A Social Media Lens on the Needs and Concerns of Information Workers

JALEHSADAT MAHDAVIMOGHADDAM*, University of Toronto, Canada
KOUSTUV SAHA, University of Illinois at Urbana-Champaign, United States
JULIE HUI, University of Michigan's School of Information, United States
TAWANNA R.DILLAHUNT, University of Michigan's School of Information, United States
EBRAHIM BAGHERI, University of Toronto, Canada

Technological advancements have greatly impacted labor market dynamics, leaving a psychological impact on workers. Although some studies have explored such labor market changes and their effects on workers, they are limited to self-reported data, such as surveys and questionnaires. In this paper, we propose a new approach for identifying information workers' challenges and their impact on workers' emotional well-being using large-scale, inexpensive, and near-real-time online social network data. Our research is among the first to utilize statistical methods, machine learning techniques, and natural language analysis to show how labor market-related issues faced by information workers can be modeled and to what extent linguistic differences in online social content can be observed depending on gender and age strata. We find and systematically report highly discussed topics by information workers in areas related to education, skill development, job applications/job search, and employment/job concerns. We show that workers from different gender and age groups disclose their needs in notably different ways. In terms of the implications of our work, we discuss how online social network data can serve as a reliable lens, allowing us to better understand workers' challenges and well-being in the knowledge economy and how online social platforms can be vehicles for providing effective peer support to improve workers' well-being.

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

Additional Key Words and Phrases: workers' wellbeing, future of work, demographic differences, social media, knowledge economy

ACM Reference Format:
Jalehsadat Mahdavimoghaddam*, Koustuv Saha, Julie Hui, Tawanna R.Dillahunt, and Ebrahim Bagheri. 2025. A Social Media Lens on the Needs and Concerns of Information Workers. 1, 1 (July 2025), 32 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Online social networks play a significant role in many aspects of people's lives, and their careers are no exception. People increasingly engage with social media for various career-related reasons, from skill development to job search [36, 64, 73, 79, 82, 96, 97, 99]. Besides professional activities, many people rely on social media to seek emotional support when going through career-driven hardships, such as unemployment [52, 92]. In online settings where there is greater anonymity, individuals are likely to disclose personal and sensitive information related to physical and psychological health, which they might not otherwise freely reveal to their friends, colleagues, or managers [50]. While people may be reluctant to self-disclose health-related matters (including mental health) for various reasons, employers and organizations need to be positively cognizant of workers' concerns about their well-being [32]. At the individual level, health issues can affect many aspects of one's life, including mental health [118], physical health [126], and lifestyle [62]. At the organizational level, workforce health issues can lead to reduced productivity, lower work quality, and turnover, which can introduce tangible and intangible costs for organizations [78].

Understanding workers' challenges and well-being in the context of the labor market has been a key topic of interest in Computer-Supported Cooperative Work (CSCW) and Human-Computer Interaction (HCI) communities. Prior research in these fields has examined the impacts of technology and Artificial Intelligence (AI) on future employment [164], workforce stress level [104], and work-life balance [67], just to name a few. Additionally, researchers have explored well-being differences across various socio-demographic groups, such as gender, age, and race [8, 9, 157]. For instance,

Authors' Contact Information: Jalehsadat Mahdavimoghaddam*, University of Toronto, Toronto, Ontario, Canada, jaleh.mahdavimoghaddam@gmail.com*CorrespondingAuthor; Koustuv Saha, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States; Julie Hui, University of Michigan's School of Information, Ann Arbor, Michigan, United States; Tawanna R.Dillahunt, University of Michigan's School of Information, Ann Arbor, Michigan, United States; Ebrahim Bagheri, University of Toronto, Toronto, Ontario, Canada.

previous research [76] found that regularly working from home can increase daily fatigue and work-to-family conflict for women more than men, while the ability to take time off during the day for personal reasons can decrease job stress for both genders. The results of another study revealed that there were differences in the levels of occupational stress experienced by Information Technology (IT) workers of different ages, with younger individuals reporting higher levels of stress than older ones [77].

The primary source of data on workers' well-being in previous studies have been self-reported data often collected through surveys and questionnaires [8, 16, 112]. The collection of such data is extremely costly and time-consuming when used frequently and periodically, which leads to a lack of temporal granularity required for effective actions [172]. Additionally, due to the cost and complexity of these methods, surveys are usually conducted on a small group of people who might not represent marginalized groups. As a result, conclusions on small samples may not be generalizable to larger populations [29].

Inspired by the growing interest in understanding workers' challenges in the fast-evolving labor market, we assess workers' well-being concerning different topics, e.g., job search and skill development, through social media data, which is large-scale, inexpensive, and near real-time. More specifically, we study the well-being of workers in the IT-related fields, also known as "information workers", who are among the most stressed [112]. Information workers are those who deal with data and information to perform their daily jobs, e.g., software engineers and business analysts, "*instead of dealing with physical objects of labor*[1]". The broader objective of our work is to show that online social network data is a ripe, reliable, and readily available source of data for understanding the challenges faced by information workers and how such challenges are impacting their well-being. On this basis and more specifically based on over 700,000 unique Reddit posts related to the IT domain, we answer the following research questions:

- **RQ1**: *What labor market-related issues are information workers expressing in online communities?* To this end, we employ a bottom-up approach to systematically develop a codebook to represent labor market topics in our collected Reddit data. We use this codebook to build annotated data, which we use to train machine learning classifiers to identify the labor market topics of online social network posts.
- **RQ2**: *What are information workers' well-being concerns that are shared in the labor market content posted in online communities, and how are they expressed, when considering gender and age groups?* For this purpose, we utilize natural language processing to infer information workers' gender and age. Then, to asses how information workers' well-being varies across different gender and age groups, we examine their linguistic variations, such as language and psycholinguistic differences, in relation to topics related to the labor market that had been discussed in online communities.

Our findings from these research questions demonstrate how online social network data may serve as a naturalistic data source to identify various issues that information workers face in the knowledge economy and their emotional well-being regarding those concerns. The findings of this work and the approach taken will help human resource policymakers of academic and industrial organizations identify up-to-date challenges that their workers experience in the knowledge economy, and whether marginalized gender and age groups are particularly affected. By understanding these challenges through online data, the IT sector can be proactive in developing policies and behavioral strategies to help workers cope with stress, increase their well-being, and enhance their job performance.

Privacy, Disclosure, and Researcher Perspectives. Our research team comprises researchers holding diverse gender, racial, and cultural backgrounds, including people of color and immigrants, and holds interdisciplinary research expertise in the areas of HCI, CSCW, and social computing. This work utilizes public data available on Reddit without direct interaction with users; however, to protect the privacy of the individuals in our dataset and prevent reverse identification, we anonymize all usernames. To reduce traceability while still providing adequate context for our findings, we also paraphrased authors' quotes and deleted any individually identifying information when reporting quotes in this paper and elsewhere. We acknowledge that in this study, we only considered users who identify themselves as female or male, disregarding a wider, more inclusive spectrum of gender identities. Further significant effort has to be made to protect marginalized communities from harm and to accommodate a wider variety of gender identities. We provide a more detailed explanation of the ethical and privacy considerations in Section 6.

[1] https://www.igi-global.com/dictionary/information-worker/14587

2 Related Work

2.1 Psychological Well-being of Workers

One of the important factors to consider when studying labor market dynamics is its impact on workers' well-being. Well-being is a diverse concept that interconnects one's physical well-being, psychological well-being, and mental health [2]. According to [68], workers' well-being is "*that part of an employee's overall health that they perceive to be determined primarily by work and can be influenced by workplace interventions*". Understanding workers' well-being is essential due to the tangible and intangible benefits that it has for employees and their employers. For example, prior studies showed a significant positive relationship between workers' well-being and their productivity and a negative relationship with employee turnover [85]. As such, a growing research direction is focused on studying workers' well-being [26, 123, 132, 136, 155]. For instance, prior studies have shown that technological advancements enable workers to have a flexible work environment that gives employees control over where and when they work, which positively correlates with their well-being [67, 80]. Other research has explored the negative effects of information and communication technologies on workers' well-being, especially work-related stress. For instance, technology can noticeably increase the mental load of individuals. For example, employees' inability to unplug themselves from work leads to a blurred boundary between work and life [156].

More specifically related to IT job settings, various researchers have focused on understanding information workers' stress, mood, and cognitive load under differing circumstances, e.g., during a technical job interview [12, 112]. For instance, [8] explored whether the well-being of IT employees is different across age and gender groups.

Considering workers well-being has received special attention in recent years [20, 22, 24, 25, 27, 28, 55, 70, 92, 124, 130, 155, 173, 174]. Notably, [69] evaluated the discrepancy between employees' subjective well-being and their objective well-being. The authors collected self-reported data related to emotions from 15 IT workers through a diary study every half an hour, and employees automatically labeled emotions using their facial expressions by utilizing an Automatic Emotion Recognition (AER) tool every microsecond. The results showed the inadequacy of AER tools for correctly detecting employees' emotions without considering self-reported emotions. These authors further analyzed emotional differences between men and women and found sadness to be the dominant emotion for women compared to men, who expressed a combination of emotions, such as happiness, sadness, and anger, in their diary self-report emotions.

2.2 Assessing Workers' Well-being via Social Media

People across the globe use social media to express their emotions, share information about their health, seek social support, and record their interests, among other activities [58, 133]. Evidence suggests that an increasing number of people rely on the Internet to gain information related to mental health and well-being [4, 117]. However, a recent systematic review states that online user-generated content is still a novel technique for evaluating people's well-being, which is receiving growing attention in recent years [138].

Research has examined the effectiveness of online social media data in understanding mental health self-disclosure concerning psychological stress [58, 129], self-harm, e.g., suicide [142], anxiety [41], and depression [59, 61], to name a few. For example, [135] utilized X (formerly Twitter) data to quantitatively measure the impact of the COVID-19 pandemic on people's mental health by exploring how one's psychological concerns have evolved during the pandemic. [163] collected tweets posted by college students in different universities located in the United States to shed light on the mental well-being of students by analyzing how they express their needs and concerns. [102] took one step further and provided insight into the topical and linguistic features that can help understand mental health by leveraging Facebook private messages. Other researchers have also validated the effectiveness of social media data for predicting the mental health of individuals [136].

In the context of the labor market, prior studies have utilized social media data to examine well-being and mental health on a variety of topics, such as unemployment [52], job satisfaction [137], organizational culture [155], and role ambiguity [132]. For example, [172] measured and predicted job burnout on a Chinese social media network called Weibo. The authors explored the relationship between job burnout and users' linguistic and semantic preferences in content posted on Weibo, the frequency and timing of the posts, and finally, the interaction between each user and their network. The authors' findings suggest that before the burst of burnout, workers generated more content, and their language styles, such as terminology selection, were different from other groups without experience of burnout. [52] applied a mixed methods approach to model topics expressed by workers on Reddit about unemployment and explored the emotions that workers revealed for each topic.

2.3 Technologies and Digital Data to Support Workers' Well-being

Social support can help people overcome difficult situations, which positively impacts individuals' well-being [44, 147]. Interviews of 28 individuals who went through major transitions, e.g., relocation, job loss, or career change, revealed that the support received from unknown networks was more effective than the support provided by known ties [176]. Through a mixed methods study consisting of a survey and semi-structured interviews, [39] showed the important role of community support in up-skilling, job search success, income improvement, and well-being enhancement of those who were experiencing financial hardships and poverty. Similarly, [63] emphasized the relationship between social support and success for those under-resourced entrepreneurs who feel the need to conduct their business using digital tools. Other scholars have designed a variety of tools to provide social and emotional support for community members in an automated way to improve their well-being [101, 151, 165].

Research has also investigated the importance of social and emotional support on workers' well-being [57, 81, 139]. An investigation including 867 IT professionals revealed that support from managers for their workers' job autonomy, e.g., to work remotely, greatly reduced workers' psychological and physical distress [103]. Other scholars have emphasized the role of social support during workers' different life stages, such as lifelong learning [84], unemployment [92], and post-college transition [55]. With the increase in the popularity of online social media, researchers have explored the utility of online communities as a source of social support [6, 75]. For example, [117] applied a mixed-method approach to explore how young people leverage the Web to seek help. [177] utilized machine learning and psycholinguistic analysis techniques to investigate veterans' needs and the type of support they receive from an online community.

2.4 Automated Detection of Social Media Users' Psychological Well-being

In contrast to the work mentioned earlier on workers' well-being, another strand of research focuses on applying machine learning methods to identify and classify online content regarding users' psychological well-being [7, 65, 66, 71]. For example, [147] explored the congruence between community mental support and users' linguistic alignment. To identify the expressed mental health issues, the authors applied a clustering technique called k-means on Reddit data to detect the overall psychological issues, followed by a semi-open coding technique to verify and label the groups. To detect tweets with self-disclosure information about schizophrenia, [147] applied natural language processing to detect posts with explicit mentions of schizophrenia disorder, i.e., being diagnosed with schizophrenia. Then, they trained machine learning models to expand their dataset by classifying more posts within these mental health categories.

Since machine learning-based methods are often data-hungry and require labeled data, another widely-used technique in this space is the utilization of psychological lexicons, such as Linguistic Inquiry and Word Count (LIWC), which allow for the curation of large-scale labeled datasets [140]. For example, [52] applied an emotion lexicon to detect emotions such as sadness, embarrassment, and rejection among Reddit users in the context of unemployment. Other researchers detected psychological stress, i.e., anxiety, frustration, and anger, using Reddit, Facebook, and X (formerly Twitter) via a lexical list of terms [55, 58]. [148] combined lexicon-based features of LIWC and N-grams (N=1,2) to train machine learning models to predict possible anxiety disorders of Reddit users. For instance, [134] used LIWC features and machine learning to study how community support affects users' psychological change on Talklife, and [177] applied LIWC to analyze Reddit posts and identify Veterans' needs and the support they received from the community.

