Research Article

Carmen Fernández-Martínez* and Alberto Fernández

AI and recruiting software: Ethical and legal implications

https://doi.org/10.1515/pjbr-2020-0030
received February 11, 2019; accepted March 12, 2020

Abstract: In this article, we examine the state-of-the-art and current applications of artificial intelligence (AI), specifically for human resources (HR). We study whether, due to the experimental state of the algorithms used and the nature of training and test samples, a further control and auditing in the research community is necessary to guarantee fair and accurate results. In particular, we identify the positive and negative consequences of the usage of video-interview analysis via AI in recruiting processes as well as the main machine learning techniques used and their degrees of efficiency. We focus on some controversial characteristics that could lead to ethical and legal consequences for candidates, companies and states regarding discrimination in the job market (e.g. gender and race). There is a lack of regulation and a need for external and neutral auditing for the type of analyses done in interviews. We present a multi-agent architecture that aims at total legal compliance and more effective HR processes management.

Keywords: domain-specific AI, ethics, human resources, recruiting

1 Introduction

In recent years, artificial intelligence (AI) has been increasingly used to support recruiting and human resources (HR) departments. AI has transformed several sectors of the society, but the impact has been largely unfelt in HR, where AI has had no more than an assistant role so far. Over decades, authors have reinforced the concept of strategic and global HR management [1]. More recently, Kapoor and Sherif [2] stated that the business environment is evolving into a more complex system with a diverse workforce and that it is essential to work with business intelligence and AI to improve HR practices. Traditionally devoid of bias, AI proved very valuable for résumé and keywords scanning and the extraction of candidate skills. There has been a recent trend towards video-interview analysis in HR. Researchers such as Strohmeier and Piazza have noted that there is a lack of comprehensive studies on the potential of AI in the field [3]. A survey by Personnel Today found that 38% of enterprises are already using AI in their workplace with 62% expected to be using it by 2018 [4].

Nowadays, the idea that AI will replace humans in the workplace is widespread. A job consists of a number of tasks. There is evidence from research on the practicality of labour displacement due to automatization and robots [5]. In HR, more and more tasks are now fully automatized and powered by algorithms, though still relying on human recruiters to make the final decision of who is interviewed on-site or who is hired. This matter has reduced candidate time and the recruitment process significantly.

The demands of a global society and the excessive recruiting time for HR managers in multinational companies have changed the traditional way of interviewing. The field is now leaning towards algorithms that produce rankings of candidates according to different features, and several commercial products are being used extensively in HR departments.

Also, large multinationals have recognized the value of several video-interview products and companies:

- *HireVue*:[1] Start-up incorporated in 2004 in Utah (USA). Its leading voice and face recognition software compares and grades candidate's word choice, tone and facial movements with the body language and vocabularies of a company's best former employees. Using models, HireVue analyses thousands of features and their correlation with the company's top performers.

* **Corresponding author: Carmen Fernández-Martínez,** CETINIA, University Rey Juan Carlos, Madrid, Spain, e-mail: carmen.urjc@gmail.com
Alberto Fernández: CETINIA, University Rey Juan Carlos, Madrid, Spain, e-mail: alberto.fernandez@urjc.es

1 https://www.hirevue.com/

Finally, it produces a ranking list and an insight score (0–100) for candidates. The final recruiting decision is in the hands of human recruiting. The system is supported by a patented deep-learning analytics engine named HireVue Iris™.

- *Montage*,[2] *SparkHire*,[3] *Wepow*:[4] Mobile and video interview software or SAAS (Software as a Service). There are preferences for the choice of a video-interview product depending on the geographic area. For example, the customer base of SparkHire is predominantly based in the United States of America (USA), whereas Wepow has a bigger market in Europe and some sectors like pharmaceuticals.[5]
- *Blizz*:[6] Cloud-based tool for video interviewing offering the possibility of secure chat and two-way HD video and audio. It was launched in 2013.
- *Human*:[7] Human, a start-up founded in London in 2016, analyses subliminal emotions in video-based job applications in order to predict human behaviour.
- *Affectiva*:[8] Worldwide emotion database with over five million faces that provides emotion measurement technologies, the basis for applications of image processing in HR. In these software products, it is essential to define the traits and facial gestures that are related to each competency assessed in a recruitment process. Some other features measured by Affectiva are listed in categories or ranks, e.g. age (18–24, 25–34, 35–44, 45–54, 55–64, 65+, under_18, unknown). Affectiva can track emotions such as *anger*, *contempt*, *disgust*, *engagement*, *joy*, *sadness*, *surprise* and *valence* (a measure of the positive or negative nature of the recorded person's experience). As an example, *engagement* implies the following set of facial gestures: *brow raise*, *brow furrow*, *nose wrinkle*, *lip corner depressor*, *chin raise*, *lip pucker*, *lip press*, *mouth open*, *lip suck* and *smile*.

This article examines the effects of new trends in AI and algorithms in HR. We analyse the implications of video-interview analysis and ranking of candidates, the inclusion of minorities in a diverse pool of candidates and, finally, the effectiveness of recruitment algorithms under different legislation.

After careful analysis of systems and techniques oriented to extract qualities and features in employees, in this work we discuss on some of them presumably controversial, both explicit features like physical appearance, voice, intonation, keywords and race and implicit traits such as optimism, customer-oriented personality, sexual identity and similarity to former employees. These advances bring a broad range of ethical issues – accuracy and privacy – for both candidates and companies who trust a particular recruiting product to assist in the recruitment process in search of efficiency. The domain-specific AI for HR research covers a whole spectrum of disciplines and backgrounds, including cognitive science and psychology and artificial vision.

The primary aim of our work is to analyse the problem of recruiting and working towards a proposal of automated multi-agent software architecture to support auditing HR. We have designed the architecture as general as possible in order to distribute different tasks of recruitment and auditing processes in different agents and also allowing the surveillance of different types of scenarios that require user's approval and legal checks.

The rest of the article is organized as follows. We first present (Section 2) some controversial features frequently measured in image and voice recognition using AI techniques. In Section 3, we present basic information about machine learning (ML) techniques to detect the above-mentioned controversial features. Section 4 focuses on the advantages and disadvantages of AI image analysis frequently used in recruiting. In Section 5, we discuss the legal and ethical issues related to the inclusion of minorities, gender equality and applicants' right to a fair and informed selection process. Then, in Section 6, we present a software architecture to support auditing recruiting processes. Section 7 includes a discussion of the analysis carried out in this article and some future lines for AI-based recruiting algorithms. We conclude the article with Section 8.

2 Controversial features of image and voice processing

Recently, there has been some critical analysis of advanced AI and image processing and face recognition. A recent Stanford study [6] claimed that it is possible to guess sexual identity and its traits through face recognition technology.[9] The study had to be reviewed and approved by the American College of Psychologists before publication, given

2 https://www.montagetalent.com/
3 https://www.sparkhire.com/
4 https://www.wepow.com/es/
5 https://siftery.com/product-comparison/spark-hire-vs-wepow
6 https://www.blizz.com/en/
7 https://citysail.co.uk/
8 https://www.affectiva.com/

the potential danger and ethical concerns it raised. This reminds us of the pseudoscience level of some of these studies that could be considered similar to *phrenology*, an outdated theory that connects skull, head and traits with behaviour, personality and criminal tendencies. The idea goes beyond morphology and questioned issues related to grooming, e.g. using less eye makeup or less revealing clothes as evidence of homosexuality in women. This idea leads us to mention the concept "soft biometrics", the analysis of body geometry, marks and tattoos that, unlike primary biometric traits, can be obtained at a distance without cooperation [7]. Several researchers [8–11] studied up to what point it is possible to infer and model race in face recognition. "How implicit, non-declarative racial category can be conceptually modelled and quantitatively inferred from the face?" as suggested in the survey on race analysis by Siyao et al. [8]. They mention implicit race detection and categorization and learning race implicitly from the face, going beyond empirical demonstration. Racial categories should not be taken for granted, so they propose a proper race categorization model following qualitative analyses (of features and so on) and not quantitative. Other significant studies concentrate on a particular geographical group [9–11].

2.1 Physical attractiveness

Facial symmetry is especially important when it comes to detecting physical attractiveness and beauty [12]. For example, a beautiful person would have an intraocular distance that is coherent with eyes size, a good ratio face width/height and a mouth width proportional to the rest of the face.

These observations pose different questions about the legitimacy of measuring the beauty of job applicants – especially in countries where CVs without photos are preferable.

Studies to evaluate facial beauty are common in the scientific community [12,13], and some apps have become popular among Internet users. ML is the key to this kind of automatic evaluation and is supported by subsets of training and test data. According to automatic evaluation of facial attractiveness, the seed training data were gathered from sites like Hotornot,[10] where users can rate females, choosing the over-average voted photos for the data set.

