Depression at Work: Exploring Depression in Major US Companies from Online Reviews

INDIRA SEN, GESIS - Leibniz Institute for Social Sciences, Germany
DANIELE QUERCIA, Nokia Bell Labs and CUSP King's College London, United Kingdom
MARIOS CONSTANTINIDES, Nokia Bell Labs, United Kingdom
MATTEO MONTECCHI, King's College London, United Kingdom
LICIA CAPRA, University College London, United Kingdom
SANJA ŠĆEPANOVIĆ, Nokia Bell Labs, United Kingdom
RENZO BIANCHI, Norwegian University of Science and Technology, Norway

Studies on depression in the workplace have mostly investigated its impact on individual employees. Little is known about its association with the company as a whole, or the state where the company is based. This is due to the lack of scalable methodologies operationalizing depression in the specific context of the workplace, and of data documenting potential distress. In this work, we adapted a work-related depression scale called Occupational Depression Inventory (ODI), gathered more than 350K employee reviews of 104 major companies across the whole US for the (2008-2020) years, and developed a deep-learning framework (called AutoODI[1]) scoring these reviews on a composite ODI score. Presence of ODI mentions manifested itself not only at micro-level (companies scoring high in ODI suffered from low stock growth) but also at macro-level (states hosting these companies were associated with high depression rates, talent shortage, and economic deprivation). This new way of applying AutoODI onto company reviews offers both theoretical implications for the literature in computational social science, occupational health and economic geography, and practical implications for companies and policy makers.

CCS Concepts: • **Human-centered computing → Collaborative and social computing**; *Empirical studies in collaborative and social computing*.

Additional Key Words and Phrases: depression, workplace, deep-learning, companies

ACM Reference Format:
Indira Sen, Daniele Quercia, Marios Constantinides, Matteo Montecchi, Licia Capra, Sanja Šćepanović, and Renzo Bianchi. 2022. Depression at Work: Exploring Depression in Major US Companies from Online Reviews. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 438 (November 2022), 21 pages. https://doi.org/10.1145/3555539

[1]https://social-dynamics.net/AutoODI

Authors' addresses: Indira Sen, GESIS - Leibniz Institute for Social Sciences, Cologne, Germany, Indira.Sen@gesis.org; Daniele Quercia, Nokia Bell Labs and CUSP King's College London, Cambridge, United Kingdom, daniele.quercia@nokia-bell-labs.com; Marios Constantinides, Nokia Bell Labs, Cambridge, United Kingdom, marios.constantinides@nokia-bell-labs.com; Matteo Montecchi, King's College London, London, United Kingdom, matteo.montecchi@kcl.ac.uk; Licia Capra, University College London, London, United Kingdom, l.capra@ucl.ac.uk; Sanja Šćepanović, Nokia Bell Labs, Cambridge, United Kingdom, sanja.scepanovic@nokia-bell-labs.com; Renzo Bianchi, Norwegian University of Science and Technology, Trondheim, Norway, renzo.bianchi@ntnu.no.

2573-0142/2022/11-ART438 $15.00
https://doi.org/10.1145/3555539

1 INTRODUCTION

Workplace depression caused by adverse working environments damages employees' well-being, motivation, and performance, leading to a significant loss of productivity and competitiveness for many organizations [58]. However, despite being a pressing organizational and societal issue, workplace depression has mainly been researched at the individual employee level, and its impact on organizations and society remains poorly understood [106]. Furthermore, the lack of a specific approach to assess the depressive symptoms that employees experience in the workplace led to overgeneralized examinations of depression, which have limited relevance to the organizational context [10]. Typically, to assess employees' mental health problems, a company would administer tailored surveys that capture work-depressive situations; due to their high cost [103], these are often restricted to a limited pool of self-selected participants [40, 49, 51].

Our work set out to identify the factors influencing workplace depression and to assess its association with organizations, industry sectors, and the broader socio-economic environment. It does this through a large scale assessment of 350K geo-referenced employees reviews about 104 S&P 500 companies that is informed by the Occupational Depression Inventory (ODI) construct and measurement scale [7].

Unlike non-workplace specific depression scales (e.g., PHQ-9 [65]) or measurements related to work stress (e.g., burnout), the ODI is designed to quantify the severity of work-attributed depressive symptoms through a nine-item scale. While ODI is a conceptually and methodologically robust measurement tool (as outlined in Section 3), it has not yet been applied to examine workplace depression at organizational and societal levels. This is due primarily to the absence of scalable methodologies operationalizing the ODI construct, and of data that honestly document the experience of employees (see Section 2 for further details).

In tackling these significant conceptual and methodological drawbacks, our research offers three substantive contributions to the multidisciplinary research on depression in the workplace:

- We developed a state-of-the-art deep-learning Natural Language Processing (NLP) framework that accurately delineates the micro foundations of depression in the workplace from the perspective of each company's employees (Section 5), which we then empirically validated (Section 6.1).
- We found that high ODI scores were associated with market performance deterioration (Section 6.2). In particular, we found that companies with high ODI scores suffered from lower online ratings and, more importantly, lower financial growth.
- We found that high ODI scores were associated with the macroeconomic context in which the organizations operate (Section 6.4). In particular, a state-level analysis revealed that US states that host companies with high ODI scores also manifested high depression rates, talent shortage, and economic deprivation.

By highlighting the importance of targeted interventions and schemes that support employees' well-being, our findings offer potential practical implications for managers and policymakers who are interested in reducing depression in the workplace and promoting broader socio-economic development (Section 7).

2 RELATED WORK

NLP Research in Computational Social Science. Social media has been shown to be a powerful tool for computational social scientists to study health outcomes [34, 63, 79, 85, 93], including the examination of depression at individual and collective levels [31, 32]. Past NLP research in social media studied aspects similar to ours: for example, sleep insomnia [57]; fatigue or loss of energy [14, 37]; social anhedonia and anxiety [17], stress levels in a variety of contexts [53, 69], from the workplace [18] to university campuses [94]; markers of depression and mental health [32, 33, 77, 83, 86]; and even suicidal ideation [24, 25]. We refer the reader to [19] for a comprehensive NLP literature review in the context of mental health.

Language is often seen as a powerful medium to verbalize one's beliefs and values [5], allowing computational social scientists to study its manifestations in the workplace. By analyzing email exchanges, Doyle *et al.* [39] found that the use of language (e.g., use of pronouns) in them reflected one's integration in a team; for example, employees who tended to use the pronoun "I" less than the pronoun "we" were found to be more integrated in their teams. Similarly, by analyzing the Enron emails data corpus [61] using LSTM and BERT, Choi et al. [21] found that certain linguistic markers were predictive of the company's fall. Researchers have aimed at capturing key workplace behavioral markers from social media as well [30]. Ehrlich and Shami [42] studied how work-related social media use contributed to the social capital of employees. Such a use has also been found to be an effective way for employers to manage their reputation [97]. For the last ten years or so, anonymized platforms such as Glassdoor have allowed employees to share their workplace experiences in the form of company reviews [13]. By analyzing these reviews, previous NLP research projects captured perceived brand personality [108]; analyzed job satisfaction factors in relation to retention and turnover [66]; modeled workplace culture [29]; and assessed perspectives on sustainability [96].

NLP methods have also been applied on text captured in clinical settings. For example, Hong-Jie et al. [28] applied state-of-the-art deep-learning NLP methods (BERT, textual convolutional neural networks, and hierarchical attention networks) on 500 patients' notes taken from psychiatrists. Similar to textual analyses, researchers employed techniques to analyze speech (e.g., acoustic features), which complement language models through the rich information modality of audio [71]. More recently, DeSouza et al. [35] argued for NLP approaches to combine acoustic and linguistic aspects of human language derived from text and speech to classify depression and its severity.

To sum up, this body of literature successfully mined the use of language in a variety of media (within- and outside-organization communication channels); however, it did not focus on work-related depressive symptoms. The main contribution of this work is a novel, principled approach to study the Occupational Depression Inventory construct at scale, by means of NLP analyses of employees' reviews about the company they work(ed) for.

Psychological Constructs of Depression. To see why we opted for ODI, consider that, over the last decades, the most popular work-contextualized construct that aims at capturing job-related suffering is the so-called "burnout" construct [74]. Yet, this construct is plagued by definitional problems [11, 27, 89]. For example, across different studies, burnout has been associated with exhaustion, psychological withdrawal, or professional inefficacy [10, 11, 15], representing a catch-all label [70], with no consensual definition available despite nearly 50 years of research. Additionally, as there is no commonly shared medical diagnosis of burnout, the most prominent scales (e.g., the Maslach Burnout Inventory, the Pines' Burnout Measure) fall short to meet stringent requirements, thus affecting their usability [99, 104]. Looking for alternatives, one would resort to PHQ-9 [65], which is a clinically-validated screening tool for diagnosing depression. While this scale has been demonstrated in workplace settings [78], the construct captures general depression by definition without any specific work-contextualized items.

Since the use of the burnout construct in occupational health research and practice became problematic [11, 27, 95] in addition to the non-workplace specific nature of PHQ-9, researchers developed a new construct named "Occupational Depression Inventory" or, simply, ODI. While

Table 1. The nine ODI dimensions. For all the dimensions, their sentence definitions (taken from the original scale in [7]) and representative reviews are provided. Depressed mood, sleep alterations, fatigue/loss of energy, and feelings of worthlessness (top part of this table) are the four dimensions upon which the AutoODI scored the reviews and generated the ODI score. The remaining five dimensions, which are in gray at the bottom of the table, were not conservatively considered in the construction of the ODI score as our validity analyses demonstrated that these dimensions could not be reliably captured by our automatic textual analysis.