Distinct contributions of this paper in relation to related work. The work in the paper provides novel contributions to the field in, at least, the following areas. We note that our work can be seen to offer synergistic impact from both methodological and findings perspectives on the strong work already undertaken by other researchers:

(1) While existing research has shed light on the role of social support in workers' well-being, such works often fall short of providing insight into the support that information workers request on social media, which would help them navigate their challenges. Building on this motivation, we contribute to the growing interest in understanding influential social support by systematically studying information workers' challenges in different career stages. We provide a complete picture of the possible support mechanisms that workers can receive, enhancing their mental health and minimizing the psychological impacts they experience in the fast-changing labor market.

(2) To the best of our knowledge, no prior research has quantitatively analyzed the challenges that the IT workforce faces and their impact on workers' well-being while considering their demographic groups. Building on this motivation, we bridge the gap in the literature by proposing a quantitative approach for understanding IT workers' challenges expressed in online social network communities across different gender and age groups.

(3) While existing literature contributes to our understanding of the workers' well-being, information workers' needs and challenges in different career stages remain unexplored. To bridge this literature gap, we offer insights into the workers' main concerns and psychological well-being across different demographic groups by studying the language by which they express their well-being in online communities. Recognizing how workers belonging to different gender and age groups disclose their challenges and well-being in online communities will enable policy and decision-makers to concentrate on designing initiatives that can enhance workers' well-being regarding their socio-demographics at different stages of their careers.

(4) Furthermore, in our work, we quantify information workers' well-being in terms of their anger, sadness, and anxiety by adopting a psychological lexicon (i.e., LIWC), and also use machine learning classifiers to identify the topics that are discussed within online communities. We do, however, note that we do not consider the use of the psychological lexicon and the machine learning classifiers to be the contributions of our work, but they are rather enabling tools that allow us to pursue the main objectives of our work in this paper.

(5) Finally, we contextualize the findings of our work in this paper within earlier findings by earlier researchers in order to examine whether (1) the findings are consistent with earlier work and hence reliable, and (2) whether social media data is a reliable reflection of information workers' challenges. In our analysis we identify areas where our findings align with earlier works, indicating that findings from our work obtained from social data can corroborate with those done through surveys, questionnaires, and focus groups. In our assessment, it is important to show this alignment in order to show that social data is a reliable source of data for the purposes presented in this paper.

3 Collecting IT-related Data from Social Media

This paper studies information workers' well-being as expressed on a community-based social network, namely Reddit. Reddit is one of the largest online communities, with more than 100 thousand active forums and over 22 billion posts and comments as of December 2024[2]. One of the contributing factors to the popularity of Reddit is that users can anonymously express their opinions on a given topic [5]. Users are afforded a lenient character limit per post (up to 40,000 characters), significantly more than other social media platforms, such as X (formerly Twitter) [148]. Additionally, Reddit data can be easily analyzed using an official API [84] or the Pushshift Reddit dataset [10]. Therefore, many scholars rely on the rich user-generated content from Reddit to answer various questions, from mental health [30, 129, 131] to the labor market [50, 52, 96, 97]. Similar to prior studies, we leverage Reddit data to address the research questions described in Section 1.

To collect Reddit data for this study, we first curated the list of IT skills from the International Standard Classification of Occupations (ISCO), which is a tool for global labor market reporting [11] (Table 1). Then, we downloaded Reddit posts from Pushshift[3], which were posted between January 2015 and August 2019. The Pushshift dataset, which is frequently updated, provides a comprehensive archive of Reddit data that contains posts and comments from almost all public subreddits since 2005 [10]. Unlike Pushshift, accessing historical data is not possible through the official Reddit API. While Pushshift also provides an API, we downloaded monthly dumps of data to avoid a delay in data collection due to the query limit imposed by the Pushshift API. Next, we conducted text pre-processing on the obtained data to optimize retrieval effectiveness based on the collected IT skills. This included lowering the case, deleting stop words, and stemming. Note that we only included Reddit posts that were written in English and contained one of the IT skills keywords in their title or body and did not have the tags *"deleted"* or *"removed"*. We also removed posts from the same author, title, and body of the post. The above steps led to a collection of posts in all subreddits that contained at least one of the IT-related keywords. Then, through an iterative process involving extensive re-evaluation, we curated a list of subreddit names primarily featuring posts related to technical skills. We then removed subreddits exclusively dedicated to news, academia, or recruitment, and non-technical topics such as "r/relationship," "r/depression," and "r/starwars". As a result, we collected 787,398 unique posts in 546 subreddits, posted by 417,983 unique users. We provide more details on the motivation behind our data collection strategy in Appendix A.

[2]https://www.redditinc.com
[3]https://files.pushshift.io/

Table 1. List of ISCO IT skills.

objective-c	cyber security	javascript	elasticsearch	git	sparql
chatbots	mobile marketing	photoshop	business intelligence	jquery	scrum
angular	tensorflow	agile	deep learning	java	cobol
tableau	sql server	xquery	software testing	iot	docker
ansible	neural networks	sass	machine learning	vagrant	xcode
data protection	3d modeling	ai	cloud computing	python	saas
joomla	coffeescript	c#	artificial intelligence	reactjs	swift
apache kafka	programming	Ruby	apache spark	php	linux
graphql	windows phone	mysql	cloud technologies	unix	c++
datamining	cassandra	django	computer vision	scala	nosql
postgresql	virtualization	robot	software development	mongodb	database
matlab	web design	css	data engineering	devops	sql
googlecloud	typescript	.net	project management	aws	haskell
vbscript	data analysis	opencv	adobe illustrator		

4 Research Question 1: Modeling and Analyzing Information Workers' Challenges on Reddit

4.1 Identifying Labor Market Topics on Reddit

To identify the topics discussed in Reddit posts, as suggested in [82, 84, 88], we created a codebook following a bottom-up process called *inductive coding* [158]. The inductive approach is a qualitative coding process where the topics are derived from raw data by a close reading of the text [158]. Using this approach, researchers can iteratively and systematically create and refine conceptual categories until they adequately represent the thematic view of the observed data [19, 160]. We randomly sampled 500 Reddit posts out of 787,398 to develop the codebook. Then, two researchers individually coded each sample in the 500 posts based on their understanding and wrote a definition for each observed code. In approximately 20% of the cases, the researchers assigned two codes to the text when they identified the presence of more than one topic. Then, the two researchers sat together and went through each of their codes in detail. As part of this process, they re-assessed their codes and grouped some conceptual categories to represent a broad category. After this step, only one code was assigned to each post. This step led to a preliminary codebook that reflected labor market-related topics.

After forming the conceptual categories, the researchers verified the obtained codebook by coding another 50 Reddit posts. A list of these codes and a detailed description of them can be found in Table 2. Finally, the two researchers coded another 450 sampled randomly posts to assess how similar their interpretations were. In doing so, a high average Cohen's Kappa coefficient of 0.91 was obtained. Using this codebook, the two researchers manually coded a random sample of 1.5% (11,854) of the entire dataset. When researchers were unsure about a specific label for a post, they debated the options. We report the frequency of observed topics in the labeled dataset in Table 2. While our data is dedicated to IT topics shared in more than 400 subreddits, we find similar topics to those found by [52] on "r/Jobs" and "r/GetEmployed" subreddits. To investigate whether IT workers' challenges remain the same or change over time, we applied the same steps to data from a different time frame. The results from this analysis are reported in Appendix B. We then used the qualitatively coded data to build machine learning classifiers to label the remaining 775,544 Reddit posts in our dataset. The details of these classifiers are described in the following section.

4.2 Building a Machine Learning Classifier of Labor Market Topics

To extract the labor market topics in Reddit posts at scale, we built machine learning classifiers on the annotated dataset. As shown in Table 2, we observed that the most frequent topics in our collected IT-related posts were "Skill Development", "Job Application/Job Search", "Education", and "Employment/Job Concerns".

To select the best-performing machine learning classifiers, we explored different classifiers, such as Logistic Regression (LR), Random Forest (RF), and Linear Support Vector Classifier (SVC). We trained these models on term frequency-inverse document frequency (TF-IDF) vectors of Reddit post titles and bodies. The GridSearchCV from the Sklearn library was used to tune the hyperparameters. Additionally, we fine-tuned bidirectional transformers, such as BERTbase (12-layer, 768-hidden, 12-heads), and XLNetbase (12-layer, 768-hidden, 12-heads), which are already trained on a large corpus of English data. To fine-tune BERTbase and XLNetbase, we set the batch size to 16, the learning rate to $2E-5$, maximum sequence length to 512, and used the ADAM optimizer. We utilized a *k-fold (k=5) cross-validation* approach on our labeled data (11,854 posts) to assess the performance of our classifiers. Table 3 reports the accuracy metrics of

Table 2. Codebook of common topics derived from Reddit posts discussing IT-related topics.

Topic	Definition	Example Posts
Skill Development	Gaining popular skills and degrees in the labor market; offering people help in skill development or picking up a new tool; posting about how endless the field is, and how much there is to learn.	The website I designed looks like shit, how do I improve it
Job Application/Job Search	Looking for resources to help with job searching, such as networking opportunities; seeking suggestions and feedback for enhancing resumes or interview techniques; things to do while searching for a job, such as volunteering.	I'm very anxious about getting my first programming job. Any advice or resources to help me get started
Education	Seeking advice for a university major/minor; Changing fields of study; discussing the cost of education; seeking advice on persuading more degrees, i.e., a Ph.D.; getting advice on getting back to school, obtaining the required skills to enter grad school; complaining about a degree/school; grades/school failure stories; discussing admission to the university; admission applications; how to improve in university to find a job.	If you're paying for graduate school in a STEM field, you're doing it wrong. I think I might be doing it very wrong, help [..]?
Employment/Job Concerns	Wages vs. work effort; hostile/unethical work environments and strategies for improving them; job burnout and decreased job satisfaction; benefits; work-life balance; underpayment; conflict with co-workers over technical ideas; people who got fired and talk about their experience (not currently doing a job search); getting hired for one position but being asked to take on other responsibilities.	Work a steady IT job but a miserable creative
Decision Making	Making career decisions and finding light when they are lost; coming up with a name for a job title if already employed; getting a job or academic degree; finding the education path that leads to the best career option in the future; making decisions on whether to leave a job for another when the job offer is received; accepting a job offer or rejecting it.	What career(s) involve problem-solving in politics?
Career/Industry Change	Thinking about changing the current job due to personal reasons; seeking advice to change the current industry; learning new skills to change a career; catching up with new skills to switch jobs; discussing resignation when there is already a job offer to change the job.	Considering a career switch, currently overwhelmed and underpaid.
Technical Challenges	The problems that learners or workers face in learning or applying a skill.	Problems booting HDD - Linux noob needs help.
Qualification	Being over- or under-educated/experienced; Not being able to apply/perform jobs because of a lack of certain skills; discussing mentioning imposter syndrome; feeling unqualified to apply for other jobs or enter the job market; worrying about getting poor performance reviews as a result of bad work; worrying about skill transfer if I switch to another job.	I failed college, am kinda lost, and have no job. Any advice?
Health	Posts related to information workers' mental and physical health.	A three-time college dropout, anxiety troubles, and feeling apprehensive about the future.

the best-performing classifier. The XLNet classifier performs the best in predicting each topic and has the best overall accuracy (F1-score = 0.84). By comparing our results to similar studies that trained machine learning classifiers to explore a variety of phenomena using online data, we noticed that the accuracy of our best model is within the range of [70-90], which is the median reported performance in the literature by prior scholars [29, 147, 172]. For example, the accuracy of the neural-based model in identifying user engagements on Amazon.com and TheGuardian were reported at 0.76 and 0.71, respectively [122]. Similarly, [131] achieved a median precision of 0.75, a median recall of 0.74, and a median F1-score of 0.75 in an attempt to train a binary Neural Network classifier to detect the Reddit posts that contained any sign of stress from the member of the LGBTQ+ group.

We use XLNet to label our entire dataset. We summarize the performance metrics of the other trained models in Table 14 (in the Appendix). By applying XLNet, 6.60% (51,987 out of 787,398) of posts were labeled as Skill Development, 2.3% (18,061 out of 787,398) of posts were labeled as Job Applications/Job Search, 2% (15,600 out of 787,398) of posts were labeled as Education, and 0.53% (4,198 out of 787,398) of the posts were labeled as Employment/Job Concerns. The rest were labeled as "other".

4.3 Detecting Information Workers' Themes around Most Discussed Topics on Reddit

As discussed in Section 4.2, most information workers posted about four major topics, namely Skill Development, Job Applications/Job Search, Education, and Employment/Job Concerns. In this section, we systematically review information workers' themes around the most discussed topics on Reddit. To get insight into the themes identified in the Reddit posts, we first conducted an exploratory data analysis. To do so, we first applied an unsupervised phrase

Table 3. Performances of the best machine learning model for the topic classification task based on the F1-score. The overall accuracy for all the topics is reported at the end of the table.

Topic	Precision	Recall	F1-Score
Job Applications/Job Search	0.77	0.78	0.77
Skill Development	0.84	0.85	0.85
Employment Concerns	0.73	0.83	0.78
Education	0.79	0.79	0.79
Other	0.91	0.87	0.89
Mean	0.85	0.85	0.85

extraction technique, namely the Rapid Automatic Keyword Extraction algorithm (RAKE) [125], to retrieve the most frequently mentioned phrases within the Reddit posts. We reported the top-mentioned phrases in each topic in Table4. Based on Table4, we noticed that some of the most frequent terms, i.e., "learn", "start", "job", and "experience", are common in all of the labor market topics. However, some terms were highly discussed in particular topics. For example, "team", "people", and "busy" were the most frequent terms in the Employment/Job Concerns topic, while "research", "university", and "graduate" were discussed the most in the Education topic.

