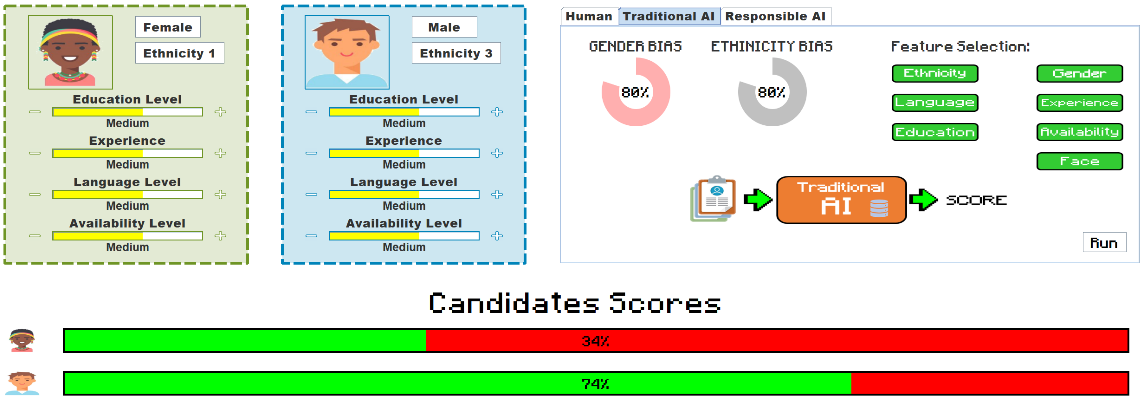
A. Peña, I. Serna, A. Morales and J. Fiérrez. FairCVtest Demo: Understanding Bias in Multimodal Learning. In ICMI '20: ACM International Conference on Multimodal Interaction, October 25-29, 2020, Utrecht, the Netherlads. ACM, New York, NY, USA.

DOI: https://doi.org/10.1145/3382507.3421165

FairCVtest Demo: Understanding Bias in Multimodal Learning with a Testbed in Fair Automatic Recruitment

Alejandro Peña, Ignacio Serna, Aythami Morales and Julian Fierrez
BiDA-Lab, Universidad Autonoma de Madrid, Spain



Female
Ethnicity 1
Education Level: Medium
Experience: Medium
Language Level: Medium
Availability Level: Medium

Male
Ethnicity 3
Education Level: Medium
Experience: Medium
Language Level: Medium
Availability Level: Medium

GENDER BIAS: 80%
ETHINICITY BIAS: 80%

Feature Selection:
Ethnicity
Gender
Language
Experience
Education
Availability
Face

Candidates Scores
34%
74%

Figure 1: Capture of the interactive demonstrator with configurable parameters: candidate skills, amount of bias introduced in the groundtruth data, features used, and method for score generation: Human, Traditional AI, and Responsible AI.

ABSTRACT

With the aim of studying how current multimodal AI algorithms based on heterogeneous sources of information are affected by sensitive elements and inner biases in the data, this demonstrator experiments over an automated recruitment testbed based on Curriculum Vitae: FairCVtest. The presence of decision-making algorithms in society is rapidly increasing nowadays, while concerns about their transparency and the possibility of these algorithms becoming new sources of discrimination are arising. This demo shows the capacity of the Artificial Intelligence (AI) behind a recruitment tool to extract sensitive information from unstructured data, and exploit it in combination to data biases in undesirable (unfair) ways. Aditionally, the demo includes a new algorithm (SensitiveNets) for discrimination-aware learning which eliminates sensitive information in our multimodal AI framework.

ACM Reference Format:
Alejandro Peña, Ignacio Serna, Aythami Morales and Julian Fierrez. 2020. FairCVtest Demo: Understanding Bias in Multimodal Learning with a Testbed in Fair Automatic Recruitment. In *ICMI '20: ACM International Conference on Multimodal Interaction, October 25–29, 2020, Utrecht, the Netherlands.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

One of the most active areas in Artificial Intelligence (AI) is around the development of new multimodal models capable of understanding and processing information from multiple heterogeneous sources of information [1]. Among such sources of information we can include structured data (e.g. in tables), and unstructured data from images, audio, and text. The implementation of these models in society must be accompanied by effective measures to prevent algorithms from becoming a source of discrimination. In this scenario, where multiple sources of both structured and unstructured data play a key role in algorithms' decisions, the task of detecting and preventing biases becomes even more relevant and difficult.