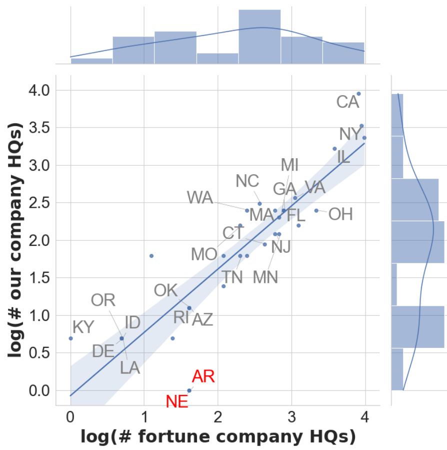
ODI Dimension	Sentence Definition	Example Review
Depressed mood	"I felt depressed because of my job."	This job made me seriously depressed!
Sleep alterations	"The stress of my job caused me to have sleep problems (I had difficulties falling asleep or staying asleep, or I slept much more than usual)."	I slept in a couple times due to extreme stress and insomnia, and they terminated me.
Fatigue/loss of energy	"I felt exhausted because of my work."	I came into work already tired many days.
Feelings of worthlessness	"My experience at work made me feel like a failure."	I constantly felt that my work was mediocre.
Anhedonia	"My work was so stressful that I could not enjoy the things that I usually like doing."	While I enjoyed the work and being busy, there was definitely very little work life balance.
Appetite alterations	"I felt my appetite was disturbed because of the stress of my job (I lost my appetite, or the opposite, I ate too much)."	I lost my job because I had to miss a lot of work when I was sick.
Cognitive impairment	"My job stressed me so much that I had trouble focusing on what I was doing (e.g., reading a newspaper article) or thinking clearly (e.g., to make decisions)."	Didn't have what I needed most of the time in order to do my job accurately.
Psychomotor alterations	"As a result of job stress, I felt restless, or the opposite, noticeably slowed down—for example, in the way I moved or spoke."	This job caused me so much stress and anxiety.
Suicidal ideation	"I thought that I'd rather be dead than continue in this job."	I would rather get up and go to work.

the ODI construct was only developed recently, its measurement scale has consistently demonstrated excellent psychometric and structural properties across studies and samples recruited in the U.S., France, South Africa, and New Zealand [8, 9, 55]. ODI meets requirements not only for "classical" characteristics such as total-score reliability, factorial validity, and convergent and discriminant validity but also for scalability, monotonicity, local independence, and invariant item responding [101].

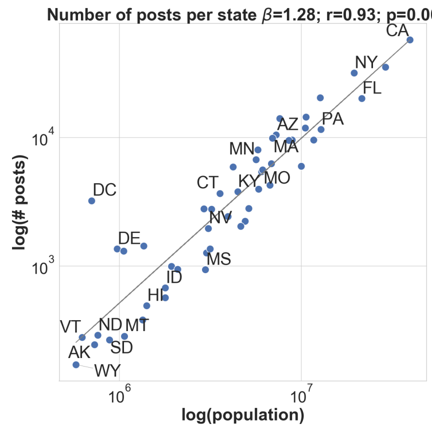
3 BACKGROUND ON ODI

As with burnout, ODI captures work-attributed suffering. As opposed to burnout, ODI: *1)* tailors work-attributed suffering to the long-established and well-defined concept of depression; and *2)* comprehensively covers depressive symptoms, including cognitive, affective, and somatic aspects. Specifically, ODI was developed with reference to the nine diagnostic criteria for major depression of the Diagnostic and Statistical Manual of mental disorders, 5^{th} edition (DSM-5) [87]. As such the ODI construct assesses nine key depressive symptoms: anhedonia, depressed mood, sleep alterations, fatigue/loss of energy, appetite alterations, feelings of worthlessness, cognitive impairment, psychomotor alterations, and suicidal ideation. Among the nine, some symptoms are more frequent than others (e.g., depressed mood), thus having higher weights in the ODI diagnostic instrument [7].

As one can see from the sentence definitions in Table 1, respondents are invited to report symptoms only when they feel able to clearly establish a link between their symptoms and their work context. In three distinct samples [7], researchers found that the most frequently endorsed ODI item was fatigue/loss of energy, and the least frequently endorsed was suicidal ideation; and that the different ODI's items were associated with a variety of work-contextualized (e.g., job satisfaction) and context-free (general health status) individual-level measures, thus further supporting the construct's external validity.



(a) # headquarters in our data *vs.* # headquarters in the Fortune 500 list



(b) # posts in our data *vs.* # residents in each state

Fig. 1. Representativeness of our company review dataset in terms of: (a) number of headquarters across the states (AR and NE deviates from the linear scaling but are within two standard deviations); and (b) the population in each state.

Despite its growing popularity, to date, there has not been any research employing the ODI construct to study depression in the workplace as an organizational- or a society-level problem, not least because it is hard to capture ODI at scale. To fix this, we propose to study the microfoundations of depression in the workplace in a bottom-up fashion, starting from employees' reviews of the company they work(ed) for.

4 DATASETS

Company Reviews. We collected data from a popular company reviewing site, where current and former employees write reviews about their own corporate experiences, ranging from job interviews to salaries to workplace culture. As of 2021, there are 50M monthly visitors on the platform, and 70M reviews of 1.3M companies. To ensure quality reviews, the site: *1)* performs both automatic and manual content moderation; *2)* allows for full access to content only to users who register on the site and write at least one review (encouraging neutral and unbiased reviews); and *3)* allows for posting maximum one review per employee per year. Our dataset consisted of reviews published over twelve years, from 2008 to 2020. Each review consists of a title; a 'pro' portion (i.e., positive aspects of the company); a 'con' portion (i.e., its negative aspects); a set of four ratings on a [0,5] scale scoring the company's *balance*, *career*, *culture*, and *management*; and a final *overall* rating of the company. Since reviewers have the option to include their location, we were able to identify the states for part of the reviews. To ensure the robustness of our text processing method, we retained companies that had at least 1,000 reviews and were present in at least 10 states, leaving us with a dataset of 358,527 reviews of 104 US-based companies. These reviews represented 88.7% of the original dataset, and 80% of these companies were S&P 500. These companies offer the same level of representativeness as the S&P 500 companies in terms of geographic distribution of headquarters across states (Figure 1(a)). Furthermore, their cumulative number of posts per state matched state population size (Figure 1(b)).

Table 2. Frequency distributions and statistics of ODI and socio-economic variables at the state level. Variables with skewed distribution (identified by the Fischer-Pearson test [62]) were log-transformed before being used in the analysis, and were denoted with (log). The distribution plots and values are reported before the log-transformation.

Variable	Distribution	Min	Max	Median	Mean	Std
ODI		0.00	100.0	35.67	38.03	21.19
Urban Population (log)		38.70	95.0	73.75	73.37	14.59
Wealth		35015.00	75258.0	52248.50	53842.48	10101.00
Creativity		17.00	188.0	103.50	103.54	46.05
Depression		12.98	25.2	19.40	19.40	2.81
Openness		21.80	65.0	49.85	49.43	9.27
Neuroticism		30.40	79.2	49.00	50.19	10.03

Statistics about Companies and States. In addition to the reviews, we collected yearly stock growth values of the 104 companies from the Yahoo Finance Portal [4]. We also collected the values of five state-level socio-economic indicators previously found to be associated with depression or, on the opposite side, well-being [46, 73]. Summary statistics of these indicators are reported in Table 2. The first indicator is *urban population* [2] and measures the percentage of the total population in urban areas from the US Census Bureau. Urbanization has been associated with high innovation levels [12] but also with mood disorders, including depression [56]. The second indicator is *wealth* and is operationalized with GDP per capita, which has been previously linked to depression [84]. The third indicator is the *creativity index* [46] and measures the ability of a state to attract the so-called 'creative class' (composed by, for example, scientists, engineers, poets, artists, architects, designers). It has been shown that even a small amount of creativity helps individuals cope with stress, leading to increased happiness and better well-being [44]. The fourth indicator is state-level depression rates obtained from [1]. The last two indicators aim to capture personality, as defined by the five-factor model of personality, or the big five [26, 52]. This model defines a comprehensive and reliable set of personality traits: openness to new experiences, conscientiousness, extraversion, agreeableness, and neuroticism. We report the results for the five traits and, based on the literature, we expect significant correlations with at least openness and neuroticsm as both have been shown to be geographically linked to well-being [90]: specifically, openness was linked to more experiences of awe, contributing to well-being, while neuroticism was linked to mental health issues, compromising one's satisfaction with life. Note that, as the two personality indicators were available for 48 U.S. states, the personality analysis had to exclude Hawaii and Alaska.

Ethical Considerations. This work used public and anonymized data from a popular company reviewing site. Despite working with public data, excerpts used to support the qualitative analysis were paraphrased, avoiding traceability and identifiability of these reviews and providing context in readership. Statistics at company and state levels were publicly available information.