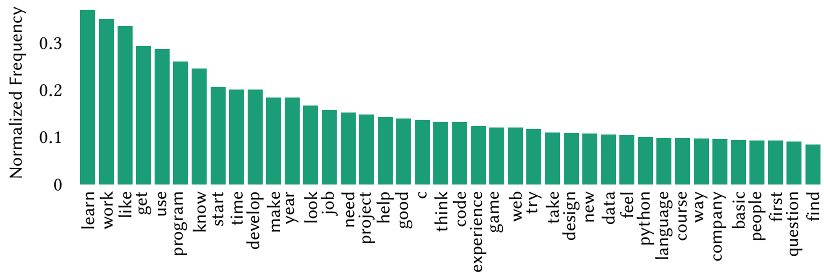
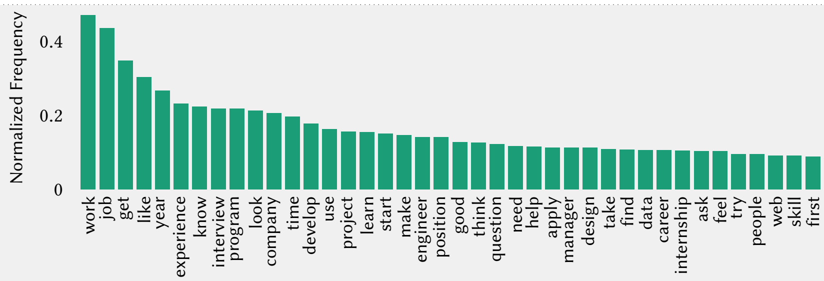
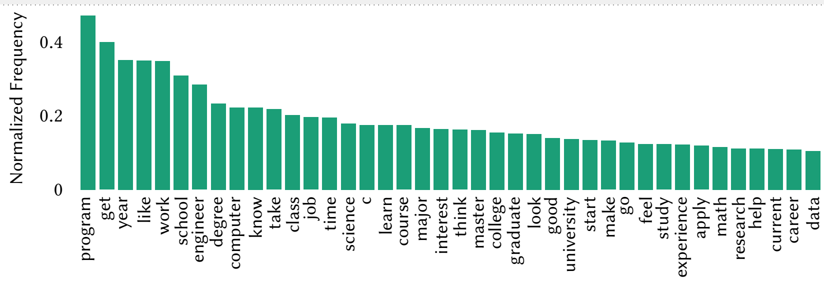
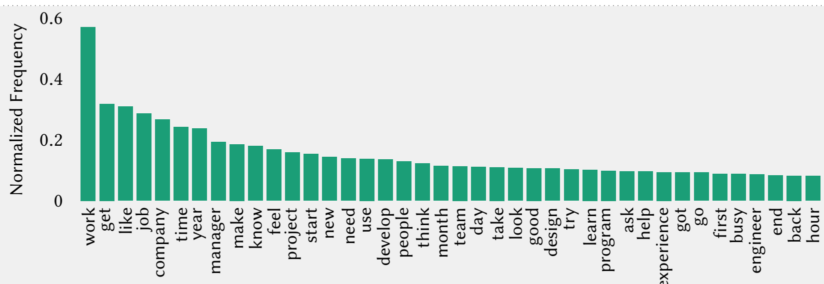
To comprehensively examine the key themes that are present in each topic, we applied a mixed methodology. First, we extracted all $n-$grams (n=2,3) from Reddit posts in each labor market topic along with their $tf-idf$ score. Then, we performed an unsupervised clustering technique, namely K-means [60], to aggregate the $n-$grams into groups based on their content similarities. Then, two researchers reviewed a 5% random sample of the Reddit posts that contained any of the $n-$grams in each group and discussed the theme that the posts presented. The final themes and the top $n-$grams that represent each theme in each labor market topic are illustrated in Table 5.

By analyzing the explored themes in Table 5, we noticed the prevalence of learning discussions in all labor market topics among information workers. This prevalence is consistent with prior research highlighting the enduring popularity of learning in this context [96]. We observed that these posts were mainly written by workers who were either in the process of entering the job market or were already employed and were looking for resources and tips to upgrade their skills. For example, one of the workers who was looking for a job as a computer programmer wrote *"I have a degree in criminology, but I am learning to program from home [..]"*. Another worker who discussed Employment/Job Concerns wrote *"I work in a company with a tight budget. My supervisor asks me to take the time to learn and understand what I'm doing, but we are low in staff, and I don't have time for learning and improvement [..]"*. Such examples highlight the importance of learning in the labor market.

As mentioned by prior researchers, workers in the knowledge economy are required to be lifelong learners who continuously learn, re-learn, and unlearn to stay employable and in demand [97]. Lifelong learning is a continuous process that happens everywhere in different contexts throughout life in a formal or informal setting [47]. Formal learning relates to a structured educational model designed and run by a predetermined curriculum in a setting that requires students, teachers, and institutes where certification or degrees will be issued upon completion of the program [34]. In our dataset, the Education topic represents content around formal learning activities. Informal learning, which is 70-90% of human learning, occurs outside institutes, e.g., through workshops or training, where students do not need to interact with teachers, duration, and goals might vary, and a certificate may or may not be given [86]. In our dataset, the Skill Development topic refers to the posts generated by information workers focused on obtaining a new skill through informal learning. Thus, even though the issues of education and skill development are unique since they both include different kinds of lifelong learning activities, it is logical that learning would be the subject that receives the most attention in both.

The reported themes in Table 5 revealed that the inquiries related to the Skill Development topic were focused on finding the next best course/topic to learn and tips on how/where to learn a particular skill. For example, one of the workers who was trying to find the best place to learn a skill wrote: "*I have a full-time job but I'd like to learn programming in my spare time. I want to start with MySQL/Python and then C. Where can I learn this? online courses? evening classes?*". Also, many workers complained about being lost when trying to find a good starting point. For instance, one of the workers in the process of learning mobile development wrote: "*I'm struggling with where to begin [..] Am I correct in thinking this is just too much to learn in too little time? [..] I've spent all morning just trying to figure out how to add a button, but I'm afraid to touch anything".*

Table 4. List of labor market topics and the most discussed terms in each topic across our Reddit dataset expressed by information workers.

Labor Market Topics	**Top 40 Terms**
Skill Development	
Job Applications/Job Search	
Education	
Employment/Job Concerns	

The themes in the **Education** topic indicated that the majority of the workers required help from online communities to find the right major/minor programs that would help workers find their desired jobs or pursue further degrees. For example, one of the workers wrote“*should I pursue a master's degree in robotics and intelligent systems alongside my wife's pursuit of a Ph.D. or prioritize job stability and remain in my current role as a controls engineer?*”. We also observed that workers engaged with online communities to ask for support and advice regarding future employment. For example, in

Table 5. $N - grams$ that represent the most discussed themes in each labor market topic.

Topic	Final Theme	Top $n - grams$
Skill Development	Finding the next best course/-topic to learn	programming language learn, best programming language, learn web development, learn new language, html cs javascript, coding boot camp, need help finding, ai machine learning, learn machine learning, improve programming skill, help getting started, asp net mvc, learn programming language, need help choosing, learn asp net, good starting point, language start learning
	Tips on how/where to learn a particular skill	best way learn, help finding good, good way start, good way learn, good stepping stone, good starting point, good resource learning, invest time learning, improving programming skill, improve python skill, improve problem-solving, important thing learn, good place learn, good programming book, need help learning, best place learn
Job Application/Job Search	Job interview process	interview coming week, got technical interview, resume cover letter, position require year, second round interview, cracking coding interview, received job offer, final round interview, big tech company, require year experience, start building portfolio
	Professional and educational background	background computer science, recently got accepted, mid-level developer, going grad school, college couple year, pursue graphic design, previous work experience, bachelor degree computer, work large company, computer science internship, went community college, work data scientist, graduated bachelor degree, graduated college degree
	Learn Skills to land a better job	data science course, learning web dev, learning basic python, started learning, computer science class, good starting salary, good stepping stone, considering going school, feel like learned, graphic design skill, new programming language, learning web development, learning html css
	Work-life balance	work-life balance, good work life, working hour week
Education	Finding the right major/minor programs that would help workers find their desired jobs	need help deciding, considering going school, entry level position, like computer science, computer engineering major, year computer science, computer science major, computer science job, computer science school, interested computer science, considering applying PhD, chance getting job, land good job, learn computer science, time finding job, bachelor degree work, time job work
	Learning a new subject	statistic machine learning, started learning programming, learn programming language, learning data structure, taking programming class, feel like learning, intro programming course, learn computer science
	Pursuing further degrees	dual degree program, going graduate school, plan getting master, engineering master degree, straight master program, master program accounting, pursuing double major, master program worth, feel like pursuing, going grad school, applying phd program, accepted master program, online master program, apply master program, second bachelor degree
Employment/Job Concerns	Decide whether they should stay in or leave their current job	feeling bit lost, finding new job, getting taken advantage, good career path, got great job, got job offer, got new job, got offered job
	Micro-management	micromanagement, title change, direct boss, rebooted, yelling, annual review, old boss, good guy, taken advantage, getting fired, deaf ear, entire office, boss say
	Working overtime	hour work week, hour week working, long period time, work overtime
	Learning	learned html cs, learning html cs, learning new language, learning new technology, learning new thing, like learning new

the **Job Search/Job Application** topic, workers mainly asked questions or shared their personal experiences about their job interview processes, their professional and educational background, the skills that they need to learn to land a better job, and the work-life balance that their current job provided versus the ones that they were looking for. We also noticed that workers who successfully landed a job discussed important hints about their career, managers, colleagues, or the projects that they were assigned in the **Employment/Job Concerns** topic. Some workers posted on Reddit when trying to decide whether they should stay in or leave their current job, and some workers disclosed information about micro-management and working overtime. For example, one worker in a post related to Employment/Job Concerns complained about the vagueness of their job description, which caused stress and anxiety, and the lack of support they received from their manager in this regard: "*I feel like not fully understanding my scope has caused a lot of stress. I have mentioned not understanding my scope to my manager and my team, but I received no help or support. Now my manager is on leave, and everything is placed on me. There is zero documentation and no help, and I am afraid of getting fired [..]*". These examples and the most discussed issues reported in Table 4 point to some of the top challenges faced by information workers in the fast-changing labor market, which are consistent with earlier research on the challenges of workers who are learning new skills through formal education or self-learning [55], those trying to enter the job market [12, 168], or workers who are already employed [132, 150].

5 Research Question 2: Well-Being Concerns and their Expression Across Information Workers' Genders and Ages

In the second research question, we examine information workers' well-being concerns as reflected in their posts and compare their linguistic characteristics across different gender and age groups. We specifically study linguistic differences in terms of two sub-research questions, namely (RQ2.a) language differences, and (RQ2.b) psycholinguistic differences. We will elaborate on how these differences are estimated and our findings in the following sections. Detailed explanations of how age and gender were extracted can be found in Section 5.1. We acknowledge that there are limitations in how we identified workers' genders and ages, which we discuss in-depth in Section 6.3.

5.1 Retrieving Information Workers' Age and Gender from Reddit

Users are not required to disclose their socio-demographic data, such as gender, age, race, or occupation, on many online social platforms. For example, users can sign up on the Reddit platform by only providing their email, a username, and a password. Researchers have applied various techniques to study the role of socio-demographic attributes in different phenomena using social media data. For example, to study whether females and males promote themselves differently on LinkedIn, Altenburger et al. [3] used predicted gender labels from users' profiles. Tifferet et al. [159] classified LinkedIn users as female or male based on their profile pictures. To study social media data about users' demographics, other researchers collected social media users' demographics such as age, gender, race, and income through a questionnaire [54, 58]. In a recent study, [120] identified gender by leveraging a set of keywords that represent females and males, such as "her", "mother", "woman", and "fiancee" for females and "he", "son", "father", "finance", and "brother" for males. [14] used labeled data as a training set to retrieve the gender of unlabeled data using deep learning classifiers, such as BERT and XLNet. According to [14], the fine-tuned BERT model can show reliable gender prediction results in the context of information retrieval. It's important to note that the proposed methodology by [14] is used when there is no explicit mention of gender in the data. Also, since the dataset from [14] consists of queries and passages taken from internet pages or social media, their suggested methodology may be used in social media studies when there is no direct mention of gender. Therefore, In our study, to retrieve information on workers' gender, we applied a similar methodology to [120] and [14] on our own data to curate a training dataset and extract gender, which we will explain in the following sections.

By analyzing Reddit posts, we noticed that while individuals have the option to withhold their gender when creating an account on Reddit, some workers decide to declare this information in the title or body of their posts when reaching out to online communities. To do so, workers include [F] or [M] in the title or phrases such as "I am a woman" and "I am a man" in the body of their posts to identify themselves as female or male, respectively. Therefore, to identify information workers' gender in our dataset, we first applied regular expressions to select the posts that explicitly mentioned one's gender either in the Reddit titles or the body of the posts. For example, we extracted posts with keywords such as "m", "male", "man", "f", "female", and "woman", where all Reddit content is lowercase. Through this approach, we extracted 512 posts as male users, 3,856 posts as female users, and 783,030 posts as others. To verify that the posts obtained through the rule-based approach indeed contained self-disclosed information about the user's gender, either male or female, one of the authors manually read the Reddit posts in the male category and confirmed that 457 posts explicitly stated the user's gender as male. However, although the remaining 55 posts contained male references, there was not enough evidence to suggest that the user revealed their gender. Therefore, those 55 posts were categorized as neutral (we use the word neutral to mean that gender was unknown). To ensure our dataset was balanced, the same author randomly sampled 1000 posts from the female category, of which 470 contained self-identified information about being female and 530 included female references, but the explicit mention of the users' gender was missing. Therefore, those 530 posts were categorized as neutral. We then used this gathered data to fine-tune a BERT model [33] with a batch size of 16, a learning rate of $3E-5$, 15 epochs, and the Adam optimizer. We assessed the performance of this model by applying a k-fold (k=5) cross-validation technique. We report the performance of this model in Table 6. As shown in Table 5, the overall performance of the model and also per gender is 91%. We used this model to retrieve gender from the 89,846 posts introduced in Section 4.2. As a result of this process, we obtained 5,847 posts submitted by female workers, 23,105 by male workers, and 60,894 posts as neutral (or undecided).

