PLOS ONE

RESEARCH ARTICLE

Schizophrenia more employable than depression? Language-based artificial intelligence model ratings for employability of psychiatric diagnoses and somatic and healthy controls

Maximin Lange [1]*, **Alexandros Koliousis**[2], **Feras Fayez**[3,4], **Eoin Gogarty**[1,3], **Ricardo Twumasi**[1]

1 Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom, 2 Northeastern University, London, United Kingdom, 3 King's College Hospital NHS Foundation Trust, London, United Kingdom, 4 Imperial College Healthcare NHS Trust, London, United Kingdom

* Maximin.lange@kcl.ac.uk

Abstract

Artificial Intelligence (AI) assists recruiting and job searching. Such systems can be biased against certain characteristics. This results in potential misrepresentations and consequent inequalities related to people with mental health disorders. Hence occupational and mental health bias in existing Natural Language Processing (NLP) models used in recruiting and job hunting must be assessed. We examined occupational bias against mental health disorders in NLP models through relationships between occupations, employability, and psychiatric diagnoses. We investigated Word2Vec and GloVe embedding algorithms through analogy questions and graphical representation of cosine similarities. Word2Vec embeddings exhibit minor bias against mental health disorders when asked analogies regarding employability attributes and no evidence of bias when asked analogies regarding high earning jobs. GloVe embeddings view common mental health disorders such as depression less healthy and less employable than severe mental health disorders and most physical health conditions. Overall, physical, and psychiatric disorders are seen as similarly healthy and employable. Both algorithms appear to be safe for use in downstream task without major repercussions. Further research is needed to confirm this. This project was funded by the London Interdisciplinary Social Science Doctoral Training Programme (LISS-DTP). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

1. Introduction

1.1 Background

Until recently, human resource (HR) professionals would manually sift through vast amounts of resumes to determine the best fitting candidate for a job opening. Judgement calls were made based on their own experience, gut feeling, or discussions with colleagues.

Artificial intelligence (AI) has revolutionised this process, assisting in extracting information about skills of applicants from resumes, social media sites and cover letters [1–3] a practice known as resume parsing. AI models that can parse resumes are so widespread that virtually all Fortune 500 firms today use some version of them [4].

Similarly, individuals in the market for a new job had to look up openings manually. Especially online, job seekers would face a plethora of available vacancies. A scenario of potentially overwhelming nature.

Job seekers now receive help by AI-powered job recommender systems (JRS). Job recommendation system are powered by machine learning algorithms, bringing together vacancies and job seekers based on behaviours, preferences or needs of the two parties [5].

Despite the excitement surrounding AI and JRS in HR, as it eases lives of recruiters and job seekers alike, such systems ought to be assessed in a broader spectrum of existing organisational transformation.

1.2 Occupational bias and allocation harm

There is a rich history of bias in general recruiting against a broad spectrum of attributes, including all protected characteristics, which have been extensively studied and summarised elsewhere [6–8].

Although more recent research suggests a changing trend [9], there is ample evidence for discrimination against people with mental health disorders in recruiting [10–13].

This leaves healthy people and people with mental illness, who might possess equivalent levels of skills, being employed at different rates. As a result, we might see job databases labelling mentally healthy people as more employable, and safe to work with.

Contrary to stigma, most people with mental health disorders report a desire to work [14–16].

It is often forgotten that people with mental illness still possess knowledge, skills, and abilities that can facilitate organisational effectiveness [17] and the World Health Organization [18] stresses that under-use of skills in the workplace can aggravate mental health disorders.

1.3 Natural language processing

Natural Language Processing (NLP) encompasses a collection of methods that transform written text passages into datasets that can subsequently be analysed using traditional statistics and machine learning models [19, 20]. As an interdisciplinary field, NLP bridges artificial intelligence, computer science, cognitive science, information processing, and linguistics, letting computers parse and process human language [21].

Real-time applications of NLP in the business field include chatbots, sentiment analysis, and speech recognition [22–25]. In the healthcare and pharmaceutical industries, NLP is used to analyse large amounts of unstructured data, such as electronic health records, to gather insights into patient behaviour, disease synthesis and prognostic predictions [26–30]. Additionally, NLP can be integrated with qualitative research methods to augment traditional text analysis approaches [20, 31, 32] include machine translation [33], or information retrieval and question answering [34, 35].

In the following two sections we address challenges related to occupational bias in NLP when used in the context of recruiting and job searching.

1.3.1 Recruiting and occupational bias in NLP. For recruiting purposes, NLP can be trained, among other things, on resumes, cover letters and social media profiles or posts [36–39]. At this stage, such data is already influenced by an individual's personal background.

Language use is heavily affected by, and dependent on, culture, age, gender and personality [40–43]. Specific word use patterns of people with mental illness have been identified: Perceptual and causal language are negatively correlated in schizophrenic individuals but positively correlated in those with mood disorders [44]. Spoken language patterns might even allow for prediction of psychosis [45, 46]. Detection of general mental illnesses and even reliable estimation of population mental health is possible through analyses of social media posts, surveys, narrative writings, and interviews [47–50]. This shows that deep learning models can detect social media users at risk for developing a mental disorder, deduced from online posts represented with linguistic features at different levels, including messages and corresponding writing style and emotions conveyed–[51] offer similar findings.

Therefore, even if candidates appear healthy or fitting for a specific role when seen by a human recruiter, the writing style of social media posts or cover letters might give detectable data of psychiatric illness for an NLP system, thus resulting in exclusion of the corresponding individual by the machine.

People from different demographics write resumes differently. Variation in resume content and textual features written by different nationalities has been observed [52]. Anonymous resume screening aims to allow for censorship of personal identifiers including socio-demographics. However, resumes stripped of non-job relevant information might still contain information about job applicants in subtle ways [53]. In an arms race to evade recruiting bias and consequently increase chance of hire, women with similar job-relevant characteristics to men write their resumes differently. This is known as social identity—based impression-management [54]. Women also tend to 'man-up' their resumes [55] while people from an ethnic minority background might 'whiten' their resumes [56]. However, these patterns can already be detected by NLP with high accuracy [57, 58] and often either do not work as intended, or work counterproductive, actively decreasing hiring chances [53, 54].

There has been, to our knowledge, no study investigating the resume writing style of people with mental health disorders. Still, we argue, given the paragraph above, that the chances of an NLP algorithm being able to detect differences between healthy people and people with mental health disorders only from writing style are extremely high.

1.3.2 Job recommender systems and occupational bias in NLP.   As argued above, even when stripped of information directly revealing key attributes (names, DOB, nationality, medical diagnosis) NLP are still able to detect these in wording structure. NLPs are often trained to screen for a specific role, which often assumes a mentally healthy individual. Many people, due to psychiatric diagnosis or other attributes, revealed or not, do not fall into the desired category, however, could still perform the job well, if they were to be hired. These individuals will be missed by NLP models that are trained on existing job descriptions looking for healthy indivudals. Furthermore, non-western or in general non-conventionally written resumes might get mistaken for ill mental health one, i.e., a false positive scenario.

Therefore, NLP models that process input language in a fit-for-all style, run danger of falsely ruling out substantial chunks of the general population. Hence, models predicting employability and job suitability from resumes, cover letters and social media sources, must factor in origins of language used, instead of inferring job unfitness.

1.3.3 NLP-based job recommender systems.   While we appreciate the fact that there is an array of JRS in the literature [59–62], often not transparently documented and being of dubious nature (For reviews: [63–65], this paper takes a special focus on NLP-based JRS. Given the extreme variety regarding model types, structure and data going into them, assessing all types of JRS for bias would not be feasible for a single paper.

Focusing on NLP allows for a simultaneous assessment of resume parsing and job recommendations in one paper. NLP has been a staple feature in JRS models [66–70]. We further chose NLP since there are now multiple plugins, e.g., Ambition (https://remoteambition.com), Mindart (https://mindart.app), Wanted Job Search (https://www.wanted.co.kr/terms), or JoPilot (https://jopilot.net/home/terms), to name a few, for ChatGPT [71] a popular large language model (LLM), allowing to search for jobs. These plugins are partially or in full powered by NLP. ChatGPT has enormous numbers of users, perhaps the most of any generative AI, which makes inner workings of NLP using JRS of special importance to be audited for bias.

1.3.3.1 Action needed now. The implementation of feedback loops allows LLMs to continuously learn from interactions, improving their performance with each input-output cycle [72, 73] These feedback loops, however, give rise to the potential for model collapse, which describes the problem of LLMs potentially delivering most of the language found online. An excess of AI-generated training data leads to irreversible defects, i.e. increased errors, and degraded performance [74, 75].

To avoid model collapse, new, clean, human-generated datasets ought to be regularly introduced into the training process. It is therefore now the time to act, working with clean datasets, free of bias and with continuous human oversight.

2. Methods

2.1 Background

Computational linguistic models used to rely on methods that interpret language by examining individual words and analysing keyword frequency in formal text analysis, which is limiting, as it overlooks the interconnected nature of word meanings [76–78]. In recent years, NLP systems have utilized deep learning and neural networks to effectively capture semantic information and contextual understanding of words within extensive text datasets [19, 79–81].

A crucial component of these techniques is the incorporation of *word embeddings*. Word embeddings represent words as vectors in a multi-dimensional space, assigning more similar vectors to words that appear in comparable contexts within the training data. These word-vectors can also be visualized as points in N-dimensional space [76, 79].

The concept of N-dimensional space in NLP word vectors refers to the number of dimensions used to represent a word in a numerical vector space. The number of dimensions can vary depending on the specific word embedding technique used, but it is typically in the range of 100–300 dimensions. The process of transforming a word into a numerical vector involves using a word embedding technique, to map the word to a vector in the N-dimensional space. The resulting vector represents the semantic meaning of the word, and words with similar meanings are located closer to each. The visualization of word vectors can be done using dimensionality reduction techniques such as PCA and t-SNE, which reduce the N-dimensional space to 2 or 3 dimensions for visualization purposes [82, 83].

Word embeddings address the aforementioned limitations by creating a consistent and continuous meaning space, where words are positioned based on their similarity to other words, as determined by their usage in natural language samples [21, 76, 84].

Since vectors define positions in space, similarity and distance become interchangeable concepts. Words with more similar vector representations are also spatially closer. This similarity, or distance, is typically measured using *cosine similarity* [79, 76, 84] The set of word-vectors can be referred to as the trained word embedding, a semantic space, or simply a word embedding.

The cosine similarity between two vectors is calculated as the cosine of the angle between them with the formula:

$$Cosine\ Similarity\ (A,\ B) = \frac{A\ \cdot B}{||A|| \times ||B||}$$

The dot product of the two vectors is divided by the product of their magnitudes to obtain the cosine similarity value, which ranges from -1 to 1. A cosine similarity of 1 indicates that the two vectors are identical, while a cosine similarity of -1 indicates that they are completely dissimilar. A cosine similarity of 0 indicates that the two vectors are orthogonal, or completely unrelated.