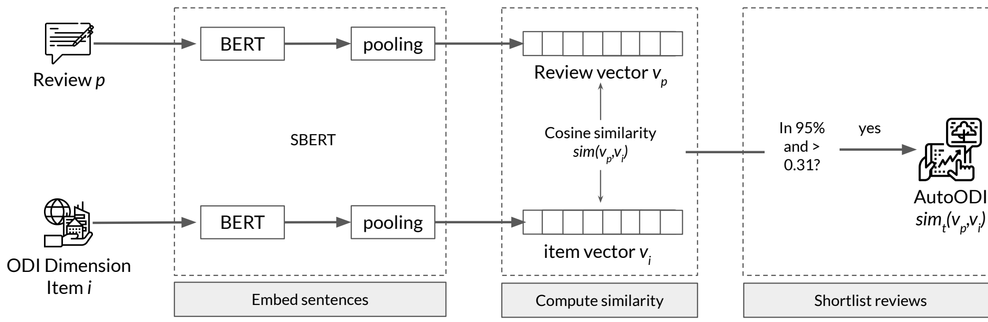


Fig. 2. Schematic of AutoODI. It consists of three blocks, which: *1)* embed both the review and the ODI dimension's sentence definition; *2)* compute the cosine similarity between the two embeddings; and *3)* shortlist reviews by discarding those that are below a given similarity level.

5 AutoODI: SELECTING AND SCORING ODI DIMENSIONS

5.1 Selecting ODI dimensions

Embed Sentences and Compute their Similarity. An unsupervised deep-learning framework based on the sentence-level BERT algorithm [88] (SBERT) was developed (first block "Embed sentences" of Figure 2).

SBERT is an advancement on the recent transformer-based deep-learning NLP technique BERT [36], which is state-of-the-art in semantic similarity tasks. It has shown strong 'zero-shot' learning capabilities, that is, it learns new language tasks or constructs with little to no labeled training data [88]. SBERT is a flexible deep-learning architecture as it can use any pre-trained language model such as BERT or RoBERTa. We used BERT because it is designed to capture the semantic meaning of reviews, not least because it: *a)* accounts for semantic aspects such as polysemy and synonyms that word-based analyses cannot account for (e.g., in the review "This job left me feeling hollow and unsatisifed", the word "depression" does not explicitly appear but SBERT correctly classifies it under 'depressed mood'); *b)* incorporates "world knowledge" by being pre-trained on a large corpus of Internet data (the whole English Wikipedia, and a large corpora of American English texts called the Brown Corpus [48]); and, as we shall see, *c)* is used within ODI to integrate not all the dimensions of the original ODI scale but only those that passed an external manual validation (we conservatively excluded five out of the nine dimensions of the original scale). We set BERT's parameters to their default values, and its two hyper-parameters (i.e., a maximum sequence length of 128 and token-based pooling mode) as per Reimers and Gurevych [88]. SBERT's ability to learn novel types of linguistic phenomenon makes it ideal for modeling ODI from unstructured reviews as opposed to traditional approaches such as fully-supervised dictionary-based techniques or topic modeling. While these approaches are still viable in certain research contexts, fully-supervised techniques often require hard-to-obtain domain-specific dictionaries, while topic modeling approaches typically require manual fine-turning of parameters (e.g., number of topics). To begin with, for each ODI dimension i, we computed the SBERT similarity score [88] (second block "Compute similarity" of Figure 2) between the SBERT vector of review p and the SBERT vector of i's sentence definition (spelled out in the second column of Table 1). As workplace depression has a negative valence, we used the 'cons' reviews rather than the 'pros' ones.

Qualitative Assessment of Reviews. Out of the nine ODI dimensions, we qualitatively ascertained which ones were correctly captured. To this end, the framework identified the 5 reviews most relevant to each dimension: for each dimension i, reviews were ranked by $sim_t(v_p, v_i)$, and three independent annotators then manually assessed the relevance of these reviews. To conservatively retain only the dimensions that were accurately identified by the framework, we discarded any dimension for which the majority of the annotators marked less than 4 out of the 5 reviews as relevant (overall, the agreement among the annotators was high, i.e., Fleiss $K = 0.75$). As a result, five dimensions were discarded. Appetite alterations, cognitive impairment, and suicidal ideation were discarded due to their lack of sufficient representation in company reviews and, as the example reviews in the bottom part of Table 1 show, the resulting inability of capturing the constructs at hand. By contrast, anhedonia and psychomotor alterations were discarded because the associated reviews were discussing general work dissatisfaction but were not necessarily discussing matters related to these two constructs. This left us with four dimensions — depressed mood, fatigue, sleep alterations, and feelings of worthlessness (listed in the top part of Table 1).

Shortlist Reviews. Based on these four dimensions, our framework then computed a *thresholded* SBERT similarity score (third block "Shortlist reviews" of Figure 2), defined as:

$$sim_t(v_p, v_i) = \begin{cases} sim(v_p, v_i), & \text{if } sim(v_p, v_i) > 0.31 \\ & \text{ AND } sim(v_p, v_i) > 95\%(i); \\ 0, & \text{otherwise.} \end{cases}$$

We chose the threshold of 0.31 as the mean SBERT similarity of the 4 dimensions left after the previously described qualitative assessment of reviews. Since the SBERT values of the four ODI dimensions were not equally distributed, we chose to pair the fixed generalized threshold of 0.31 with an ODI dimension-specific threshold. Based on our experiments, this latter threshold value (denoted as $95\%(i)$) was the 95% percentile of the dimension's distribution, which is in line with the threshold found in previous studies [22]. Note that, by review, we mean the con portion of the review. That is because we were mostly interested in shortcomings (cons) rather than positive initiatives (pros). We indeed found that, if we were to instead take pros (or combine pros with cons together), our deep-learning framework would perform worse in the previously described qualitative assessment of reviews.

5.2 Scoring ODI dimensions

AutoODI scored reviews for the four remaining ODI dimensions at two units of analysis u (i.e., company and state), and did so to test whether the impact of work-related depression manifested itself at both micro-level (e.g., in a company's performance deterioration) and macro-level (e.g., in a state's economic deprivation). More specifically, the overall ODI score for company/state u is defined as the sum of the individually quantified ODI dimensions:

$$ODI(u) = \sum_{i \in D} \frac{\sum_{p \in R(u)} sim_t(v_p, v_i)}{|R(u)|} \tag{1}$$

where D is the set of the ODI dimensions being quantified (i.e., D = {depressed mood, sleep alterations, fatigue, and feelings of worthlessness}), $R(u)$ is the set of u's reviews, v_i is the SBERT vector of dimension i's sentence definition (reported in Table 1), and $sim_t(v_p, v_i)$ is the *thresholded* SBERT similarity score between the SBERT vector of review p and the SBERT vector of i's sentence.

Table 3. Cross-correlations of the four ODI's dimensions plus the composite ODI score at company level.

	Depressed mood	Sleep alterations	Fatigue/loss of energy	Feelings of worthlessness	ODI
Depressed mood	–				
Sleep alterations	0.499***	–			
Fatigue/loss of energy	0.565***	0.944***	–		
Feelings of worthlessness	0.720***	0.270**	0.310***	–	
ODI	0.854***	0.833***	0.865***	0.706***	–
Note:				*p<0.05; **p<0.01; ***p<0.005	

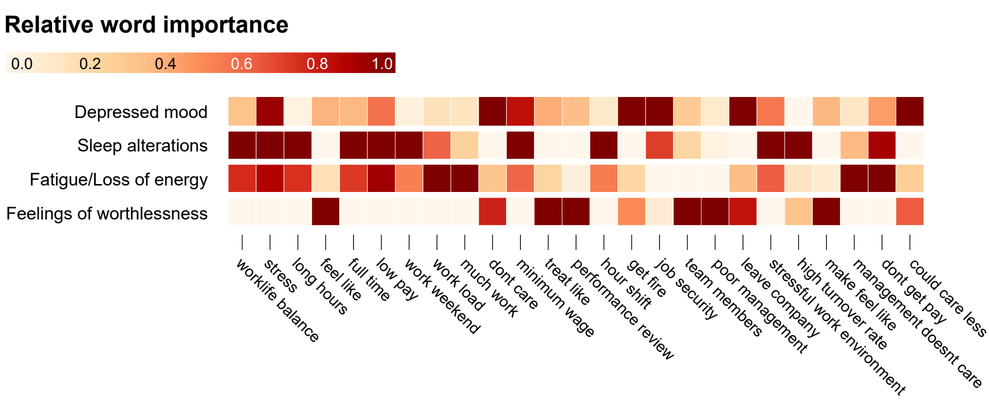


Fig. 3. **Top n-grams in sentences expressing ODI items.** Darker colors (higher normalized *TF-IDF* score) indicate greater relative relevance to an item.

We chose to sum up the scores of the four ODI dimensions based on what previous survey-based studies found: in [7], for example, factor analysis of the ODI dimensions indicated that they can be used as a unidimensional measure with an explained variance as high as 0.891 (indicative of essential unidimensionality); furthermore, the factor analysis indicated that ODI dimensions contributed *homogeneously* to the unidimensionality of the measure. That is the case also in our study: all the four dimensions are indeed highly correlated with the ODI score itself (Table 3), with values ranging from 0.70 up to 0.86.

6 RESULTS

6.1 Does our machine learning method capture ODI?

We validated our deep-learning method for detecting the ODI items by establishing its face validity. To that end, we took the three-step linguistic approach proposed in [29]. For each of the ODI dimensions, we obtained the most frequent keywords, that is, 1, 2, 3, and 4-grams from the reviews deemed relevant by our framework. Second, we computed the *TF-IDF* scores for such *n*-grams, where each document was comprised of all shortlisted reviews for each ODI item. Finally, we ranked keywords for each item based on their *TF-IDF* score. This allowed us to find the keywords judged to be important for a certain ODI item by our embedding-based method. The top-ranked keywords for the ODI items are visualized as a heatmap in Figure 3.