2.2 Age

Regardless of age, there should be some inspection of enterprises' policies to avoid a preference for employees of a particular age group. Ageing follows the same patterns in every individual. Some of them, like the appearance of wrinkles, can be delayed as long as the person follows some cosmetic procedure, e.g. Botox or fillers. However, some changes like thinning and the decrease in dental arcs cannot be avoided easily. It is easy to study morphologic changes in faces and determine an approximate age. The issue of querying candidate age is forbidden in many countries to promote the inclusion of all citizens.

Next, we include an analysis of the basic features, lines, wrinkles and morphologic changes in face measured in image processing to detect age, as appear in recent research such as Hayashi (Table 1) and colleagues [14,15] and research on age and gender using convolutional neural networks [16]. Recently, Yoti [17] published a White Paper related to the application of facial recognition technology and age tracking with AI. It was already suggested for its use in retail and supermarkets in the UK and possibly hiring. The system relies on indicators of age and screens out applicants according to a 10-year error margin. For example, screening out over 28's when it comes to detecting underage customers and proceeding with the request of ID cards for alcohol drinks purchase.

2.3 Race

One of the most controversial characteristics that could be analysed with video interviewing is race. It is unfair, and it must be said that most algorithms and classifiers are trained with images of White people, not performing well with Black people. As Buolamwini and Gebru [18] pointed out, there is enormous controversy regarding the inclusion of racial discrimination in algorithms. It has been shown that there are soap dispensers that detect easier White people than others [19]. Recent research and theories on analysis of race, like Siyao et al. [8], recall about the

9 https://www.washingtonpost.com/news/morning-mix/wp/2017/09/12/researchers-use-facial-recognition-tools-to-predict-sexuality-lgbt-groups-arent-happy/?utm_term=.524d8912daab

10 www.hotornot.com

Table 1: Wrinkle progression with age

Event type	Description
10 years male	None
10 years female	None
20 years male	Under the eyes
20 years female	None
30 years male	Under the eyes, at the corner of the eye
30 years female	Under the eyes
40 years male	Under the eyes, at the corner of the eye, on the cheek
40 years female	Under the eyes, at the corner of the eye
50 years male	Under the eyes, at the corner of the eye, on the cheek, on the brow
50 years female	Under the eyes, at the corner of the eye, on the cheek, on the brow

The table was extracted from Hayashi et al. (2002) [14].

complexities of facial recognition of non-White individuals, especially women, during the night or with bad light conditions. This survey argues that the development of systematic methods for race detection is more difficult in comparison with other attributes. Collecting sample images for races is not that easy on the grounds of prejudice. The analysis is more qualitative than quantitative, and the definition is ambiguous for the race. How many racial categories do exist? Whereas for other facial analysis, the categories are known beforehand: one-to-one matching in face detection, two genders and a closed list of six universal emotion types [8]. The lack of accuracy in a subgroup of dark-skinned women stands out in the study by Muthukumar et al. [20]. Other authors [21,22] have noted the difficulties in face detection and race in unconstrained captures, especially the blurred ones. We recall briefly that ML techniques were used for these systems. They were initially designed for White people or trained predominantly with images of White individuals, both training and test data sets. We need to carry out further studies to include accuracy concerning race for law enforcement algorithms, mainly when law enforcement authorities are excessively controlling the Afro-American and non-White community in countries like the USA. Past works have claimed inaccuracies in the face recognition systems used by police [23] and surveillance cameras [24].

2.4 Analysis of gender and sexual orientation in images and voice

As an illustration of the advances of sexual orientation recognition both in images and in sound, it is important to mention some recent experiments carried out in Italy and the USA. Both studies needed ethical supervision.

Kosinski and Wangs' work [6] shows theories based on a morphologic study of homosexuality. The study argues that there is a relationship between homosexuality and exposure to a particular concentration of hormones in the womb and that morphological features can determine the sexual orientation (for example, more masculine or feminine jawline and forehead). This research was contrasted with what users disclose in dating applications. The opaque and too invasive nature of this research required advice and permission of the American College of Psychologists before publication.

The voice-based categorization study [25] reminds us, however, that every language has its particularities and that a voice that sounds more feminine in a man or *vice versa* may be due to anatomical reasons or more exposure to feminine voices during childhood. The sexual orientation guessing proved to be more explicit for listeners in Italian or Romanic languages, but the categorization works similarly both in German and in Italian, where some gay speakers were consistently misidentified as heterosexual and *vice versa*.

2.5 Use of keywords in speech-affirmative/negative connotation of words

AI systems have been able to detect the frequent words that could identify an optimistic and convincing salesperson. Since every language has its connotations, and the use of more or less optimistic phrasal constructions, proverbs and popular expressions could lead to misinterpretations of what candidates want to express or generalize to the English language.

According to recent sources [26], Google translator still kept gender bias until the end of 2018 and is currently working towards removing it. The last updates include multiple translations for neutral meanings.[11] For instance, so far it links "she" to nouns and adjectives traditionally feminine (e.g. "nurse", "teacher"), "he" with "hardworking", the Chinese pronoun to refer to "they" is naturally translated into English as masculine "he", etc. According to Leavy [27], the presence of a majority of men in the design of these technologies is the root of the problem. These male-dominated environments could unbalance the percentage of masculine and feminine

11 https://www.blog.google/products/translate/reducing-gender-bias-google-translate/

words. There is a gender perspective of translation [28]. Individuals tend to translate neutral words according to their own gender. This is very often the case of automatic translations involving Chinese language.

3 AI techniques for identifying controversial features

Following, we give an insight into the existing literature and the main ML techniques used for face/patterns recognition in domain HR, and more specifically video-interview systems over the last few years. Going into detail, we analyse traits and characteristics, race, gender and so on, mentioned in the previous sections.

The initial interest in the analysis of surveillance of video images and streaming has promoted the current state-of-the-art in facial recognition related to video-interview systems. However, current systems give weight to the specific responses and subtle details from candidates, even considering controversial features like race or age. While techniques for facial feature analyses multiply and the video-interview systems gather more data points, related to image or audio analysis, survey and review become important. Before starting, it is important to acknowledge the heterogeneity and duplicity of studies that call into the question of which ML technique is the most efficient.

First, it is worth mentioning the research conducted by Viola and Jones [29]. Actually, they do not present an ML technique but a rapid method to detect a face in an image, namely, an image-processing technique. The article sets the ground for new research making use of the Adaboost classifier for this purpose. The next step would be identifying traits and contours. Masood et al. [30] proposed a normalized forehead area calculation using the Sobel edge detection method [31]. The following prediction of human ethnicity study [30] acknowledges that some races, Black for instance, have higher forehead ratios.

It is important to keep in mind that we address an experimental field. We have given a summary of the main controversial features in image processing in the previous sections. Even though there is extensive literature on the matter, we should not generalize on which ML methods are the best for every characteristic. According to the previous literature related to the recognition of human gender, the Viola–Jones is considered an important milestone. Viola and Jones [29] introduced rectangle or Haar-like features for rapid face detection and was used by other authors for real-time gender and ethnicity classification of videos in Shi and Tomasi [32].

A study mentioned above targeting sexual orientation [32] gives importance to the ratio of the forehead area, whereas the aforementioned works apply this analysis for race detection. Presumably, the ML techniques used could be interchangeable.

Concerning physical attractiveness and according to relevant authors [13], some main ML algorithms used to train classifiers are K-nearest neighbours (k-NN), artificial neural networks (ANN) and Adaboost. k-NN is dependent on choosing the class according to a local neighbourhood and offers better results than ANN, where it is difficult to understand the interpretation. The philosophy behind Adaboost is combining several weak classifiers to make a strong classifier and offers positive results.

With detecting race, as with physical attractiveness, k-NN, ANN and Adaboost are some of the ML algorithms used in race detection. Siyao et al. [8] acknowledged in their main paper about race recognition having difficulties in detecting mixed race.

Recognition of emotion has always been crucial in domain human–computer interaction, interactive video games or artificial vision for health. As of today, an essential part of the video-interview analyses is the feature extraction and emotion recognition tool. Abdulsalam et al. [33] acknowledges that an influential milestone in the analysis of facial expressions was the work of Paul Ekman [34] who described a set of six basic emotions (anger, fear, disgust, happiness, sadness and surprise) that are universal in terms of expressing and understanding them. According to the comparative study [35], it is important to differentiate among techniques used for feature extraction and those used for classification. The main ML techniques used for features are Gabor filters, a histogram of oriented gradients and local binary pattern for feature extraction, and for classification, support vector machine, random forest and nearest neighbour algorithm (k-NN).

