Forschungsinstitut
Cyber Defence
Universität der Bundeswehr München

Universität der Bundeswehr München

Freie Universität Berlin

MAMMOth

Funded by the European Union

STELAR

Co-funded by the European Union

EWAF'24

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

Third European Workshop on Algorithmic Fairness (EWAF'24)

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

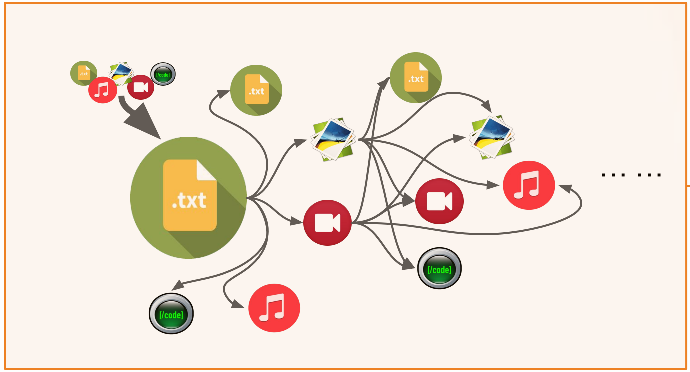
[1]Research Institute CODE, University of the Bundeswehr Munich, Germany,

[2]Institute of Computer Science, Free University Berlin, Germany

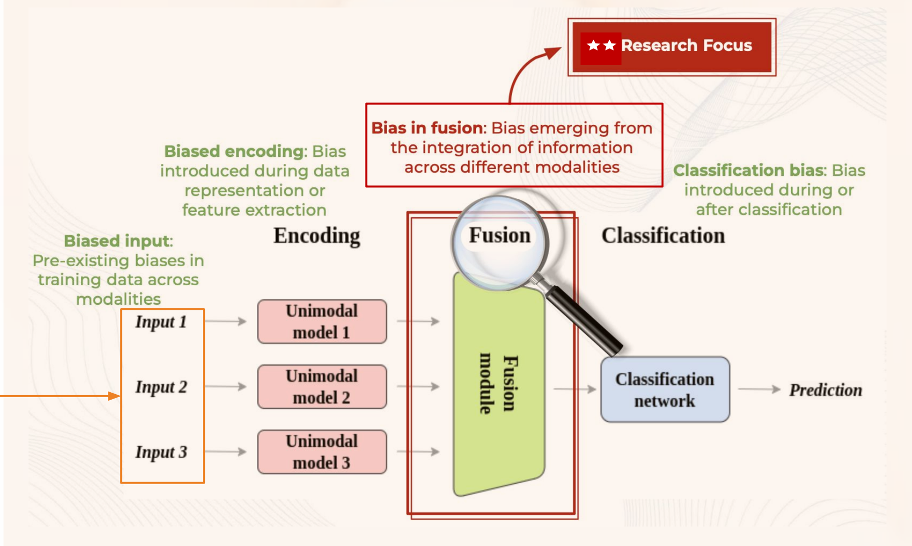
Speaker: Swati Swati
swati.swati@unibw.de

Introduction

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



Multimodal learning integrates data from different modalities.



Bias across stages of multimodal learning.

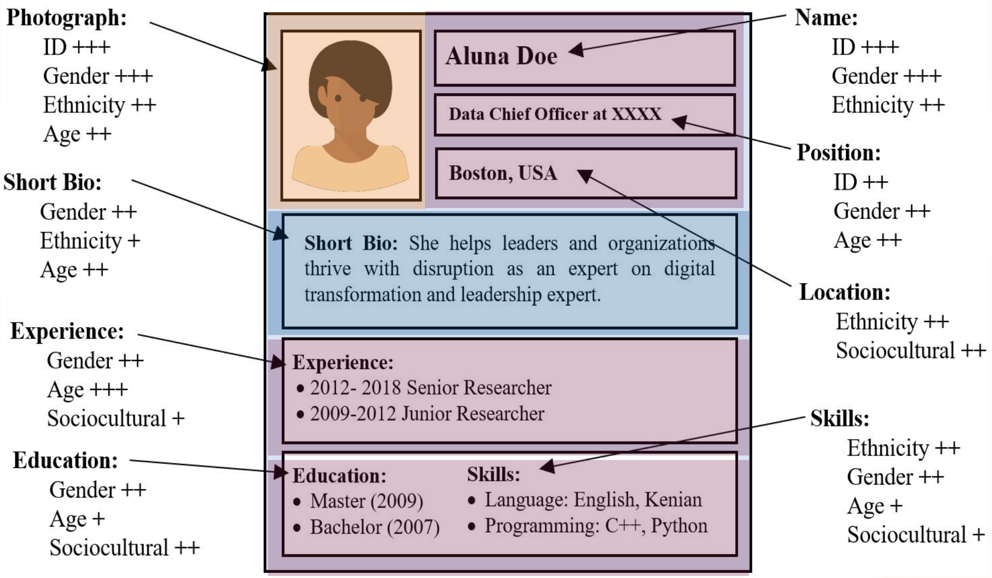
Experimental Setup 1/2

Dataset: FairCVdb[1] for fairness study:

- **Synthetic** research dataset: **24,000** profiles which contain **rich multimodal information** tailored to assess fairness and bias aspects in AI-driven recruitment algorithms.
- **Modalities**: **Visual** (image), **Tabular** (attributes 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.

Evaluation Metrics: Mean Absolute Error (**MAE**) and Kullback-Leibler (**KL**) divergence.

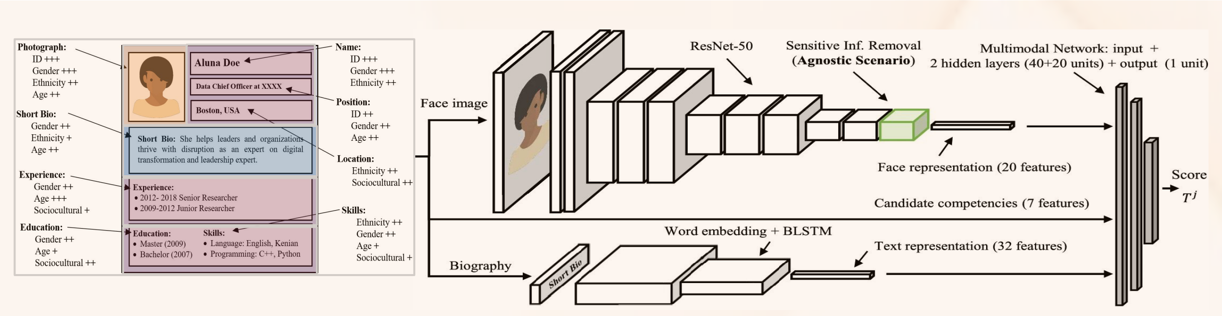


Dataset: FairCVdb. The number of crosses represent the level of sensitive information (+++ = high, ++ = medium, + = low)

[1] 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.

Experimental Setup 2/2

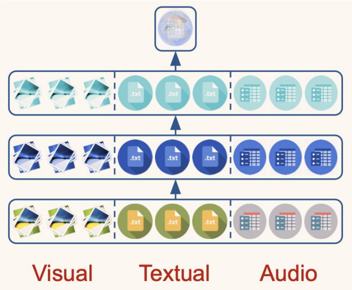
Methodology: Recruitment model to predict scores based on candidate resumes, following the methodology from Peña et al. (2023)[2].



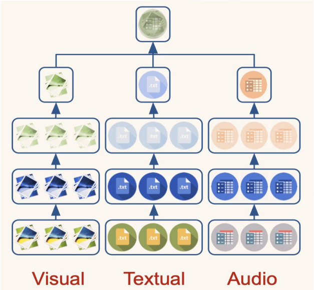
Recruitment model to predict scores based on candidate resumes.

Multimodal Fusion Strategies:

- **Early Fusion (Feature-Level Fusion):** typically occurs before the data is fed into the network.
- **Late Fusion (Classifier-Level Fusion)**: typically occurs at the final decision-making stage, after each modality has been processed separately and the decision scores have been calculated.