We observed many keywords to be highly discriminant of the specific item they were associated with. For example, keywords 'get fired', 'let go', and 'felt like' ranked highly only for item 'Depression'. Keywords 'full-time', 'workload', and 'work-life balance' were uniquely strongly associated with item Fatigue. Keywords 'minimum wage', 'hour shift', 'part time' were ranked highly for item Sleep alteration. Keywords 'don't care' 'treat like' and 'performance review' were highly discriminant of Feeling of worthlessness. Lastly, keywords such as 'turnover rate', 'stressful work environment', and 'management doesn't care' were associated with all the four items.

Additionally, we manually evaluated our method, and did so by extracting a representative sample of reviews among those that AutoODI assessed to be containing markers of workplace depression. To make the sample representative, we randomly selected two reviews per company and two reviews per state, ensuring no duplicates. We ended up with 208 reviews stratified by company, and 100 reviews stratified by state, giving us a total of 308 reviews. Then, three annotators read those reviews, and independently evaluated whether the review included mentions of workplace depression or not. We resolved disagreements in a joint discussion round. With these binary labels at hand, we computed: the inter-annotator reliability score, which showed a moderate to strong agreement with a Fleiss Kappa = 0.68, and the true positive rate (i.e., the proportion of the reviews that were singled out to contain markers of depression by both ODI and the set of annotators), which amounted to 73%.

We evaluated these manual annotations as binary rather than with four ODI dimension labels, since a single review can express multiple dimensions. For example, consider the following review: "Management wasn't the greatest. *Feedback on performance was poor and poorly executed, as well as not totally truthful.* Merchandise was hardly in stock. I was asked to setup displays that were heavily labor intensive. *Hated working Sundays and was required to most of the time.*". As the text suggests, this review expresses both 'feelings of worthlessness' and 'fatigue'.[2] An alternative to the binary evaluation is a multi-label setup wherein an annotator would evaluate each dimension independently. However, this kind of setup typically has more complicated accuracy and less interpretable metrics [42, 72] such as Hamming distance and label density [42]. As we were interested in studying the workplace depression construct as a whole, assessing whether our automated AutoODI method captures all dimensions equally well is left for future work.

While AutoODI's performance is highly satisfactory, our method also classified 72 posts out 308 (23%) as manifesting mentions of workplace depression, while the three annotators labeled them otherwise. We manually revisited these cases to unpack why the automated method fell short, and found that these cases comprised of cons that were not practically cons such as "good company to work I don't have any cons to mention about this company", and "more flexible scheduling hours, shorter shifts". In particular, the last example mentions working hours (associated with fatigue), but in a positive sense. These cases largely reflected the fact that embedding-based methods are not accurate in terms of polarity at times [20].

6.2 Are high ODI scores associated with company performance deterioration?

There are several ways to measure a company's success. We considered two complementary indicators: the online ratings the company received from its employees (available from the company reviewing site), and its financial position (measured as stock growth).

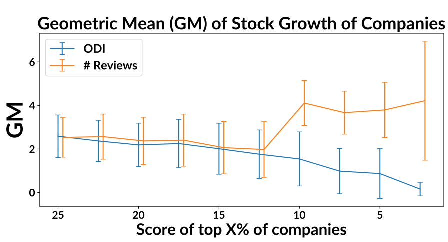
ODI and Company Online Ratings. Employees have the option to rate the company they are reviewing based on four different facets—balance, career, culture, management, plus a fifth company's overall score. We thus investigated to what extent a company's success across these five facets could be predicted based on the company's ODI score. We did so by first aggregating ODI scores and ratings at company level, and by then conducting an OLS regression using our ODI score. To ensure that the association between ratings and ODI scores is not confounded by company popularity or general ill-feeling towards the company, we controlled for a company's

[2]Though reviews on the platform are anonymous, we paraphrased them to protect privacy.

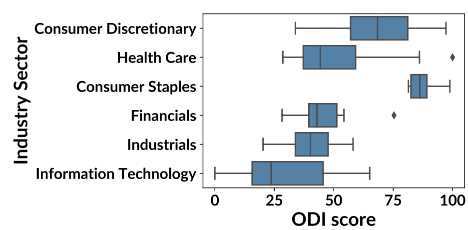
Table 4. OLS model predicting company ratings from the ODI score, while controlling for the total number of reviews and the length of negative reviews (cons).

	Balance	Career	Culture	Management	Company
const	74.723***	60.126***	65.001***	50.006***	65.527***
	(3.637)	(3.699)	(4.027)	(3.709)	(3.825)
ODI	-0.650***	-0.324***	-0.344***	-0.256***	-0.450***
	(0.083)	(0.084)	(0.091)	(0.084)	(0.087)
Con Length	0.195	-0.309	-0.684*	-0.493	0.180
	(0.327)	(0.332)	(0.362)	(0.333)	(0.344)
Total Reviews	-0.216	0.322	0.667*	0.481	-0.145
	(0.321)	(0.327)	(0.356)	(0.327)	(0.338)
Observations	104	104	104	104	104
R^2	0.416	0.177	0.213	0.148	0.227
Adjusted R^2	**0.398**	**0.152**	**0.189**	**0.123**	**0.204**
Residual Std. Error	16.467(df = 100)	16.746(df = 100)	18.232(df = 100)	16.792(df = 100)	17.315(df = 100)
F Statistic	23.710***	7.149***	9.018***	5.795***	9.775***
	(df = 3.0; 100.0)	(df = 3.0; 100.0)	(df = 3.0; 100.0)	(df = 3.0; 100.0)	(df = 3.0; 100.0)

Note: *p<0.1; **p<0.05; ***p<0.01



(a) Geometric mean of stock growth values for the top x% of companies ranked by ODI scores, and by number of reviews. The lower a company's ranking, the more reviews and the higher ODI score the company has.



(b) Boxplots of the distribution of the ODI score by industry sector.

Fig. 4. Workplace depression and companies (a) association between stock performance and ODI scores; and (b) ODI score distributions of industry different sectors calculated with Formula (1).

total number of reviews and the length of negative reviews (cons). As reported in Table 4, we found that ODI could explain up to: 40% of the variance in a company's balance, 15% of culture, 19% of career, 12% of management, and, 20% of the overall company rating. If one were to assume external validity for the ODI construct, it would come as no surprise that the facet most related to ODI is balance rather than the more general career, culture, or management. Balance reflects the ability of employees to balance work and life outside the office, and, more generally, to cope with stress, depression and anxiety [54]. Additionally, the beta coefficients for the number of reviews were low, suggesting that a company's score on each of the facets could be far better predicted from ODI than from the company's online popularity.

ODI and Company Stock Growth. We obtained stock data for 84 of the 104 companies in our dataset, from 2009 to 2019, using the Yahoo Finance portal [4]. For each company, we calculated the geometric mean of its stock growth during such period; we used the geometric mean since the distribution of stock growth values across companies was heavy-tailed. The geometric mean of stock growth is calculated as $GM(\text{stock growth}_{[09-19]}) = \Pi(\text{stock growth}_{[09-19]}(c))^{1/n}$, where

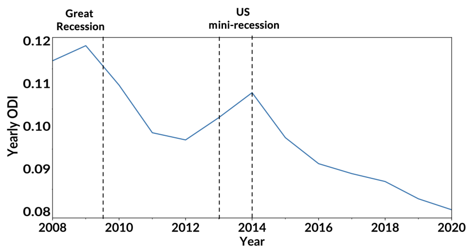
c is a company in a specific *(ODI or review score, percentile)* bin, and n is the number of the companies in such a bin. Error bars represent geometric standard error $GSE(stock\_growth_{[09-19]}) = \bar{GM}(stock\_growth_{[09-19]})/\sqrt{N} \cdot \sigma(log(stock\_growth_{[09-19]}))$. To inspect whether a company's financial success (measured as stock growth) was negatively associated with its ODI scores, we plotted the geometric mean of its stock growth (y axis) against its ranking in terms of the ODI score (x axis) in Figure 4(a). We also included the total number of reviews in the figure to check whether stock growth was merely associated with the company's popularity rather than its well-being practices. In the range top [13%, 25%] companies, the average differences in growth were not statistically significant. Yet, those at the top 10% experienced significantly different rates. The top 10% most popular companies (in terms of number of online reviews), on average, grew by a factor of 60.9 (GM = 4.11), while those in the top 10% most 'depressed' companies (in terms of ODI scores), on average, grew by a factor of only 4.66 (GM = 1.54) in ten years. To ease the interpretation of such values, consider the example of Home Depot, the largest home improvement retailer in the US, which is in the top 10% popular companies. Its stock price traded at 19.04$ in 2009 and increased by a factor of 10.48 (its $GM(\text{stock growth}_{[09-19]})$ was 2.35), trading at 200.8$ ten years later. On the other hand, Kroger, the second-largest general retailer in the US, is in the top 10% most depressed companies. Its stock price traded at 8.9$ in 2009 and increased by a factor of only 2.73 (its $GM(\text{stock growth}_{[09-19]})$ was 1.01), trading at 24.4$ ten years later. Figure 4 (a) suggests that companies with high ODI scores have significantly lower growth compared to those with low ODI scores — the top 5% have lower stock growth compared to the top 25%, and the difference between these two sets is highly significant, as the error bars show. Indeed, the average growth difference in the two sets is GM = 1.02, which practically translates into, say, KMart, at the top 5% (company with high ODI score), plummeting from $41.2 in 2009 to $0.4 in 2019; and Amazon, at the top 25% (company with low ODI score), growing from $87.2 to $1789.2.