4 Pros and cons of image analysing in HR

We begin a description of the primary and controversial features that could be measured through image and video analyses. An AI-based video interview system could be programmed to gather the following types of features – *lighting*, *eye contact*, *tone of voice*, *cadence*, *keywords* used (substantial conversation), etc. and infer

features such as *age*, or states, e.g. *mood* and *behaviour* (eccentric, movement or quite calmed and not talkative). Nowadays, there is much emphasis on emotions. They could detect specific personality traits that employers want in their teams, sign of customer oriented or salesperson. Recently, both companies and researchers have worked on emotions and sentiment detection, e.g. company Affectiva, Dehghan et al. [36] or algorithm Emotionet [37].

Some benefits of domain-specific AI for HR and more specifically AI-powered video interviewing, are as follows:

- *Time*. Reducing selection process time and candidate time/travel distances.
- *Recruiting task easier*. Possibility of reviewing videos and interviews several times.
- *Customized candidate experience and customized questions*. Possibility of repeating and clarifying questions. HR managers can customize questions and answers and set an individual score for each trait too.
- *Attention to detail*. Systems like HireVue can detect up to 25,000 human feature data points in each interview.[12] It can outperform the HR recruiter in the number of characteristics to be analysed simultaneously.
- *Attention to eye time*, emotions/intonation and body language crucial for customer-facing roles.
- *Lack of human bias related to physical appearance, accessories, tattoos and regional accents*. Recruiter bias has been empirically examined in previous works [38]. If well designed, the system should favour a focus on responses and keywords rather than being influenced by such bias.

Concerning problems that accompany the use of these technologies, it is important to highlight the *imprecision of this technology* and *gender and racial bias*. That prevents of hiring intuitively or causes selection of individuals of similar characteristics to the initial training set. These observations pose challenges if the predicting AI is fed with data based on, for instance, top performers, blue-eyed blonde candidate residents in London and close-to-native English. Over time, companies and start-ups will develop their own AI algorithms. For the time being, they are dependent on limited public databases that do not stand out for their diversity. Database Affectiva,[13] for instance, was initially fed with images from North American, more concretely Super Bowl[14] viewers, expressing different emotions, in English, as a matter of fact, rich in intonations too. Thereby, it is difficult to guarantee that other races and cultures – like Asians – show the same level of expressiveness as North American, traditionally shown as more expressive and talkative than other cultures. This could be mistaken by inaction or lack of drive – e.g. necessary for commercial roles. There are recent works related to emotions detection such as Deghan et al. [36] that labelled photos with up to seven emotional levels, “happy”, “sad”, “neutral”, “disgusted”, “fearful”, “angry”, “surprised”. It is understood that the categories are a subset not generalizable to every culture. Table 2 shows the main problems identified.

5 Ethical and legal implications of AI in recruiting

In Section 4, we approached the pros and cons of recruiting software. The pros are related to technical limitations, user cognitive experience and HR business processes. Next, we focus mainly on ethical and legal implications of software, not only caused by software limitations but also related to companies’ deliberated choices for interview design.

The analysis of some features poses challenges for governments and regulatory entities. In certain countries, it is illegal to ask some questions to the candidates (e.g. age) and the candidates should not include personal details, like the birthdate, in their resumes. They are asked to fill informative questionnaires asking about ethnicity, gender, sexual orientation or upbringing at the end of the selection process. It is advisable to look at frequently asked questions in the USA interview settings that are illegal [39]. HR managers often ask questions on issues such as marital status, children, previous arrests, country of origin, first language, debt status, social drinking and military discharges. The anti-discrimination laws cover different sectors, from employment to insurance, housing and banking [40].

Although law scholars have used the term *discrimination by proxy* extensively for years, we are seeing a trend on the analysis of intentional and unintentional proxy discrimination related to the effects of AI and big data analysis [40,41]. Next, we describe in brief the possibility of *discrimination by proxy*, but it is

12 https://www.hirevue.com/blog/assessments-and-big-data-can-we-predict-better-hires

13 https://www.affectiva.com/

14 https://www.nfl.com/super-bowl

Table 2: Problem types of image analysis in video interviewing

Problem type	Description
Candidates are unfamiliar with video-interviewing analysis	In selection processes, there are two central values for candidates: trust and information. The candidates are usually unfamiliar with this type of technology and the settings to perform well in the interview and maximize success rate, even though they are given the option of exploring the tool beforehand. Some gestures like looking away from camera are certainly mistaken for global performance. Self-confidence is compromised due to a distorted image in smartphones
Imprecisions of technology and imperfect training sets	The technology does not reflect reality at a maximum. AI is imprecise because a limited and somehow biased data set supports it. The first bias could influence the algorithm in subsequent iterations
Gender and nationality bias	Google is deemed to be biased towards men's voices. When detecting gay/straight, the cues and sounds of some languages/accents are more prone to be considered homosexual by listeners and classifiers
Training classifiers with too generalist and broad data sets	Very generalist data sets were used for the training test, for instance, human. The companies use "proprietary data" from former employees who performed well for contrasting profiles, leading to endogamy and unfair results. Racially biased learning algorithms constitute a fundamental problem in computer science too. Problems of race and mixed-race recognition
Technical limitations	The inaccuracy of race detection (or skin colour) in AI analysis if the flashlight falls upon directly on the face or with night conditions. Discrimination based on actual races without ever knowing it

understandable that it also leads to unfair decisions in HR. Proxy discrimination means to discriminate in favour or against a protected class or a protected group (e.g. race, people with higher incomes and so on) but based first on legitimate or innocuous grounds (e.g. zip code). These reasons for discrimination are highly correlated with belonging to a protected class (e.g. racial minorities living in the same areas or ghettos). So this segregation finishes affecting the protected groups and being the reason for more discrimination. When there is a human actor behind this favourable treatment or less favourable, we speak of "intentional" discrimination by proxy [40]. However, the nature and mix of the data sets used for testing and training could also lead to "unintentional" discrimination by proxy. Relying on external sources often causes more algorithm bias. It is often the case that the company holds that they can justify a practice as a business necessity if it is a predictor of future performance and therefore is not infringing equal employment legislation. Even though the practices resemble historic patterns of discrimination or use limited data sets [42].

It is noted that algorithms can detect, and sometimes protect against, indirect discrimination resulting from automated analyses. Pedreschi et al. [43] describe well in their work the possibility of finding evidence of discrimination in automatic decision support systems (DSS), used for screening purposes in social-sensitive tasks, including access to credit or mortgage, tracking if the input (innocuous candidate data like ZIP) could lead to a wrong yes/no decision. Hajian and Domingo-Ferrer [44] propose techniques which are effective at removing direct and/or indirect discrimination biases. So they reinforce Pedreschi's proposal, in their paper, of inducting patterns that avoid discriminatory decisions despite drawing from training-biased data sets.

These cases lead us to think how future generations of students and candidates should prepare for these interviews and how the HR field will develop? Companies are given the option of customizing their systems and have their own agency for selecting the features of their future employees. Would it be ethical to classify men or another subset in the first place? This idea of black box assessment does not help in the empowerment of some collectives, which could be marginalized or be given unfair treatment in the round of interviews.

Even though reports have shown that the results of HireVue systems are highly diverse and show even more presence of non-White candidates and gender equality, there is an element of disinformation about what exactly

this software targets for every position, junior, managerial and so forth. This is in line with the arguments discussed in the previous section. Not only does technology bring with it drawbacks related to the technical limitations and biased data sets, but also the disinformation and the prejudices worsen the problem. Legal protection related to information gathering from minorities or protected groups causes more inaccuracy. Typical and classic algorithms in computer science have a better effect in White people, just because the technology was initially created in their origins by White men and it was tested initially with a majority of White people.

Another issue of debate is the *legal implications* of some analysis. Image recognition technology is so widely developed to the point that it could even identify age or sexual orientation of a candidate through the analysis of their facial features [6,45,46]. As mentioned earlier, Google is deemed to be biased towards men voices, which questions the legal validity of most of these advances. Algorithms could be tuned to detect real age according to the biological factors, skin type, baldness and so on. Under Spanish law, it is possible to pay fewer taxes for under 30-year-olds and over 40-year-olds. For cost-saving factors, AI could be tuned to discriminate against senior people or young people, could seek for women or minorities or reject them at employer discretion or based on arbitrary or temporary employment. Sexual orientation detection follows similar patterns. In this light, some critics – including the American society of psychologists – have issued a review before permitting the publication of the work of Wang and M. Kosinski [6], arguing that it is illegal and against the fundamental civil rights to use this kind of software.