Similar to the gender retrieval task, scholars have tried to detect the age of users in social media posts using a variety of methods, such as traditional machine learning models and deep learning techniques [1, 51, 106]. However, a literature review conducted by [110] on age and gender prediction methods revealed that age prediction using language features

Table 6. Performance of the fine-tuned BERT uncased model for the gender retrieval task.

Gender	Precision	Recall	F1-Score
Female	0.90	0.92	0.92
Male	0.91	0.91	0.91
Neutral	0.92	0.90	0.91
Mean	0.91	0.91	0.91

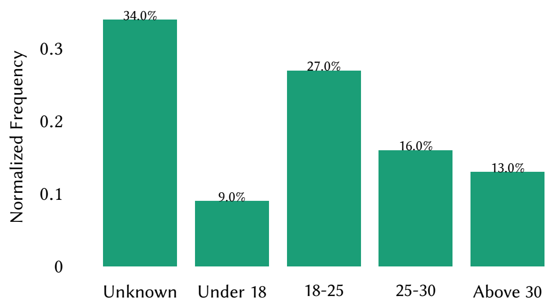


Fig. 1. Distribution of age in different groups.

of social media leads to a lower performance than gender prediction. For instance, by applying deep learning algorithms on X (formerly Twitter) data, [166] could only achieve a 0.52 accuracy on the multi-class age prediction task, while the performance of their model was over 0.91 for the gender prediction task on the same data. Therefore, to retrieve the age of the information workers who talked about one of the labor market topics in our 89,846 posts dataset, we applied regular expressions to only retrieve those Reddit posts that explicitly used phrases, such as “25 years old”, “M/23”, “I’m 45 Y”, and “I turned 30”, to name a few. In other words, instead of predicting the users’ age, we only relied on the posts with explicit mention of the users’ age to avoid fabricated prediction results in our analysis.

According to a report by Agrawal [4], 8% of the Reddit population are individuals between the ages of 13 and 17, 45% are between 18 and 29, 40% are between the ages of 30 and 49, and 7% are above the age of 50. Inspired by these statistics, for our analysis, we used similar groups except for breaking the 18–29 age bracket into two groups: individuals between 18 and 24 in one group and individuals between 25 and 29 in another. The PwC Global Workforce Hopes and Fears survey[5], which revealed that employees between the ages of 18 and 24 had distinct demands and expectations compared to workers belonging to other age groups, had a significant impact on the decision to divide the 18-29 age group into two units. Additionally, we noticed that in our dataset, only 0.2% of the posts were written by users in the 50-64 age category, and 0.02% were written by users above the age of 65. As a result of insufficient data quantity for older information workers, we merged those categories, and we only reported them as above 30. Using this approach, we extracted 11,910 posts containing age information. We report a breakdown of the age in each group in Figure 1. To validate the gathered data, one of the researchers randomly selected 500 posts, 100 posts in each age group, and manually labeled them. As a result, we obtained an overall average accuracy of 88.2% (below 18: 76%, 18-25: 91%, 25-30: 89%, above 30: 88%, and unknown: 97%).

5.2 RQ2.a: Language Differences Across Gender and Age Groups

To characterize language differences in labor market content across different demographic groups, we apply an unsupervised language modeling technique, namely the Sparse Additive Generative Model (SAGE) [42]. This technique retrieves the salient keywords in a given set of documents by measuring log-frequencies to distinguish between the lexical distribution of the document of our interest (i.e., posts written by women workers[6]) compared to a baseline

[4]https://medium.com/@sm_app_intel/the-user-demographics-of-reddit-the-official-app-7e2e18b1e0e1
[5]https://://www.pwc.com/gx/en/issues/workforce/hopes-and-fears-2022.html
[6]We embrace “women” as an adjective in this work to reach a broader meaning that places gender over biological sex.

Table 7. Top 30 salient keywords as per distinctness scores [42]. Bar lengths indicate the magnitude of scores; The **positive scores** show greater distinctness of phrases among women workers, and the **negative scores** show greater distinctness of phrases among men workers.

Women Workers			
Keyword	**Score**	**Keyword**	**Score**
total-code	1.27	mother	0.73
woman	1.17	cry	0.72
female	1.17	girlfriend	0.71
home-mom	1.10	sunday	0.69
beginner	1.04	sister	0.69
complete-beginner	0.99	tablet	0.68
girl	0.96	therapist	0.67
lady	0.93	adhd	0.67
daughter	0.91	nurse	0.66
receptionist	0.89	cashier	0.66
pissed	0.81	angry	0.60
park	0.81	adult	0.60
absolute-beginner	0.81	next-day	0.60
mom	0.79	medication	0.60
secretary	0.75	ashamed	0.59

Men Workers			
Keyword	**Score**	**Keyword**	**Score**
cyber-security	-1.06	second-bachelor	-0.78
thesis	-1.06	bachelor-degree	-0.78
undergraduate-degree	-0.96	canada	-0.78
double-major	-0.95	edu	-0.78
electrical-engineering	-0.94	master	-0.77
master-degree	-0.94	phd-program	-0.76
bioinformatics	-0.90	hey-everyone	-0.76
neuroscience	-0.87	graduate-degree	-0.75
diploma	-0.83	minor	-0.75
master-program	-0.83	job-prospect	-0.74
recently-graduated	-0.82	electrical	-0.74
computer-science-degree	-0.82	public-acounting	-0.72
phd	-0.81	gre	-0.71
mba	-0.80	undergrad	-0.70
germany	-0.80	chemical	-0.70

document (the entire corpus) [42]. We refer to these log-frequencies as the "*distinctness score*". This technique has been widely used as an effective approach among researchers who have studied lexical differences in social media, such as X (formerly Twitter) [43, 137, 142], and Reddit [136, 147, 177]. We applied SAGE to the extracted phrases using the RAKE model (introduced in Section 4.3) to retrieve distinct phrases between the content posted by different gender and age groups.

5.2.1 Language Differences by Gender. Table 7 represents the results of our analysis to distinguish the unique phrases used by workers discussing labor market topics. The scores reported in this figure indicate the degree of word uniqueness in each gender group. A positive score shows greater distinctness of phrases among women, and a negative score shows greater distinctness of phrases among men workers. By examining the results in Table 7, we identify the following findings:

Women' focus on social connections. As illustrated in Table 7, distinct keywords used by women were dominated by social references, such as women references (e.g., "female", "woman", "lady", "girl") and family/interpersonal relationships references (e.g., "daughter", "mom", "sister"). Unlike women workers and aligned with the prior work which studied job satisfaction across demographics [137], we found that the salient keywords for men were centered around occupations and job titles (project management, years of experience, software engineer, start working), education ("bachelor's degree", "master degree", "Ph.D.", "MBA"), and growth ("online course", "double major").

We observed that while the main purpose of Reddit posts generated by both women and men information workers was around obtaining answer/support to a variety of inquiries related to recruitment and learning, many women provided more detailed information about their background, family, and motivation behind the expressed question compared to men. For example, several workers used phrases that referred to social and interpersonal relationships to express the influence that those relationships had on the situation they described in their posts. This influence was expressed as fear of failing others by unsuccessful results, having the support of others to make the right decisions, or concerns around the outcome of particular actions on others. For example, one of the women workers who wanted to start a freelancing job wrote: "*I have a full-time job but I am feeling like a robot trying to finish my to-do list every day and I feel a little depressed. My husband travels every two weeks [..] My son spends most of his summer in camps [..] I have a mom who lives abroad [..] I like to spend more time with my family [..] Is freelancing a good option for me?*". As portrayed in the above example, the workers presented their situation and priorities (references to the mom, child, and spouse) to seek support from the community on a job change. Our findings are consistent with prior research conducted by [109] who reported that women were more likely to form their conversations around people and their internal processes as opposed to men who prioritize their discussions around their occupations and external events.

Table 8. Top 30 salient keywords as per distinctness scores [42]. Bar lengths indicate the magnitude of scores; The **positive scores** show greater distinctness of phrases among workers in the specified age range.

Under 18			
Keyword	Score	Keyword	Score
year-old-student	3.17	chemistry	1.25
love-computer	2.20	steam	1.21
young-age	1.98	father	1.20
grade	1.70	wanna	1.18
son	1.67	mom	1.17
age	1.66	game-engine	1.17
sister	1.60	dad	1.15
first-game	1.59	politics	1.10
head-start	1.59	max	1.07
young	1.32	gaming	1.07
game-developer	1.29	video-game	1.05
career-choice	1.29	game-development	1.04
mother	1.26	learn-c	1.04
high-school	1.25	lesson	1.03
pc	1.25	first-language	1.03

18-25			
Keyword	Score	Keyword	Score
vet-school	2.16	absolutely-hated	1.62
year-old-living	2.13	peace-corp	1.62
year-old-kid	2.02	enjoy-writing	1.62
year-old-student	1.84	sibling	1.61
mum	1.80	finished-high-school	1.61
still-living	1.74	grandmother	1.56
career-guidance	1.73	third-time	1.53
decent-money	1.73	older-brother	1.52
start-university	1.72	nursing	1.51
dental	1.66	bad-job	1.51
mental-health-issue	1.64	teach-english	1.51
pic	1.64	outdoors	1.51
firefighter	1.62	creative-writing	1.50
many-interest	1.62	like-computer	1.48
nursing-school	1.62	athlete	1.48

25-30			
Keyword	Score	Keyword	Score
lending-stream	2.89	barista	1.91
high-percentile	2.85	monotony	1.91
uncharted	2.69	moved-back-home	1.86
quarter-life-crisis	2.57	open-university	1.80
investigator	2.55	year-old-male	1.78
extremely-unhappy	2.43	poor-choice	1.75
camping	2.36	office-manager	1.74
considered-trying	2.28	two-kid	1.74
smoking	2.22	grocery-store	1.71
life-around	2.20	still-living	1.71
misdemeanor	2.18	theatre	1.70
year-old-man	2.13	making-enough-money	1.70
bartender	2.10	male	1.70
creative-job	1.97	life-together	1.70
oilfield	1.97	counseling	1.66

Above 30			
Keyword	Score	Keyword	Score
mortgage	1.95	old	1.37
pension	1.90	nursing	1.33
retirement	1.87	son	1.28
work-history	1.81	career-switch	1.27
change-career	1.78	lived	1.25
retire	1.74	mental-health	1.24
age	1.74	mid	1.22
married	1.69	killing	1.22
fortune	1.62	saving	1.21
changing-career	1.59	miserable	1.16
early	1.56	local-community-college	1.15
wife	1.55	younger	1.15
career-change	1.49	teenager	1.14
student-debt	1.48	husband	1.12
older	1.44	math-skill	1.12

Women express more psychological words. In addition to the differences in how women workers form their conversations and priorities compared to men workers, we observed a noticeable difference in the use of psychological words. By psychological, we refer to words that express emotional and cognitive experiences. Overall, our results reported in Table 7, revealed that women are, on average, more expressive of their emotions and well-being compared to men. For example, frequently used salient words by women workers were “pissed”, “cry”, “therapist”, “ADHD”, “angry”, and “ashamed”, while none of the top 30 distinct words expressed by men workers were around well-being. This finding is consistent with prior research that found when disclosing information, women use words to describe their emotional states, especially negative emotions, more than men [13, 157, 161]. We explore these differences in the expressed emotions and well-being across gender with statistical models in the next section.

5.2.2 Language Differences by Age. In Table 8, we present the distinct phrases used by workers in different age groups discussing the labor market topics. The scores reported in this Table indicate the degree of word uniqueness in each age group. We yield the following observations:

Career advice for younger workers. We find that among all workers, younger people rely more on online communities to seek career advice. We noticed that individuals under the age of 25 requested career advice by using keywords such as “career choice”, “career guidance”, and “decent money”, which could suggest that these young people rely on online communities to shape their future decisions or paths. For example, a user posting “*I am a new college graduate and I am desperately trying to get a job closer to my family. Although I have done one decent internship, my resume is on the weaker side. I have 2 questions: 1) how to deal with negative emotions after rejections, 2) how to set myself apart from more experienced candidates (like online certifications, YouTube lectures, etc.)?*”

Career switch for older workers. We found that some of the unique words used by workers above 30 are around career change, such as “change career”, “career change”, and “career switch”. For example, one worker mentioned a plateau in their career, which motivated them to further their learning, which would eventually lead to a job change to one that would be more challenging and have room for growth: “*I have an MBA and no technical background with several years of experience in analytics in various tech companies. We can assume there are few to no learning opportunities at my current job (outside of learning on my own). We can also assume there is no room to move up in my current company. At this point in my career, would it be worth it to get some sort of statistics/analytics Bootcamp credential?*”. In addition to career change, this age group had references to financial commitments and long-term goals through keywords such as “mortgage”, “retirement”, and “pension”. Overall, our findings could suggest that while workers in other age groups try to enter the workforce and upgrade their skills or establish themselves more in their current occupations, older workers’ priority is to change their careers or fields to break a plateau, gain financial advantage, or find jobs that are more fulfilling than their current ones. Our findings align well with earlier studies on behavioral patterns of older job seekers [90, 175]. For instance, [90] reports that job-seeking motivations for older adults could be categorized based on factors addressing their income, among others. This finding can be important for employers to understand the needs of their employees better and work on solutions to help the workforce with growth and learning plans to prevent attrition, increase their productivity, and find fulfillment while employed. We discuss these findings in more detail in Section 6.