ODI and Temporal Shifts. As work cultures and market labor evolve over time [41], it is important to determine the temporal evolution of workplace depression over the fairly long period analyzed in this research. To do so, we computed the proportion of reviews that express workplace related depression for each year:

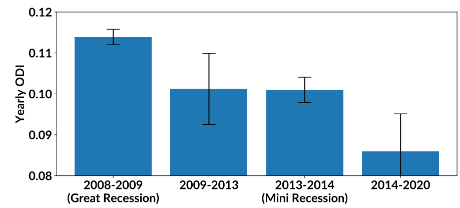
$$ODI(y) = \frac{|[p \in R(y) \ni sim_t(v_p, v_i) > 0, \forall i \in D]|}{|R(y)|} \tag{2}$$

where y represents a year, $R(y)$ refers to all reviews in that year, D is the set of the ODI dimensions being quantified (i.e., D = {depressed mood, sleep alterations, fatigue, and feelings of worthlessness}), and $sim_t(v_p, v_i)$ is the *thresholded* SBERT similarity score between the SBERT vector of review p and the SBERT vector of the sentence reflecting the i^{th} ODI dimension (the similarity value above zero in the formula means that the review has been flagged by AutoODI to contain mentions of workplace depression).

As observed in Figure 5(a), workplace depression (measured by our AutoODI method) followed a downward trend over the years except for the period 2013-2014. The overall downward trend reflects the growing organizational focus on employee wellbeing [102] and of the diffusion of workplace policies designed to promote respect and dignity [59], a healthy work-life balance [107], and diversity and inclusion [100]. Conversely, the growth in workplace depression in the years 2013 and 2014 is consistent with the mini-recession that followed the global financial crisis of 2008 [68]. The identified trends and the possible explanations offered here provide additional evidence of the robustness of our ODI scores, thus further confirming the negative association between workplace depression and companies' financial growth previously identified.



(a) Yearly ODI decreased over time, indicating improving workplace conditions.



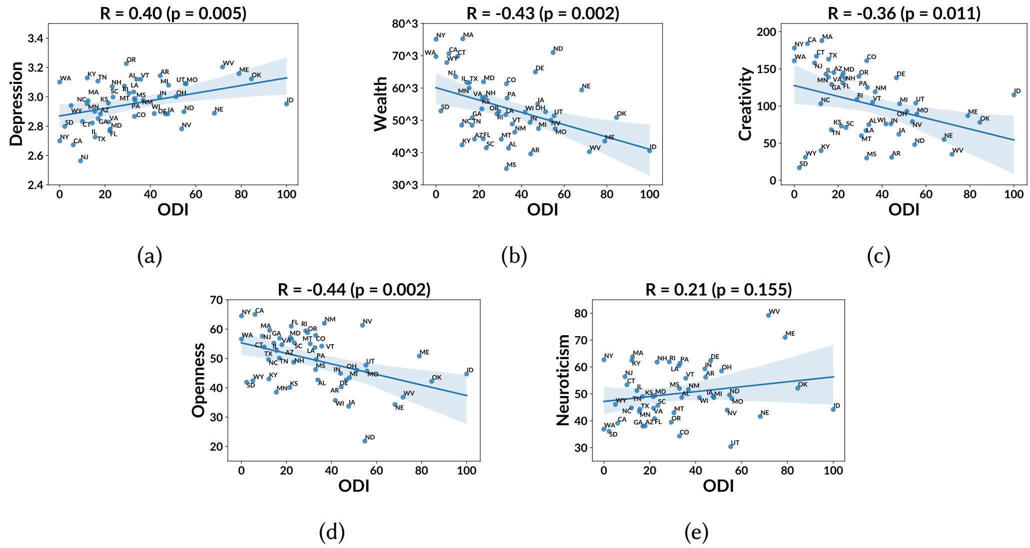
(b) Binned yearly ODI with statistical significance. The error bar represents the standard deviation of the bin's average.

Fig. 5. Yearly ODI. It was worse during the Great Recession in 2008-2009. Overall it decreased over time, yet improvements were not consistent over the years but stalled during the mini recession in 2013-2014. The two recessions have narrow error bars (limited deviation from the average values).

We binned ODI scores (which are individual numbers) by time intervals marking key financial events (i.e., [2008-2009], [2009-2013], [2013-2014], [2014-2020]). We do so to compute statistical significance. From the binned plot (Figure 5), we see that ODI was worse during the Great Recession in 2008-2009. Furthermore, ODI decreased over time, yet improvements were not consistent over the years but stalled during the mini recession in 2013-2014. Also, the error bars are narrowest during both great and mini recessions (the standard deviation is extremely small), suggesting that recession has a negative impact on mental health/depression for everyone, regardless of job sector or state. This is in line with the literature linking economic downturn and job instability during recession to depression prevalence [75].

6.3 Are high ODI scores associated with specific industry sectors?

To examine whether work depression was mostly associated with certain industry sectors, we conducted an ANOVA test and found that the differences in ODI scores across sectors were statistically significant ($F = 11.71, p < 0.001$). We also plotted the distribution of the ODI scores for each industry sector in Figure 4(b). Consumer Discretionary and Consumer Staples had the highest ODI median scores. Both sectors include direct-to-consumer retailers characterized by irregular hours and poor work-life balanced [16, 105]. Health Care was also characterized by high scores, but with high variability. According to [92], this is explained by: *1)* the whole sector chronically suffering from time pressure and skill shortages; and *2)* the high diversity of its sub-sectors[3], which are characterized by diverse economic benefits and work-life balance expectations. A similar analysis applies to the IT sector. The IT sector was the one scoring the lowest in ODI but, again, with high variability, which may be explained by: *1)* the high presence in this sector of what Richard Florida called the "creative class" (*"the primary job function of its members is to be creative and innovative."* [45]); and *2)* at the same time, the high diversity of its workforce, which indeed includes "creatives" but also enlists IT support workers up to factory workers in the supply chain (e.g., in the e-commerce business).



(a) (b) (c)

(d) (e)

Fig. 6. Correlations between the ODI score and US states socioeconomic variables: *(a)* Depression, *(b)* Wealth, *(c)* Creativity, *(d)* Openness to Experience, and *(e) Neuroticism*.

6.4 Are high ODI scores associated with negative state outcomes?

Lastly, we investigated whether the detrimental effects of depression in the workplace manifested itself also at state level. We did so by computing the Pearson's correlations between our ODI score and three state-level socioeconomic indicators—depression rates [1], wealth [3], and creativity [46] — plus the two personality traits of openness and neuroticism [38, 52]. In line with the expectations detailed in Section 4, ODI was: positively related with a state's depression rates ([$r = 0.40, P = 0.005$]); negatively with the state's GDP per capita ([$r = -0.43, P = 0.002$]); negatively with the state's ability to attract the creative class ([$r = -0.36, P = 0.001$]); negatively with openness ([$r = -0.44, p = 0.002$]); and positively but not significantly with neuroticism ([$r = 0.21, p = 0.155$]). As hypothesized based on the well-being literature, the correlations for the remaining three personality traits were low and statistically non-significant, i.e., for agreeblenss ([$r = 0.095, p = 0.521$]), extraversion ([$r = -0.092, p = 0.532$]), and conscientiousness ([$r = -0.026, p = 0.860$]). These results suggest that, in line with the literature on individual-level [43, 67] and geographic-level effects [64, 90], ODI was related to the socio-economic variables and to the personality trait of openness as hypothesized: states hosting companies with high ODI scores tend to suffer from depressive symptoms themselves (suggesting a diffusion of mental health issues among the general population), from a diminishing economic output per person, the inability to attract talent in their industries, and the low prevalence of individuals open to new experiences (who were found to be those who are more satisfied with their lives [90], and are more likely to be employed in the creative industries [50]).

Since these socio-economic and personality factors interacted with each other, we then considered them separately by predicting each of them from our ODI score, while controlling for urban population [2]. We controlled for urban population because urban areas have been repeatedly found to be wealthier, more attractive for talent, and characterized by individuals with high openness to

[3]Biotechnology Industry, Health Care Equipment & Supplies Industry, Health Care Providers & Services Industry, Health Care Technology Industry, Life Sciences Tools & Services Industry, and Pharmaceuticals Industry [60].

Table 5. OLS models predicting state depression rates, wealth (GDP per capita), creativity index, openness, and neuroticism from a stepAIC analysis where the dependent variables are the ODI score, urban population (logged), and their interaction term. All variables were scaled between 0 and 100 to aid interpretation.

	Depression	Wealth	Creativity	Openness	Neuroticism
Intercept	85.291***	15.357*	-39.678**	40.246***	37.842***
	(8.390)	(8.932)	(15.074)	(7.758)	(4.738)
ODI			1.037***		0.641***
			(0.303)		(0.175)
ODI x Urban Population	0.238*	-0.391***	-1.273***	-0.366***	-0.590***
	(0.122)	(0.130)	(0.289)	(0.113)	(0.170)
Urban Population	-0.584***	0.633***	1.392***	0.509***	
	(0.118)	(0.125)	(0.197)	(0.109)	
Observations	48	48	48	48	48
R^2	0.356	0.386	0.623	0.366	0.245
Adjusted R^2	**0.328**	**0.359**	**0.598**	**0.338**	**0.212**
Residual Std. Error	18.880(df = 45)	20.100(df = 45)	17.081(df = 44)	17.457(df = 45)	18.250(df = 45)
F Statistic	12.448***	14.145***	24.279***	12.984***	7.313***
	(df = 2.0; 45.0)	(df = 2.0; 45.0)	(df = 3.0; 44.0)	(df = 2.0; 45.0)	(df = 2.0; 45.0)

Note: *p<0.1; **p<0.05; ***p<0.01

new experiences [12, 46]. We used OLS regression models and added the interaction term between ODI and urban population. To ease interpretation, all variables were scaled between 0 and 100. As shown in Table 5, ODI was predictive of state-level measures of depression (Adj. R^2 of 0.328), wealth (Adj. R^2 of 0.359), creativity (Adj. R^2 of 0.598), openness (Adj. R^2 of 0.338), and neuroticism (Adj. R^2 of 0.212), but in different ways. More precisely, states with higher depression rates tend to be less urbanized ($\beta = -0.58$) and, after controlling for urbanization, to have higher ODI scores ($\beta = 0.23$). On the contrary, wealthier states tend to be more urbanized ($\beta = 0.63$) and, after controlling for urbanization, to have lower ODI scores ($\beta = -0.39$). Similarly, the states high in openness tend to be more urbanized ($\beta = 0.51$) and, even after controlling for urbanization, to have lower ODI scores ($\beta = -0.36$). Finally, states high in neuroticism tend to have higher ODI scores ($\beta = 0.64$), as one could hypothesize. To sum up, wealthy, creative, and open states were associated with lower incidences of workplace depression.