2.2 Investigating occupational bias and mental health bias in NLP

Cosine similarities between word-vectors often mirror human-rated similarities between words [85–88]. This supports that word embeddings can reflect and investigate cultural phenomena in ways that otherwise (e.g. using surveys or implicit observations) would not be practical or at all possible, while demonstrating biases and inadequacies in human language [79] Word embeddings thus deliver a stable, reliable and valid estimate of biases [89].

This has led to a common research approach in which latent semantic dimensions (e.g., gender, ethnicity, minority) are paired with how words of interest (e.g., jobs, stereotypes) are located within a dimension.

Reviews on biases in word embeddings of NLP models have been done [90–97]. Papers on specific NLP occupational bias and potential allocation harms because of word embeddings exist primarily related to gender and ethnicity [98–103].

First papers relating to mental health bias in NLP models have come up [85, 104, 105]. No paper has yet combined occupational bias and mental health bias in NLP research. Our paper is the first research examining occupational bias against psychiatric diagnoses in NLP models.

2.3 Word2Vec

Word2Vec [106] is an algorithm to produce word embeddings, pre- trained on a Google News dataset, containing roughly 100 billion words, which gives a model of 300-dimensional vectors for 3 million words and phrases, available at: GoogleNews-vectors-negative300. bin.gz.

Word2Vec is an established tool to investigate bias in word embeddings [107–109], especially using analogies [105, 110–115]. Word2Vec is frequently used in resume parsing [116–120] and job recommendation [121–125].

We conducted analogies of Word2Vec embeddings in Python version 3.8.17 using the *pymagnitude* package version 0.1.120. We took the Word2Vec model that was pre-trained on GoogleNews Word2Vec model from the *Magnitude* library as it is and did not fine-tune, retrain or adjusted it in any way.

We examine the results of word analogies as follows.

First, as with most other analogy papers, we look at the top-1 finding. We further examine the similarity score, if there are other biased terms, and at which position they are.

We do not combine the upper- and lower-case versions of an identifier term i.e. "Lawyer" and "lawyer", to leave open the possibility of same word returns.

An analogy would be biased if words are returned that would not be of equal standing in society. We assume a common-sense approach for judging these—for further information see our section *Examining the Validity of Word Analogies as Indicators of Bias*

Since *doctor* and various terms regarding mental health conditions could be associated closer to each other, mainly since they are often used together in a diagnostic scenario, we

included enquiries regarding other medical professions. When quiring job titles analogies, we only asked for psychosis to not exceed the frame of this paper.

We selected the employability attributes of general employability, reliability, competence, and resilience, since these are common concerns voiced by employers to not hire people with mental health disorders [13, 126–129].

We select job titles and fields that are associated with prestige and status, e.g., medicine, law, engineering, finance. Competition to access such occupations is higher than in most jobs. This, we argue, would result in potential bias against mental illness, which is often aimed at ability, and hence being the easiest to detect.

2.4 Global vectors for word representation (GloVe)

GloVe [130] is a weighted least squares model that was trained on global word-word co-occur-rences from a dataset of 200 million words from Wikipedia pages, available at: https://nlp.stanford.edu/projects/glove/.

GloVe just as Word2Vec, is frequently used to parse resumes [116, 131–133] and recommend jobs [134–136]. GloVe is also used to investigate bias in NLP models; this is commonly done using graphical representations of words [91, 105, 134, 137–140].

Bias in GloVe word embeddings can be illustrated by plotting terms onto graphs to examine relations in terms of cosine similarity—in our case psychiatric diagnoses and employability. A pair of opposing words is used on the X axis (e.g. 'employable' and 'unemployable'), another pair of opposing words is used on the Y axis (e.g. 'healthy' and 'ill'). Space between diagnoses terms within the graph mirror mathematical distance between vector points in the word embeddings.

We conducted graphical representations of GloVe embeddings in Python version 3.8.17 using the *matplotlib* package version 3.7.1. We downloaded the GloVe model from the *Magnitude* library as it is and did not fine-tune, retrain, or adjusted it in any way.

There are different dimensional models for GloVe embeddings. The dimensionality represents the total number of features that the vector encodes. The larger the dimensionality, the more information the vector can encode [21]. We used the 300-dimension model, which is the maximum number of dimensions available for GloVe embeddings.

We used synonyms to assess for the consistency of our findings. In Graph 2 we use 'reliable' and 'unreliable' as well as 'normal' and 'abnormal'. We further added physical diagnoses and very healthy control attributes.

2.5 Word analogies

Analogies are equations formulated as *A*: *B*:: *C*: *D*. In plain speech: *A is to B as C is to D*. When supplied with words representing *A*, *B*, *C*, the model returns a word that it deems representative of *D* in the analogy. Embeddings of analogies, or word relationships enables analogical questions to be solved by vector addition and subtraction [141].

Example: Tokyo (A) is to Japan (B) as London (C) is to X (D). We expect X to be *England* (D), or some variation of this term, e.g., *Great Britain*, *United Kingdom*. There is more than one possibility for X (D). The model can return an arbitrary number of items, determined in descending order of similarity to C. We ran this analogy on the Word2Vec algorithm. The result (X-1) is *Britain*, with a vector similarity of 0.72. X-2, the second most similar item to London (given the analogy), is *UK*, vector similarity of 0.68, hence close to X-1. Jumping further down the line, at position X-9, we get *Scotland*, vector similarity 0.51, and at X-10 *continental_Europe*, vector similarity 0.51. We can thus see, the further away from X-1, the smaller the vector similarity gets, accompanied by a plethora of wrong answers.

We use this approach to explore the relationships between psychiatric diagnoses, occupations, and perceived occupational fitness/employability. Examining relationships within word embeddings through vector arithmetic word analogies are a popular research approach [94, 141, 142], as they have become a proxy for bias [89, 94, 105, 137, 143, 144].

2.5.1 Alternative methods.   We appreciate the fact that there are other quantitative analyses methods of biases in word embeddings, such as the WEAT [85] or MAC [145] which all have their own shortcomings which have been discussed in detail by Schroeder and colleagues [146] SAME is the latest method described to overcome limitations of the prior two [146].

Still, we side with Straw & Callison-Burch [105] in that, when starting an initial investigation into the combination of mental health biases within occupations and consequent allocation harms, an open-ended analogy approach allows for a wider scope of discovery, since analogies can demonstrate bias with simple examples [89, 92, 94] and have become a benchmark method of examining word embeddings [143, 147, 148].

2.5.2 Examining the validity of word analogies as indicators of bias.   The utilisation of word analogies for investigating linguistic bias has garnered contention within the research community. Varied interpretations of the same results have led to ambiguous conclusions, accentuating the need for a thorough examination of this methodology.

Petreski & Hashim [149] critique the use of analogies for bias detection in word embeddings, terming them "inaccurate and incompetent diagnostic tools for bias in word embeddings" (pp. 978). Conversely, Ushio et al. [143] argue that analogies "misguide or hide the real relationships existing in the vector space" (pp.3). Despite the difference in assertions, both sets of authors cite the same seminal works—Gonen & Goldberg [150] and Nissim et al. [151]—to support their arguments.

In the following, we look at both these publications, Gonen & Goldberg and Nissim et al.

Nissim et al. align with the scepticism surrounding the efficacy of analogies as bias diagnostics. They themselves also reference Gonen & Goldberg to substantiate their claims.

However, a closer inspection reveals that Gonen & Goldberg were primarily focused on the limitations of existing bias removal techniques, particularly in the context of gender-neutral modelling.

Schröder et al. [146] offer insights into the findings of Gonen & Goldberg, proposing two plausible explanations: either the debiasing methods were inadequately executed or the stereotypical groups identified were reflective of other relations unrelated to the bias attributes, thus misguiding the classification task. They argue that these factors could potentially account for the observed persistence of bias, thereby challenging the assertion that cosine-based metrics are ineffective for investigating bias in word embeddings, based on findings by Gonen & Goldberg alone.

Moreover, Gonen and Goldberg acknowledge that, while bias direction can facilitate the measurement of a word's bias association, it doesn't conclusively determine it. This suggests that bias direction serves as a tool for detecting bias association, but its efficacy in revealing the true extent of bias remains under question: Bias can be detected, the magnitude of it might be hidden.

Considering the above discussions, it becomes imperative to further scrutinise the arguments presented by Nissim et al. against the use of word analogies as proxies for bias.

Nissim et al. did not present empirical evidence undermining the accuracy of using analogy as a proxy for bias; their critique is primarily methodological. They posit that the customary formation of analogies could skew the results, as the model is compelled to generate a distinct concept from the input terms. They label this phenomenon a "dangerous artefact" (pp. 488) when words are desired to be the same—exemplified by the analogy 'man is to doctor as woman is to doctor'.

The argument of Nissim et al. may not hold in scenarios where ample synonyms exist, which could provide alternative, yet equally valid, outputs when investigating bias through analogies. Our empirical examination, as illustrated in Tables 1–3 (utilizing a Word2Vec model which was discussed in detail above), demonstrates that reversing the analogy (e.g.,

Table 1. Top-10 returns from the word2vec algorithm when asked the word analogy query "*woman* is to doctor what *man* is to X".

Position	Return Item	Similarity
1	physician	0. 64
2	doctors	0.58
3	surgeon	0. 57
4	dentist	0.55
5	cardiologist	0.54
6	neurologist	0. 52
7	neurosurgeon	0.52
8	urologist	0.52
9	Doctor	0.52
10	internist	0.51

https://doi.org/10.1371/journal.pone.0315768.t001

Table 2. Top-10 returns from the word2vec algorithm when asked the word analogy query "*man* is to lawyer what *woman* is to X".

Position	Return Item	Similarity
1	lawyer	0.68
2	attorneys	0.67
3	Attorney	0.62
4	Attorneys	0.56
5	prosecutor	0.55
6	counsel	0.54
7	solicitor	0.54
8	Attorney_Stephen_Houze	0.53
9	lawyers	0.53
10	Attorney_Ralph_Capitelli	0.53

https://doi.org/10.1371/journal.pone.0315768.t002

Table 3. Top-10 returns from the word2vec algorithm when asked the word analogy query "*man* is to doctor what *woman* is to X".

Position	Return Item	Similarity
1	gynecologist	0.70
2	nurse	0.64
3	doctors	0.64
4	physician	0.64
5	pediatrician	0.62
6	nurse practitioner	0.62
7	obstetrician	0.60
8	ob gyn	0.59
9	midwife	0.59
10	dermatologist	0.57

https://doi.org/10.1371/journal.pone.0315768.t003

'woman is to doctor what man is to X') as well as the original, and another control within the field of Law, yields desirable and 'correct' terms, underscoring the potential for mitigating the identified issue through methodological adjustments due to the availability of synonyms.

