM values into the PBG Framework at each AI Video Analysis redevelopment stage. This approach prioritises fairness, accountability, transparency, ethical use of data, and a strong connection with candidates throughout the hiring process, fostering trust and inclusivity while mitigating ethical risks.

4.1 Problem Formulation

Using the FAST Track principles, equity and fairness must be a foundation, including transparency and accountability in decision-making. The interviews are structured to ask all applicants the same questions, ensuring equality regarding the evaluation Suen, Hung, and Lin 2020. The different cultural and ethnic groups may not have been explored in depth. Specifically, the data sets can lack diverse representations from different sources and biases are developed

Problem Formulation

- Principle 1: Transparency in decision-making
- Principle 2: Equity and fairness should be the foundation
- Respect: Ensure privacy and dignity through the interview process
- Connect: Connect with candidates in a respectful and meaningful way
- Best Practice 1: Establish fairness and diversity objectives from the start
- Best Practice 2: Involve experts in ethics to guide the problem formulation
- Guideline 1: Perform a fairness and bias assessment during the problem formulation
- Guideline 2: Include an applicant feedback process for continuous feedback

Table 1: PBG Framework for AI Video Analysis: Problem Formulation

at the developmental stage Zixun, 2020 as cited by Chen 2023, hence why AI expresses unconscious bias. To maintain the respect of others using the SUM Values means to ensure the interview process respects the applicant's privacy and dignity. Protecting data and ensuring security is crucial. It should also strive to connect with applicants respectfully and meaningfully. Applicants can apply for jobs no matter where they are. The benefits of eliminating self-reported personality can remove the social desirability bias Suen, Hung, and Lin 2020. Having a user-friendly interface is also crucial in that connection for accessibility. Feedback should also be given to the applicants for them to understand how they performed and improve. The ability to give feedback will empower the applicants because the feedback can be challenged if they disagree.

4.2 Data Extraction and Acquisition

Data Extraction and Acquisition

- Principle 3: Transparent data collection
- Principle 4: Safeguarding security and data privacy
- Care: Handle applicant data with care and security
- Protect: Follow data protection laws and safeguard applicant's data
- Best Practice 3: Gain consent and clearly explain the usage of data
- Best Practice 4: Implement strong data encryption and access controls
- Guideline 3: Update consent and management of data regularly
- Guideline 4: Hire a data protection officer to oversee data compliance and handling

Table 2: PBG Framework for AI Video Analysis: Data Extraction and Acquisition

Unsurprisingly, organisations must follow the Data Protection Act 2018 when using personal data, ensuring the information is lawfully, fairly and transparently used. Data must be used in a way that is relevant, adequate and limited to what is necessary, be it training, testing and real-time data. The purpose of this data must be specified explicitly to the applicant and not kept for longer than necessary. Acquired data must be securely stored and access controlled. Not following these laws will lead to loss of reputation, compensation and fines gov.UK 2018. Facebook was fined 1.2 billion euros by unlawfully transferring personal data to the US Board 2023. In 2022, a 98-page document was produced titled Data Protection and Digital Information Bill to assess the use of data to be more effective. Under Article 22, individuals have the right not to make decisions based solely on automated processing that affects their legal rights, which infers AI. However, decisions for a contract between an individual and the organisation or the person's consent are permissible when there is human intervention Woodhouse 2022 and Brkan 2017. Hiring a data protection officer to oversee data compliance and handling is crucial for the company's reputation that uses AI and the creators themselves Chen 2023.

4.3 Data Processing

Data Processing

- Principle 5: Pre-processing of data to mitigate bias and ensure transparency
- Principle 6: Document all data processing methods for transparency
- Care: Document data processing to demonstrate accountability
- Connect: Data processing needs to respect the applicant's experience and backgrounds
- Best Practice 5: Put in place techniques to anonymise data and bias protection algorithms
- Best Practice 6: Document all data cleaning procedures and transformation
- Guideline 5: An independent review board should be established to ensure objective evaluation of data processing.
- Guideline 6: Regularly assess and review the data processing procedures for ethical compliance

Table 3: PBG Framework for AI Video Analysis: Data Processing

The data processing stage is crucial in significantly reducing bias and should take up most of the time before feeding to the model. A wide range of techniques should be used, such as resizing, brightness trimming and contrast, among others, for an effective model Bovik 2010 to enable it to reduce bias. There are instances where not enough candidates are participating in the training model, which will impact the model's predictive power and a more diverse pool is needed Suen, Hung, and Lin 2020.