One application scenario of multimodal AI that is gaining in importance and where data biases can be very harmful is automated or semi-automated recruitment. The usage of AI is growing significantly in human resources departments, with video- and text-based screening software becoming increasingly common in the hiring pipeline [2].

In this environment of desirable fair and trustworthy AI, the main contribution of this demo is:

- to enable interactive experimentation with a testbed for Fair Automatic Recruitment, FairCVtest[1], consisting of: 1) 24,000 resumes (including structured and unstructured data from face images and text), 2) various experimental protocols, and 3) popular neural network multimodal learning architectures.

This short demo paper also reports a key experiment showing the risks of traditional AI for a fair treatment of gender in automated recruitment (see Section 3) over a large population based on FairCVtest. The companion paper [4] reports additional experiments around bias effects in multimodal learning and how to overcome them.

ICMI '20, October 25–29, 2020, Utrecht, the Netherlands
ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

This work has been supported by projects: BIBECA (RTI2018-101248-B-I00 from MINECO/FEDER), TRESPASS-ETN (H2020-MSCA-ITN-2019-860813), PRIMA (H2020-MSCA-ITN-2019-860315); and by Accenture.

[1] https://github.com/BiDAlab/FairCVtest

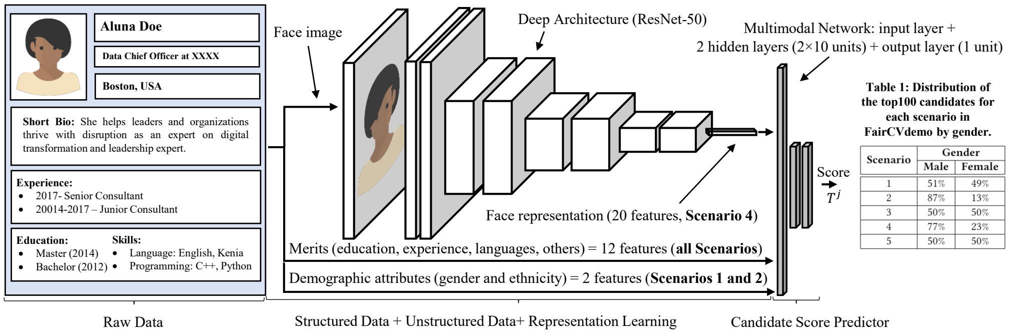

Aluna Doe
Data Chief Officer at XXXX
Boston, USA
Short Bio: She helps leaders and organizations thrive with disruption as an expert on digital transformation and leadership expert.
Experience:
2017- Senior Consultant
20014-2017 – Junior Consultant
Education:
Master (2014)
Bachelor (2012)
Skills:
Language: English, Kenia
Programming: C++, Python
Face image
Deep Architecture (ResNet-50)
Multimodal Network: input layer + 2 hidden layers (2×10 units) + output layer (1 unit)
Face representation (20 features, Scenario 4)
Merits (education, experience, languages, others) = 12 features (all Scenarios)
Demographic attributes (gender and ethnicity) = 2 features (Scenarios 1 and 2)
Score
T^j
Raw Data
Structured Data + Unstructured Data+ Representation Learning
Candidate Score Predictor

Table 1: Distribution of the top100 candidates for each scenario in FairCVdemo by gender.

Scenario	Gender	
	Male	Female
1	51%	49%
2	87%	13%
3	50%	50%
4	77%	23%
5	50%	50%

Figure 2: Multimodal learning architecture composed by a Convolutional Neural Network (ResNet-50) and a fully connected network used to fuse the features from different domains (image and structured data).