7 DISCUSSION AND CONCLUSION

As workplace depression becomes a widespread phenomenon, the early identification of the organizational factors that can adversely affect staff mental health becomes paramount to safeguard employees' wellbeing and productivity [98]. By adopting the ODI construct and measure, this research examines the micro-foundations of workplace depression and assesses its potential consequences at organizational and societal levels.

7.1 Implications

Theoretical Implications. Our work offers two substantive theoretical contributions. First, while the existing literature on the use of machine learning to assess depression often presents this phenomenon as context-free [31, 33], our research operationalizes workplace depression as the combination of critical aspects of the organizational environment which have a direct impact on employees' mental health. Typical of computational social science, this operationalization of workplace depression, which is both theoretically sound and methodologically robust, enabled us to identify how workplace depression shapes employee perceptions and is associated with financial performance. Second, this research contributes to the economic geography literature by examining the impact of workplace depression beyond corporate boundaries [109]. Specifically, we delineated the negative relationships between US states harbouring companies with high ODI scores and their depression rates, unfavourable personality traits, and deficiencies in important socio-economic factors such as state wealth or ability to attract the so-called creative class.

Practical Implications. Our work also has research, policy-making, and managerial implications. First, we proposed and validated a framework to analyze mentions of workplace depression that is underpinned by the ODI construct and measurement scale. Our framework will enable researchers to conduct large scale, text-analysis investigations of workplace depression in a wide range of organizational and sectorial contexts. Second, by establishing a link between workplace depression and the wider socio-economic environment, our research offers support to policy-making interventions directed towards promoting more flexible working environments and a healthier work-life balance. At a macroeconomic level, states that incentivize working conditions conducive to positive mental health could potentially enjoy advantages in terms of economic growth and attractiveness for the creative workforce. Third, our framework and empirical insights can guide managers to develop corporate policies that support employees and prevent those working environment conditions that are associated with depression [47]. For example, the provision of mental-health training for managers can have important occupational outcomes, such as a reduction in employees absence due to sickness [76]. Furthermore, managers can deploy employee assistance programs to address employees work-related struggles and mental health issues as soon as these manifest in order to safeguard employees well-being and improve work satisfaction and productivity [91].

7.2 Limitations and Future Work

This research operationalizes the workplace depression construct on a large-scale dataset and outlines how it relates to important organizational and societal outcomes. While these results shed light on the antecedents and consequences of workplace depression, our research comes with five main limitations that could inspire future research.

Misclassifications. While our text analysis framework is methodologically robust, it inevitably suffered from some text classification errors, mainly due to polarity assessments in sentences. To gain confidence in our results' validity and reliability, we followed guidelines from previous work [29] and assessed their accuracy with a principled mix-method assessment. Due to the black-box nature of this model, recent work has focused on the so-called *probing classifiers* [6, 23], which AutoODI could eventually integrate. Probing classifiers represent a way of interpreting and analyzing deep neural network models of natural language processing. They identify which parts of a neural model are responsible for certain linguistic properties (e.g., syntax, semantics), or which parts of the input led the model to make a certain decision.

AutoODI dimensions. While our AutoODI method was able to correctly detect workplace depression with a true positive rate of 74%, the method does not include all nine ODI's dimensions. Despite ensuring a fairly representative sample of reviews, dimensions such as suicidal ideation were less prominent in our sample. That is because individuals do not necessarily disclose every aspects of their lives online. In Islands of Privacy [80], Christena Nippert-Eng offers an intimate view into people's efforts to preserve the border between themselves and the rest of the world (e.g., how we manage our secrets, our email conversations, or even what we disclose online in the Internet). Despite the growing interest in online therapy services [81] and an increased awareness of mental health issues in the workplace, many employees are less likely to disclose severe mental health conditions on a review website. An employee who feels overwhelmed and stressed may use words such as overworked or burnout, but not words pointing to suicidal ideation (e.g., 'my life is miserable and I want an end'). However, as online conversations about mental health become less stigmatized and more data on a wider range of mental health symptoms become available, our AutoODI model could be retrained with more and diverse samples, including all the nine ODI's dimensions.

Data biases. While our dataset is extensive (350K geo-referenced employees reviews), the sample of major companies that met the minimum threshold for automatic processing is limited to 104. Future research could extend this sample by adopting alternative methodological approaches (e.g., organizational ethnography) that do not rely on published employee reviews. Additionally, these public employee reviews suffer from population biases (e.g., imbalance in gender, race, ethnicity, or educational background) and self-representation biases (e.g., the desire to portray one's self in a particular way, thus not necessarily disclosing 'the whole truth') [82]. However, despite such biases, these publicly available reviews allowed us, for the first time, to operationalize the ODI construct at scale.

Data representativeness. At an aggregate level, some sectors (e.g., IT) may attract more employee reviews, thus potentially influencing the representativeness of our dataset. To address this issue, we examined the industry sector distribution of the S&P 500 companies and demonstrated that our data does not over-represent any given sector. We also examined the issue of data representativeness at the state level and, in Figure 1, showed that: *1)* the official state population size scaled linearly with the number of employees aggregated at state level in our data; and *2)* the number of officially registered companies in each state is almost perfectly correlated with the number of state headquartered companies extrapolated from our data.

Causality. Given the structure of our dataset, we could not probe the causal relationship between workplace depression and the wider socio-economic outcomes examined. Future research should assess whether limiting the insurgence of workplace depression directly produces desirable socio-economic results (e.g., state wealth), whether more favorable socio-economic conditions reduce the likelihood of workplace depression, or whether the two causal effects operate concurrently in a self-reinforcing cycle.

Reproducibility. To facilitate the reproducibility of our methodology, all the company-level and state-level ODI scores plus the code have been made publicly available.[4]

REFERENCES

[1] [n. d.]. 2021 Depression Prevalence in US States. https://worldpopulationreview.com/state-rankings/depression-rates-by-state. Accessed: 2021-08-06.

[2] [n. d.]. Iowa Community Indicators Program. https://www.icip.iastate.edu/tables/population/urban-pct-states. Accessed: 2021-08-06.

[3] [n. d.]. US Bureau of Economic Analysis. https://www.bea.gov/data/gdp/gdp-state. Accessed: 2021-08-06.

[4] [n. d.]. Yahoo Finance portal. https://finance.yahoo.com.. Accessed: 2021-08-02.

[5] Stephen R Barley, Gordon W Meyer, and Debra C Gash. 1988. Cultures of culture: Academics, practitioners and the pragmatics of normative control. *Administrative science quarterly* (1988), 24–60.

[6] Yonatan Belinkov. 2022. Probing classifiers: Promises, shortcomings, and advances. *Computational Linguistics* 48, 1 (2022), 207–219.

[7] Renzo Bianchi and Irvin Sam Schonfeld. 2020. The Occupational Depression Inventory: A new tool for clinicians and epidemiologists. *Journal of Psychosomatic Research* 138 (2020), 110249.

[4]https://github.com/Indiiigo/autoODI

[8] Renzo Bianchi and Irvin Sam Schonfeld. 2021. Occupational Depression, Cognitive Performance, and Task Appreciation: A Study Based on Raven's Advanced Progressive Matrices. *Frontiers in psychology* (2021), 4276.

[9] Renzo Bianchi and Irvin Sam Schonfeld. 2022. Is the Occupational Depression Inventory predictive of cognitive performance? A focus on inhibitory control and effortful reasoning. *Personality and Individual Differences* 184 (2022), 111213.

[10] Renzo Bianchi, Irvin Sam Schonfeld, and Eric Laurent. 2017. Can we trust burnout research? *Annals of Oncology* 28, 9 (2017), 2320–2321.

[11] Renzo Bianchi, Irvin Sam Schonfeld, and Eric Laurent. 2019. Burnout: Moving beyond the status quo. *International Journal of Stress Management* 26, 1 (2019), 36.

[12] Moreno Bonaventura, Luca Maria Aiello, Daniele Quercia, and Vito Latora. 2021. Predicting urban innovation from the US Workforce Mobility Network. *Humanities and Social Sciences Communications* 8, 1 (2021), 1–9.

[13] Danah M Boyd and Nicole B Ellison. 2007. Social network sites: Definition, history, and scholarship. *Journal of Computer-mediated Communication* 13, 1 (2007), 210–230.

[14] Laura F Bright, Susan Bardi Kleiser, and Stacy Landreth Grau. 2015. Too much Facebook? An exploratory examination of social media fatigue. *Computers in Human Behavior* 44 (2015), 148–155.