5.3 RQ2.b: Psycholinguistic Differences Across Gender and Age Groups

We next discuss results related to psycholinguistic differences disclosed in information workers’ content across different aspects of the labor market based on workers’ gender and age groups. To address this, influenced by several earlier works that utilized a lexicon to identify the psycholinguistic characteristics of social media language [134, 148, 177], we used the well-validated Linguistic Inquiry and Word Count 2022 (LIWC-22) lexicon [15] to understand how information workers express their psychological well-being. The LIWC lexicon has 50 categories that include words from the interpersonal focus, lexical density and awareness, personal and social concerns, cognitive processes, affect, health and behavioral states, motives, perception, and temporal reference categories representing a range of psychological attributes. Prior research has emphasized the importance of lexical analysis for understanding the psychological well-being and processes of individuals [131, 134]. For example, it has been shown that variation in using articles and pronouns can reveal critical information on people’s reactions to complex psychological stimuli [29, 128].

After extracting psycholinguistic characteristics of workers’ labor market content, we statistically test the normality of our data by performing the Shapiro-Wilk test on each category. A p-value of less than 0.05 was found by the Shapiro-Wilk test on each of the psycholinguistic characteristics of our four labor market topics (Skill Development, Job Applications/Job Search, Education, and Employment/Job Concerns), indicating that our data on the labor market is not normally distributed. To determine whether there are statistically significant differences between two or more groups that are not normally distributed, we use the Kruskal-Wallis H-test to analyze the psychological well-being of information workers. Then, we measured the effect size using Cohen’s d to estimate the differences between the groups’ means. Note that Cohen’s d was only measured for significant psychological categories based on the Kruskal-Wallis H-test.

5.3.1 *Psycholinguistic Differences by Gender.*

Table 9 represents the distinctive psycholinguistic characteristics across the labor market topics and gender groups. By examining the results in Table 9, we offer the following observations:

Affective and health attributes. As discussed in Section 5.2.1, we found a greater prevalence of keywords related to emotional well-being in content posted by women workers. Similarly, our analysis of workers’ psycholinguistic attributes, illustrated in Table 9, revealed that negative affective categories, such as “negative tone” were more prevalent in content posted by women workers. Specifically, we found a greater usage of negative attributes, such as “anxiety”, “anger”, and “sadness” in women workers’ content when disclosing information about the Education and Skill Development topics, and a higher prevalence of positive tone among men workers when discussing Job Applications/Job Search topic. However, we found no significant differences between affective attributes used by women and men who discussed Employment/Job Concerns. Our finding on the difference between disclosure of negative affective attributes is aligned with [157], which showed women are more expressive of their negative emotions compared to men. Among health attributes, women are more expressive about health and behavioral states, e.g., “*I dropped out of college due to depression and drug addiction, but now I am trying to resume my education.*” While these categories have negative implications, prior

Table 9. Comparing psycholinguistic differences across gender groups for the labor market topics by reporting effect size (Cohen's d) in each age group. Following Bonferroni correction, statistical significance is reported as follows: *** p < .001, **.001 < p < .01, *.01 < p < .05.

	Education		Learning		Job Concerns		Job application	
Category	**Gender**	**H-stat.**	**Gender**	**H-stat.**	**Gender**	**H-stat.**	**Gender**	**H-stat.**
Interpersonal Focus								
1st person singular	0.32	87.24***	-	-	0.23	15.62***	-	-
1st person plural	0.18	136.52***	0.17	59.95***	-	-	0.13	58.88***
3rd person singular	0.29	365.9***	0.24	161.69***	-	-	0.27	203.12***
3rd person plural	0.15	88.78***	0.22	74.37***	0.33	36.02***	0.3	111.46***
Lexical Density and Awareness								
Articles	-0.2	17.55***	-	-	-	-	-	-
Adverbs	0.33	66.04***	-	-	-	-	0.16	14.68***
Negations	0.41	120.1***	0.12	15.73***	-	-	0.15	16.9***
Quantities	0.21	29.72***	0.1	13.56***	-	-	-	-
Personal and Social Concerns								
Affiliation	0.41	198.38***	0.27	102.08***	-	-	0.17	34.35***
Achievement	-0.33	30.21***	-0.17	23.24***	-	-	-0.26	44.91***
Power	0.17	54.58***	-	-	-	-	0.22	39.69***
Prosocial behavior	-0.26	13.44***	-	-	-	-	-0.21	19.04***
Interpersonal conflict	0.11	141.95***	0.02	22.19***	0.14	28.76***	0.06	20.66***
Communication	-0.17	15.02***	-	-	-	-	-	-
Family	0.27	263.46***	0.19	122.5***	0.22	30.94***	0.16	89.02***
Friends	0.13	170.87***	0.1	52.49***	-	-	0.08	33.61***
Politeness	-0.35	47.41***	-0.13	13.51***	-	-	-0.3	63.35***
Female references	0.4	652.88***	0.62	1523.77***	0.6	329.44***	0.48	748.31***
Male references	-	-	-0.93	240.03***	-0.89	77.48***	-0.49	55.15***
Technology	-0.54	107.51***	-0.17	16.47***	-0.34	23.18***	-	-
Leisure	0.1	95.66***	-	-	-	-	-	-
Home	0.19	223.78***	0.07	36.03***	-	-	0.07	25.45***
Work	-0.43	100.78***	-0.43	159.16***	-0.24	12.91***	-	-
Religion	0.08	128.12***	0.03	24.17***	-	-	0.04	34.14***
Cognitive processes								
All-or-none	0.35	120.81***	-	-	-	-	-	-
Causation	0.25	58.1***	0.19	40.99***	-	-	-	-
Tentative	-0.22	16.65***	-	-	-	-	-	-
Certitude	0.13	38.92***	-	-	-	-	-	-
Affect								
Negative Tone	0.52	243.82***	0.14	26.95***	0.24	19.83***	0.14	12.32***
Positive Tone	-	-	-	-	-	-	-0.15	12.32***
Anxiety	0.23	180.02***	0.01	12.37***	-	-	0.1	33.76***
Anger	0.13	139.23***	0.05	26.34***	-	-	-	-
Sadness	0.16	168.52***	0.06	33.41***	-	-	-	-
Health and Behavioral States								
Illness	0.1	102.75***	0.05	33.32***	-	-	0.01	12.34***
Wellness	0.03	33.31***	0.05	30.07***	-	-	0.04	27.34***
Mental health	0.18	200.77***	-	-	0.12	13.84***	0.07	29.24***
Death	0.05	84.84***	0.05	33.32***	-	-	-	-
Need	0.04	25.81***	0.05	30.07***	0.18	12.21***	-	-
Fatigue	0.11	115.21***	0.02	19.88***	-	-	-	-
Motives								
Risk	-0.1	28.64***	-	-	-	-	-	-
Curiosity	-0.34	45.31***	-0.16	15.73***	-0.17	14.15***	-0.18	16.32***
Allure	0.25	42.54***	-	-	-	-	-	-
Perception								
Motion	0.22	51.42***	-	-	-	-	0.11	12.15***
Space	-0.22	17.85***	-0.26	47.37***	-	-	-0.19	21.83***
Feeling	0.27	128.33***	0.11	21.71***	-	-	-	-
Temporal References								
Time	0.33	76.49***	-	-	-	-	0.16	13.75***
Past focus	0.38	86.02***	-	-	-	-	0.26	38.05***

research suggests that disclosing information regarding emotions and health can improve psychological well-being and help to cope with hardship [114].

Motivations, personal and social concerns. Among personal and social attributes, we find that men workers have a higher prevalence of attributes related to prosocial behavior (tendency to help or care about others), communication (e.g., talk, explain, or ask), and politeness (e.g., please or thanks), e.g., "*Hey, I have my interview soon. They'll ask me about scripting, and I told them I am going to do it in Python. Can you help me to prepare for it? Thank you so much!*". Also, we observed that men had greater usage of keywords around their careers, such as "achievement", "technology", and "work", followed by high usage of keywords related to curiosity as their motives; e.g., "*I want to be a great programmer, so I've done online learning for python and JavaScript. My plan in life would be software engineering, and I want to learn more!*". Unlike men workers, we find a greater prevalence of keywords about family, friends, and home in the content

Table 10. Comparing psycholinguistic attributes across age groups in the Job Application/Job Search theme, by reporting effect size (Cohen's d) in each age group. Following Bonferroni correction, statistical significance is reported as follows: *** p < .001, **.001 < p < .01, *.01 < p < .05.

Category	Under 18	18-25	25-30	Above 30	H-stat.
Interpersonal Focus					
1st person singular	-0.01	0.17	-0.01	-0.22	20.74*
Personal and Social Concerns					
Power	-0.28	-0.03	0.06	0.06	18.68*
Money	-0.45	-0.03	0.10	0.06	28.68***
Cognitive processes					
Discrepancy	0.34	0.10	-0.06	-0.19	20.08*
Tentative	0.32	0.08	-0.05	-0.17	19.46*
Health and Behavioral States					
Illness	-0.07	-0.07	0.04	0.08	20.30*
Perception					
Space	-0.30	-0.08	0.13	0.06	18.82*
Temporal References					
Time	-0.31	-0.00	-0.08	0.20	21.26*

posted by women. For example, *I am interested in Architecture and Animation, but my parents forced me to take computer science because my mom thinks I am bad at drawing and my stepdad is against Animation.* . This finding is aligned with our earlier observation in Section 5.2.1 where we reported that women workers have a higher tendency to focus on social connections while men are more focused on career growth.

5.3.2 Psycholinguistic Differences by Age. We identified distinctive psycholinguistic characteristics across the labor market topics and age groups by examining the psycholinguistic attributes indicated in workers' Reddit posts. We found that although there are significant differences across genders when discussing **Employment/Job Concerns**, there is no significant difference in the psycholinguistic attributes of workers in different age groups. This section reports our findings on labor market topics with a statistically significant difference in the expressed keywords across age groups. We also report our findings in Table 11,Table 12, and Table 10.

Affective and cognitive attributes. Among cognitive categories, we observe that workers under the age of 18 show the highest prevalence of words related to discrepancy (Table 11, Table 12, and Table 10) and tentativeness (Table 11 and Table 10), which represent a person's uncertainty and insecurity over the circumstance [136]. For instance, "*I'm 16 and I think I might be getting my first job!*" or "*I'm a 15 yo sophomore. I was just wondering what I should actively be doing if I want a career in mechanical engineering*". Among the affective categories, we find that workers under the age of 18 revealed the highest usage of positive tone in **Skill Development** (Table 12), which might suggest that learning for workers under 18 is an overall positive experience. In contrast, workers between the ages of 18 and 25 used the most keywords around negative tone, anxiety, and anger in **Education** (Table 11). This might be associated with the majority of individuals between the ages of 18 and 25 being college students, who disclose negative emotions about their campus life or trying difficulties in trying to get some professional work experience, e.g., "*College grad, lost, frustrated [..] I'm just lost as to what I should be doing/what I'm doing wrong. I've been applying nonstop for a while now to all kinds of jobs with no luck.*" Our observations about the negativity expressed by college-age workers can also be explained by the findings of [56], which show that the expression of negative emotions by college students is strongly correlated with positive support received from their peers, e.g., their roommates.

Temporal references, personal and social concerns. Among personal and social concerns, we find that workers above the age of 25 use the most number of terms around "achievement", "work", and "money" when discussing **Education** (Table 11) while prevalent keywords among workers under the age of 25 were around "leisure" with a focus on the present time. This finding could mean that as workers become older, their priorities for pursuing education or getting a formal degree shift toward following a passion, becoming more successful, or finding a better career with more income, e.g., "*I am 32 yrs. and have been a nurse for a few years. I am considering going back to school for Computer Engineering because I think it's fascinating and want to learn how to create and develop hospital EMRs, since I have so many ideas working with different systems.*" When discussing **Skill Development**(Table 12), we observe a higher focus on the present time in workers' content under the age of 25 and a higher focus on the past time in workers' content above

Table 11. Comparing psycholinguistic attributes across age groups in the Education theme, by reporting effect size (Cohen's d) in each age group. Following Bonferroni correction, statistical significance is reported as follows: *** p < .001, **.001 < p < .01, *.01 < p < .05.