Therefore, including multiple 'correct' answers might render this first critique point superfluous. Suggestions similar to ours were made by Newman-Griffis et al. [152].

Second, Nissim et al. argue, morphosyntactic and semantic levels are not always distinct, i.e., there is a correct, grammatical answer to be expected when asking "man is to actor what woman is to X"–the term *actress* is morphosyntacticly correct. The same applies when asking "London is to England what Tokyo is to X", *Japan* is *factually* the correct answer. Therefore, Nissim et al's argument goes,

"querying man:doctor:: woman:X, is one after a morphosyntactic or a semantic answer, and what would be the correct one?" (pp. 490). As they state themselves, morphosyntactically *doctor*, should be returned, which, however, violates the all-terms-different-constraint. This problem has been discussed in our previous paragraph; It remains the semantics: Nissim et al. see no single predefined term that "correctly" completes the analogy.

We argue, some answers are more or less biased than others, whereby some even appear plain wrong or not applicable. When asking 'man:doctor:: woman:X', and the model returned *tree*, this would be non-applicable. If the model returns *nurse*, it is biased, if it returns some variation or synonym for *doctor*, it is not biased. Same for 'man:attorney:: woman:X'. If the model gave back *table*, it would be non-applicable, if it gave back *paralegal*, it would be biased, if it gave back some variation of *attorney*, it would not be biased.

Nissim et al. further note, "In order to claim bias, one should also conceive the expected unbiased term" (pp. 490). This is easily done, and we will abide to this in this paper. In this, we assume common sense in readers to know what qualifies as biased, as discussed in the paragraph above.

To sum up, Nissim et al, when subtracting the points made by us, rather than showing analogies to be inaccurate, explained best practices in how to use analogies to detect bias. Analogies remain a sound method for diagnosing bias in word embeddings. They have been getting a tainted reputation in the literature through flawed assessment as well as citation of statements taken out of context.

3. Results

3.1 Word2Vec word analogies

Tables 4 & 5 show cosine similarity scores for top-10 returned words when querying two analogies regarding employability and mental health in general. The psychosis related analogy

Table 4. Words to complete the analogy 'healthy is to employable as psychosis is to X'.

Position	Return Item	Similarity
1	psychotic symptom	0.5148
2	psychotic disorders	0.5094
3	mental illness	0.4869
4	**unemployable**	**0.4844**
5	schizophrenia	0.4670
6	psychiatric disorder	0. 4611
7	psychopathology	0.4603
8	psychotic	0.4587
9	mental disorder	0.4532
10	psychiatrist	0.4456

Table 5. Words to complete the analogy 'healthy is to employable as depression is to X'.

Position	Return Item	Similarity
1	**unemployable**	**0. 4899**
2	mental illness	0. 4846
3	mental disorders	0. 4711
4	mental illnesses	0. 4683
5	depressive illness	0. 4643
6	depressive episode	0. 4612
7	undergone electroshock therapy	0. 4581
8	attempters	0. 4533
9	psychotic disorders	0. 4520
10	bipolar disorder	0. 4495

features one biased item in total, which is on position 4 (*unemployable*, Cosine Similarity = 0.4844); the depression related analogy features one biased item in total, which is on position 1 (*unemployable*, Cosine Similarity = 0.4899). This suggests bias to be present in both queries, however, strongest for the depression related analogy.

3.1.1 Psychosis. See Table 4.

3.1.2 Depression. See Table 5.

Tables 6–8 show cosine similarity scores for top-10 returned words when querying analogies regarding employability, mental health and high earning/prestige professions. No analogy features a biased word in the top-10 returned items. This indicates that the model might see people with mental illness as equally able to enter and/or perform in high earning professions as healthy individuals.

3.1.3 Law. See Table 6.

3.1.4 Medicine. *3.1.4.1 Doctor*. See Table 7.

3.1.4.2 Surgeon. See Table 8.

More queries can be found in the S1 –S15 Boxes in S1 File. There is no order to which analogy results were included in the main manuscript and which went to the supporting information, except the analogy regarding employability and depression, since this was the only analogy with a biased item on position 1.

Table 6. Words to complete the analogy 'healthy is to lawyer as psychosis is to X'.

Position	Return Item	Similarity
1	lawyer	0. 5526
2	attorney	0. 5462
3	psychiatrist	0. 5153
4	attorney Norm Pattis	0. 5069
5	barrister	0. 5054
6	lawyers	0. 5029
7	prosecutor	0. 4914
8	forensic psychologist	0. 4887
9	solicitior	0. 4878
10	attornies	0. 4766

Table 7. Words to complete the analogy 'healthy is to doctor as psychosis is to X'.

Position	Return Item	Similarity
1	psychiatrist	0. 6376
2	psychiatric	0. 5388
3	psychiatrists	0. 5316
4	neurologist	0. 5170
5	Psychiatrist	0. 5103
6	physician	0. 5057
7	psychotic_episodes	0. 5037
8	paranoid_psychosis	0. 4934
9	forensic_psychologist	0. 4880
10	psychologist	0. 4863

https://doi.org/10.1371/journal.pone.0315768.t007

Table 8. Words to complete the analogy 'healthy is to surgeon as psychosis is to X'.

Position	Return Item	Similarity
1	psychiatrist	0. 572
2	neurosurgeon	0. 550
3	neurologist	0. 496
4	psychiatrists	0. 475
5	Psychiatrist	0. 5103
6	'forensic_psychologist'	0. 467
7	'forensic_psychiatry'	0. 456
8	'Surgeon'	0. 452
9	'psychiatric'	0. 451
10	'urologist'	0. 450

https://doi.org/10.1371/journal.pone.0315768.t008

3.2 GloVe

3.2.1 Reading the chart. Figs 1 and 2 display the t-SNE cosine proximities of the multi-dimensional embeddings produced from the GloVe algorithm. The x-axis of the graph represents the "healthiness" of a concept. As an example, the fact that the word *depression* is being displayed on the x-axis around -.10 means that it is closer to the *ill* end of the spectrum than to the *healthy* end. This is likely because depression negatively affects moods and thoughts. It's important to note again that the position of a word on the x-axis of this graph is based on its similarity to the vector obtained by subtracting the vector for *ill* from the vector for *healthy* using GloVe embeddings. Therefore, the position of a word on the x-axis does not necessarily reflect its actual, real world, epidemiological healthiness or illness. Rather, it reflects how similar the word is to the vector that represents the concept of "healthiness" versus "illness" in the GloVe embedding space. The same applies to employability on the y-axis.

3.2.2 Diagnoses, health, and employability. *3.2.2.1 Diagnoses and health.* From the list we supplied to the model, *depression* is seen as the most ill psychiatric diagnosis (similarity to *healthy* is -0.097983), *ADHD* as the healthiest (similarity to *healthy* is 0.088889). *Depression*, together with *bipolar* (-0.082122) and *psychosis* (-0.069086) are visibly the least healthy, shown in the bottom left corner. *Schizophrenia* shows up in the lower mid-field (-0.022633). *Eating disorder* (0.040484), *OCD* (0.035088) and *anxiety disorder* (0.033978) are all seen as similarly healthy in the midrange. Most of physical control diagnoses are also in the mid-healthy section. An outlier is *back pain*, which is seen as ill as *bipolar* and *psychosis* and only
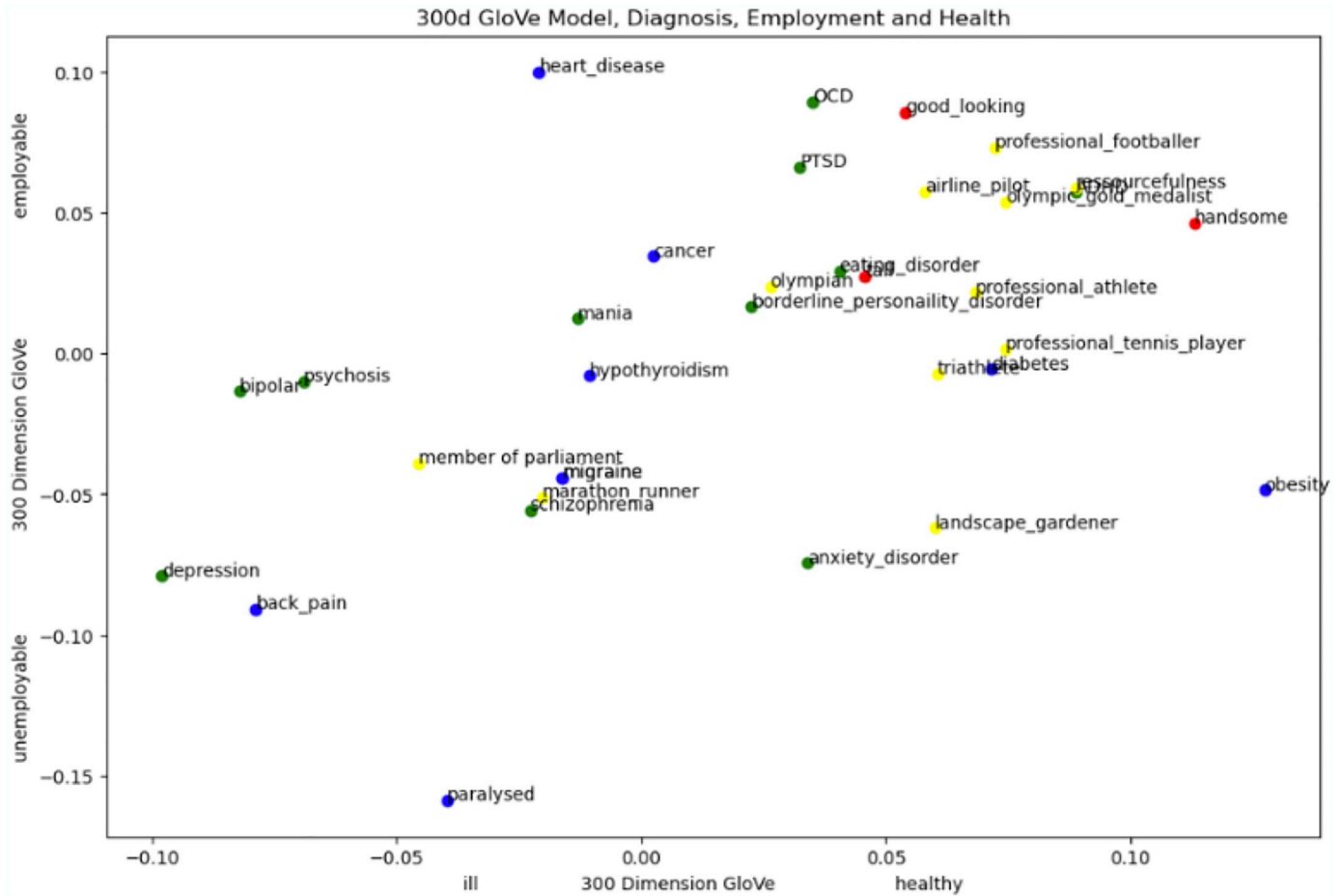
Fig 1. 300 dimensions GloVe word vectors. Vectors of words 'employable' and 'unemployable' as poles on the Y axis. The X axis contains words 'ill' and 'healthy'. Psychiatric diagnoses (green) physical diagnoses (blue) and favourable physical attribute control terms (red) as well as very healthy control terms (yellow). This examines how different psychiatric labels correspond to concepts of employability and health.

https://doi.org/10.1371/journal.pone.0315768.g001

slightly healthier than *depression*. *Obesity* (0.127438) is seen as the healthiest attribute, even healthier than favourable physical attribute control terms *handsome*, *good looking* or *tall* and even healthier than *olympic gold medallist*, *professional footballer*, and *professional tennis player*.