[15] Romain Brisson and Renzo Bianchi. 2017. On the inconsistency of burnout conceptualization and measurement. *Journal of the American College of Surgeons* 224, 1 (2017), 87.

[16] Adelina Broadbridge. 2003. The appeal of retailing as a career 20 years on. *Journal of Retailing and Consumer Services* 10, 5 (2003), 287–296.

[17] Leslie H Brown, Paul J Silvia, Inez Myin-Germeys, and Thomas R Kwapil. 2007. When the need to belong goes wrong: The expression of social anhedonia and social anxiety in daily life. *Psychological Science* 18, 9 (2007), 778–782.

[18] Eliane Bucher, Christian Fieseler, and Anne Suphan. 2013. The Stress Potential of Social Media in the Workplace. *Information, Communication & Society* 16, 10 (2013), 1639–1667.

[19] Rafael A Calvo, David N Milne, M Sazzad Hussain, and Helen Christensen. 2017. Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering* 23, 5 (2017), 649–685.

[20] Zhigang Chen, Wei Lin, Qian Chen, Xiaoping Chen, Si Wei, Hui Jiang, and Xiaodan Zhu. 2015. Revisiting word embedding for contrasting meaning. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing*. 106–115.

[21] Minje Choi, Luca Maria Aiello, Krisztián Zsolt Varga, and Daniele Quercia. 2020. Ten social dimensions of conversations and relationships. In *Proceedings of The Web Conference 2020*. 1514–1525.

[22] Minje Choi, Luca Maria Aiello, Krisztián Zsolt Varga, and Daniele Quercia. 2020. Ten Social Dimensions of Conversations and Relationships. In *Proceedings of The ACM Web Conference (WWW)*. 1514–1525. https://doi.org/10.1145/3366423.3380224

[23] Kevin Clark, Urvashi Khandelwal, Omer Levy, and Christopher D Manning. 2019. What Does BERT Look at? An Analysis of BERT's Attention. In *Proceedings of the ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. 276–286.

[24] Glen Coppersmith, Ryan Leary, Patrick Crutchley, and Alex Fine. 2018. Natural language processing of social media as screening for suicide risk. *Biomedical informatics insights* 10 (2018), 1178222618792860.

[25] Glen Coppersmith, Kim Ngo, Ryan Leary, and Anthony Wood. 2016. Exploratory analysis of social media prior to a suicide attempt. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*. 106–117.

[26] Paul T Costa Jr and Robert R McCrae. 2008. *The Revised Neo Personality Inventory (neo-pi-r)*. Sage.

[27] Tom Cox, Mary Tisserand, and Toon Taris. 2005. The conceptualization and measurement of burnout: questions and directions. (2005).

[28] Hong-Jie Dai, Chu-Hsien Su, You-Qian Lee, You-Chen Zhang, Chen-Kai Wang, Chian-Jue Kuo, and Chi-Shin Wu. 2021. Deep learning-based natural language processing for screening psychiatric patients. *Frontiers in psychiatry* 11 (2021), 1557.

[29] Vedant Das Swain, Koustuv Saha, Manikanta D Reddy, Hemang Rajvanshy, Gregory D Abowd, and Munmun De Choudhury. 2020. Modeling organizational culture with workplace experiences shared on glassdoor. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–15.

[30] Munmun De Choudhury and Scott Counts. 2013. Understanding affect in the workplace via social media. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. 303–316.

[31] Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Social media as a measurement tool of depression in populations. In *Proceedings of the Annual ACM Eeb Science Conference*. 47–56.

[32] Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Eighth International AAAI Conference on Weblogs and Social Media*.

[33] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *Seventh International AAAI Conference on Weblogs and Social Media*.

[34] Munmun De Choudhury, Meredith Ringel Morris, and Ryen W White. 2014. Seeking and sharing health information online: comparing search engines and social media. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. 1365–1376.
[35] Danielle D DeSouza, Jessica Robin, Melisa Gumus, and Anthony Yeung. 2021. Natural language processing as an emerging tool to detect late-life depression. *Frontiers in Psychiatry* (2021), 1525.
[36] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
[37] Amandeep Dhir, Yossiri Yossatorn, Puneet Kaur, and Sufen Chen. 2018. Online social media fatigue and psychological wellbeing—A study of compulsive use, fear of missing out, fatigue, anxiety and depression. *International Journal of Information Management* 40 (2018), 141–152.
[38] Rui Dong and Shi G Ni. 2020. Openness to experience, extraversion, and subjective well-being among Chinese college students: The mediating role of dispositional awe. *Psychological Reports* 123, 3 (2020), 903–928.
[39] Gabriel Doyle, Amir Goldberg, Sameer Srivastava, and Michael C Frank. 2017. Alignment at work: Using language to distinguish the internalization and self-regulation components of cultural fit in organizations. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 603–612.
[40] Mark Damian Duda and Joanne L Nobile. 2010. The fallacy of online surveys: No data are better than bad data. *Human Dimensions of Wildlife* 15, 1 (2010), 55–64.
[41] Francis Duffy and Patrick Hannay. 1992. *The changing workplace*. Phaidon London.
[42] Kate Ehrlich and N Sadat Shami. 2010. Microblogging inside and outside the workplace. In *Fourth International AAAI Conference on Weblogs and Social Media*.
[43] Lynn Elinson, Patricia Houck, Steven C Marcus, and Harold Alan Pincus. 2004. Depression and the ability to work. *Psychiatric Services* 55, 1 (2004), 29–34.
[44] Daisy Fancourt, Claire Garnett, Neta Spiro, Robert West, and Daniel Müllensiefen. 2019. How do artistic creative activities regulate our emotions? Validation of the Emotion Regulation Strategies for Artistic Creative Activities Scale (ERS-ACA). *PLoS One* 14, 2 (2019), e0211362.
[45] Richard Florida. 2002. *The rise of the creative class*. Vol. 9. Basic Books.
[46] Richard Florida. 2005. *Cities and the creative class*. Routledge.
[47] Kayla B Follmer and Kisha S Jones. 2018. Mental illness in the workplace: An interdisciplinary review and organizational research agenda. *Journal of Management* 44, 1 (2018), 325–351.
[48] W Nelson Francis and Henry Kucera. 1979. Brown corpus manual. *Letters to the Editor* 5, 2 (1979), 7.
[49] Ronald D Fricker and Matthias Schonlau. 2002. Advantages and disadvantages of Internet research surveys: Evidence from the literature. *Field methods* 14, 4 (2002), 347–367.
[50] Daniel Fujiwara, Paul Dolan, and Ricky Lawton. 2015. *Creative occupations and subjective wellbeing*. Nesta.
[51] Larry M Gigliotti. 2011. Comparison of an internet versus mail survey: A case study. *Human Dimensions of Wildlife* 16, 1 (2011), 55–62.
[52] Lewis R Goldberg, John A Johnson, Herbert W Eber, Robert Hogan, Michael C Ashton, C Robert Cloninger, and Harrison G Gough. 2006. The international personality item pool and the future of public-domain personality measures. *Journal of Research in personality* 40, 1 (2006), 84–96.
[53] Sharath Chandra Guntuku, Anneke Buffone, Kokil Jaidka, Johannes C Eichstaedt, and Lyle H Ungar. 2019. Understanding and Measuring Psychological Stress Using Social Media. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 214–225.
[54] Jarrod M Haar, Marcello Russo, Albert Suñe, and Ariane Ollier-Malaterre. 2014. Outcomes of work–life balance on job satisfaction, life satisfaction and mental health: A study across seven cultures. *Journal of Vocational Behavior* 85, 3 (2014), 361–373.
[55] C Hill, L. T. de Beer, and R Bianchi. 2022. Validation and measurement invariance of the Occupational Depression Inventory in South Africa. *PloS one* (2022).
[56] Erin Hoare, Felice Jacka, and Michael Berk. 2019. The impact of urbanization on mood disorders: an update of recent evidence. *Current Opinion in Psychiatry* 32, 3 (2019), 198–203.
[57] Sue Jamison-Powell, Conor Linehan, Laura Daley, Andrew Garbett, and Shaun Lawson. 2012. " I can't get no sleep" discussing# insomnia on twitter. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI)*. 1501–1510.
[58] Melanie K Jones, Paul L Latreille, and Peter J Sloane. 2016. Job anxiety, work-related psychological illness and workplace performance. *British Journal of Industrial Relations* 54, 4 (2016), 742–767.
[59] Clare Kelliher, Julia Richardson, and Galina Boiarintseva. 2019. All of work? All of life? Reconceptualising work-life balance for the 21st century. *Human Resource Management Journal* 29, 2 (2019), 97–112.
[60] Joshua Kennon. [n. d.]. What Are the Sectors and Industries of the S&P 500? https://www.thebalance.com/what-are-the-sectors-and-industries-of-the-sandp-500-3957507

[61] Bryan Klimt and Yiming Yang. 2004. The enron corpus: A new dataset for email classification research. In *European Conference on Machine Learning*. Springer, 217–226.

[62] Stephen Kokoska and Daniel Zwillinger. 2000. *CRC standard probability and statistics tables and formulae.* Crc Press.

[63] Michal Kosinski, Sandra C Matz, Samuel D Gosling, Vesselin Popov, and David Stillwell. 2015. Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist* 70, 6 (2015), 543.

[64] Roman Kotov, Wakiza Gamez, Frank Schmidt, and David Watson. 2010. Linking "big" personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. *Psychological bulletin* 136, 5 (2010), 768.

[65] Kurt Kroenke, Robert L Spitzer, and Janet BW Williams. 2001. The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine* 16, 9 (2001), 606–613.