6 Towards automated support for auditing HR

It is evident that the legal issues concerning candidates, such as the identification of race and sexual orientation in the selection process due to the advances in image processing, entail ethical questions that cannot be ignored. Careful analysis from auditing bodies, governments, ethics committees and psychologies is needed.

In this section, we describe a proposal of a multi-agent software architecture for auditing, which is depicted in Figure 1. Coordination [47] and interoperability of agents in heterogeneous domains have been widely used in different domains, such as healthcare [48], emergencies due to natural disasters [49], smart cities [50], etc. Multi-agent systems (MASs) have also been used to improve business processes in complex organizations. Practicalities

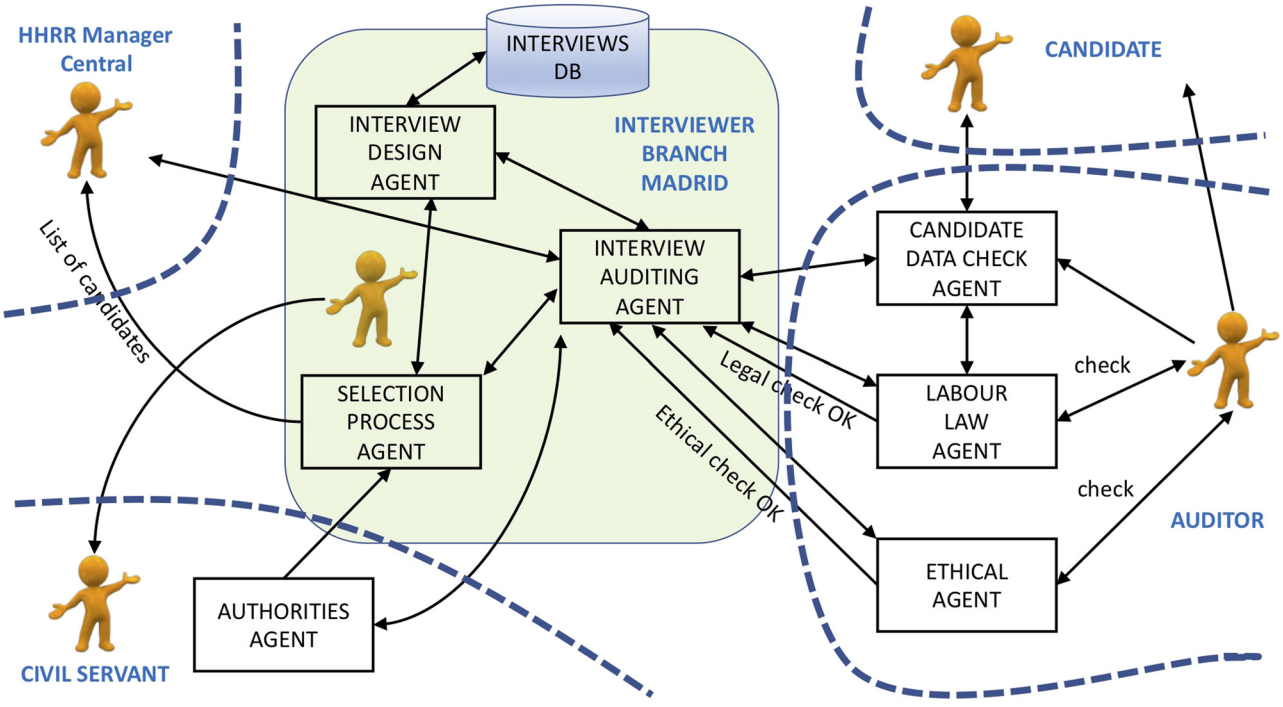

Figure 1: Multi-agent system architecture for HR auditing.

of big multinational corporations and the difference in formats and regulations make it necessary to create architectures and models to rule over changing specific agent populations. Architectures have been proposed to deal with the random and quick changes in a particular productive section, such as the manufacturing industry, but could be applied to other dynamic corporate environments [51]. The recruiting scenario in a multinational context is also quick and complex and needs models. So far there is not much literature related to applications of HR and MAS as an enabling technology, but these types of architectures have been extensively used, as noted above, in manufacturing and corporate control production (e.g. Ciortea et al. [52]). The domain of law is open to MAS applications and has been addressed over the last decade by law scholars such as Walker [53].

Figure 1 shows an abstract MAS architecture that must be adaptable to different international corporate environments and by recruiters of different nationalities in search of international compliance. The challenge of storing legal knowledge and doing sound checks and reasoning becomes complicated in cross-jurisdictional cases.

It requires full interoperability of all the agents to achieve a functional system. The core of the architecture comprises three different parties that must collaborate: (i) a recruiter/company, (ii) an external auditor and (iii) (if necessary) the government/authorities of the country where the company is incorporated or recruiting and the country of origin of the candidate. Our vision of auditing system targets varied selection processes and works the same way for different job categories, from technical to managerial positions. Nevertheless, it should be noted that video-interview systems are specially designed to reduce candidate time and shortlist candidates in selection processes with thousands of applicants. These situations occur with more frequency in the lowest levels of the business hierarchy, e.g. selection of interns of graduates. Managers are often pointed out by headhunters and directly selected by the board of directors. AI in HR can be useful to create models that predict successful promotions among the working staff though. There is no specific hierarchy among all the agents and no co-dependencies. However, the *interview design agent* seems responsible for starting the selection process and taking first steps by making public an offer. It is a precursor to an interview in the same way the selection process agent is responsible for closing one. However, some doubts might arise about how human recruiters and agents interact. Human recruiters and auditors delegate some of their duties and checks on agents, and additional reasoning or test could be done by agents on their own. The additional auditing checks are requested on demand by internal auditing checks or carried out by agents without prior request. We intend it to consider normative reasoning to regulate human and artificial agents. Arguably, we assumed that agents behave as they should. It is highly recommendable to consider normative MASs or normative reasoning for cases like the one we handle.

The *interview auditing agent* requires the approval of candidates in case of law infringement. Once all necessary checks are completed, the information flows back to the *selection process agent* responsible for closing the selection process or cancelling it on legal/ethical grounds. The candidate data agent could conduct some additional checks at their own risk, though.

The workflow of data analysis related to agents, managers, auditors and authorities (e.g. legal and nationality checks) is fundamental to guarantee international legal compliance. The final aim is keeping the system as efficient, transparent and neutral as possible, i.e. reduce or mitigate racial and gender discrimination and cases where the company wants to profit certain legal benefits by applying laws from the country where it is incorporated and not the laws of the country where it is recruiting. An interesting fact of employment law is that some countries, like Brazil, keep certain jobs for their nationalities. It is understandable that there are many necessary checks to guarantee a fair process.

6.1 Interview design agent

In order to design and compare different and fair interviews, an *interview design agent* is required. The design agent is located in the company's main headquarters. According to the company requirements, some interview profiles could be designed to be recruited after that internationally (e.g. software engineer, marketing expert). The input of the agent would be the specific characteristics of the open position, controversial or not (e.g. experience, commercial role, age > 40, gender = woman, English/Spanish native speaker). The output would be the interview questions in the natural language as well as a structured format for a particular role together with scores for possible answers and rankings expected for the role (optimism 50%, commercial profile 60%, etc.). It is expected that the complete implementation will require the use of technologies such as Resource Description Framework (RDF) or ontologies as a starting point to define and harmonize the formats. Once the interview has been designed, it is submitted to the database to be reviewed thereafter by the interview in-company auditing agent.

6.2 Interview in-company auditing agent

The *interview in-company auditing agent* manages the selection processes in different branches. Human recruiters select a particular generalist interview from the corporate database for a particular role, for instance, a software analyst. It applies the general format to every regional scenario, e.g. Europe, Asia Pacific and the Middle East, adapting the prospective candidate criteria to be selected to the culture of the recruiting company's country and local culture. This agent controls interactions among human recruiters and decides if it is essential to pass interviews to additional stages if they handle controversial data. This agent receives an interview in a structured format (later redistributed among the HR team) and returns whether the process requires auditing. If the information needs further validation, it is passed to other agents. All the legal and ethical transactions are hosted outside of the company, in external auditing bodies and governmental premises. The recruiting process could end at this point if the questions are simple and do not need legal handling.

6.3 Selection process agent

The *selection process agent* is the central element of the proposed MASs. It is hosted in the company branch where the selection process is being held. It is in charge of processing the different events that occur in the system (new interviews, new auditing and legal checks, authorities, etc.), triggering new events when necessary and providing a ranking of candidates to the human interviewer.

It could occasionally reuse interviews from other companies; it coordinates different companies, external neutral auditors and authorities, but only if necessary. In case no external auditing is needed – no controversial issues are found – the selection process agent provides direct feedback of the candidate.