2 FAIRCVTEST DEMO: DESCRIPTION

The testbed visualized with this demo consists of 24,000 synthetic resume profiles, as described in more detail in [4]. These resumes include 12 features obtained from 5 information blocks, 2 demographic attributes (gender and ethnicity), a face photograph, and a candidate score (target score). The demo allows to analyze how multimodal algorithms are influenced by biases that are present in the target function over that testbed. Note that the testbed (FairCVtest) is available in GitHub for further research including the resumes, protocols, and baseline results on traditional and responsible AI multimodal learning methods. Fig. 1 shows a capture of the interactive demo where the user selects the following attributes for 2 candidates: gender, ethnicity, and 4 skill levels. Additionally, the user specifies the amount of bias in the training data and the inputs of the recruitment AI (see Fig. 2). The demo shows the candidate score for each specific combination of attributes, biases, and inputs. Finally, the user can select between three models based on Human scoring, Traditional AI, or Responsible AI trained according to the discrimination-aware method SensitiveNets [3, 5].

3 FAIRCVTEST DEMO: EXPERIMENTS

In the experiments summarized here we aim at showing over a large population of the FairCVtest testbed how gender and ethnicity biases can be harmful in recruitment tools. For that we generated 5 different versions (Scenarios, *S*) of the recruitment tool with different Training (T) and additional Inputs (I). *S1:* (T) Unbiased scores, (I) Gender/Ethnicity attributes. *S2:* (T) Gender-Biased scores, (I) Gender attribute. *S3:* (T) Gender-Biased scores, (I) No Gender attribute. *S4:* (T) Gender-Biased scores, (I) Feature embedding from the Face image. *S5:* (T) Discrimination-aware training with Gender-Biased scores, (I) Feature embedding from the Face image.

Note that other scenarios can be generated interactively in the demo by changing online some parameters, see Fig. 1. In all 5 Scenarios, we designed the candidate score predictor as a feedforward neural network with two hidden layers, both of them composed by 10 neurons with ReLU activation, and only one neuron with sigmoid activation in the output layer, treating this task as a regression problem (see Fig. 2).

The recruitment tool was trained with the 80% of the synthetic profiles (19,200 CVs) described in [4], and retaining 20% as validation set (4,800 CVs), each set equally distributed among gender and ethnicity, using Adam optimizer, 10 epochs, batch size of 128, and mean absolute error as loss metric.

We assume that the recruitment tool will be used to realize a first screening among a large list of candidates including the 4,800 resumes used as validation set in our experiments. We simulate the candidates screening by choosing the top 100 scores among them (i.e. scores with highest values). We present the distribution of these selections by gender in Table 1 (see Fig 2). In Scenarios 1 and 3, where the classifier shows no gender bias, we have almost no difference in the percentage of candidates selected from each gender group. On the other hand, in Scenarios 2 and 4 the impact of the bias is notorious, with a difference of 74% in Scenario 2. However, when the sensitive features removal technique is applied [3], the demographic difference drops from 54% to 0%, effectively correcting the gender bias in the training data. These results demonstrate the potential hazards of these recruitment tools in terms of fairness, and also serve to show possible ways to solve them.

4 CONCLUSIONS

We present FairCVtest, a testbed consisting of 24,000 resumes (structured and unstructured text and face images) helpful to study how multimodal machine learning is affected by biases present in the training data. Using FairCVtest, we can study the capacity of common deep learning algorithms to expose and exploit sensitive information from commonly used structured and unstructured data.

This demo allows to evaluate recent methods to prevent undesired effects of these biases, and to improve fairness in this AI-based recruitment framework.

REFERENCES

[1] T. BaltruÅąaitis, C. Ahuja, and L. Morency. 2019. Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41 (2019), 423–443.

[2] S. Chandler. 2017. The AI chatbot will hire you now. *Wired* (Sep. 2017). https://www.wired.com/story/the-ai-chatbot-will-hire-you-now/

[3] A. Morales, J. Fierrez, R. Vera-Rodriguez, and R. Tolosana. 2020. SensitiveNets: Learning Agnostic Representations with Application to Face Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).

[4] A. Peña, I. Serna, A. Morales, and J. Fierrez. 2020. Bias in Multimodal AI: Testbed for Fair Automatic Recruitment. In *Proc. of IEEE CVPR Workshop on Fair, Data Efficient and Trusted Computer Vision.*

[5] A. PeÃśa, J. Fierrez, A. Lapedriza, and A. Morales. 2021. Learning Emotional Blinded Face Representations. In *Proc. of IAPR Intl. Conf. on Pattern Recognition.*