Category	Under 18	18-25	25-30	Above 30	H-stat.
Interpersonal Focus					
1st person singular	0.13	0.27	-0.18	-0.37	62.76***
Impersonal pronouns	0.24	0.19	-0.18	-0.29	52.7***
Lexical Density and Awareness					
Articles	-0.32	-0.15	0.21	0.21	52.96***
Numbers	0.10	-0.24	0.12	0.20	49.67***
Adverbs	0.27	0.15	-0.13	-0.29	47.88***
Negations	-0.01	0.24	-0.16	-0.23	38.50***
Quantities	-0.36	-0.02	0.10	0.15	29.30***
Personal and Social Concerns					
Achievement	-0.58	-0.09	0.23	0.23	101.75***
Friends	0.07	0.00	0.09	-0.25	18.54*
Female References	-0.19	-0.10	0.16	0.06	21.83**
Technology	0.31	-0.11	-0.10	0.12	27.13***
Leisure	0.13	0.10	-0.07	-0.24	25.93***
Work	0.07	-0.20	0.11	0.16	21.68**
Money	-0.52	-0.04	0.14	0.20	85.69***
Cognitive processes					
Insight	0.14	0.19	-0.22	-0.15	38.21***
Discrepancy	0.20	0.07	-0.04	-0.23	21.72**
Tentative	0.25	0.07	-0.09	-0.20	24.25**
Certitude	0.23	0.13	-0.16	-0.22	25.80***
Affect					
Negative Tone	-0.07	0.25	-0.11	-0.30	44.06***
Anxiety	-0.10	0.18	-0.01	-0.29	37.97***
Anger	-0.05	0.14	-0.08	-0.11	19.15*
Swear Words	-0.19	-0.05	0.00	0.18	18.56*
Health and Behavioral States					
Mental Health	-0.24	0.09	0.02	-0.10	18.94*
Want	0.14	0.11	-0.04	-0.27	28.09***
Fulfilled	-0.27	0.00	0.08	0.05	29.59***
Perception					
Space	-0.10	-0.17	0.15	0.19	30.04***
Feeling	-0.21	0.12	-0.01	-0.07	27.34***
Temporal References					
Time	-0.39	-0.10	0.18	0.18	41.18***
Present focus	0.16	0.10	-0.09	-0.18	19.03*

the age of 25. We also found a distinct usage of keywords in revealed personal and social concerns among different age groups. For example, we observed the highest prevalence of keywords related to "family" and "communication" for workers under the age of 18, "achievement", "work", and "home" for workers between the ages of 25 and 30, and "power", "interpersonal conflicts", and "politics" for workers above the age of 30. The difference in the use of keywords in the personal and social categories could be associated with the fact that, while in the rapidly changing labor market, workers are demanded to be lifelong learners who continuously update their skills [96, 97], the learning process takes different shapes depending on the age group.

Linguistic style attributes. We find a significant difference between pronoun usage across age groups, especially among the workers under the age of 25, which is an indication of different social attentions [136]. For example, we observed a high prevalence of first-person singular pronouns (e.g., me, my, myself) in workers' content under the age of 25 who discussed **Education** and **Skill Development**(Table 11 and Table 12), which can indicate a higher focus on personal experiences and self-reflection [115]. For example, "*I started my career in graphic design last year, but I am concerned about my choice. This job makes me feel overwhelmed, and I don't get motivated by it. Should I continue with this career?*" where the workers disclose their challenges with their career. However, we only found a high usage of keywords related to first-person singular pronouns in the workers' content aged between 18 and 25 who discussed **Job Application/Job Search** (Table 10). For instance: "*I have been out of college for 2 years, but I can't even get an internship. I have done a few projects to build my portfolio, but I still think I am unqualified [..]*". Prior studies show that

Table 12. comparing psycholinguistic attributes across age groups in the Skill Development theme, by reporting effect size (Cohen's d) in each age group. Following Bonferroni correction, statistical significance is reported as follows: *** p < .001, **.001 < p < .01, *.01 < p < .05.

Category	Under 18	18-25	25-30	Above 30	H-stat.
Interpersonal Focus					
1st person singular	0.19	0.04	-0.02	-0.26	47.41***
3rd person singular	0.26	-0.14	-0.13	-0.03	20.71*
3rd person plural	-0.04	0.04	0.01	-0.02	21.45**
Lexical Density and Awareness					
Articles	-0.36	0.04	0.11	0.22	87.79***
Prepositions	-0.27	0.02	0.15	0.11	53.82***
Auxiliary verbs	0.24	-0.00	-0.11	-0.14	45.65***
Adverbs	0.12	0.05	-0.09	-0.12	19.79*
Common verbs	0.37	-0.06	-0.10	-0.22	93.62***
Common adjectives	-0.16	0.02	0.00	0.14	20.02*
Personal and Social Concerns					
Achievement	-0.40	0.09	0.26	0.03	133.61***
Female References	0.12	-0.10	-0.06	0.03	28.11***
Power	-0.30	0.03	0.11	0.14	97.65***
Interpersonal conflict	-0.14	-0.03	0.01	0.16	56.47***
Communication	0.21	-0.02	-0.13	-0.08	20.45*
Family	0.14	-0.14	-0.08	0.10	25.92**
Politic	-0.13	-0.01	0.02	0.12	46.79***
Home	-0.13	-0.02	0.11	0.04	48.28***
Work	-0.24	0.08	0.18	-0.03	59.25***
Money	-0.28	0.10	0.06	0.07	127.97***
Cognitive processes					
All-or-none	-0.19	0.02	0.10	0.08	51.31***
Insight	0.31	-0.03	-0.16	-0.16	48.08***
Discrepancy	0.36	-0.07	-0.14	-0.18	67.87***
Affect					
Negative Tone	-0.12	0.04	0.08	-0.01	37.42***
Positive Tone	0.24	-0.06	-0.06	-0.12	28.95***
Anxiety	-0.18	0.04	0.10	0.01	46.96***
Anger	-0.08	-0.03	0.07	0.05	26.50***
Sadness	-0.05	0.02	0.02	-0.00	22.37**
Swear Words	-0.13	0.05	-0.01	0.06	22.55**
Health and Behavioral States					
Illness	-0.08	-0.02	0.08	0.02	22.32**
Wellness	-0.11	0.06	0.03	-0.04	20.39*
Food	-0.10	0.05	-0.03	0.04	23.32**
Want	0.26	-0.08	-0.00	-0.19	27.23***
Need	-0.05	0.00	-0.01	0.07	23.73**
Lack	-0.11	0.01	0.10	-0.01	47.51***
Fulfilled	-0.06	0.03	0.00	0.01	31.64***
Motives					
Reward	-0.22	0.06	-0.02	0.15	72.16***
Risk	-0.02	-0.02	0.06	0.00	52.25***
Allure	0.21	-0.05	-0.03	-0.13	21.15*
Perception					
Motion	-0.33	0.02	0.16	0.15	103.39***
Space	-0.45	0.02	0.17	0.27	144.36***
Feeling	-0.11	0.03	0.01	0.05	39.41***
Temporal References					
Past focus	-0.13	-0.04	0.11	0.08	27.45***
Present focus	0.24	0.00	-0.10	-0.18	38.19***

understanding people's social attention is an important factor in effective social interactions [31], which can be an essential component when addressing workers' needs.

6 Discussion

Rapid advancement of digital technologies has led to profound changes in the labor market and has impacted the meaning of work(places) [46, 53, 108]. In addition to technological advancement, ongoing crises such as the global COVID-19 pandemic have altered employment and the way we work [152]. As such, there has been a growing interest among CSCW scholars to explore the impact of these continuous changes on occupations, employment, and workers [92, 104, 164, 169]. As highlighted earlier in this paper, most existing research uses qualitative methods and data collected through surveys and questionnaires to study the well-being of information workers in the evolving labor market. Our work is among the first to shed light on the importance of data from online social networks to understand the challenges and well-being of workers in the knowledge economy. Here, we briefly cover our findings and discuss the contributions of our work.

In response to our first research question, we found that information workers have a high likelihood of disclosing information about Education, Learning/Skill Development, Job Application/Job Search, and Employment/Job Concerns on online social networks. In the second research question, we showed clear differences in the self-disclosure of workers from various genders and age groups. Looking more closely at the gender differences, we discovered that women workers were more expressive of their psychological well-being using keywords such as "anxiety", "anger", and "sadness". Furthermore, we found that social and personal concerns vary between gender groups, where women workers focused on social connections, e.g., family, friends, and home. In contrast, men focused on their careers and growth opportunities. We also found clear linguistic differences between age groups. We observed a high prevalence of keywords that expressed a positive tone among workers under the age of 18 and a high usage of keywords related to negative emotions in the self-disclosed content of workers between the ages of 18 and 25. In addition, we observed the highest prevalence of keywords related to "family" and "communication" for workers under the age of 18, "achievement", "work", and "home" for workers between the ages of 25 and 30, and "power", "interpersonal conflicts", and "politics" for workers above the age of 30. Although we used self-disclosed information on social media, our findings are consistent with prior qualitative research that showed women express more negative affect compared to men, and younger adults are more at risk of depleted well-being compared to older adults [77, 157].

In the following subsections, we further contextualize our work in the body of the existing research. We discuss how the Research communities can benefit from the subjective disclosure of well-being reflected on social networks at scale to study the labor market and information workers. We also explain how our work contributes to designing effective support systems for information workers.

6.1 Online Communities as a Lens to Understand Information Workers' Needs

In the fast-evolving labor market, the significant role of social media in workers' lives is highlighted more than ever. CSCW scholars showed that fast occupational changes, such as the emergence of new skills and jobs, have led individuals to rely heavily on their personal networks and online communities for skill and professional development [82, 83, 99]. Aligned with prior research, our findings revealed that Skill Development is the highest discussed topic among all information workers. Interestingly, we found that whether learning was being discussed directly or indirectly, it remained an important theme across all labor market topics. For example, workers asked online experts about what skills to learn and where to learn them to remain relevant in the knowledge economy. We also showed that when workers are trying to enter the job market or change careers, they consider skill upgrades beneficial. By providing examples from user-generated posts, we also demonstrated how a lack of learning opportunities could lead to employee dissatisfaction and, eventually, employee turnover. Therefore, motivated by the existing literature that shed light on the importance of user-generated social content for gaining insight into the future job market changes, hiring, and talent acquisition [73, 79, 96, 97], we propose online social networks as a reliable source to proactively determine the demands and challenges of workers.

Beyond the reflection of workers' voices on social media, our results revealed the adequacy of online communities for understanding workers' psychological well-being. We showed that when workers discuss topics such as learning and education, they express the most negative affect, i.e., anger, anxiety, and sadness. The prevalence of these emotions on online social networks can be explained by the unique characteristics of social media. Social media enables users to form diverse relationships, disclose self-initiated content, and express their emotions as they experience them. For instance, on social media sites, such as X (formerly Twitter) and Reddit, individuals with differing ages, genders, ethnic origins, and social power create virtual communities based on weak ties to engage with one another [178]. Because of these weak relationships, users can openly exchange their experiences and express their emotions and concerns regarding a variety of topics, such as unemployment, which they might not freely disclose with their intimate relationships [52]. Moreover, many people, particularly young people, rely on social media to express their psychological well-being and get health-related information due to the possibility of generating content anonymously on various social media platforms [117]. On the one hand, the inexpensive, near real-time, and unobtrusive social media data can expand our understanding of workers' needs and psychological well-being beyond conventional processes, e.g., traditional surveys and administrative sources [95, 137]. In contrast, other ethnographic tools such as surveys have low response rates [137] and are conducted on self-selected/self-reporting participants often with lower sample sizes [121]; therefore, making social data a valuable source of information. On the other hand, Reddit users might view this type of data collection to be a form of surveillance and choose to circumvent the technology by moving to private forums as Chinese citizens did after the implementation of the XueXi QiangGuo government platform [93]. Thus, informing impacted members and communities of the analysis conducted and incorporating an opt-in/opt-out consent mechanism for individuals or perhaps forums is warranted before risking the invasion of user privacy.

6.2 Designing Effective Support for Information Workers

With the rise of a skilled labor shortage in Science, Technology, Engineering, and Mathematical (STEM) fields [45], the priority of many information technology organizations has been to enhance employee experience and increase retention. Previous studies argue that factors such as providing growth opportunities, engaging with meaningful work, and ensuring work-life balance can prevent attrition more than high compensation for many information workers [18]. As such, more institutions have invested in employee training and upskilling to prevent employee attrition, improve productivity, and keep workers up-to-date [100]. There is also increasing investment from research communities, such as CSCW and HCI, to explore workers' needs to design effective support [37, 96, 107, 127]. Building on these endeavors, we showed how workers' challenges in the job market are associated with their backgrounds and life stages. For example, in Section 5.2.2 we indicated that workers above the age of 30 are focused on a career change and consider future financial goals, such as pension and retirement, more than any other age groups. On the other hand, younger workers' focus is on entering the job market and career development. Also, we demonstrated that workers' needs, especially women's, are more connected with their social affinities compared to men's. Therefore, there is a need to focus on creating environments that support workers in all stages of their careers with various backgrounds and life experiences/expectations.

The importance of designing person-centered support is also highlighted in prior work [171]. For instance, to support the older workers' motivation to switch jobs, organizations can design rotation programs with predefined learning objectives, tasks, and outcomes which can lead to acquiring new/broader skills, finding jobs that are closer to employees' interests, increasing their income by switching to a better-paying role within the company, and improving employee satisfaction [40, 74, 154]. Lifelong learning possibilities through personalized training, seminars, and courses based on employees' abilities, interests, and career objectives are another type of assistance that organizations may offer to help younger workers improve and flourish in their professions [87, 116].

The importance of social support in workplaces and introducing interventions that help workers in the evolving labor market cannot be overstated. As mentioned by [141], workers who receive more social support at work experience lower anxiety and therefore have less depleted well-being compared to workers who are not supported by their managers, supervisors, or co-workers. [49] argue that representatives at workplaces, e.g., unions, who act as the voice of employees, raise their concerns, and defend their rights, can decrease sadness and depression and, overall, lead to enhanced well-being and happiness. Prior research also shed light on the importance of labor market policies on individuals' well-being in the new world of work [105]. For example, during the COVID-19 pandemic, some countries, such as Canada and the USA, provided incentives to families and their children to protect them from the fear of job loss and anxiety around unemployment [23]. Considering all the changes that work and employment have been undergoing and the important role of organizations and policies on workers' well-being, now is an important time for the CSCW community to systematically understand the factors that influence employees' well-being and play a role in addressing employees' needs under these new working conditions. While being mindful of the limitations of our approach as discussed in Section6.5, we hope policymakers can leverage our proposed approach to identifying IT workers' requirements, considering their demographics to proactively design influential support mechanisms. Similarly, tailored support systems can be offered to unemployed individuals with respect to their personal concerns and background as opposed to providing them with a *"one-size-fits-all"* approach [94].