3.2.2.2 Diagnoses and employability. The least employable psychiatric diagnosis is *depression* (similarity to *employable* is -0.07907), slightly less employable than *anxiety disorder* (-0.074401) and *schizophrenia* (-0.055556). The most employable psychiatric diagnosis is OCD (0.089468). *Paralysed* is the least employable physical ailment as well as overall item (-0.158639), *heart disease* is the most employable physical diagnosis and overall item (0.099992).

3.2.3 GloVe, diagnoses, normality and reliability. To estimate stability of findings, we repeated our analyses with two comparable words, i.e., normal abnormal and reliable unreliable.

3.2.3.1 Diagnoses and normality. Whereas *depression* was in the previous figure seen as the most ill (similarity to healthy is -0.097983), in this figure, *depression* is seen as the most normal of all mental health diagnoses (similarity to *normal* 0.002462). *Psychosis* is seen as the most
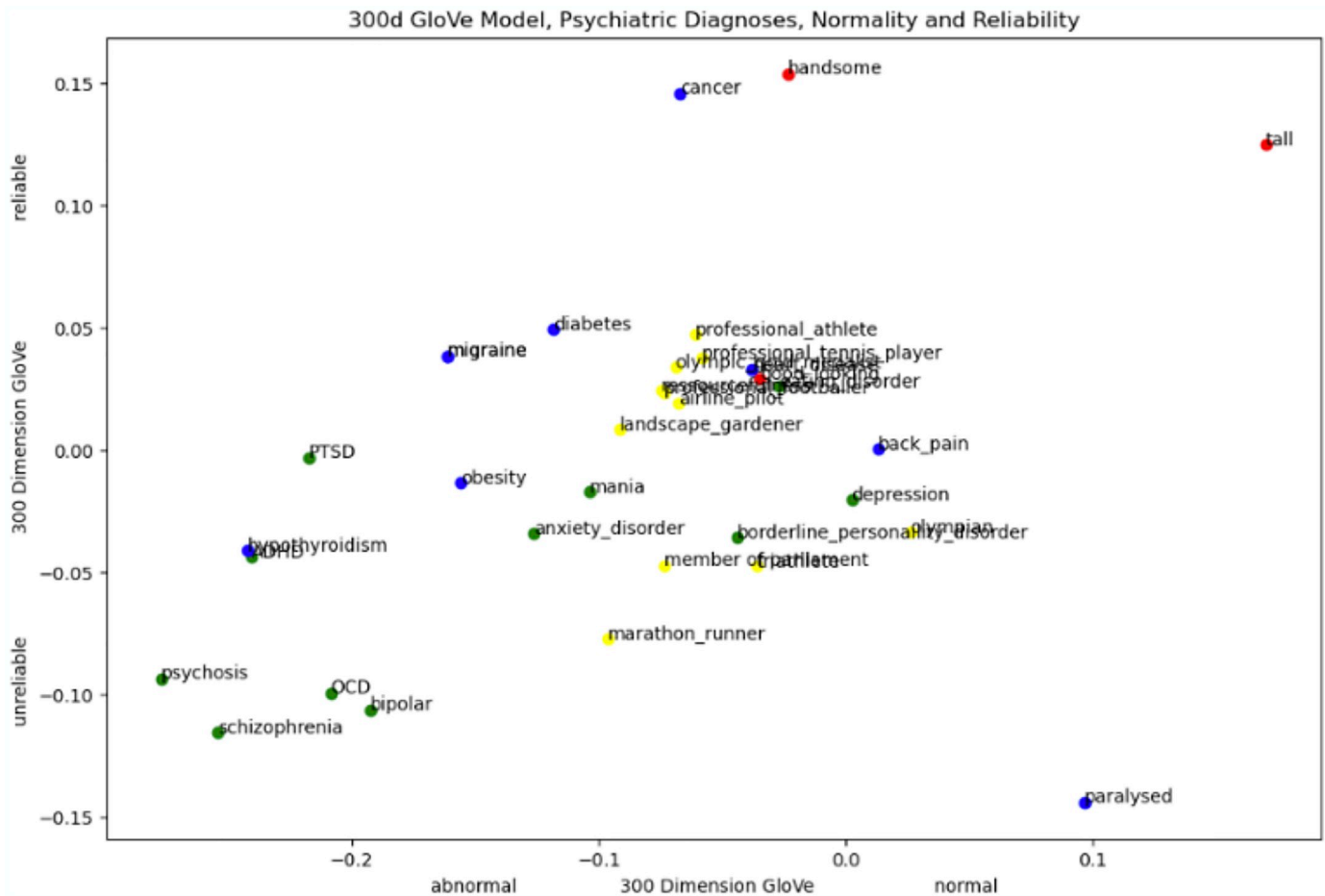
Fig 2. 300 dimensions GloVe word vectors. Vectors of words 'reliable' and 'unreliable' as poles on the Y axis. The X axis words 'normal' and 'abnormal'. Psychiatric diagnoses (green) physical diagnoses (blue) and favourable physical attribute control terms (red) as well as very healthy control terms (yellow). This examines how different labels vary in the relationship to the concepts of 'unemployable' and 'healthy'.

abnormal (similarity to *normal* -0.276995) *OCD*, *PTSD*, *ADHD*, *Schizophrenia* and *bipolar* are all similarly seen as abnormal, shown in the bottom left corner. *Mania* and *anxiety* are similarly in the lower midfield, whereas *borderline* and *eating disorder* are closer to *depression* and therefore seen as normal.

Most of physical control diagnoses are also in the mid-normal section. An outlier is *paralysed*, which is seen as considerably more normal than any mental or physical ailment. Physical attributes like *handsome* or *good-looking* are seen about as normal as *eating disorders* and *borderline*. The most normal seen by far is *tall*, interestingly followed by *paralysed*, which is the most normal physical ailment by far.

3.2.3.2 Diagnoses and reliability. Physical control conditions are seen as similarly, but slightly more reliable than *anxiety disorders*, *depression* and *borderline* as well as *anxiety disorder* and *ADHD*. *Cancer* is the most reliable physical health condition. It is even mor reliable than *tall* and only slightly worse than *handsome*. *Psychosis* is seen as the least reliable, close to *OCD*, *bipolar* and *psychosis*. Favourable physical control terms *handsome* and *tall* are seen as considerably more reliable than most other items supplied to the model. *Paralysed*, while being the most normal physical condition, is the least reliable term overall.

4. Discussion

To our knowledge, this is the first investigation of NLP models and bias against people with mental health disorder in the context of employability.

Out of eleven analogies to investigate mental health bias and employability and corresponding attributes, none but one exhibits a bias in the top-1 returned item is an applicable, biased term (Analogy: *"healthy (A) is to employable (B) as depression (C) is to _ (D)?"* D = *unemployable*).

Some analogies yield discriminatory items at second returned item (Analogy: *"healthy (A) is to employable (B) as anxiety disorder (C) is to _ (D)?"* D at Top-2 = *unemployable*; Analogy: *"healthy (A) is to reliable (B) as psychosis (C) is to _ (D)?"* D at Top-2 = *unreliability*), and some at the third returned (Analogy: *"healthy (A) is to reliable (B) as anxiety disorder (C) is to _ (D)?"* D at Top-3 = *unreliable)* and fourth (Analogy: *"healthy (A) is to employable (B) as psychosis (C) is to _ (D)?"* D at Top-4 = *unemployable)*. These are in all cases very close to the first hit. Nissim et al. [151] suggest considering at least the top-5 up to the top-10 returned items. However, even when looking at perfect analogies, top-10 returns contain unrelated words, *London is to England what Tokyo is to X*, *Ronaldo* (.512) and *rooney* (.502) come positions two and three, not far from Japan (.547), with many others not related or incorrect, such as *America* (.495) at five or *juan* at six (.490). It is therefore questionable how much weight should be given to second or further down positions in analogy answers. There remains no gold-standard or rule for this situation. It is furthermore expected that accuracy would decrease, the further one goes away from the top-1.

We do, however, note an interesting point: When querying the reverse, i.e. *Tokyo is to Japan as London is to X*, the model gives more sensible answers. This might speak for the fact that more is being written about London and the UK than there is about Tokyo and Japan, thus resulting in those words featuring more in the training text corpus, resulting in more accurate embeddings. In the same notion, there might be more written about depression, however less about schizophrenia, thus resulting in more embeddings for depression, thus leaving more room for discrimination.

This would leave embeddings for schizophrenia be less accurate, however, at the same time, less discriminating. Individuals with schizophrenia would therefore, by accident due to fewer text and word embeddings, be discriminated less.

No analogy investigating bias against specific job titles shows evidence of bias. All returned top-1 items are of equal desirability to (B). Furthermore, while in the analogies investigating employability attributes biased terms were found in the top-2 or top-10, when investigating high earning professions, in most cases all top 10 returned items are of equal desirability.

This is in stark contrast to most other papers using similar methodology to ours, who found profound evidence of bias against marginalised groups [105, 110, 145].

Straw & Callison-Burch [105] are the only study investigating bias within NLP against mental health, looking at demographic categories, not at occupations. Our absence of evidence for much bias against NLP might therefore be explained by investigating a different section of mental health bias. Furthermore, Straw & Callison-Burch's queries were in the format *A is to mental health disorder (B) as C is to D*, example: *Grandparent is to Depression, as Adolescent is to _ (W4)?* Or *British is to Depression, as Irish is to _ (W4)* or *Christian is to Depression, as Atheist is to _ (W4)*. This was done since they were looking at clinical misuse of NLP, i.e., which demographics were most likely to be associated with which disorder and therefore might be under or over diagnosed due to their demographic group. This framing is assuming a difference in magnitude, not in classification, as a pathology was expected and even all but forced to be returned, which is fine for their kind of framing of the research question. Another way of

thinking about this is, if they substituted diagnoses for fruit, they would be asking *Christian is to liking apples*, *as atheist is to X*, the return would be very likely what fruit or at least what general food atheist would like.