Early fusion



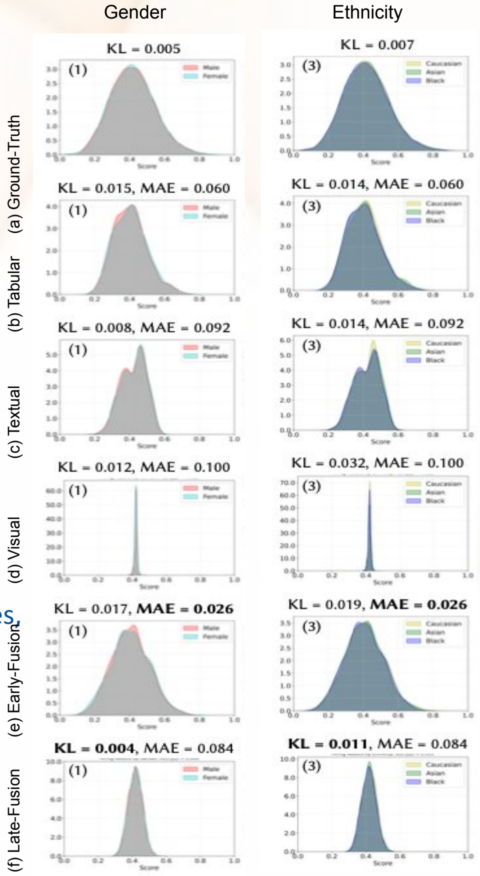
Late fusion

[2] 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.

Experimental Results 1/2

Neutral: Unbiased Ideal-World Scenario:

- Ground-truth: closely aligned for both demographics.
- Tabular: lower score distribution centered around a mean of 0.4 with a negatively-skewed distribution, underestimating the ground-truth.
- Textual: bimodal distribution, differentiates between high and low scores.
- Visual: narrow range [0.39–0.44], over-generalizes mean score.
- Late-fusion: least biased, but influenced by visual extremity, higher MAEs.
- Early-fusion: most accurate, lowest MAEs, effectively resolves modality-specific issues, closely matches ground-truth.

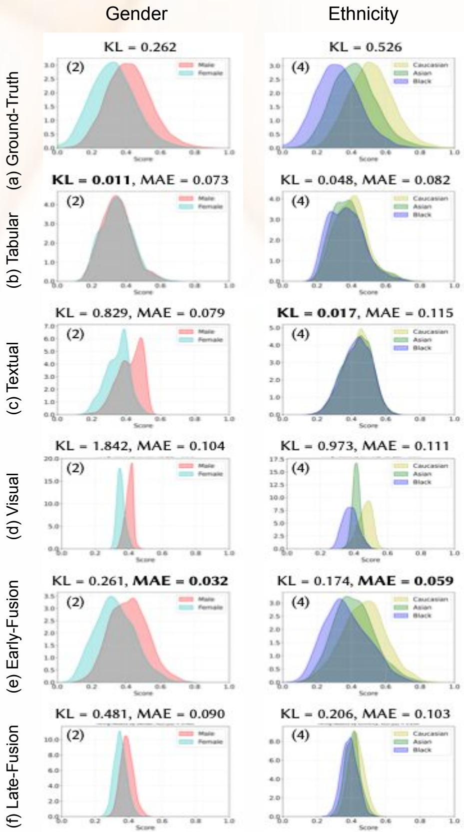


Neutral: KL-divergence, MAE, and score distributions.

Experimental Results 1/2

Gender/Ethnicity Biased: Biased Real-World Scenario:

- Ground-truth: unaligned for both demographics.
- Tabular: underestimates across all demographics, closely aligns demographic-specific distributions.
- Textual: favorably skewed for males in job-related words, no bias in ethnicity, indicating higher gender-skewness than ethnicity-skewness.
- Visual: most extreme bias for both demographics; favors males, overgeneralizes Asians, discriminates against Blacks, and favors Caucasians.
- Early-fusion: mimics ground-truth for both demographics, lowest MAEs, maintains fairness.
- Late-fusion: over-generalizes mean score, higher MAEs and KL scores.



Gender/Ethnicity Biased: KL-divergence, MAE, and score distributions.

Conclusions and Future Directions

Key Conclusions:

- *Fusion techniques* play a crucial role in addressing fairness and bias in multimodal AI. Nonetheless, they have the potential to amplify biases from individual modalities, and blindly fusing them may not lead to optimal results.
- *Early fusion* closely mimics ground truth for both demographics and achieves lowest MAEs by incorporating unique characteristics of each modality effectively. It yields fairer solutions even in the presence of demographic biases.
- *Late fusion* leads to highly over-generalized mean scores, resulting in higher MAEs.

Future Directions:

- Bias-aware fusion strategies: Mid-fusion may enhance fairness and accuracy by strategically selecting and combining modalities.
- Test the applicability of these findings across diverse datasets and domains beyond hiring for broader impact and relevance.

Ethics statement: Understanding the risks of using simulated or synthetic data is crucial for fairness, transparency, and effectiveness in automated hiring processes.

Thank you for your attention!

For code and additional insights, visit: https://github.com/Swati17293/Multimodal-AI-Based-Recruitment-FairCVdb

Swati Swati
swati.swati@unibw.de

0000-0002-7637-6640

@swati17293