Exploring Fusion Techniques in Multimodal AI-Based Recruitment: Insights from FairCVdb

Paper ID: 68

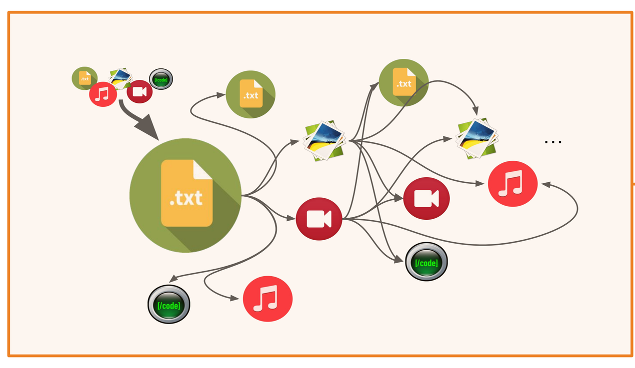
www.mammoth-ai.eu

Swati Swati[1] , **Arjun** Roy[1,2] and **Eirini** Ntoutsi[1]

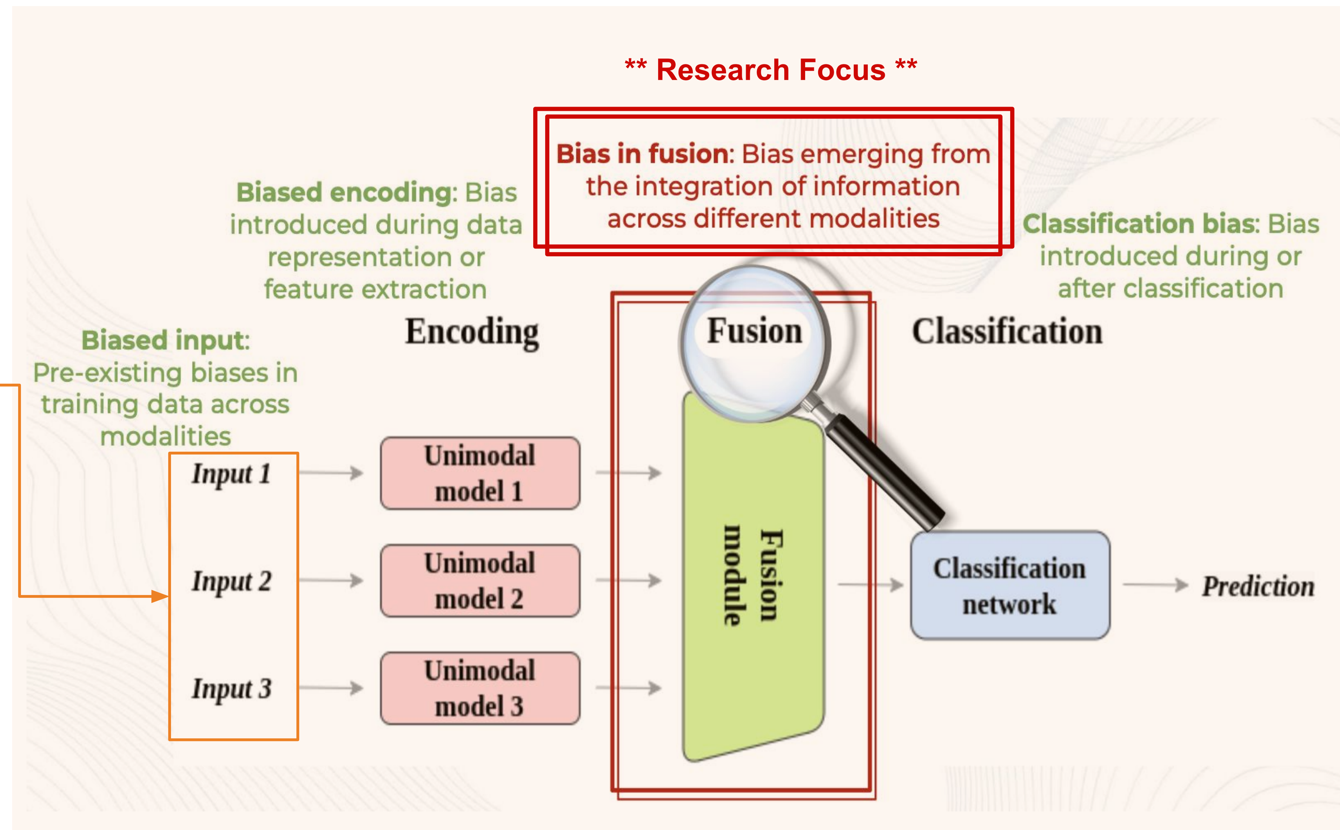
[1]Research Institute CODE, University of the Bundeswehr Munich, Germany, [2]Institute of Computer Science, Freie University Berlin, Germany

INTRODUCTION & MOTIVATION

Research Objective: Investigate the **fairness and bias implications of Fusion Approaches** in multimodal AI-based systems.