4.4 Modeling

When creating a model that respects the applicant's dignity and diversity, using the right features for fair and non-biased models during testing and validation is essential. Some models use emotions as a measure but are unreliable as they differ from race, country and culture Singhania, Unnam, and Aggarwal 2020. Models need to recognise that there is a human face being analysed and need to be able to recognise this no matter the skin colour. Google experienced a scandal where it labelled two images of dark-skinned black men as gorillas Prates, Avelar, and Lamb 2020. The professional view from the Industrial and

Modeling

- Principle 7: To develop AI models that are interpretable, fair and accountable
- Principle 8: Review and refine AI models for ethical standards regularly
- Respect: Create models that respect applicant's dignity and diversity
- Care: Conduct regular audits and assessments
- Best Practice 7: Ensure an AI committee oversees the model development and assesses the fairness
- Best Practice 8: Diverse training data must be used to reduce bias
- Guideline 7: Regular AI audits and impact assessments should be enforced
- Guideline 8: Enforce regular ethical AI audits and impact statement

Table 4: PBG Framework for AI Video Analysis: Modeling

Organisational Psychology community, whereby structured interviews are more reliable and valid than unstructured interviews, should be adhered to when designing the model. For fairness, applicants should be measured on the same questions Suen, Hung, and Lin 2020 and have the same set of time to answer the questions unless they have a disability, in which case, more time should be given. Psychology literature has said that the big five personality traits are difficult to assess Suen, Hung, and Lin 2019 and should not be used in the model. Other features to ignore are people's names, as researchers from Princeton University used an off-the-shelf ML software that favoured European names as opposed to those from other nations Lee, Resnick, and Barton 2019. Using diverse training data, as mentioned in the Problem Formulation section, to reduce bias makes the model worthwhile as an off-the-shelf product. The model will not be able to give a fair score if it cannot accurately use NLP to turn speech to text to sum the keywords said by the applicant; this feature should contain a diverse amount of data for robustness.

4.5 Testing and Validation

Testing and Validation

- Principle 9: Thorough testing and audits from third parties for fairness and accuracy
- Principle 10: Maintain an external audit board to scrutinise everything regularly
- Protect: Ensure the audits safeguard the applicant
- Connect: Use the results revealed by the audit to improve applicant's experiences
- Best Practice 9: Perform cross-validation and fairness assessments
- Best Practice 10: Collaborate with external organisations who specialise in ethical AI validation
- Guideline 9: Publish audit results to the public regularly
- Guideline 10: Annually update and review the audit board structure and authority

Table 5: PBG Framework for AI Video Analysis: Testing and Validation

The guidelines set in the Modelling stage must be tested and validated externally to identify and rectify bias. Audits allow for processes and methods to be tracked and hold the developers accountable for safeguarding applicants to ensure transparency and fairness Porter, Simon, and Hatherly 2014 and Lee, Resnick, and Barton 2019. It is a good idea to use the applicant's feedback revealed by the audit to improve the experience since it will replace human interaction and the interface must be engaging and easy to use. A regular review of the audit board structure and authority, ensuring that those on the board are diverse, will further help to reduce bias.

4.6 Deploy, Monitor and Release

A disclosure policy for hiring decisions based on AI is essential for transparency. Applicants need to know how their data is being used, as it is a legal requirement under the Data Protection Act 2018. Applicants and organisations should clearly understand how the AI works and its role in the process, including how applicant's responses are scored and evaluated. Full disclosure can further help during the monitoring process to ensure nothing has been missed and any feedback made is reviewed and updated if it infringes the SUM values and FAST principles.

Deploy, Monitor and Release

- Principle 11: Continually monitor the AI model performance, applicant experience and ethical impacts
- Principle 12: Be ready to adapt and iterate it based on new insights
- Care: Have a way for applicants to give feedback to address concerns and gather insights
- Connect: Have a disclosure policy for hiring decisions based on the AI
- Best Practice 11: Create a disclosure policy for AI decisions
- Best Practice 12: Review and update the AI oversight member's structure and authority
- Guideline 11: Maintain an external oversight board of members to ensure continuous ethical AI implementation
- Guideline 12: Create an AI ethics committee to address applicant concerns and improve the process

Table 6: PBG Framework for AI Video Analysis: Deploy, Monitor and Release

5 Conclusion

Integrating AI Video Analysis in human resources while offering potential benefits in efficiency and cost reduction entails a careful and responsible approach to addressing ethical and privacy concerns. The proposed Process-Based Governance (PBG) Framework, incorporating the FAST Track Principles and SUM Values, provides an inclusive structure for developing and deploying AI technology in assessing applicants. By prioritising fairness, transparency, accountability, and respect, the framework ensures that AI applications in HR align with ethical standards. It emphasises the importance of fair interview processes, user privacy, and clear communication with applicants regarding data usage. The framework also highlights the need for continuous monitoring, external validation, and regular audits to identify and rectify biases to aid in a culture of accountability. The framework advocates for a balanced approach that improves the AI's strengths while mitigating its limitations when acknowledging the potential pitfalls, such as algorithmic biases and technical challenges. Furthermore, the framework highlights the significance of legal compliance, emphasising adherence to data protection laws and appointing data protection officers. Indeed, the PBG Framework provides a roadmap for organisations to

responsibly handle AI's ambiguousness in HR, promoting trust, inclusivity, and ethical practices in the recruitment process. Striking this balance is crucial for realising the transformative potential of AI while safeguarding individual rights and maintaining the integrity of the hiring process.