In contrast, we put emphasis on the diagnoses, not the demographic, i.e. we did not ask which demographic is associated with which diagnosis, as Straw & Callison-Burch did. We asked which diagnoses are most associated with which profession, hence allowing for an unbiased return, i.e., psychosis is as much associated with being a CEO as healthy is associated with being a CEO. We could have also played the same way as Straw & Callison-Burch, asking which profession is associated with which condition, however, in the context of occupational bias, this is not as insightful as in the context of clinical diagnostic bias. These queries would however definitely help when investigating which professions are more associated to which diagnoses when clinical NLP models are used.

Word2Vec is trained on a large corpus of Google News articles. There is some evidence suggesting that reporting in the news about mental health has become more positive and liberating in recent years [153–155], which might contribute to lesser associations of mental health disorder and stigma related to employability mirrored in word embeddings.

In GloVe embeddings, we first see no clear clusters, i.e. there is a large amount of overlap between points belonging to physical and mental health, as well as controls.

Furthermore, severe mental health disorders such as bipolar and psychosis/schizophrenia and PTSD are seen more employable than common mental health disorders such as depression and anxiety. Psychiatric diagnoses and physical control conditions are overall similarly seen as employable. Physical conditions are both the most and the least employable.

At the same time, bipolar, psychosis and mania are more employable than very healthy controls such as member of parliament, landscape gardener and marathon runner. *Professional tennis player* appears slightly more employable than *psychosis*. *OCD* is more employable than *professional footballer*, *PTSD* is more employable than *airline pilot* and *Olympic gold medallist*.

Therefore, GloVe embeddings do exhibit bias against some mental health disorders, primarily depression, painting it as less employable than other, more severe, and rare conditions such as schizophrenia/psychosis or bipolar. This does not reflect actual real-world data, as sufferers of psychosis/schizophrenia and bipolar are less often in employment than people with common mental disorders or physical disorders such as back pain, while schizophrenia sufferers are more employed than people with bipolar [11, 156, 157]. In fact, back pain is one of the most common physical health disorders, especially in the workforce [158, 159].

GloVe embeddings, therefore, like Word2Vec, show limited bias against people with mental health disorders, mostly seeing them similar to somatic and very healthy control terms. Both algorithms appear thus safe for downstream use in job and/or candidate recommending or resume parsing, as the threat of allocation bias against the disorders we investigated appears low.

4.1 Limitations

Absence of evidence does not indicate evidence of absence of bias against mental health disorders in recruiting and job recommending in word embeddings. We did not find much occupational bias in the analogies we queried, however, we only looked at three diagnoses for Word2Vec analogies (psychosis, anxiety, depression), only a limited number of professions, and only high earning ones at that. Furthermore, we only looked at a handful of employability attributes. Other diagnoses, professions at lower salary levels or other employability attributes might contain bias.

4.2 Future research

There are more sophisticated bias investigation tools than word embeddings. A natural extension of our work is to repeat investigations into occupational and mental health bias using WEAT [85] MAC [145] or SAME [146] methods. Context-aware embeddings like BERT [160] and ELMo [161] outperform context-independent embeddings such as Word2vec and GloVe across various NLP tasks [162, 163]. Hence, there is a chance that downstream task developers might switch to BERT and ELMo for resume parsing or job recommending. This would call for a repeating of this study using such embeddings; these results might be different than ours, as not only co-occurrences are captured by more advanced models, but complex relationships between words within sentences.

5. Conclusion

Word2Vec embeddings perceive psychosis, anxiety disorder and depression as similarly employable to healthy controls. GloVe embeddings perceive some mental health disorder as being less healthy and less employable when compared to more severe mental health disorders and most physical health conditions. Overall, as with Word2Vec embeddings, GloVe appears to perceive a parity in physical and psychiatric disorders in terms of healthiness and employability. Our findings should make job seekers with mental health disorders hopeful, as our findings support the notion that they could openly disclose their condition to employers *without* facing discrimination. For future research, the use of sophisticated bias investigation tools and context-aware embeddings holds promise for a more nuanced discernment of occupational and mental health bias. This, in turn, could significantly bolster the robustness and fairness of intelligent applications within occupational recruitment, helping to build a fairer hiring landscape that provides better opportunities and uses human capital more efficiently.

Supporting information

S1 File. Further results from analogies investigating employability and associated attributes for people with mental health disorders.
(DOCX)

S2 File. Source code.
(DOCX)

Author Contributions

Conceptualization: Maximin Lange, Alexandros Koliousis, Ricardo Twumasi.

Data curation: Maximin Lange, Feras Fayez, Eoin Gogarty.

Formal analysis: Maximin Lange.

Funding acquisition: Maximin Lange.

Investigation: Maximin Lange.

Methodology: Maximin Lange.

Project administration: Maximin Lange, Ricardo Twumasi.

Resources: Maximin Lange.

Software: Maximin Lange, Feras Fayez, Eoin Gogarty.

Supervision: Ricardo Twumasi.

Visualization: Maximin Lange.

Writing – original draft: Maximin Lange.

Writing – review & editing: Alexandros Koliousis, Ricardo Twumasi.

References

1. Deepak G, Teja V, Santhanavijayan A. A novel firefly driven scheme for resume parsing and matching based on entity linking paradigm. Journal of Discrete Mathematical Sciences and Cryptography. 2020; 23: 157–165. https://doi.org/10.1080/09720529.2020.1721879

2. Mittal V, Mehta P, Relan D, Gabrani G. Methodology for resume parsing and job domain prediction. Journal of Statistics and Management Systems. 2020; 23: 1265–1274. https://doi.org/10.1080/09720510.2020.1799583

3. Sajid H, Kanwal J, Bhatti SUR, Qureshi SA, Basharat A, Hussain S, et al. Resume Parsing Framework for E-recruitment. 2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM). Seoul, Korea, Republic of: IEEE; 2022. pp. 1–8.

4. Myers S. 2023 Applicant Tracking System (ATS) Usage Report: Key Shifts and Strategies for Job Seekers. Jobscan. 2023. https://www.jobscan.co/blog/fortune-500-use-applicant-tracking-systems/

5. Mhamdi D, Moulouki R, El Ghoumari MY, Azzouazi M, Moussaid L. Job recommendation based on job profile clustering and job seeker behavior. Procedia Computer Science. Elsevier B.V.; 2020. pp. 695–699.

6. Lippens L, Vermeiren S, Baert S. The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments. Eur Econ Rev. 2023; 151: 104315. https://doi.org/10.1016/j.euroecorev.2022.104315

7. Spence JL, Hornsey MJ, Stephenson EM, Imuta K. Is Your Accent Right for the Job? A Meta-Analysis on Accent Bias in Hiring Decisions. Pers Soc Psychol Bull. 2024; 50: 371–386. https://doi.org/10.1177/01461672221130595 PMID: 36326202

8. Baert S. Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments Since 2005. In: Gaddis SM, editor. Audit Studies: Behind the Scenes with Theory, Method, and Nuance. Cham: Springer International Publishing; 2018. pp. 63–77.

9. Lange M, Twumasi R. Anxious People, Please Apply! No Evidence for Decreased Perceptions of Employability in Individuals with Mental and Physical Illness. N Am J Psychol. 2022; 24: 319–336. Available: https://www.proquest.com/scholarly-journals/anxious-people-please-apply-no-evidence-decreased/docview/2661589535/se-2?accountid=11862

10. Bjørnshagen V. The mark of mental health problems. A field experiment on hiring discrimination before and during COVID-19. Soc Sci Med. 2021; 283: 114181. https://doi.org/10.1016/j.socscimed.2021.114181 PMID: 34216884

11. Brouwers EPM. Social stigma is an underestimated contributing factor to unemployment in people with mental illness or mental health issues: position paper and future directions. BMC Psychol. 2020; 8: 36. https://doi.org/10.1186/s40359-020-00399-0 PMID: 32317023

12. Voldby KG, Hellström LC, Berg ME, Eplov LF. Structural discrimination against people with mental illness; a scoping review. SSM—Mental Health. 2022; 2: 100117. https://doi.org/10.1016/j.ssmmh.2022.100117

13. Østerud KL. Mental illness stigma and employer evaluation in hiring: Stereotypes, discrimination and the role of experience. Sociol Health Illn. 2023; 45: 90–108. https://doi.org/10.1111/1467-9566.13544 PMID: 36103320

14. McAlpine DD. Barriers to employment among persons with mental illness: A review of the literature. Institute for Health, Health Care Policy, and Aging Research; 2015.

15. Hennekam S, Follmer K, Beatty J. Exploring mental illness in the workplace: the role of HR professionals and processes. The International Journal of Human Resource Management. 2021; 32: 3135–3156. https://doi.org/10.1080/09585192.2021.1960751

16. Twamley EW, Narvaez JM, Becker DR, Bartels SJ, Jeste D V. Supported Employment for Middle-Aged and Older People with Schizophrenia. Am J Psychiatr Rehabil. 2008; 11: 76–89. https://doi.org/10.1080/15487760701853326 PMID: 19212460

17. Follmer KB, Jones KS. Mental Illness in the Workplace: An Interdisciplinary Review and Organizational Research Agenda. J Manage. 2018; 44: 325–351. https://doi.org/10.1177/0149206317741194

18. Organization WH. Mental health at work. WHO News Room. 2024. https://www.who.int/news-room/fact-sheets/detail/mental-health-at-work

19. Khurana D, Koli A, Khatter K, Singh S. Natural language processing: state of the art, current trends and challenges. Multimed Tools Appl. 2023; 82: 3713–3744. https://doi.org/10.1007/s11042-022-13428-4 PMID: 35855771

20. Guetterman TC, Chang T, DeJonckheere M, Basu T, Scruggs E, Vydiswaran VV. Augmenting Qualitative Text Analysis with Natural Language Processing: Methodological Study. J Med Internet Res. 2018; 20: e231. https://doi.org/10.2196/jmir.9702 PMID: 29959110

21. Asudani DS, Nagwani NK, Singh P. Impact of word embedding models on text analytics in deep learning environment: a review. Artif Intell Rev. 2023; 56: 10345–10425. https://doi.org/10.1007/s10462-023-10419-1 PMID: 36844886

22. Ni X, Dai H, Ren Z, Li P. Multi-Source Multi-Type Knowledge Exploration and Exploitation for Dialogue Generation. https://github.com/Patrick-Ni/KnowEE

23. Wan R, Etori N, Badillo-Urquiola K, Kang D. User or Labor: An Interaction Framework for Human-Machine Relationships in NLP. arXiv; 2022. http://arxiv.org/abs/2211.01553

24. Jim JR, Talukder MAR, Malakar P, Kabir MM, Nur K, Mridha MF. Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. Natural Language Processing Journal. 2024; 6: 100059. https://doi.org/10.1016/j.nlp.2024.100059

25. Olujimi PA, Ade-Ibijola A. NLP techniques for automating responses to customer queries: a systematic review. Discover Artificial Intelligence. 2023; 3: 20. https://doi.org/10.1007/s44163-023-00065-5

26. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. Journal of the American Medical Informatics Association. 2019; 26: 364–379. https://doi.org/10.1093/jamia/ocy173 PMID: 30726935

27. Alemzadeh H, Devarakonda M. An NLP-based cognitive system for disease status identification in electronic health records. 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). Orland, FL, USA: IEEE; 2017. pp. 89–92.