6.4 Candidate data check agent

Once a candidate dossier arrives at the external auditors, it may happen that, due to the sensitivity or inaccuracy of the data, some confirmation or permission from the candidate is required to proceed with the data handling. Since the main purpose of this MAS is to assess the selection process in each phase, the *candidate data check agent* would inform the company side promptly if a candidate denies confirmation or does not give consent to the data being used in a particular way.

6.5 Labour law external auditing agent

The human auditor or the internal auditing agent could optionally proceed with legal tests. The analyses are very convenient since the *labour law agent* runs different checks in different legislation, specifically focusing on employment laws. For the proposed architecture, we focus on the US and Spain Labour Law. There will be a database storing the laws, conveniently represented in a rule-based format. In particular, we chose an if-then rule format *condition* $\Rightarrow$ *decision*, where conditions is a logical formula and decision can be (i) *include*, (ii) *warning* or (iii) *exclude*. If a rule is fired (conditions = true) with decision = exclude, then it means the exclusion of the selection process, while the other two possible decisions mean continuity. Warning decision is useful for future reporting, auditing and statistical purposes. In the first example below (R1), a warning is used to signal cases that qualify for benefits (e.g. younger unemployed candidates).

We assume that the set of rules in the knowledge base is consistent. This means there is no possibility that different rules could be fired (e.g. with conditions "age < 16" and "age < 60"), obtaining different outcomes with the same input data. However, we allow that situation if they get the same conclusion. Therefore, the knowledge engineer in charge of filling the knowledge bases must be careful of keeping consistency (she might be supported by a consistency checking or editing tool).

The following rules consider both fiscal benefits for enterprises according to citizen groups and laws infringement in Spain. Warnings signalize situations likely to qualify for benefit or not benefits that should be reviewed afterwards:

R1: Age $< 30 \wedge$ *first job* $\Rightarrow$ *warning*
R2: Age $> 45 \wedge$ *unemployed* $\Rightarrow$ *warning*
R3: Terrorism victim $\Rightarrow$ *warning*
R4: Gender violence victim $\Rightarrow$ *warning*
R5: Age $< 30 \wedge$ *first job* $\Rightarrow$ *warning*

The *labour law agent* allows to adapt to new legal scenarios considering multiple countries and jurisdictions. In the case of law infringement, it would inform the main agent, i.e. the *selection process agent*. In the case of American legislation, which is particularly protective of racial discrimination, any subtle reference

to race in a hiring process and exclude in one iteration could be marked to be watched (warning), e.g.:

R1: Non-White ∧ rejection ⇒ warning

R2: Mixed race ∧ rejection ⇒ warning

R3: Minority ∧ rejection ⇒ warning

One of the main objectives of an auditing will be the legal check carried out by the labour law agent when necessary. The in-company auditing agent passes information to the labour law agent only when there are suspicions of legal infringements in the recruitment processes.

Typically, one of the first analyses to be carried out is checking the legal age of candidates and then the analysis proceeds more accurately on advanced criteria to avoid discrimination. The first cut will be based on valid reasons. For example, checking if the candidate has the minimum age to access the job market in that country or work permit.

The Spanish Labour Law is more lenient for other age prerequisites. In Spain, it is possible to be working past the retirement age (active retirement, a formula for both being a pensioner and earning a reduced salary simultaneously). We formalize this fragment of law with the following rules:

R1: Age > = 18 ∧ Age < = 60 ⇒ include

R2: Age < 18 ∧ emancipated ∧ consent parents/tutor ⇒ include

R3: Age < 18 ∧ emancipated ∧ authorized ⇒ exclude

R4: Age < 18 ∧ emancipated ∧ authorized mother ∧ consent father ⇒ include

R5: Age < 18 ∧ emancipated ∧ authorized father ∧ consent mother ⇒ include

R6: Age < 18 ∧ emancipated ∧ total orphan ∧ authorized legal tutor ⇒ include

R7: Age < 18 ∧ emancipated ∧ total orphan ∧ authorized legal tutor ⇒ exclude

R8: Age < 18 ∧ emancipated ∧ partial orphan ∧ authorized mother ⇒ include

R9: Age < 18 ∧ emancipated ∧ partial orphan ∧ authorized father ⇒ include

R10: Age < 18 ∧ emancipated ∧ partial orphan ∧ authorized mother ⇒ exclude

R11: Age < 18 ∧ emancipated ∧ partial orphan ∧ authorized father ⇒ exclude

R12: Age > 60 ∧ legally retired ⇒ exclude

R13: Age > 60 ∧ legally retired ⇒ include

R14: Age > 60 ∧ legally retired ∧ active retirement status ⇒ include

The legal agent could trigger warnings as well. It could proceed with integrity checks in order to detect the granting of benefits, illegal or not. As an illustration, it would detect the preselection of individuals by gender or age merely based on fiscal benefit for the employers and positive discrimination on economic grounds. Likewise, the following typology of rules could have statistical and reporting purposes, that is, recording the number of candidates eligible for each type of benefit or reduction of tax rate:

R1: Age < 30 ∧ first job ⇒ warning

R2: Age > = 45 ∧ long-term unemployment ⇒ warning

R3: Gender = Woman ∧ domestic violence victim ⇒ warning

R4: Terrorism victim ⇒ warning

In our work, we have resolved to include a rule engine to implement the legal classifier with up to 100 rules so far organized in rules' files specifically for two countries at first: Spain and the USA. When it comes to preselecting which legislation could be considered being the most practical subset. The USA and Spain are two major examples. These countries present some of the most extensive ranges of codes, statutes and formal sources that consider "protected attributes" [41]. Spain offers subsidies for business owners that hire workers according to protected attributes, such as age. It constitutes positive discrimination to easy access to the job market to some disadvantaged collectives, like women and young unemployed people. In the complex case of the USA, several instances of discrimination are prohibited by law. The case of race must be especially highlighted. Title VII of the US Civil Rights Act prohibits making employment decisions based on race, sex and other protected attributes. However, it does not prohibit the interviewer from knowing beforehand the race of the candidate, only making discriminatory decisions. Luckily, the use of protected attributes for decisions in wider society, beyond the workplace, like housing and credit decisions, is being restricted progressively by law [41]. In the UK, the Equality Act 2010 and further legislation have a protective role too.

Therefore, we narrowed down our analysis to the US and Spain cases, with an intent to cover the Brazil case soon, which makes up one of the most interesting cases in labour law worldwide, with several national laws that give preference to Brazilian nationals for employment.

7 Ethical agent

In the case the user data are very controversial and address ethical issues, e.g. the interview format contains terms or analyses traits designed to check age or sexual orientation, the internal auditing agent would pass data to the *ethical agent*. This would only happen if enough

evidence is found to complete an ethical test. The ethical agent carries out a similar analysis to the *labour law agent*, following a rule-based approach. The ethical agent could additionally support the analysis with ethical committees and experts to make a verdict on the excessive invasion of interviewee privacy.

The input of this agent would be a request put forward by the internal auditing agent and it returns the ethical approval or disapproval (include/exclude or include with warnings). The ethical agent would be addressed by the mediation of the interview auditing agent. In case it checks any subtle irregularity in the candidate criteria database, both positive and negative discriminations, the auditing agent launches a check with the ethical agent. The candidate could be accepted for legal reasons, the algorithm would include the rule and the possible bias would be detected by means of warning, as the following example:

R1: Candidate $\wedge$ *homosexual* $\Rightarrow$ *warning*

R2: Candidate $\wedge$ *heterosexual* $\Rightarrow$ *warning*

R3: Candidate $\wedge$ *transgender* $\Rightarrow$ *warning*

These three rules would select similarly. First, the interview content is evaluated. If the auditing result shows that the terms "sexual orientation", "heterosexual", "transgender", "homosexual" come up in the interview – excluded survey or statistical treatment of confidential data – then the candidate is likely to be preselected or excluded on grounds of sexuality. The system launches a warning to signal evidence of inappropriate interview content.

Sexual orientation and religion could be considered more serious cases of discrimination than other rules of the ethical rule base. Tracking the term "race" in an interview would be less negligent, given that companies retain the right to hire candidates of a particular race, for example, in the fashion industry or for statistical or survey purposes, for example, representation of minorities in their workforces.

7.1 Authorities agent

In case a company is expatriated and recruiting international candidates, an *authorities agent* coordinates different governments' decisions and must decide whether a company must comply with additional proceedings or if it is needed to check the register of a foreign candidate in a city census. The civil servant or company could additionally contact with the candidate and request the visa, passport or refugee status necessary to move forward to other phases in the selection process.