6.3 Implications of Automatic Gender Retrieval

In this study, we included gender to highlight that interpersonal differences between individuals should be considered when studying the needs and well-being of workers to design more effective support systems. Yet gender identity is a social construct and people's gender experiences are constantly evolving [146]. However, our extracted data from Reddit did not contain a dedicated field for gender expressed by the users themselves. One of the common procedures for including self-identified genders in research is through user input, which can be obtained through surveys and interviews [143]. Nevertheless, applying those approaches is not feasible when conducting research on large-scale social media data, i.e., online social networks. With technological advancements and the widespread use of natural language processing (NLP) and machine learning, researchers have increasingly focused on detecting an individual's gender from social data [3, 14, 159], a similar approach that we inherited in this work. While this development offers potential for a variety of applications, it also raises significant implications and ethical concerns that merit a closer examination.

One of the important concerns with automatic gender detection algorithms is the potential for algorithmic bias and stereotyping [89, 145]. Algorithms trained on existing text data might inadvertently reinforce gender stereotypes, as they learn from historical language usage patterns that reflect societal biases [149]. This can lead to unfair categorizations, discrimination, or exclusion of individuals who do not conform to traditional gender norms. To identify the existence of such biases in our gender detection algorithm, we conducted a qualitative error analysis technique on 1% (241) of the posts where the author explicitly mentioned their gender, i.e., "I am a female" or "I am a man". By comparing the text and the automatic gender labels assigned by the algorithm we witnessed interesting cases. For example, we noticed that in posts where the author self-identified themselves as a female trans person, the algorithm correctly categorized them in the "Female" group. For example:

- Is this job feasible for a trans person (trans female, male to female)?
- I am 31, a trans woman from South America.
- I am a 32-year-old, queer, middle eastern woman seeking a mentor.

In other cases where the gender was falsely identified as a "Female" or "Male", we noticed that the posts contained many "male" or "female" references or were describing experiences of a "female" or "male" subject. For example, one of the authors dedicated their post to the 22 most successful male workers in the industry and a few more lines about their biography. Although the mentioned example did not contain any self-identified gender of the authors, the algorithm labeled the post as "Male", potentially due to the high usage of "male" references. In another post, a worker described how they are overworked in their current position and mentioned that they are looking for a job to be able to spend more time with their wife. In this example, although the post was missing the self-expressed gender of the author, this post was categorized into the "Male" group, potentially because of the existence of the word "wife" in the text. While the correct identification of trans-gender authors by the algorithm was heartwarming and the overall performance of the model, as reported in Table 6, was satisfactory compared to previous literature, we acknowledge the potential harm of analysis on falsely classified data because of the societal stereotypes.

Moreover, consideration of gender as a binary construct oversimplifies the rich spectrum of gender identities that exist and can result in a lack of inclusivity, alienating individuals who identify as non-binary, genderqueer, or trans, thus failing to recognize their authentic identities [17]. Also, in many contexts, a binary view of gender reinforces stereotypical gender roles and expectations, limiting individual freedom, expression, and opportunities. As society becomes more aware of the limitations and injustices tied to a binary understanding of gender, there is a growing movement to recognize and respect a wider spectrum of gender identities [170]. However, there is still a wide gap between the research community's efforts to acknowledge nuances in gender identities and the existing algorithmic approaches and guidelines that allow social computing researchers to identify gender. We recognize that significant further work must be done to accommodate a more comprehensive and inclusive range of gender identities [91] and to prevent unwanted harm to marginalized populations. In the context of our work, one possible way to include non-binary individuals would be conducting research on non-binary advocacy groups, such as "r/NonBinary" subreddit on the Reddit platform. These individuals and groups frequently exchange information, resources, and personal stories about non-binary experiences. Another way to include non-binary individuals in social media research is obtaining the pronouns and terminologies being used by non-binary groups, such as the informal term "enby", which is frequently used to describe people who identify as non-binary genders [72]. While such efforts require further expertise, for the current study, we reluctantly categorized users into binary groups based on self-identification of women or men. We discuss the ethical considerations of our decision. [144] argues that "If gender must be used, consider the context of its application". Women have reported experiencing gender discrimination in the U.S. labor force and have had higher prevalence rates of gender discrimination in male-dominated fields [21]. We felt that, despite not having adequate means of categorizing non-women or men identities, understanding whether and how these inequities manifest themselves in online forums is an essential initial investigation.

6.4 Ethical Considerations

In this data-driven age, the use of open data in research has become increasingly prevalent. Publicly accessible data sources, including social media posts and open-access scientific databases, provide researchers with a wealth of information for investigations across various disciplines [111]. While it is clear that having access to such data has broadened research possibilities, scholars argue that the blurred line between public and private data can lead to potential harm and negative implications for individuals [119]. For instance, although some online social platforms, such as X (formerly Twitter), warn users about the usage of public data by researchers under the Privacy Policy section, some online research subjects are unaware that their data is being used by academics [48]. Also, usually, online users do not have any control over the type of data that is collected for research purposes [162]. Additionally, users may have written identifiable private information in public data as a means of seeking social support during a crisis; hence, failing to restrict the revealing of people's identities in research may endanger users [153].

To this end, while this study primarily utilizes public social media data to help organizations better understand the information workers' challenges and needs, especially those who belong to marginalized groups, such as women, we recognize that some users may consider social media analytics to constitute surveillance. To limit the unwanted harm to the users, we made sure to anonymize any quotations that could trace the content to the users. Also, we did not keep any usernames or names of the users in our dataset to protect the privacy of the users. However, we acknowledge that while individuals might decide to share information about their career challenges or overall well-being in a workplace in a public forum, they might be reluctant to provide their employers with their online social content [137]. One possible way to overcome this challenge is to engage with the social platform under study, i.e., forums administrative, to inform users about the collection of their data and provide them with an opportunity to opt out of the research.

6.5 Limitations and Future Directions

We acknowledge that our research is not without limitations. First, while online social media data can assist in capturing and analyzing timely labor market issues at a large scale, previous research has emphasized *selection bias* as a potential problem in social media studies [137]. Selection bias may occur because of over or under-representation of individuals with particular demographics such as race, age, gender, level of education, and income on different social media platforms [138]. For instance, Pew Research found that Reddit's audience is skewed toward younger (65% of users between the ages of 18-29) and men (67%) with a college degree[7]. Also, our dataset may likely lean toward the viewpoints of those who reside in the Global North and belong to a certain ethnic group. Therefore, we note that the self-disclosed content in our Reddit data may not necessarily be representative of a diverse labor market. Moreover, we only analyzed English Reddit data, and hence our work is not representative of the rich lived experiences of those information workers who communicate in languages other than English. To address these limitations, further studies are required to assess information workers' concerns and well-being using data from various social media platforms to paint a more complete picture.

We also acknowledge that more factors beyond gender and age can influence the challenges that workers experience in the job market. For example, socioeconomic status, education level, marital status, family size, personality traits, and lifestyles such as alcohol consumption, smoking, exercising, or prior health issues, and disabilities, among others, can also have an impact on the reported challenges and worker well-being [35, 36, 38, 98]. Despite limitations, we hope that our work inspires future research to examine the societal factors that impact the experience and emotions of information workers across broader demographic groups.

7 Concluding Remarks

This paper leverages social media data to identify the challenges and the psychological well-being of information workers in the evolving labor market. Using machine learning techniques, we identified Skill Development, Job Applications/Job Search, Employment/Job Complaints, and Education as the most discussed topics among information workers. We also looked at linguistic variations among different genders and age groups in online communities. Concerning gender, we found that, on average, women workers report more psychological issues and negative affect than men. While men express more concerns about their jobs, job titles, education, and progress, women are more likely to express concerns about their health, families, and interpersonal connections. For age, we found that younger workers rely on online communities to seek career advice, while older workers disclose/require information about career changes. We also

[7]https://www.alphr.com/demographics-reddit/

found high uncertainty in all labor market topics and a positive affect regarding learning topics among youth workers (under 18) and a higher negative affect among workers between the ages of 18 and 25 who discussed the Education topic. This paper is among the first to use social media data, statistical methods, and machine learning techniques to shed light on information workers' challenges and the areas in which workers require the most support, and it advances our understanding of workers' well-being in the emerging labor market, considering their age and gender.

Appendix

A Data Collection Strategy

In this section, we discuss our data collection strategy and explain our choice to selectively curate posts explicitly referencing IT skills from various subreddits instead of opting for an all-encompassing approach by collecting all posts solely from a dedicated IT subreddit. First and foremost, opting to select all posts from a dedicated IT subreddit introduces a potential source of selection bias. This bias arises from the inherent differences in tone, language, and context across various subreddits [167]. Each subreddit develops its unique community norms, communication styles, and jargon, which can significantly affect the content shared within it. For example, in a study conducted by [113], the authors explored the commonalities and distinctions in content discussed by Reddit users across three mental health-related subreddits. Employing an unsupervised technique, namely K-means, the authors found that while a few themes like expressions of gratitude and positive emotions were shared among the subreddits, the majority of topics were unique to each subreddit. Therefore, by exclusively focusing on particular IT-dedicated subreddits, there is a risk of overlooking valuable insights and discussions occurring in other communities where IT skills are discussed in a context-specific manner. Thus, this selection approach risks favoring posts that align with the tone and language of the IT subreddit while excluding potentially relevant contributions from the broader Reddit community.

Furthermore, our quantitative analysis of the subreddit titles and their content showed that in addition to communities dedicated to particular skills, i.e., “r/Javascript”, “r/Programming”, and “r/PostgreSQL”, by applying our data collection approach explained in Section 3, we were also able to include a wider range of subreddits, such as “r/InformationTechnology”, and “r/Redditdev”, where more general IT-related topics were discussed. We analyzed all the posts that were posted within a 30-day period in our dataset for each subreddit and computed the frequency of the posts that contained one of our IT skills in their body or title. We found that while the majority of the posts in skill-related subreddits contained one of the IT skills in the posts, i.e., 64% in “r/CSS”, 75% in “r/SQL”, and 93% in “r/docker”, a noticeable portion of the posts in more general IT-related subreddits also contained at least one of our IT-skills. For example, 33% of the posts in “r/computerscience”, 30% in “r/ComputerEngineering”, and 29% in “r/TechInsightreports” mentioned one of our IT skills in them. This observation demonstrates that while by focusing on the posts based on a list of IT skills, we might overlook some of the issues that IT workers discuss without mentioning IT skill terms, our data collection method was able to capture a more comprehensive and diverse range of relevant discussions that may not surface within the confines of a specialized IT-focused subreddit. This approach allows for a richer and more holistic understanding of how IT skills are employed and discussed across Reddit’s vast and multifaceted landscape of communities.

B Codebook Validity (Extended Analysis with Additional Data)

To investigate whether IT workers’ challenges remain the same or change over time, we obtained additional Reddit posts from Pushshift, posted between March 2020 and December 2021. Based on the additional data, we followed the same data pre-processing steps explained in Section 3 to prepare the collected data for further analysis. As a result, we collected an additional 459,156 unique posts in 516 subreddits, posted by 259,286 unique users. To identify the topics discussed in Reddit posts, as discussed in Section 4.1, we created a codebook by employing an inductive coding technique. To find the conceptual categories in the new data, we randomly sampled 1,000 Reddit posts out of 459,156 to develop the codebook. To avoid potential biases, we recruited and trained three coders familiar with labor market topics who were not involved in the original coding process to individually code each sample in the 1,000 posts. After an iterative process in which the three coders met regularly to form the conceptual categories, the coders verified the obtained codebook. We found that 83.5% of the identified codes were completely overlapping with the findings reported in the results in this paper. The list of additional unique codes (remaining non-overlapping 17% of codes) and a detailed description of them can be found in Table 13. Please note that in Table 13, we only reported the new codes that the coders identified in the new data that were not mentioned in Table 2.

By analyzing the frequency of the posts that were categorized under each topic in Table 13, we noticed that 6% of the data belonged to the “Community Building”, 4.7% to the “Community Appreciation”, and only 5.8% of the posts were coded into “Anecdote”, “Advertisement”, “General Curiosity”, “App Promotion and User Engagement”, and “Investment and Financial Analysis” topics. However, it is important to note that 83.5% of the posts were labeled with the codes in Table 2. This observation demonstrated that while some of the challenges of IT workers might evolve or new ones might arise over time, their main concerns around “Skill Development”, “Job Applications/Job Search”, “Education”, and “Employment/Job Concerns” remained the same over time.

Table 13. Codebook of new topics derived from Reddit posts extracted between March 2020 and December 2021 that were different from those observed between January 2015 and August 2019. We note that 83.5% of the data had the same codes.

Topic	Definition
Community Appreciation	Demonstration of gratitude for support during the difficult COVID-19 era.
Community Building	Reaching out to individuals of common interests.
Anecdote	User shares anecdotal lived experience(s) of a certain circumstance/challenge.
Advertisement	An ad or coupon code promoting traffic to a service or product.
General Curiosity	User demonstrates a general curiosity for the topic and/or its contextual relevance.
App Promotion and User Engagement	A set of strategies, activities, and techniques used by app developers and marketers to increase the visibility and usage of a mobile application (app) among its target audience. This involves promoting the app to attract new users and engaging existing users to ensure they continue using the app regularly.
Investment and Financial Analysis	Refers to the process of evaluating the financial performance, viability, and potential risks and rewards associated with investments. It involves using various financial metrics, techniques, and models to assess the attractiveness of an investment opportunity, make informed investment decisions, and manage financial assets effectively.