28. Kormilitzin A, Vaci N, Liu Q, Nevado-Holgado A. Med7: A transferable clinical natural language processing model for electronic health records. Artif Intell Med. 2021; 118: 102086. https://doi.org/10.1016/j.artmed.2021.102086 PMID: 34412834

29. Chilman N, Song X, Roberts A, Tolani E, Stewart R, Chui Z, et al. Text mining occupations from the mental health electronic health record: a natural language processing approach using records from the Clinical Record Interactive Search (CRIS) platform in south London, UK. BMJ Open. 2021; 11: e042274. https://doi.org/10.1136/bmjopen-2020-042274 PMID: 33766838

30. Hauser TU, Skvortsova V, De Choudhury M, Koutsouleris N. The promise of a model-based psychiatry: building computational models of mental ill health. Lancet Digit Health. 2022; 4: e816–e828. https://doi.org/10.1016/S2589-7500(22)00152-2 PMID: 36229345

31. Gamieldien Y, Case JM, Katz A. Advancing Qualitative Analysis: An Exploration of the Potential of Generative AI and NLP in Thematic Coding. SSRN Electronic Journal. 2023. https://doi.org/10.2139/ssrn.4487768

32. Abram MD, Mancini KT, Parker RD. Methods to Integrate Natural Language Processing Into Qualitative Research. Int J Qual Methods. 2020; 19: 160940692098460. https://doi.org/10.1177/1609406920984608

33. Khan NS, Abid A, Abid K. A Novel Natural Language Processing (NLP)–Based Machine Translation Model for English to Pakistan Sign Language Translation. Cognit Comput. 2020; 12: 748–765. https://doi.org/10.1007/s12559-020-09731-7

34. Dörpinghaus J, Darms J, Jacobs M. What was the Question? A Systematization of Information Retrieval and NLP Problems. 2018. pp. 471–478.

35. Jun C, Jang H, Sim M, Kim H, Choi J, Min K, et al. ANNA":" Enhanced Language Representation for Question Answering. Proceedings of the 7th Workshop on Representation Learning for NLP. Stroudsburg, PA, USA: Association for Computational Linguistics; 2022. pp. 121–132.

36. Bhoir N, Jakate M, Lavangare S, Das A, Kolhe S. Resume Parser using hybrid approach to enhance the efficiency of Automated Recruitment Processes. 2023 Apr.

37. Bhatia V, Rawat P, Kumar A, Shah RR. End-to-End Resume Parsing and Finding Candidates for a Job Description using BERT. 2019.

38. Bhor S, Gupta V, Nair V, Shinde H, Kulkarni S. M. Resume Parser Using Natural Language Processing Techniques. International Journal of Research in Engineering and Science. 2021;9: 1–6.

39. Kecht C, Kurschilgen M, Strobel M. Revival of the Cover Letter? Experimental Evidence on the Performance of AI-driven Personality Assessments. Copenhagen; 2022. https://n2t.net/ark:/51647/srd1323044

40. Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, Agrawal M, et al. Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. PLoS One. 2013; 8: e73791. https://doi.org/10.1371/journal.pone.0073791 PMID: 24086296

41. Allik J, Realo A, McCrae RR. Conceptual and methodological issues in the study of the personality-and-culture relationship. Front Psychol. 2023; 14. https://doi.org/10.3389/fpsyg.2023.1077851 PMID: 37057156

42. Tripp A, Munson B. Perceiving gender while perceiving language: Integrating psycholinguistics and gender theory. WIREs Cognitive Science. 2022; 13. https://doi.org/10.1002/wcs.1583 PMID: 34716654

43. Lewis M, Lupyan G. Gender stereotypes are reflected in the distributional structure of 25 languages. Nat Hum Behav. 2020; 4: 1021–1028. https://doi.org/10.1038/s41562-020-0918-6 PMID: 32747806

44. Fineberg SK, Deutsch-Link S, Ichinose M, McGuinness T, Bessette AJ, Chung CK, et al. Word use in first-person accounts of schizophrenia. British Journal of Psychiatry. 2015; 206: 32–38. https://doi.org/10.1192/bjp.bp.113.140046 PMID: 24970770

45. Corcoran CM, Mittal VA, Bearden CE, E. Gur R, Hitczenko K, Bilgrami Z, et al. Language as a bio-marker for psychosis: A natural language processing approach. Schizophr Res. 2020; 226: 158–166. https://doi.org/10.1016/j.schres.2020.04.032 PMID: 32499162

46. Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr. 2015; 1: 15030. https://doi.org/10.1038/npjschz.2015.30 PMID: 27336038

47. Zhang T, Schoene AM, Ji S, Ananiadou S. Natural language processing applied to mental illness detection: a narrative review. NPJ Digit Med. 2022; 5: 46. https://doi.org/10.1038/s41746-022-00589-7 PMID: 35396451

48. Mangalik S, Eichstaedt JC, Giorgi S, Mun J, Ahmed F, Gill G, et al. Robust language-based mental health assessments in time and space through social media. arXiv; 2023. http://arxiv.org/abs/2302.12952

49. Guntuku SC, Yaden DB, Kern ML, Ungar LH, Eichstaedt JC. Detecting depression and mental illness on social media: an integrative review. Curr Opin Behav Sci. 2017; 18: 43–49. https://doi.org/10.1016/j.cobeha.2017.07.005

50. Uban A-S, Chulvi B, Rosso P. An emotion and cognitive based analysis of mental health disorders from social media data. Future Generation Computer Systems. 2021; 124: 480–494. https://doi.org/10.1016/j.future.2021.05.032

51. Thorstad R, Wolff P. Predicting future mental illness from social media: A big-data approach. Behav Res Methods. 2019; 51: 1586–1600. https://doi.org/10.3758/s13428-019-01235-z PMID: 31037606

52. TIchich E. A Cross Cultural Comparison of Resume Content and Textual Features. Masters Thesis, University of Minnesota. 2005.

53. Derous E, Decoster J. Implicit Age Cues in Resumes: Subtle Effects on Hiring Discrimination. Front Psychol. 2017; 8. https://doi.org/10.3389/fpsyg.2017.01321 PMID: 28848463

54. He JC, Kang SK. Covering in Cover Letters: Gender and Self-Presentation in Job Applications. Academy of Management Journal. 2021; 64: 1097–1126. https://doi.org/10.5465/amj.2018.1280

55. Cugno M. Talk Like a Man: How Resume Writing Can Impact Managerial Hiring Decisions for Women. Graduate School Southern Illinois University Edwardsville. 2020. https://www.proquest.com/openview/4ccca1251597530d2c497df3f9274ad2/1?pq-origsite=gscholar&cbl=18750&diss=y

56. Kang SK, DeCelles KA, Tilcsik A, Jun S. Whitened Résumés: Race and Self-Presentation in the Labor Market. Adm Sci Q. 2016; 61: 469–502. https://doi.org/10.1177/0001839216639577

57. Parasurama P, Sedoc J, Ghose A. Gendered Information in Resumes and Hiring Bias: A Predictive Modeling Approach. SSRN Electronic Journal. 2022.

58. Marti Marcet S. Natural language processing for gender and ethnic differences in self-presentation strategies and career implications. Utrecht University. 2023. https://studenttheses.uu.nl/handle/20.500.12932/44294

59. de Ruijt C, Bhulai S. Job Recommender Systems: A Review. 2021.

60. Tran M-L, Nguyen A-T, Nguyen Q-D, Huynh T. A comparison study for job recommendation. 2017 International Conference on Information and Communications (ICIC). Hanoi, Vietnam: IEEE; 2017. pp. 199–204.

61. Hunkenschroer AL, Luetge C. Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda. Journal of Business Ethics. 2022; 178: 977–1007. https://doi.org/10.1007/s10551-022-05049-6

62. Köchling A, Wehner MC. Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. Business Research. 2020; 13: 795–848. https://doi.org/10.1007/s40685-020-00134-w

63. Lange M, Koutsouleris N, Twumasi R. Could We Prescribe Jobs? Recommendation Accuracy of Job Recommender Systems Using Machine Learning: A Systematic Review and Meta-Analysis. 2023.

64. Konstan JA, Adomavicius G. Toward identification and adoption of best practices in algorithmic recommender systems research. Proceedings of the International Workshop on Reproducibility and Replication in Recommender Systems Evaluation. Hong Kong China: ACM; 2013. pp. 23–28. https://doi.org/10.1145/2532508.2532513

65. Beel J, Breitinger C, Langer S, Lommatzsch A, Gipp B. Towards reproducibility in recommender-systems research. User Model User-adapt Interact. 2016; 26: 69–101. https://doi.org/10.1007/s11257-016-9174-x

66. Appadoo K, Soonnoo MB, Mungloo-Dilmohamud Z. Job Recommendation System, Machine Learning, Regression, Classification, Natural Language Processing. 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE). IEEE; 2020. pp. 1–6.