References

Ahmed, Owais (Dec. 2018). "Artificial Intelligence in HR". In: *International Journal of Research and Analytical Reviews* 5.4, pp. 971–978. DOI: 10.31221/osf.io/cfwvm.

Board, European Protection (2023). *1.2 billion euro fine for Facebook as a result of EDPB binding decision*. URL: https://edpb.europa.eu/news/news/2023/12-billion-euro-fine-facebook-result-edpb-binding-decision_en#:~:text=Brussels%2C%2022%20May%20%2D%20Following%20the,Protection%20Authority%20(IE%20DPA).. (accessed: 16th November 2023).

Bovik, Alan C (2010). *Handbook of image and video processing*. Academic press.

Brkan, Maja (2017). "AI-supported decision-making under the general data protection regulation". In: *Proceedings of the 16th edition of the International Conference on Articial Intelligence and Law*, pp. 3–8.

Cardon, Melissa S and Christopher E Stevens (2004). "Managing human resources in small organizations: What do we know?" In: *Human resource management review* 14.3, pp. 295–323.

Chen, Zhisheng (2023). "Ethics and discrimination in artificial intelligence-enabled recruitment practices". In: *Humanities and Social Sciences Communications* 10.1, pp. 1–12.

Escalante, Hugo Jair et al. (2018). "Explaining first impressions: Modeling, recognizing, and explaining apparent personality from videos". In: *arXiv preprint arXiv:1802.00745*.

gov.UK (2018). *Data protection*. URL: https://www.gov.uk/data-protection. (accessed: 16th November 2023).

Jatobá, Mariana et al. (2019). "Evolution of artificial intelligence research in human resources". In: *Procedia Computer Science* 164, pp. 137–142.

Kammerer, Brittany (2021). "Hired by a Robot: The Legal Implications of Artificial Intelligence Video Interviews and Advocating for Greater Protection of Job Applicants". In: *Iowa L. Rev.* 107, p. 817.

Kaufmann, Andreas (2014). "Using Natural Language Processing to Support Interview Analysis". PhD thesis. Master Thesis]. Freiderich-Alexander Univsersität, Erlangen-Nürnberg.↑ p 43.

Köchling, Alina et al. (2021). "Highly accurate, but still discriminatory: A fairness evaluation of algorithmic video analysis in the recruitment context". In: *Business & Information Systems Engineering* 63, pp. 39–54.

Lee, Nicol Turner, Paul Resnick, and Genie Barton (2019). "Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms". In: *Brookings Institute: Washington, DC, USA* 2.

Leslie, David (2019). "Understanding artificial intelligence ethics and safety". In: *arXiv preprint arXiv:1906.05684*.

Pena, Alejandro et al. (2020). "Bias in multimodal AI: Testbed for fair automatic recruitment". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 28–29.

Porter, Brenda, Jon Simon, and David Hatherly (2014). *Principles of external auditing*. John Wiley & Sons.

Prates, Marcelo OR, Pedro H Avelar, and Luís C Lamb (2020). "Assessing gender bias in machine translation: a case study with google translate". In: *Neural Computing and Applications* 32, pp. 6363–6381.

Singhania, Abhishek, Abhishek Unnam, and Varun Aggarwal (2020). "Grading video interviews with fairness considerations". In: *arXiv:2007.05461*.

Suen, Hung-Yue, Kuo-En Hung, and Chien-Liang Lin (2019). "TensorFlow-based automatic personality recognition used in asynchronous video interviews". In: *IEEE Access* 7, pp. 61018–61023.

— (2020). "Intelligent video interview agent used to predict communication skill and perceived personality traits". In: *Human-centric Computing and Information Sciences* 10, pp. 1–12.

Vivek, Ramakrishnan (2023). "Enhancing diversity and reducing bias in recruitment through AI: a review of strategies and challenges". In: *. . -Informatics. Economics. Management* 2.4, pp. 0101–0118.

Woodhouse, John (2022). *Data Protection and Digital Information Bill*. URL: `https://researchbriefings.files.parliament.uk/documents/CBP-9606/CBP-9606.pdf`. (accessed: 16th November 2023).