[66] Jongseo Lee and Juyoung Kang. 2017. A study on job satisfaction factors in retention and turnover groups using dominance analysis and LDA topic modeling with employee reviews on Glassdoor.com. (2017).

[67] Debra Lerner, David A Adler, Hong Chang, Leueen Lapitsky, Maggie Y Hood, Carla Perissinotto, John Reed, Thomas J McLaughlin, Ernst R Berndt, and William H Rogers. 2004. Unemployment, job retention, and productivity loss among employees with depression. *Psychiatric Services* 55, 12 (2004), 1371–1378.

[68] Suzan Lewis, Deirdre Anderson, Clare Lyonette, Nicola Payne, and Stephen Wood. 2017. *Work-life balance in times of recession, austerity and beyond.* Routledge, Taylor & Francis Group.

[69] Huijie Lin, Jia Jia, Liqiang Nie, Guangyao Shen, and Tat-Seng Chua. 2016. What Does Social Media Say about Your Stress?.. In *IJCAI*. 3775–3781.

[70] Philip M Liu and David A Van Liew. 2003. Depression and burnout. (2003).

[71] Daniel M Low, Kate H Bentley, and Satrajit S Ghosh. 2020. Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investigative Otolaryngology* 5, 1 (2020), 96–116.

[72] Gjorgji Madjarov, Dragi Kocev, Dejan Gjorgjevikj, and Sašo Džeroski. 2012. An extensive experimental comparison of methods for multi-label learning. *Pattern recognition* 45, 9 (2012), 3084–3104.

[73] Alan Marshall, Stephen Jivraj, James Nazroo, Gindo Tampubolon, and Bram Vanhoutte. 2014. Does the level of wealth inequality within an area influence the prevalence of depression amongst older people? *Health & place* 27 (2014), 194–204.

[74] Christina Maslach, Wilmar B Schaufeli, and Michael P Leiter. 2001. Job burnout. *Annual Review of Psychology* 52, 1 (2001), 397–422.

[75] Kaushal Mehta, Holly Kramer, Ramon Durazo-Arvizu, Guichan Cao, Liping Tong, and Murali Rao. 2015. Depression in the US population during the time periods surrounding the great recession. *The Journal of Clinical Psychiatry* 76, 4 (2015), 4221.

[76] Josie S Milligan-Saville, Leona Tan, Aimée Gayed, Caryl Barnes, Ira Madan, Mark Dobson, Richard A Bryant, Helen Christensen, Arnstein Mykletun, and Samuel B Harvey. 2017. Workplace mental health training for managers and its effect on sick leave in employees: a cluster randomised controlled trial. *The Lancet Psychiatry* 4, 11 (2017), 850–858.

[77] Megan A Moreno, Lauren A Jelenchick, Katie G Egan, Elizabeth Cox, Henry Young, Kerry E Gannon, and Tara Becker. 2011. Feeling bad on Facebook: Depression disclosures by college students on a social networking site. *Depression and Anxiety* 28, 6 (2011), 447–455.

[78] RD Newcomb, MW Steffen, LE Breeher, GM Sturchio, MH Murad, Z Wang, and RG Molella. 2016. Screening for depression in the occupational health setting. *Occupational Medicine* 66, 5 (2016), 390–393.

[79] Mark W Newman, Debra Lauterbach, Sean A Munson, Paul Resnick, and Margaret E Morris. 2011. It's not that I don't have problems, I'm just not putting them on Facebook: challenges and opportunities in using online social networks for health. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. 341–350.

[80] Christena E Nippert-Eng. 2010. *Islands of privacy.* University of Chicago Press.

[81] Amy Novotney. 2017. A growing wave of online therapy. *Monitor on Psychology* 48, 2 (2017), 48.

[82] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kıcıman. 2019. Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data* 2 (2019), 13.

[83] Minsu Park, David W McDonald, and Meeyoung Cha. 2013. Perception differences between the depressed and non-depressed users in twitter. In *Seventh International AAAI Conference on Weblogs and Social Media.*

[84] Vikram Patel, Jonathan K Burns, Monisha Dhingra, Leslie Tarver, Brandon A Kohrt, and Crick Lund. 2018. Income inequality and depression: a systematic review and meta-analysis of the association and a scoping review of mechanisms. *World Psychiatry* 17, 1 (2018), 76–89.

[85] Michael J Paul and Mark Dredze. 2011. You are what you tweet: Analyzing twitter for public health. In *Fifth International AAAI Conference on Weblogs and Social Media.*

[86] Umashanthi Pavalanathan and Munmun De Choudhury. 2015. Identity management and mental health discourse in social media. In *Proceedings of the International Conference on World Wide Web*. 315–321.

[87] Darrel A Regier, Emily A Kuhl, and David J Kupfer. 2013. The DSM-5: Classification and criteria changes. *World Psychiatry* 12, 2 (2013), 92–98.

[88] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084* (2019).

[89] Steven P Reise and Niels G Waller. 2009. Item response theory and clinical measurement. *Annual Review of Clinical Psychology* 5 (2009), 27–48.

[90] Peter J Rentfrow, Charlotta Mellander, and Richard Florida. 2009. Happy states of America: A state-level analysis of psychological, economic, and social well-being. *Journal of Research in Personality* 43, 6 (2009), 1073–1082.

[91] Melissa K Richmond, Fred C Pampel, Randi C Wood, and Ana P Nunes. 2017. The impact of employee assistance services on workplace outcomes: Results of a prospective, quasi-experimental study. *Journal of Occupational Health Psychology* 22, 2 (2017), 170.

[92] Jani H Ruotsalainen, Jos H Verbeek, Albert Mariné, and Consol Serra. 2014. Preventing Occupational Stress In Healthcare Workers. *Cochrane Database of Systematic Reviews* 11 (2014).

[93] Koustuv Saha, Ayse E Bayraktaroglu, Andrew T Campbell, Nitesh V Chawla, Munmun De Choudhury, Sidney K D'Mello, Anind K Dey, Ge Gao, Julie M Gregg, Krithika Jagannath, et al. 2019. Social Media As A Passive Sensor In Longitudinal Studies of Human Behavior And Wellbeing. In *Extended Abstracts of the ACM CHI Conference on Human Factors in Computing Systems (CHI)*. 1–8.

[94] Koustuv Saha and Munmun De Choudhury. 2017. Modeling stress with social media around incidents of gun violence on college campuses. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–27.

[95] Thomas L Schwenk and Katherine J Gold. 2018. Physician burnout—a serious symptom, but of what? *Jama* 320, 11 (2018), 1109–1110.

[96] Indira Sen, Daniele Quercia, Licia Capra, Matteo Montecchi, and Sanja Šćepanović. 2022. Insider Stories: Analyzing Internal Sustainability Efforts of Major US Companies from Online Reviews. *arXiv preprint arXiv:2205.01217* (2022).

[97] N Sadat Shami, Jeffrey Nichols, and Jilin Chen. 2014. Social media participation and performance at work: a longitudinal study. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. 115–118.

[98] Clare Shann, Angela Martin, Andrea Chester, and Scott Ruddock. 2019. Effectiveness and application of an online leadership intervention to promote mental health and reduce depression-related stigma in organizations. *Occupational Health Psychology* 24, 1 (2019), 20.

[99] Y Shoman, SC Marca, R Bianchi, L Godderis, HF van der Molen, and I Guseva Canu. 2021. Psychometric properties of burnout measures: a systematic review. *Epidemiology and psychiatric sciences* 30 (2021).

[100] Lynn M Shore, Jeanette N Cleveland, and Diana Sanchez. 2018. Inclusive workplaces: A review and model. *Human Resource Management Review* 28, 2 (2018), 176–189.

[101] Klaas Sijtsma and L Andries van der Ark. 2017. A tutorial on how to do a Mokken scale analysis on your test and questionnaire data. *Brit. J. Math. Statist. Psych.* 70, 1 (2017), 137–158.

[102] Karina Van De Voorde, Jaap Paauwe, and Marc Van Veldhoven. 2012. Employee well-being and the HRM–organizational performance relationship: a review of quantitative studies. *International Journal of Management Reviews* 14, 4 (2012), 391–407.

[103] Jerry J Vaske. 2011. Advantages and Disadvantages of Internet Surveys: Introduction to the Special Issue. *Human Dimensions of Wildlife* 16, 3 (2011), 149–153.

[104] Denna L Wheeler, Matt Vassar, Jody A Worley, and Laura LB Barnes. 2011. A reliability generalization meta-analysis of coefficient alpha for the Maslach Burnout Inventory. *Educational and Psychological Measurement* 71, 1 (2011), 231–244.

[105] Anna Wirtz, Friedhelm Nachreiner, and Katharina Rolfes. 2011. Working on Sundays–effects on safety, health, and work-life balance. *Chronobiology international* 28, 4 (2011), 361–370.

[106] Stephen Wood, Karen Niven, and Johan Braeken. 2016. Managerial abuse and the process of absence among mental health staff. *Work, Employment and Society* 30, 5 (2016), 783–801.

[107] Patrick M Wright. 2021. Rediscovering the "Human" in strategic human capital. *Human Resource Management Review* 31, 4 (2021), 100781.

[108] Anbang Xu, Haibin Liu, Liang Gou, Rama Akkiraju, Jalal Mahmud, Vibha Sinha, Yuheng Hu, and Mu Qiao. 2016. Predicting perceived brand personality with social media. In *Tenth International AAAI Conference on Web and Social Media*.

[109] Wei Yang and Lan Mu. 2015. GIS analysis of depression among Twitter users. *Applied Geography* 60 (2015), 217–223.

Received January 2022; revised April 2022; accepted May 2022