67. Kumari TS, Sagar K. A Semantic Approach to Solve Scalability, Data Sparsity and Cold-Start Problems in Movie Recommendation Systems. International Journal of Intelligent Systems and Applications in Engineering. 2023; 11: 825–837. Available: https://ijisae.org/index.php/IJISAE/article/view/2917

68. Alsaif SA, Sassi Hidri M, Ferjani I, Eleraky HA, Hidri A. NLP-Based Bi-Directional Recommendation System: Towards Recommending Jobs to Job Seekers and Resumes to Recruiters. Big Data and Cognitive Computing. 2022; 6: 147. https://doi.org/10.3390/bdcc6040147

69. Alsaif SA, Sassi Hidri M, Eleraky HA, Ferjani I, Amami R. Learning-Based Matched Representation System for Job Recommendation. Computers. 2022; 11: 161. https://doi.org/10.3390/computers11110161

70. Parida B, KumarPatra P, Mohanty S. Prediction of recommendations for employment utilizing machine learning procedures and geo-area based recommender framework. Sustainable Operations and Computers. 2022; 3: 83–92. https://doi.org/10.1016/j.susoc.2021.11.001

71. OpenAI. ChatGPY(July 2023 Version). 2023. https://chat.openai.com/chat

72. Ge Y, Hua W, Mei K, Ji J, Tan J, Xu S, et al. OpenAGI: When LLM Meets Domain Experts. arXiv; 2023. http://arxiv.org/abs/2304.04370

73. Okada Y, Mertens M, Liu N, Lam SSW, Ong MEH. AI and machine learning in resuscitation: Ongoing research, new concepts, and key challenges. Resusc Plus. 2023; 15: 100435. https://doi.org/10.1016/j.resplu.2023.100435 PMID: 37547540

74. Shumailov I, Shumaylov Z, Zhao Y, Gal Y, Papernot N, Anderson R. The Curse of Recursion: Training on Generated Data Makes Models Forget. arXiv; 2023. http://arxiv.org/abs/2305.17493

75. Alemohammad S, Casco-Rodriguez J, Luzi L, Humayun AI, Babaei H, LeJeune D, et al. Self-Consuming Generative Models Go MAD. 2023.

76. Stoltz DS, Taylor MA. Cultural Cartography with Word Embeddings. 2020.

77. Stevenson S, Merlo P. Beyond the Benchmarks: Toward Human-Like Lexical Representations. Front Artif Intell. 2022; 5: 796741. https://doi.org/10.3389/frai.2022.796741 PMID: 35685444

78. Tsujii J. Natural Language Processing and Computational Linguistics. Computational Linguistics. 2021; 1–21. https://doi.org/10.1162/coli_a_00420

79. Arseniev-Koehler A. Theoretical Foundations and Limits of Word Embeddings: What Types of Meaning can They Capture? Sociol Methods Res. 2022; 004912412211401. https://doi.org/10.1177/00491241221140142 PMID: 39554804

80. Mukhamediev RI, Popova Y, Kuchin Y, Zaitseva E, Kalimoldayev A, Symagulov A, et al. Review of Artificial Intelligence and Machine Learning Technologies: Classification, Restrictions, Opportunities and Challenges. Mathematics. 2022; 10: 2552. https://doi.org/10.3390/math10152552

81. Rezaeenour J, Ahmadi M, Jelodar H, Shahrooei R. Systematic review of content analysis algorithms based on deep neural networks. Multimed Tools Appl. 2023; 82: 17879–17903. https://doi.org/10.1007/s11042-022-14043-z PMID: 36313481

82. Malmqvist L, Yuan T, Manandhar S. Visualising Argumentation Graphs with Graph Embeddings and t-SNE. arXiv; 2021. http://arxiv.org/abs/2107.00528

83. Young JC, Rusli A. Review and Visualization of Facebook's FastText Pretrained Word Vector Model. 2019 International Conference on Engineering, Science, and Industrial Applications (ICESI). Tokyo, Japan: IEEE; 2019. pp. 1–6.

84. Kozlowski AC, Taddy M, Evans JA. The Geometry of Culture: Analyzing Meaning through Word Embeddings. 2018.

85. Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. Science (1979). 2017; 356: 183–186. https://doi.org/10.1126/science.aal4230 PMID: 28408601

86. Kapetanios E, Alshahrani S, Angelopoulou A, Baldwin M. What Do We Learn from Word Associations? Evaluating Machine Learning Algorithms for the Extraction of Contextual Word Meaning in Natural Language Processing. 2018 May.

87. Toney A. A Large-Scale, Automated Study of Language Surrounding Artificial Intelligence. arXiv; 2021. http://arxiv.org/abs/2102.12516

88. Periñán-Pascual C. Measuring associational thinking through word embeddings. Artif Intell Rev. 2022; 55: 2065–2102. https://doi.org/10.1007/s10462-021-10056-6

89. Durrheim K, Schuld M, Mafunda M, Mazibuko S. Using word embeddings to investigate cultural biases. British Journal of Social Psychology. 2023; 62: 617–629. https://doi.org/10.1111/bjso.12560 PMID: 35871272

90. Blodgett SL, Barocas S, Daumé III H, Wallach H. Language (Technology) is Power: A Critical Survey of "Bias" in NLP. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics; 2020. pp. 5454–5476.

91. Caliskan A, Ajay PP, Charlesworth T, Wolfe R, Banaji MR. Gender Bias in Word Embeddings: A Comprehensive Analysis of Frequency, Syntax, and Semantics. 2022.

92. Hovy D, Prabhumoye S. Five sources of bias in natural language processing. Lang Linguist Compass. 2021; 15: e12432. https://doi.org/10.1111/lnc3.12432 PMID: 35864931

93. Sun T, Gaut A, Tang S, Huang Y, ElSherief M, Zhao J, et al. Mitigating Gender Bias in Natural Language Processing: Literature Review. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics; 2019. pp. 1630–1640.

94. Garrido-Muñoz I, Montejo-Ráez A, Martínez-Santiago F, Ureña-López LA. A Survey on Bias in Deep NLP. Applied Sciences. 2021; 11: 3184. https://doi.org/10.3390/app11073184

95. Dev S, Sheng E, Zhao J, Amstutz A, Sun J, Hou Y, et al. On Measures of Biases and Harms in NLP. 2021.

96. Cheng L, Ge S, Liu H. Toward Understanding Bias Correlations for Mitigation in NLP. 2022.

97. Rogers A, Augenstein I. What Can We Do to Improve Peer Review in NLP? arXiv; 2020. http://arxiv.org/abs/2010.03863

98. Zhao J, Zhou Y, Li Z, Wang W, Chang K-W. Learning Gender-Neutral Word Embeddings. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: Association for Computational Linguistics; 2018. pp. 4847–4853.

99. Zhao J, Wang T, Yatskar M, Ordonez V, Chang K-W. Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods. arXiv; 2018. http://arxiv.org/abs/1804.06876

100. Lu K, Mardziel P, Wu F, Amancharla P, Datta A. Gender Bias in Neural Natural Language Processing. 2018.

101. Garg N, Schiebinger L, Jurafsky D, Zou J. Word embeddings quantify 100 years of gender and ethnic stereotypes. Proceedings of the National Academy of Sciences. 2018; 115. https://doi.org/10.1073/pnas.1720347115 PMID: 29615513

102. Kirk H, Jun Y, Iqbal H, Benussi E, Volpin F, Dreyer FA, et al. Bias Out-of-the-Box: An Empirical Analysis of Intersectional Occupational Biases in Popular Generative Language Models. arXiv; 2021. http://arxiv.org/abs/2102.04130

103. Raza S, Garg M, Reji DJ, Bashir SR, Ding C. NBIAS: A Natural Language Processing Framework for Bias Identification in Text. 2023.

104. Hutchinson B, Prabhakaran V, Denton E, Webster K, Zhong Y, Denuyl S. Social Biases in NLP Models as Barriers for Persons with Disabilities. arXiv; 2020. http://arxiv.org/abs/2005.00813

105. Straw I, Callison-Burch C. Artificial Intelligence in mental health and the biases of language based models. Danforth CM, editor. PLoS One. 2020; 15: e0240376. https://doi.org/10.1371/journal.pone.0240376 PMID: 33332380

106. Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space. arXiv; 2013. http://arxiv.org/abs/1301.3781

107. Chen Y, Mahoney C, Grasso I, Wali E, Matthews A, Middleton T, et al. Gender Bias and Under-Representation in Natural Language Processing Across Human Languages. Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society. New York, NY, USA: ACM; 2021. pp. 24–34.

108. Chen L, Sugimoto T. Quantifying and Debiasing Gender Bias in Japanese Gender-specific Words with Word Embedding. 2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS\&{ISIS}). Ise, Japan: IEEE; 2022. pp. 1–4.

109. Pair E, Vicas N, Weber AM, Meausoone V, Zou J, Njuguna A, et al. Quantification of Gender Bias and Sentiment Toward Political Leaders Over 20 Years of Kenyan News Using Natural Language Processing. Front Psychol. 2021; 12: 712646. https://doi.org/10.3389/fpsyg.2021.712646 PMID: 34955949

110. Bolukbasi T, Chang K-W, Zou JY, Saligrama V, Kalai AT. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. Adv Neural Inf Process Syst. 2016; 29.

111. Curto G, Jojoa Acosta MF, Comim F, Garcia-Zapirain B. Are AI systems biased against the poor? A machine learning analysis using Word2Vec and GloVe embeddings. AI Soc. 2022. https://doi.org/10.1007/s00146-022-01494-z PMID: 35789618

112. Sahlgren M, Olsson F. Gender Bias in Pretrained Swedish Embeddings. Proceedings of the 22nd Nordic Conference on Computational Linguistics. Turku, Finland; 2019. ttps://aclanthology.org/W19-6104

113. An H, Liu X, Zhang D. Learning Bias-reduced Word Embeddings Using Dictionary Definitions. Findings of the Association for Computational Linguistics: ACL 2022. Stroudsburg, PA, USA: Association for Computational Linguistics; 2022. pp. 1139–1152.

114. Kumar V, Bhotia TS, Kumar V, Chakraborty T. Identifying and Mitigating Gender Bias in Hyperbolic Word Embeddings. arXiv; 2021. http://arxiv.org/abs/2109.13767

115. Chen X, Li M, Yan R, Gao X, Zhang X. Unsupervised Mitigating Gender Bias by Character Components: A Case Study of Chinese Word Embedding. Proceedings of the 4th Workshop on Gender Bias in Natural Language Processing (GeBNLP). Stroudsburg, PA, USA: Association for Computational Linguistics; 2022. pp. 121–128.

116. Rawat A, Malik S, Rawat S, Kumar D, Kumar P. A Systematic Literature Review (SLR) On The Beginning of Resume Parsing in HR Recruitment Process & SMART Advancements in Chronological Order. 2021 Jul.

117. Zu S, Wang X. Resume Information Extraction with A Novel Text Block Segmentation Algorithm. International Journal on Natural Language Computing. 2019; 8: 29–48. https://doi.org/10.5121/ijnlc.2019.8503

118. Ayishathahira CH, Sreejith C, Raseek C. Combination of Neural Networks and Conditional Random Fields for Efficient Resume Parsing. 2018 International CET Conference on Control, Communication, and Computing (IC4). Thiruvananthapuram: IEEE; 2018. pp. 388–393.

119. Liu J, Shen Y, Zhang Y, Krishnamoorthy S. Resume Parsing based on Multi-label Classification using Neural Network models. 2021 6th International Conference on Big Data and Computing. Shenzhen China: ACM; 2021. pp. 177–185.