7.2 System operation

In this section, we include a more detailed explanation of the choices towards a full legal audit of interviews so as to give a brief description of the inputs and outputs of the systems. The final outcome would be an auditing report. The warning messages and other results of agents' reasoning will be displayed for human and agent auditors during the whole process and used for reporting purposes afterwards.

The system is composed of two different kinds of agents: *corporate agents* and *auditing and authorities agents*. Following this, we offer an insight on how the agents interact. It is important to mention that all agents work at the same level and are not codependent, and although there is no hierarchy, two agents stand out. The *interview design agent*, the one that comes first, opening the selection process and the *selection process agent*, responsible for finalizing the hiring process. The positive intrinsic characteristic of MASs is the possibility of initializing different agents without the need of restarting the others. A human auditor or recruiter could request a check to the agents and they could also carry out reasoning or automated tasks on their own. The ethical and legal analyses are integrated with the rest of the system, but there is a clear separation of roles: corporate analysis on one hand and external auditing on the other hand. The legal checks will be independent of the corporate procedures and only carried out in specific circumstances.

We propose a rule-based system for automatized legal and ethical analyses. The bias is suitably detected by "warnings". Candidate exclusion for valid reasons results in an immediate firing of exclusion rules and notification for posterior dismissal of the recruitment process. Inclusion with a warning would contribute to negative reporting. After this analysis, the internal auditing agent would determine if there is enough evidence for data requests or negative auditing results.

The central elements of the architecture are the *selection process agent* – responsible for closing the selection process – and the *internal auditing agent*. The auditing agent will request an ethical and legal check if necessary. In the rare case that unethical keywords appear in the interview template (e.g. "religion"), the internal auditing would proceed to request ethical agent analysis.

For example, the identification of the word "sexual orientation" or "religion" in the designed interview will suppose the inclusion for valid reasons but the

automatic firing of a warning or negative auditing, whereas other words like race will mean firing a warning for later reviews. For example, it could be possible that a company actually seeks a Black or Asiatic person, for instance, in the modelling or fashion industry, but supposedly no company should recruit in terms of religion and sexual orientation. As a result, the ethical auditing would end with the result "non-ethical". The main goal of discrimination-aware data mining research is to avoid unlawful discrimination. According to Calders and Verwer's work [54], the topic of the discrimination-aware classification was first introduced due to the observation of unwanted dependencies among the attributes. Additionally, the article suggests that splitting a group of candidates according to a particular feature would favour one group over another. The so-called stratification and red-lining effect cause first to prioritize minority groups and problems in future iterations; the classifier then favours one group and uses features that correlate with these preselected features to apply the same type of rules.

Sometimes, mechanical application of discrimination-aware data mining to avoid overall discrimination could lead to miss out genuine occupational requirements for fashion or companies needs, so, understandably, it could be convenient to track sensitive attributes for business purposes occasionally. Pedreschi et al. [43] deal in-depth with cases of supposed direct discrimination that could be refuted by the respondent (company) to be lawful. In the examples provided, they assume a case that claims for discrimination against women among applicants for a job position (e.g. $sex = female \wedge city = NYC \Rightarrow exclude$).

The company could support the defence arguing that the rule is an instance of a more general rule: $drive\_truck \wedge city = NYC \Rightarrow exclude$. This is a legitimate rule and indeed a genuine occupational requirement for a specific job and as a consequence, exclude women lawfully.

The main objective of the auditory will be indeed the legal check carried out by the labour law agent and ethical check when necessary. It is assumed that the interviews will have a great amount of confidential information and private responses and often, they will be processed by the external law agent to make sure there are no infringements of labour law of the countries carrying out the process.

Likewise, if the scanning of interview format or candidate data seems to be noncompliant with the basic regulations, the *internal corporate auditing agent* would put forward a request for external legal testing.

Concerning the proposed corrections related to rules, the use case includes a specification of rules for Spanish Labour Law in check. However, the proposed architecture and system are developed for implementation in different legislation and settings, with sharp differences in anti-discrimination laws. Therefore, the rules included are a limited subset for an illustration of the system versatility in future use cases.

Below, we globally describe the most relevant steps in the procedure and workflow, starting with the company's decision to open a new position in a branch. Following a step-by-step timeline, the initial entry of the system is an interview format in plain text associated with the particular position. As for the details of the interview, there is an important record field: *language of the interview* and the definition of *STAC* (*skills, traits, abilities* and *characteristics*) *questions* related to skills.

1. The central HR manager *creates a position* (general specification) [internal activity] to be applied in different branches, for example:

Example 1.
Level: Analyst.
Division: Investment banking.
Situation: Asia-Pacific.
Position: Risk analyst.

Example 2.
Level: Manager.
Division: Wealth management.
Situation: EMEA.
Position: Wealth manager.

2. The regional branch recruiter *adapts the general specification* to their specific context and needs. She is assisted by the interview design agent to *create an interview script*. The interview comprises STAC questions. The interview will be defined for a specific language and country, and there will be three clear subcategories in the questions: behavioural questions (body language and response time), leadership style (suitable for commercial role) and optional corporate questions (willingness to travel, dream job and economic compensation). As described before, these are possible and generalistic categories, but the design of the interview is at the discretion of the company.
3. The company publishes the call.
4. The selection process agent analyses the interview with the aim of detecting major discriminations (e.g. religion).

5. Candidate interviews take place and are stored in the interviews' database.
6. The selection process agent carries out reasoning based on interview questions and candidate answers. It follows an analysis to exclude candidates for valid reasons (e.g. non-legal working age). The shortlisted candidates are passed to the next step.
7. In-company auditing agent passes information to labour law auditing agent and ethical agent in case of warning or irregularities of the shortlisted candidates. If the candidate data concerning citizenship status need to be checked against the analysis, it would be passed on to the *candidate data check agent*. If there is no need for ethical or legal processing, the internal auditing agent closes the analysis and passes the shortlisted candidates to the selection process agent.
8. The *labour law agent*, ethical agent and candidate data request to carry out a sound check if necessary at the request of a human external auditor. The candidate data check agent contacts the candidate self or authorities to confirm citizen or expatriate status, if necessary. The tests are independent. They produce an acknowledge document and give back information about the citizen and legal and ethical warning to the selection process agent.
9. The selection process agent is responsible for closing the selection process, giving back a report with a ranking of candidates to the human recruiter. The human recruiter decides the hire depending on the scores. The final say is at his/her own discretion.

8 Discussion

AI will be a turning point in the previous way of recruiting, especially when combined with cognitive psychology. It saves candidate time and reduces the geographic distance and time of face-to-face interviews.

It is questionable if scientists and, in particular, companies can make gradual improvements to the capabilities of algorithms or ride them to the limit according to their agenda. It is essential to foster proper auditing and supervision of these systems, which, in any way, perform well. Technology for recruiting enables the detection of over 11 movements/minute surpassing human recruiters. Thereby, wearables or cognitive help like microphones, and augmented reality glasses will be needed to give candidates the chance to perform well and make up for their individual handicaps.

The future of AI and its consequences in organizational change management should be more inclusive and informative. Likewise, ML classification methods remind us of the importance of the attention to detail. To achieve good hires using ML algorithms training, data sets should be diverse to support the idea behind learning, i.e. generalization and subsequent application to new individuals.

In this article, we have presented the pros and cons of domain-specific AI for HR. It is noteworthy to outline some present challenges such as the *lack of reasoning and inference ability of current domain-specific AI* and the fact that it remains in *the experimental state*. As mentioned above, the training *data sets* are *imperfect* as they lead to an imperfect AI.

Some future lines regarding domain-specific AI for HR are the following:

- Ensuring that ML and domain-specific AI work equally for many people and are neutral, guaranteeing diverse training sets.
- Conditions to classify well are uncertain, but there are conditions to record and perform well technically in an interview that could avoid false positives (e.g. good microphone and distance to camera).
- Enhancing candidate consciousness of the machine hardware and wearables for better results. Informing candidates of language requirements of the company. Well-informed users would avoid misclassification problems.
- Fostering proper auditing in domain-specific AI for HR to avoid manipulative uses. Systems are slightly opaque so far and not officially audited.
- It is preferably reducing the importance of body language in candidates to focus on answers. Technical characteristics, such as light and behaviour, are not that important. Tracking how many times they look away from the camera disadvantages the chances of candidates who are natural and spontaneous and not used to direct gaze. Another suggestion for interviewing is experimenting with different rankings, not just preselecting candidates based on top performers.
- Addressing the issue of gender bias. Blind auditions have proved to be useful.
- Controlling the pace of digital change for small- and medium-size companies. Some are concerned with not being innovative enough, but unnecessary adoption of video interviewing could cause business disruptions: "A 2013 research from Oxford University states that

almost 43% of jobs will be automatized in the USA and up to 70% in developing countries by 2033".[15]

Despite so many expectations in the field of HR, it should be noted that AI is a tool and the ethics of it will always depend on the person creating the tool. Image processing is very powerful and looks beyond ordinary things.