The concept of Multimodal learning involves data from different modalities.



Bias across stages of multimodal learning.

The Real-World Application: Multimodal AI-based recruitment systems:

700+ companies are using AI-based recruitment systems[1].

- Increasing use of decision-making algorithms raises concerns about transparency and discrimination, especially affecting specific social groups.
- Investigating how sensitive elements and internal biases influence current multimodal algorithms is crucial.

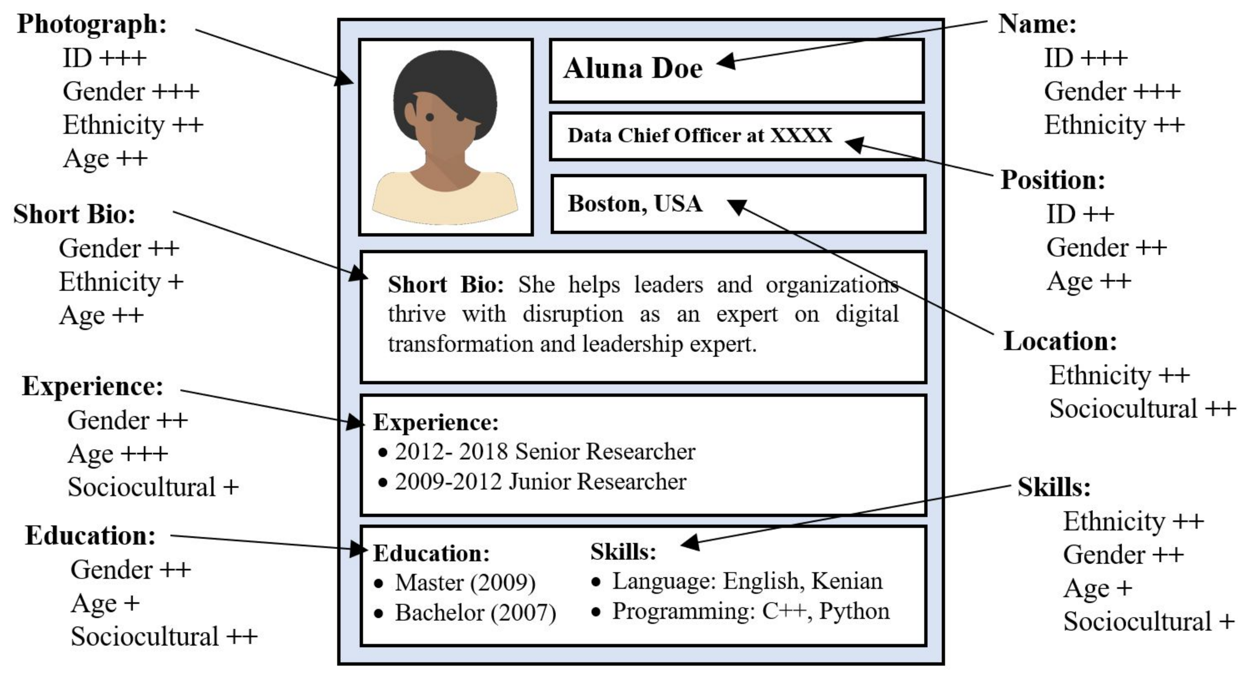
[1] Harwell, Drew. "A face-scanning algorithm increasingly decides whether you deserve the job." *Ethics of Data and Analytics*. Auerbach Publications, 2022. 206-211.

EXPERIMENTAL SETUP

Dataset: FairCVdb[2] for fairness study:

- **Synthetic** research dataset: **24,000** profiles.
- Contains **rich multimodal information** tailored to assess fairness and bias aspects in AI-driven recruitment algorithms.
- **Modalities**: **Visual** (Image), **Tabular** (attributes generated from US Census 2018 Education Attainment data), **Textual** (Short Bio).
- **Protected attributes**:
  - **Gender**: Female, Male.
  - **Ethnicity**: Asian, Caucasian, African-American.