120. Pudasaini S, Shakya S, Lamichhane S, Adhikari S, Tamang A, Adhikari S. Scoring of Resume and Job Description Using Word2vec and Matching Them Using Gale—Shapley Algorithm. In: Jeena Jacob I, Gonzalez-Longatt FM, Kolandapalayam Shanmugam S, Izonin I, editors. Expert Clouds and Applications. Singapore: Springer Singapore; 2022. pp. 705–713.

121. Rus C, Luppes J, Oosterhuis H, Schoenmacker GH. Closing the Gender Wage Gap: Adversarial Fairness in Job Recommendation. arXiv; 2022. http://arxiv.org/abs/2209.09592

122. Joshi RK, Riti, Setru S, Srinivasaiah PT. Analysis On Word2Vec For User Recommendations. 2022 Fourth International Conference on Cognitive Computing and Information Processing (CCIP). Bengaluru, India: IEEE; 2022. pp. 1–6.

123. Dhameliya J, Desai N, Department of Information Technology Dharmsinh Desai University Nadiad India. Job Recommendation System using Content and Collaborative Filtering based Techniques. International Journal of Soft Computing and Engineering. 2019; 9: 8–13. https://doi.org/10.35940/ijsce.C3266.099319

124. Zhu G, Chen Y, Wang S. Graph-Community-Enabled Personalized Course-Job Recommendations with Cross-Domain Data Integration. Sustainability. 2022; 14: 7439. https://doi.org/10.3390/su14127439

125. Bothmer K, Schlippe T. Investigating Natural Language Processing Techniques for a Recommendation System to Support Employers, Job Seekers and Educational Institutions. In: Rodrigo MM, Matsuda N, Cristea AI, Dimitrova V, editors. Artificial Intelligence in Education Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium. Cham: Springer International Publishing; 2022. pp. 449–452.

126. Janssens KME, van Weeghel J, Dewa C, Henderson C, Mathijssen JJP, Joosen MCW, et al. Line managers' hiring intentions regarding people with mental health problems: a cross-sectional study on workplace stigma. Occup Environ Med. 2021; 78: 593–599. https://doi.org/10.1136/oemed-2020-106955 PMID: 33542095

127. Shankar J, Liu L, Nicholas D, Warren S, Lai D, Tan S, et al. Employers' Perspectives on Hiring and Accommodating Workers With Mental Illness. Sage Open. 2014; 4: 215824401454788. https://doi.org/10.1177/2158244014547880

128. Shahwan S, Yunjue Z, Satghare P, Vaingankar JA, Maniam Y, Janrius GCM, et al. Employer and Co-worker Perspectives on Hiring and Working with People with Mental Health Conditions. Community Ment Health J. 2022; 58: 1252–1267. https://doi.org/10.1007/s10597-021-00934-2 PMID: 35098388

129. Fukuura Y, Shigematsu Y. The Work Ability of People with Mental Illnesses: A Conceptual Analysis. Int J Environ Res Public Health. 2021; 18: 10172. https://doi.org/10.3390/ijerph181910172 PMID: 34639474

130. Pennington J, Socher R, Manning C. Glove: Global Vectors for Word Representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA, USA: Association for Computational Linguistics; 2014. pp. 1532–1543.

131. Brahushi G, Ahmad U. Empirical Evaluation of Word Representation Methods in the Context of Candidate-Job Recommender Systems. 2022 9th International Conference on Soft Computing & Machine Intelligence (ISCMI). Toronto, ON, Canada: IEEE; 2022. pp. 183–187.

132. Dolci T, Azzalini F, Tanelli M. Improving Gender-Related Fairness in Sentence Encoders: A Semantics-Based Approach. Data Sci Eng. 2023; 8: 177–195. https://doi.org/10.1007/s41019-023-00211-0

133. Trinh T-T-Q, Chung Y-C, Kuo RJ. A domain adaptation approach for resume classification using graph attention networks and natural language processing. Knowl Based Syst. 2023; 266: 110364. https://doi.org/10.1016/j.knosys.2023.110364

134. Raut S, Rathod A, Sharma P, Bhosale P, Zope B. Best-Fit: Best Fit Employee Recommendation. 2022 IEEE Pune Section International Conference (PuneCon). Pune, India: IEEE; 2022. pp. 1–6.

135. Kara A, Daniş FS, Orman GK, Turhan SN, Özlü ÖA. Job Recommendation Based on Extracted Skill Embeddings. In: Arai K, editor. Intelligent Systems and Applications. Cham: Springer International Publishing; 2023. pp. 497–507. https://doi.org/10.1007/978-3-031-16075-2_35

136. Schlippe T, Bothmer K. Skill Scanner: An AI-Based Recommendation System for Employers, Job Seekers and Educational Institutions. International Journal of Advanced Corporate Learning (iJAC). 2023; 16: 55–64. https://doi.org/10.3991/ijac.v16i1.34779

137. Ding L, Yu D, Xie J, Guo W, Hu S, Liu M, et al. Word Embeddings via Causal Inference: Gender Bias Reducing and Semantic Information Preserving. Proceedings of the AAAI Conference on Artificial Intelligence. 2022; 36: 11864–11872. https://doi.org/10.1609/aaai.v36i11.21443

138. Rathore A, Dev S, Phillips JM, Srikumar V, Zheng Y, Yeh C-CM, et al. VERB: Visualizing and Interpreting Bias Mitigation Techniques for Word Representations. 2021.

139. Ravfogel S, Vargas F, Goldberg Y, Cotterell R. Adversarial Concept Erasure in Kernel Space. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics; 2022. pp. 6034–6055.

140. Llorens M, Llorens Salvador M. Text Analytics Techniques in the Digital World: Word Embeddings and Bias. Irish Communications Review. 2018; 16: 76. https://doi.org/10.21427/D7TJ05

141. Allen C, Hospedales T. Analogies Explained: Towards Understanding Word Embeddings. 2019.

142. Ethayarajh K, Duvenaud D, Hirst G. Towards Understanding Linear Word Analogies. 2018.

143. Ushio A, Espinosa-Anke L, Schockaert S, Camacho-Collados J. BERT is to NLP what AlexNet is to CV: Can Pre-Trained Language Models Identify Analogies? arXiv; 2022. http://arxiv.org/abs/2105.04949

144. Abid A, Farooqi M, Zou J. Persistent Anti-Muslim Bias in Large Language Models. Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society. Virtual Event USA: ACM; 2021. pp. 298–306.

145. Manzini T, Lim YC, Tsvetkov Y, Black AW. Black is to Criminal as Caucasian is to Police: Detecting and Removing Multiclass Bias in Word Embeddings. arXiv; 2019. http://arxiv.org/abs/1904.04047

146. Schröder S, Schulz A, Kenneweg P, Feldhans R, Hinder F, Hammer B. The SAME score: Improved cosine based bias score for word embeddings. 2022.

147. Knipper A, Hassan M, Sadi M, Santu S. Analogy-Guided Evolutionary Pretraining of Binary Word Embeddings. Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing. Association for Computational Linguistics; 2022. pp. 683–693. https://aclanthology.org/2022.aacl-main.52/

148. Koehl D, Davis C, Nair U, Ramachandran R. Analogy-based Assessment of Domain-specific Word Embeddings. IEEE SoutheastCon. 2020. https://ntrs.nasa.gov/citations/20200001890

149. Petreski D, Hashim IC. Word embeddings are biased. But whose bias are they reflecting? AI Soc. 2023; 38: 975–982. https://doi.org/10.1007/s00146-022-01443-w

150. Gonen H, Goldberg Y. Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings But do not Remove Them. arXiv; 2019. http://arxiv.org/abs/1903.03862

151. Nissim M, Van Noord R, Van Der Goot R. Fair Is Better than Sensational: Man Is to Doctor as Woman Is to Doctor. Computational Linguistics. 2020; 46: 487–497. https://doi.org/10.1162/coli_a_00379

152. Newman-Griffis D, Lai AM, Fosler-Lussier E. Insights into Analogy Completion from the Biomedical Domain. arXiv; 2017. http://arxiv.org/abs/1706.02241

153. Goulden R, Corker E, Evans-Lacko S, Rose D, Thornicroft G, Henderson C. Newspaper coverage of mental illness in the UK, 1992–2008. BMC Public Health. 2011; 11: 796. https://doi.org/10.1186/1471-2458-11-796 PMID: 21992410

154. Whitley R, Wang J. Good News? A Longitudinal Analysis of Newspaper Portrayals of Mental Illness in Canada 2005 to 2015. The Canadian Journal of Psychiatry. 2017; 62: 278–285. https://doi.org/10.1177/0706743716675856 PMID: 27777273

155. Chen M, Lawrie S. Newspaper depictions of mental and physical health. BJPsych Bull. 2017; 41: 308–313. https://doi.org/10.1192/pb.bp.116.054775 PMID: 29234506

156. Luciano A, Meara E. Employment Status of People With Mental Illness: National Survey Data From 2009 and 2010. Psychiatric Services. 2014; 65: 1201–1209. https://doi.org/10.1176/appi.ps.201300335 PMID: 24933361

157. Holm M, Taipale H, Tanskanen A, Tiihonen J, Mitterdorfer-Rutz E. Employment among people with schizophrenia or bipolar disorder: A population-based study using nationwide registers. Acta Psychiatr Scand. 2021; 143: 61–71. https://doi.org/10.1111/acps.13254 PMID: 33155273

158. Ferreira ML, de Luca K, Haile LM, Steinmetz JD, Culbreth GT, Cross M, et al. Global, regional, and national burden of low back pain, 1990–2020, its attributable risk factors, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Lancet Rheumatol. 2023; 5: e316–e329. https://doi.org/10.1016/S2665-9913(23)00098-X PMID: 37273833

159. Melman A, Lord HJ, Coombs D, Zadro J, Maher CG, Machado GC. Global prevalence of hospital admissions for low back pain: a systematic review with meta-analysis. BMJ Open. 2023; 13: e069517. https://doi.org/10.1136/bmjopen-2022-069517 PMID: 37085316

160. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2018.

161. Peters M, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, et al. Deep Contextualized Word Representations. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). Stroudsburg, PA, USA: Association for Computational Linguistics; 2018. pp. 2227–2237.

162. Wang C, Nulty P, Lillis D. A Comparative Study on Word Embeddings in Deep Learning for Text Classification. Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval. Seoul Republic of Korea: ACM; 2020. pp. 37–46.

163. Wei H, Lin G, Li L, Jia H. A Context-Aware Neural Embedding for Function-Level Vulnerability Detection. Algorithms. 2021; 14: 335. https://doi.org/10.3390/a14110335