In this context, more governance is needed. In a broader perspective, governments should track selection processes if there is an infringement of fundamental employment laws and human rights. In particular, due to the diverse nature of the global job markets, an ever-growing mobility of employees exposed to different approaches in regulations (for instance, some countries keep some jobs only for their nationals) is expected.

Concerning the idea of identifying sexual orientation, it reminds us of the difficulty level. The cues and sounds in every language vary. The same applies to gender, and males produce shorter vowel duration than females.

Problems could arise not just with the technicalities of the sounds for females, males and nationalities. It could be potentially dangerous that a company or state can predict the sexual orientation of employees. It is critical when keeping in mind the number of countries that condemn homosexuality.

9 Conclusion

The conflict behind AI for recruiting is that it relies on proprietary products trained with limited data sets. Even though they offer accuracy to look at certain characteristics, they were not thought of as mainstream recruiting tools in their beginnings. As a matter of fact, they are progressively adopted by large corporations with thousands of candidates for efficiency reasons. The software could not control potentially discriminatory outcomes if recruitment is carried out by the company under the wrong reasons or controlled by non-democratic state, e.g. being selective against minorities, women, people under or over a certain age, senior citizens, immigrants or customer with accents. Image processing could even filter candidates by appearance reasons.

This reality is ultimately in a clash with employment laws in most jurisdictions. For example, the US Law is especially protective of racial discrimination. The Civil Rights Act 1964 gathers the idea that is forbidden "improperly classifying or segregating employees or applicants by race" in selection processes. There are still many territories where the US Employment Law applies, like American Samoa, Guam, the Commonwealth of the Northern Mariana Islands, Puerto Rico and the US Virgin Islands. Even though the recruiter and company are foreigners, they should comply with the US Law, which is especially protective of discrimination by race and age over 40 years.

We have illustrated in the previous sections that current technology enables to detect race easily in images. The discrimination could be subtle, especially when the selection process is about promoting an employee. The same goes with discrimination related to sexual orientation. The technology has no limits of privacy. Even though a minimum level of fairness is reached concerning the main attributes or protected classes, like race and gender, it is difficult to assess the side effects and the interaction among other subclasses and combination of them in ML analyses. The overall fairness is somehow unreachable. Thus, there is growing concern about the limitations of AI technology and the effects of "intersectionality", especially when it becomes mainstream and its use spreads over different and miscellaneous fields, from criminal justice to finances [55].

What ultimately needs to be discussed is if we should rely entirely on private companies and AI solutions instead of seeing it for what it is today: a tool with limitations, computer-aided recruitment. Technology in the HR has evolved enormously but often has lost the human aspect.

In this article, we analysed controversial characteristics that are measured by commonly used recruiting software. This article serves as a ground and summarizes our research following both a technical and an informative approach. It is precise to mention the increment with respect to our previously published work that offers a brief insight of the proposal [56,57] or focuses more on semantic technologies and technicalities of legal reasoning [58]. Here we explore in greater detail the advantages and disadvantages of image analysis during interviews and discuss their legal and ethical implications. To overcome these problems, we think auditing should be carried out on recruiting processes. We have proposed an MAS architecture, so as to support humans on those auditing procedures. The most notable issue raised by the prototype is, being self-critical enough, the limitations of the proposal here mentioned, a consequence of differences in formats and the need of interoperability, proprietary HR software products that do not favour auditing and lack of access to real corporate HR business scenarios. The proposal is therefore limited to basic agents' checks and reasoning. The auditing of AI poses

15 https://www.oxfordmartin.ox.ac.uk/publications/view/1314

challenges that are in line with the current state of the art of AI technology.

Taking due account of the inherent limitations of this proposal and the difficulty to explain a complex system in just one paper, we include a comprehensive description of the system functioning. Open issues are remaining that could be beneficial, such as the full description of knowledge representation, e.g. formalizing interviews and job description with ontologies and fostering interoperability and a variety of formats. Relying on a common vocabulary could smooth the interaction among agents.

In this line, we may need to use some extensions to the rule format approach we are currently using or even adopt a different paradigm. For example, priorities can be useful to solve rule inconsistencies while keeping rules as simple as possible. Another option that we are considering is having several warning degrees, e.g. the higher the number of rules supporting a warning, the stronger the belief that a warning should be raised.

We continue analysing which agents' duties could be automatized and semi-automatized to further strengthen full automatization. At this point, the system is arranged in some instances as a decision-making tool or a tool to assist decision for HR managers. Some analyses remain semi-automatized. In the future, we plan to investigate the development of the proposed architecture and test it in operation.

Acknowledgements: This work was partially supported by the Spanish Ministry of Science, Innovation and Universities and co-funded by EU FEDER Funds through Grants TIN2015-65515-C4-4-R and RTI2018-095390-B-C33 (MCIU/AEI/FEDER, UE).

References

[1] N. J. Adler and E. Ghadar, "Strategic human resource management: a global perspective," *Human Resource Management in International Comparison*, Berlin: Walter de Gruyter, 1990.

[2] B. Kapoor and J. Sherif, "Human resources in an enriched environment of business intelligence," *Kybernetes*, vol. 41, no. 10, pp. 1625–1637, 2012.

[3] S. Strohmeier and F. Piazza, "Artificial intelligence techniques in human resource management – a conceptual exploration," in *Intelligent Techniques in Engineering Management*, C. Kahraman and S. Çevik Onar, Eds, Intelligent Systems Reference Library, vol. 87, Cham: Springer, 2015.

[4] J. Min, *Ten ways HR tech leaders can make the most of artificial intelligence*, 2017, https://www.personneltoday.com/hr/ten-ways-hr-tech-leaders-can-make-artificial-intelligence/ [accessed March 14, 2020].

[5] E. P. Bruun and A. Duka, "Artificial intelligence, jobs and the future of work: racing with the machines," *Basic Income Stud.*, vol. 13, no. 2, Art. 20180018, 2018, https://doi.org/10.1515/bis-2018-0018.

[6] Y. Wang and M. Kosinski, "Deep neural networks are more accurate than humans at detecting sexual orientation from facial images," *J. Abnorm. Psychol. Soc. Psychol.*, vol. 114, no. 2, pp. 246–257, 2018.

[7] D. A. Reid, S. Samangooei, C. Chen, M. S. Nixon, and A. Ross, "Soft biometrics for surveillance: an overview," *Handbook of Statistics*, 2013, pp. 327–352.

[8] F. Siyao, H. Haibo and H. Zeng-Guang. "Learning race from face: a survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 12, pp. 2483–2509, 2014.

[9] Y. Wang, Y. Feng, H. Liao, J. Luo, and X. Xu, "Do they all look the same? Deciphering Chinese, Japanese and Koreans by fine-grained deep learning," in *2018 IEEE Conference on Multimedia Information Processing and Retrieval* (*MIPR*), Apr. 2018.

[10] M. C. Roh and S. W. Lee, "Performance analysis of face recognition algorithms on Korean face database," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 21, no. 6, pp. 1017–1033, Sep. 2007.

[11] A. Bastanfard, M. A. Nik, and M. M. Dehshibi, "Iranian face database with age, pose and expression," *2007 International Conference on Machine Vision*, Dec. 2007.

[12] D. Sutic, I. Breskovic, R. Huic, and I. Jukic, "Automatic evaluation of facial attractiveness," in *The 33rd International Convention MIPRO*, 2010, pp. 1339–1342.

[13] J. Gan, L. Li, Y. Zhai, and Y. Liu, "Deep self-taught learning for facial beauty prediction," *Neurocomputing*, vol. 144, pp. 295–303, Nov. 2014.

[14] J. Hayashi, M. Yasumoto, H. Ito, Y. Niwa, and H. Koshimizu, "Age and gender estimation from facial image processing," in *Proceedings of the 41st SICE Annual Conference*, 2002, pp. 13–18.

[15] S. E. Padme and P. S. Desai, "Estimation of age from face images," *Int. J. Sci. Res.*, vol. 4, no. 12, pp. 1927–1931, 2015.

[16] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," *in 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), 2015, pp. 34–42.

[17] Yoti, "White paper. Yoti age scan – public version," https://www.yoti.com/wp-content/uploads/2019/09/Age-Scan-White-Paper-Executive-Summary-December19.pdf [accessed March 14, 2020].

[18] J. Boulamwini and T. Gebru, "Gender shades: intersectional accuracy disparities in commercial gender classification," in *Conference on Fairness, Accountability and Transparency*, 2018, pp. 77–91.