Task: Determining whether the subject should be invited for a job interview.



Information blocks in a resume and personal attributes that can be derived from each one. The number of crosses represent the level of sensitive information (+++ = high, ++ = medium, + = low)

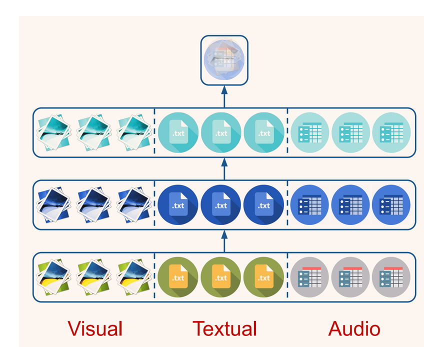
Evaluation Metrics: Mean Absolute Error (MAE) to measure accuracy. Kullback-Leibler (KL) divergence to measure biases between demographic distributions. Gender: compare and report the score distributions for males and females. Ethnicity: compute pairwise comparisons and report the average divergence.

Methodology: Recruitment model to predict scores based on candidate resumes, following the methodology from Peña et al. (2023)[3].

[2] Pena, Alejandro, et al. "Bias in multimodal AI: Testbed for fair automatic recruitment." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020.
[3] A. Peña, I. Serna, A. Morales, J. Fierrez, A. Ortega, A. Herrarte, M. Alcantara, J. Ortega-Garcia, Human-centric multimodal machine learning: Recent advances and testbed on ai-based recruitment, SN Computer Science 4 (2023) 434.

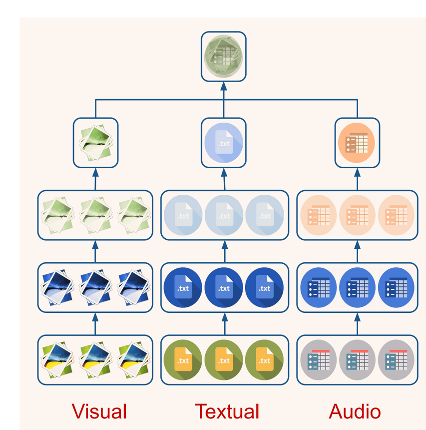
Fusion Strategies: Early and **Late**:

- **Early Fusion** (Feature-Level Fusion): Early fusion occurs at beginning, typically before the data is fed into a neural network.
- Advantageous when the relationships between different modalities are simple.



Early fusion

- **Late Fusion** (Classifier-Level Fusion): Late fusion occurs at the final decision-making stage, after each modality has been processed separately and the decision scores have been calculated.
- Advantageous when modalities have very different data characteristics.