[19] S. Buell, "MIT researcher: artificial intelligence has a race problem and we need to fix it," *The Boston Magazine*, Jan. 24 2019, https://www.bostonmagazine.com/news/2018/02/23/artificial-intelligence-race-dark-skin-bias/ [accessed March 14, 2020].

[20] V. Muthukumar, et al., "Understanding unequal gender classification accuracy from face images," *arXiv preprint arXiv:1812.00099*, 2018.

[21] S. P. Zafeiriou, C. Zhang, and Z. Zhang, "A survey on face detection in the wild: past, present and future," *Comput. Vis. Image Und.*, vol. 138, pp. 1–24, 2015.

[22] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem, "Finding tiny faces in the wild with generative adversarial network," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018.
[23] C. Garvie, *The perpetual line-up: unregulated police face recognition in America*, Georgetown Law, Center on Privacy & Technology, Washington, DC, 2016.
[24] S. Cosar, G. Donatiello, V. Bogorny, C. Garate, L. O. Alvares, and F. Bremond, "Toward abnormal trajectory and event detection in video surveillance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 3, pp. 683–695, 2017.
[25] S. Sulpizio, et al., "The sound of voice: voice-based categorization of speakers' sexual orientation within and across languages," *PLOS One*, vol. 10, no. 7, pp. 1–38, Jul. 2015.
[26] J. Kuczmarski, "Reducing gender bias in google translate," *Google Blog*, Dec. 6 2018, https://www.blog.google/products/translate/reducing-gender-bias-google-translate/ [accessed March 14, 2020].
[27] S. Leavy, "Gender bias in artificial intelligence: the need for diversity and gender theory in machine learning," in *2018 IEEE/ACM 1st International Workshop on Gender Equality in Software Engineering* (*GE*), Gothenburg, 2018, pp. 14–16.
[28] L. Yang, "A gender perspective of translation: taking three chinese versions of the purple color as an example," *J. Lang. Teach. Res.*, vol. 5, no. 2, pp. 371–375, 2014.
[29] P. Viola and M. Jones, "Robust real-time face detection," *Int. J. Comput. Vision*, vol. 57, no. 2, pp. 137–154, 2004.
[30] S. Masood, S. Gupta, A. Wajid, S. Gupta, and M. Ahmad, "Prediction of human ethnicity from facial images using neural networks," in *Data Engineering and Intelligent Computing*, S. Satapathy, V. Bhateja, K. Raju, and B. Janakiramaiah, Eds., Advances in Intelligent Systems and Computing, vol. 542, Springer, Singapore, 2017, pp. 217–226.
[31] Z. Jin-Yu, C. Yan, and H. Xian-Xiang, "Edge detection of images based on improved Sobel operator and genetic algorithms," in *2009 International Conference on Image Analysis and Signal Processing*, 2009, pp. 31–35.
[32] J. Shi and C. Tomasi, "Good features to track," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, 1994.
[33] W. H. Abdulsalam, R. S. Alhamdani, and M. Najm Abdullah, "Facial emotion recognition from videos using deep convolutional neural networks," *Int. J. Mach. Learn. Comput.*, vol. 9, no. 1, pp. 14–19, 2019.
[34] P. Ekman, "Darwin, deception, and facial expression," *Ann. N. Y. Acad. Sci.*, vol. 1000, no. 1, pp. 205–221, 2003.
[35] D. Mehta, M. F. Siddiqui, and A. Y. Javaid, "Recognition of emotion intensities using machine learning algorithms: a comparative study," *Sensors*, vol. 19, no. 8, Art. 1897, 2019, https://doi.org/10.3390/s19081897.
[36] A. Dehghan, E. G. Ortiz, G. Shu, and S. Z. Masood, "Dager: deep age, gender and emotion recognition using convolutional neural network," *arXiv preprint arXiv:1702.04280*, 2017.
[37] C. F. Benitez-Quiroz, R. Srinivasan, Q. Feng, Y. Wang, and A. M. Martinez, "EmotioNet challenge: recognition of facial expressions of emotion in the wild," *arXiv preprint arXiv:1703.01210*, 2017.
[38] S. L. S. Purkiss, P. L. Perrewé, T. L. Gillespie, B. T. Mayes, and G. R. Ferris, "Implicit sources of bias in employment interview judgments and decisions," *Organ. Behav. Hum. Decis. Process.*, vol. 101, no. 2, pp. 152–167, 2006.
[39] V. Ging, "11 common interview questions that are actually illegal, business insider," https://www.businessinsider.com/11-illegal-interview-questions-2013-7?IR=T [accessed March 14, 2020].
[40] A. Prince and D. Schwarcz, "Proxy discrimination in the age of artificial intelligence and big data," *Iowa Law Review*, vol. 105, no. 3, pp. 1257–1318, 2020.
[41] A. Datta, M. Fredrikson, G. Ko, P. Mardziel, and S. Sen, "Proxy discrimination in data-driven systems," *arXiv preprint arXiv:1707.08120*, 2017.
[42] S. Barocas and A. Selbst, "Big data's disparate impact," *Calif. Law Rev.*, vol. 104, no. 3, pp. 671–732, 2016.
[43] D. Pedreschi, S. Ruggieri, and F. Turini, "Integrating induction and deduction for finding evidence of discrimination," in *Proceedings of the 12th International Conference on Artificial Intelligence and Law* (*ICAIL '09*), 2009.
[44] S. Hajian and J. Domingo-Ferrer, "A methodology for direct and indirect discrimination prevention in data mining," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 7, pp. 1445–1459, 2013.
[45] D. Hawkins, "Researchers use facial recognition tools to predict sexual orientation. LGBT groups aren't happy," *The Washington Post*, September 12, 2017, https://www.washingtonpost.com/news/morning-mix/wp/2017/09/12/researchers-use-facial-recognition-tools-to-predict-sexuality-lgbt-groups-arent-happy/?utm_term=23020459ddcd [accessed March 14, 2020].
[46] H. Murphy, "Why Stanford researchers tried to create a "gaydar" machine," *The New York Times*, Oct. 9, 2017, https://www.nytimes.com/2017/10/09/science/stanford-sexual-orientation-study.html [accessed March 14, 2020].
[47] S. Ossowski and A. Omicini, "Coordination knowledge engineering," *Knowl. Eng. Rev.*, vol. 17, no. 4, pp. 309–316, 2002.
[48] S. Mahmoud, et al., "Multi-agent system for recruiting patients for clinical trials," in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, 2014, pp. 981–988.
[49] E. Mas, A. Suppasri, F. Imamura, and S. Koshimura, "Agent-based simulation of the 2011 great east Japan earthquake/tsunami evacuation: an integrated model of tsunami inundation and evacuation," *J. Nat. Disaster Sci.*, vol. 34, no. 1, pp. 41–57, 2012.
[50] H. Billhardt, A. Fernández, M. Lujak, and S. Ossowski, "Agreement technologies for coordination in smart cities," *Appl. Sci.*, vol. 8, no. 5, Art. 816, 2018, DOI: https://doi.org/10.3390/app8050816.
[51] J. Debenham, "A multi-agent architecture for business process management adapts to unreliable performance," *Adaptive Computing in Design and Manufacture V*, Springer, London, 2002, pp. 369–380.
[52] A. Ciortea, S. Mayer, and F. Michahelles, "Repurposing manufacturing lines on the fly with multi-agent systems for the web of things," in *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems*, 2018, pp. 813–822.
[53] V. R. Walker, "A default-logic framework for legal reasoning in multiagent systems," *AAAI Fall Symposium*, 2006, pp. 88–95.

[54] T. Calders and S. Verwer, "Three naive Bayes approaches for discrimination-free classification," *Data Min. Knowl. Discov.*, vol. 21, no. 2, pp. 277–292, 2010.
[55] M. Kearns, S. Neel, A. Roth, and Z. S. Wu, "Preventing fairness gerrymandering: auditing and learning for subgroup fairness," *arXiv preprint arXiv:1711.05144*, 2017.
[56] C. Fernández and A. Fernández, "Ethical and legal implications of ai recruiting software," *ERCIM News*, vol. 116, pp. 22–23, 2019.
[57] C. Fernández and A. Fernández, "AI in recruiting. Multi-agent systems architecture for ethical and legal auditing," in *Proceedings of the Twenty-Eight International Joint Conference Artificial Intelligence* (*IJCAI*), 2019.
[58] C. Fernández and A. Fernández, "Ontologies and AI in recruiting. A rule-based approach to address ethical and legal auditing," in *Proceedings of the International Semantic Web Conference* (*ISWC*), 2019.