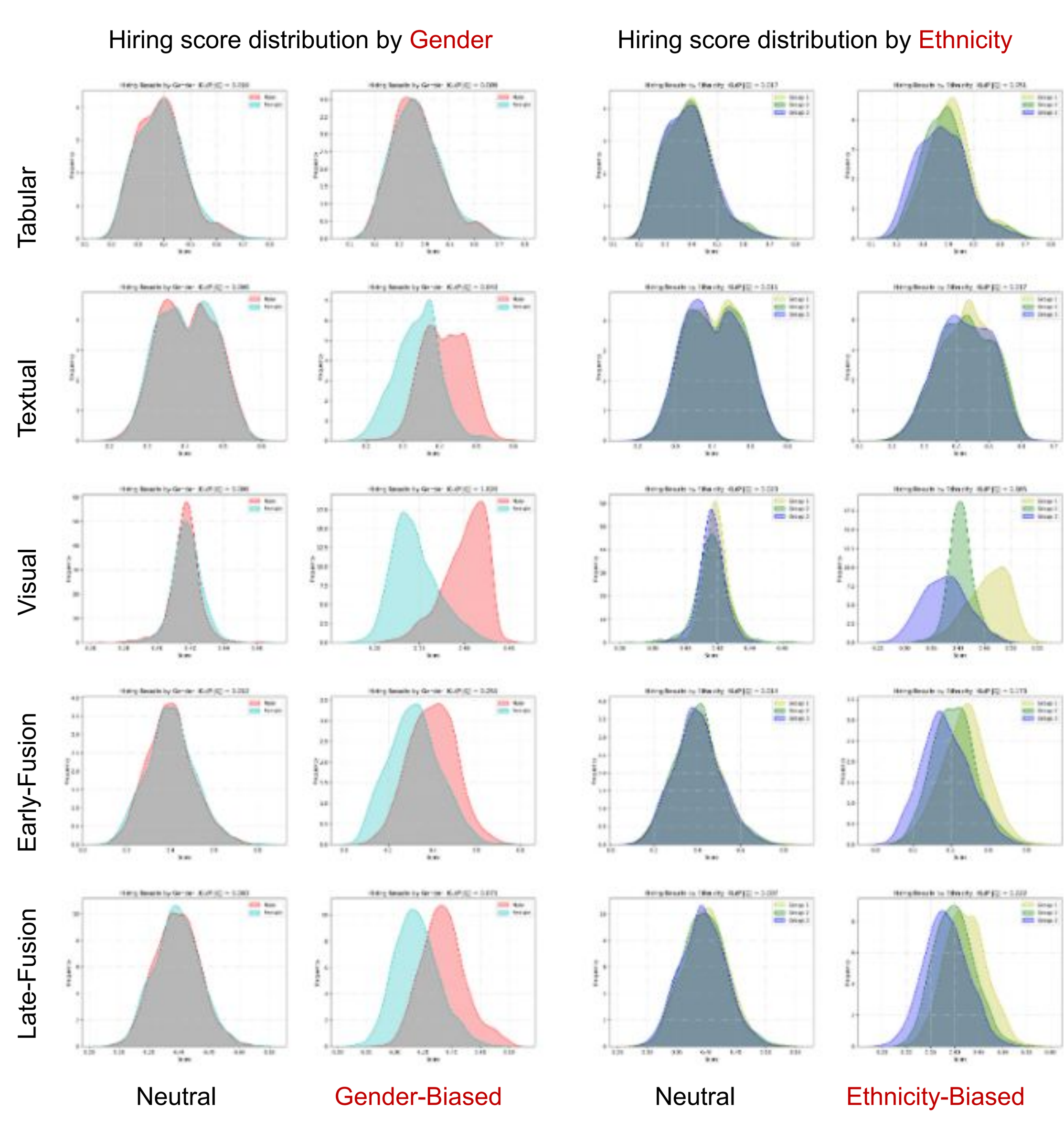


Late fusion

RESULTS & CONCLUSION

KL-divergence between score distributions across Gender and Ethnicity demographics:

- **Neutral**: models trained on unbiased neutral scores.
- **Gender/Ethnicity-Biased**: models trained on Gender/Ethnicity biased scores.
- **Tabular, Textual,** and **Visual**: models trained exclusively on tabular data, textual data, and visual data, respectively.
- **Early-Fusion** and **Late-Fusion**: models that integrate all three modalities at the data representation level and the decision level, respectively.
- **Interpretation**: Lower KL-divergence signifies closer alignment between distributions across various demographics, implying more fairness in hiring.



KL-divergence between score distributions across Gender and Ethnicity demographics.

Key Conclusions:

- Fusion techniques play a crucial role in addressing fairness and bias concerns in AI-based systems.
- Late fusion excels in unbiased ideal conditions, fostering fair decisions. However, it may exacerbate biases under biased conditions by independently learning biased models for each modality fostering undesired correlations.
- Early fusion often leads to fairer outcomes, especially with carefully chosen modalities.
- Not all fusion strategies are universally effective. Models trained solely on tabular data outperformed those using late fusion of multiple modalities in both accuracy and fairness.
- Fusing textual and visual modalities provides richer context and information, resulting in more accurate and fairer outcomes compared to using these modalities individually.
- Assessing the inherent risks of using simulated or synthetic data is crucial to ensuring fairness, transparency, and effectiveness in automated processes.

Future Directions:

- Diversify Fusion Strategies: Investigate beyond traditional early and late fusion techniques to uncover new insights for enhancing fairness in multimodal AI systems.
- Examine Generalisability: Test the applicability of these findings across various datasets and domains beyond hiring to broaden the study's impact and relevance.

For code and additional insights, visit: https://github.com/Swati17293/Multimodal-AI-Based-Recruitment-FairCVdb **or write to:** swati.swati@unibw.de

Funded by the European Union

Funded by the European Union under Grant Agreement ID: 101070285. UK participant in Horizon Europe Project MAMMOth is supported by UKRI grant number 10041914 (Trilateral Research LTD). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission or UKRI. Neither the European Union nor the granting authority nor UKRI can be held responsible for them.