C Performance Comparison of Machine Learning Classifiers

In this section, we provide a performance comparison of the five trained machine learning models for the topic classification task. This comparison can be found in Table 14.

Table 14. Comparison of the performances of the 5 trained machine learning models for the topic classification task. The overall accuracy for all the topics is reported at the end of the table. The outperforming model based on F1-score is XLNet.

Topic	Classifier	Precision	Recall	F1-Score
Job Applications/Job Search	BERT (base uncased)	0.76	0.77	0.76
	XLNet	**0.77**	**0.78**	**0.77**
	Logistic Regression	0.75	0.68	0.72
	LinearSVC	0.74	0.68	0.71
	Random Forest	0.77	0.18	0.29
Skill Development	BERT (base uncased)	0.82	0.84	0.83
	XLNet	**0.84**	**0.85**	**0.85**
	Logistic Regression	0.81	0.81	0.81
	LinearSVC	0.83	0.78	0.80
	Random Forest	0.57	0.85	0.68
Employment/Job Concerns	BERT (base uncased)	0.73	0.74	0.73
	XLNet	**0.73**	**0.83**	**0.78**
	Logistic Regression	0.66	0.73	0.70
	LinearSVC	0.64	0.71	0.67
	Random Forest	0.71	0.06	0.12
Education	BERT (base uncased)	0.77	0.78	0.77
	XLNet	**0.79**	**0.79**	**0.79**
	Logistic Regression	0.73	0.75	0.74
	LinearSVC	0.70	0.75	0.73
	Random Forest	0.82	0.13	0.23
Other	BERT (base uncased)	0.90	0.87	0.88
	XLNet	**0.91**	**0.87**	**0.89**
	Logistic Regression	0.87	0.86	0.87
	LinearSVC	0.84	0.88	0.86
	Random Forest	0.77	0.88	0.82
Mean	BERT (base uncased)	0.834	0.832	0.833
	XLNet	**0.848**	**0.845**	**0.846**
	Logistic Regression	0.807	0.806	0.806
	LinearSVC	0.801	0.797	0.797
	Random Forest	0.698	0.667	0.602

References

[1] Abdullah M Abu Nada, Eman Alajrami, Ahmed A Al-Saqqa, and Samy S Abu-Naser. 2020. Age and gender prediction and validation through single user images using cnn. (2020).

[2] Iffat Ali Aksar, Mahmoud Danaee, Amira Firdaus, and Jiankun Gong. 2022. Moderating effect of gender between social media needs and psychological wellbeing: a case study of Pakistani adolescents. *Journal of Human Behavior in the Social Environment* 32, 3 (2022), 383–401.

[3] Kristen M Altenburger, Rajlakshmi De, Kaylyn Frazier, Nikolai Avteniev, and Jim Hamilton. 2017. Are there gender differences in professional self-promotion? an empirical case study of linkedin profiles among recent mba graduates. In *Eleventh International AAAI Conference on Web and Social Media*.

[4] Miguel Angel Alvarez-Mon, Cesar Ignacio Fernandez-Lazaro, Cristina Vidal, Rosa M Molina-Ruiz, Miguel A Ortega, Melchor Alvarez De Mon, and Miguel Ángel Martinez-Gonzalez. 2022. Analyzing Psychotherapy on Twitter: An 11-Year Analysis of Tweets From Major US Media Outlets. *Frontiers in Psychiatry* (2022), 924.

[5] Ashley Amaya, Ruben Bach, Florian Keusch, and Frauke Kreuter. 2019. New data sources in social science research: things to know before working with reddit data. *Social science computer review* (2019), 0894439319893305.

[6] Nazanin Andalibi, Pinar Ozturk, and Andrea Forte. 2017. Sensitive Self-disclosures, Responses, and Social Support on Instagram: the case of# depression. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 1485–1500.

[7] Ashutosh Anshul, Gumpili Sai Pranav, Mohammad Zia Ur Rehman, and Nagendra Kumar. 2023. A multimodal framework for depression detection during COVID-19 via harvesting social media. *IEEE Transactions on Computational Social Systems* 11, 2 (2023), 2872–2888.

[8] Richa Aryan and Deepika Kathuria. 2017. Psychological wellbeing at workplace:-An analytical study on IT sector. *International Journals of Advanced Research in Computer Science and Software Engineering* 7, 6 (2017), 223–228.

[9] Cassondra Batz-Barbarich, Louis Tay, Lauren Kuykendall, and Ho Kwan Cheung. 2018. A meta-analysis of gender differences in subjective well-being: Estimating effect sizes and associations with gender inequality. *Psychological science* 29, 9 (2018), 1491–1503.

[10] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The pushshift reddit dataset. In *Proceedings of the international AAAI conference on web and social media*, Vol. 14. 830–839.

[11] Miroslav Beblavỳ, Mehtap Akgüc, Brian Fabo, and Karolien Lenaerts. 2016. What are the new occupations and the new skills? And how are they measured.

[12] Mahnaz Behroozi, Shivani Shirolkar, Titus Barik, and Chris Parnin. 2020. Does stress impact technical interview performance?. In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 481–492.

[13] Naiara Berasategi Santxo, Nahia Idoiaga Mondragon, Naiara Ozamiz-Etxebarria, and Maria Dosil-Santamaria. 2021. Well-Being of Adolescents in De-Escalation Situation: Physical, Emotional, Social, and Academic Impact. *Frontiers in Psychology* 12 (2021), 646027.

[14] Amin Bigdeli, Negar Arabzadeh, Morteza Zihayat, and Ebrahim Bagheri. 2021. Exploring gender biases in information retrieval relevance judgement datasets. In *European Conference on Information Retrieval*. Springer, 216–224.

[15] Ryan L Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W Pennebaker. 2022. The development and psychometric properties of LIWC-22. *Austin, TX: University of Texas at Austin* (2022).

[16] Jenna Butler and Sonia Jaffe. 2021. Challenges and gratitude: A diary study of software engineers working from home during covid-19 pandemic. In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. IEEE, 362–363.

[17] Jessica J Cameron and Danu Anthony Stinson. 2019. Gender (mis) measurement: Guidelines for respecting gender diversity in psychological research. *Social and personality psychology compass* 13, 11 (2019), e12506.

[18] Karen Carleton. 2011. How to motivate and retain knowledge workers in organizations: A review of the literature. *International Journal of Management* 28, 2 (2011), 459.

[19] Kathy Charmaz and Linda Liska Belgrave. 2007. Grounded theory. *The Blackwell encyclopedia of sociology* (2007).

[20] Shreya Chowdhary, Anna Kawakami, Mary L Gray, Jina Suh, Alexandra Olteanu, and Koustuv Saha. 2023. Can workers meaningfully consent to workplace wellbeing technologies?. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 569–582.

[21] Adrienne Colella and Eden B King. 2018. *The Oxford handbook of workplace discrimination*. Oxford University Press.

[22] Shanley Corvite, Kat Roemmich, Tillie Ilana Rosenberg, and Nazanin Andalibi. 2023. Data subjects' perspectives on emotion artificial intelligence use in the workplace: A relational ethics lens. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–38.

[23] Maria Cotofan, Jan-Emmanuel De Neve, Marta Golin, Micah Kaats, and George Ward. 2021. Work and well-being during COVID-19: Impact, inequalities, resilience, and the future of work. *World Happiness Report* (2021), 153–190.

[24] Vedant Das Swain, Lan Gao, Abhirup Mondal, Gregory D Abowd, and Munmun De Choudhury. 2024. Sensible and sensitive AI for worker wellbeing: factors that inform adoption and resistance for information workers. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–30.

[25] Vedant Das Swain, Lan Gao, William A Wood, Srikruthi C Matli, Gregory D Abowd, and Munmun De Choudhury. 2023. Algorithmic power or punishment: Information worker perspectives on passive sensing enabled ai phenotyping of performance and wellbeing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[26] Vedant Das Swain, Javier Hernandez, Brian Houck, Koustuv Saha, Jina Suh, Ahad Chaudhry, Tenny Cho, Wendy Guo, Shamsi Iqbal, and Mary P Czerwinski. 2023. Focused time saves nine: Evaluating computer–assisted protected time for hybrid information work. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.

[27] Vedant Das Swain and Koustuv Saha. 2024. Teacher, trainer, counsel, spy: How generative AI can bridge or widen the gaps in worker-centric digital phenotyping of Wellbeing. In *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 1–13.

[28] Vedant Das Swain, Koustuv Saha, Hemang Rajvanshy, Anusha Sirigiri, Julie M Gregg, Suwen Lin, Gonzalo J Martinez, Stephen M Mattingly, Shayan Mirjafari, Raghu Mulukutla, et al. 2019. A multisensor person-centered approach to understand the role of daily activities in job performance with organizational personas. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–27.

[29] 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*.

[30] Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2098–2110.

[31] Martina De Lillo, Rebecca Foley, Matthew C Fysh, Aimée Stimson, Elisabeth EF Bradford, Camilla Woodrow-Hill, and Heather J Ferguson. 2021. Tracking developmental differences in real-world social attention across adolescence, young adulthood and older adulthood. *Nature human behaviour* 5, 10 (2021), 1381–1390.

[32] Jan-Emmanuel De Neve, Ed Diener, Louis Tay, and Cody Xuereb. 2013. The objective benefits of subjective well-being. *World happiness report* (2013).

[33] 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).

[34] Claudio Zaki Dib. 1988. Formal, non-formal and informal education: concepts/applicability. In *AIP conference proceedings*, Vol. 173. American Institute of Physics, 300–315.

[35] Tawanna R Dillahunt, Matthew Garvin, Marcy Held, and Julie Hui. 2021. Implications for Supporting Marginalized Job Seekers: Lessons from Employment Centers. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–24.

[36] Tawanna R Dillahunt, Aarti Israni, Alex Jiahong Lu, Mingzhi Cai, and Joey Chiao-Yin Hsiao. 2021. Examining the Use of Online Platforms for Employment: A Survey of US Job Seekers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–23.

[37] Tawanna R Dillahunt and Alex Lu. 2019. DreamGigs: Designing a Tool to Empower Low-resource Job Seekers. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.

[38] Tawanna R Dillahunt, Alex J Lu, and Afton Branche-Wilson. 2021. The Promises and Perils of Technology for Marginalized Job Seekers. *UM Poverty Solutions* (2021).

[39] Tawanna R Dillahunt, Alex Jiahong Lu, Aarti Israni, Ruchita Lodha, Savana Brewer, Tiera S Robinson, Angela Brown Wilson, and Earnest Wheeler. 2022. The Village: Infrastructuring Community-based Mentoring to Support Adults Experiencing Poverty. In *CHI Conference on Human Factors in Computing Systems*. 1–17.

[40] Indi Djastuti. 2015. The influence of job characteristics on job satisfaction, organizational commitment and managerial performance a study on construction companies in Central Java. *International Research Journal of Business Studies* 3, 2 (2015).

[41] Sarmistha Dutta, Jennifer Ma, and Munmun De Choudhury. 2018. Measuring the Impact of Anxiety on Online Social Interactions.. In *ICWSM*. 584–587.

[42] Jacob Eisenstein, Amr Ahmed, and Eric P Xing. 2011. Sparse additive generative models of text. In *Proceedings of the 28th international conference on machine learning (ICML-11)*. 1041–1048.

[43] Mai ElSherief, Vivek Kulkarni, Dana Nguyen, William Yang Wang, and Elizabeth Belding. 2018. Hate lingo: A target-based linguistic analysis of hate speech in social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12.

[44] Sindhu Kiranmai Ernala, Tristan Labetoulle, Fred Bane, Michael L Birnbaum, Asra F Rizvi, John M Kane, and Munmun De Choudhury. 2018. Characterizing Audience Engagement and Assessing Its Impact on Social Media Disclosures of Mental Illnesses.. In *ICWSM*. 62–71.

[45] Elaine Farndale, Mohan Thite, Pawan Budhwar, and Bora Kwon. 2021. Deglobalization and talent sourcing: Cross-national evidence from high-tech firms. *Human Resource Management* 60, 2 (2021), 259–272.

[46] Anne-Laure Fayard. 2021. Notes on the meaning of work: Labor, work, and action in the 21st century. *Journal of Management Inquiry* 30, 2 (2021), 207–220.

[47] John Field. 2011. Lifelong learning. *Adult learning and education* (2011), 20–28.

[48] Casey Fiesler and Nicholas Proferes. 2018. "Participant" perceptions of Twitter research ethics. *Social Media+ Society* 4, 1 (2018), 2056305118763366.

[49] Richard Freeman, David Blanchflower, and Alex Bryson. 2020. Unions raise worker wellbeing. *VoxEU. org* 11 (2020).

[50] Kausalya Ganesh and Amanda Lazar. 2021. The Work of Workplace Disclosure: Invisible Chronic Conditions and Opportunities for Design. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–26.

[51] Roberto Garcia-Guzman, Yair A Andrade-Ambriz, Mario-Alberto Ibarra-Manzano, Sergio Ledesma, Juan Carlos Gomez, and Dora-Luz Almanza-Ojeda. 2020. Trend-based categories recommendations and age-gender prediction for pinterest and twitter users. *Applied Sciences* 10, 17 (2020), 5957.

[52] Radhika Garg, Yash Kapadia, and Subhasree Sengupta. 2021. Using the Lenses of Emotion and Support to Understand Unemployment Discourse on Reddit. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–24.

[53] Camilla Gjellebæk, Ann Svensson, and Catharina Bjørkquist. 2020. The Dark Sides of Technology-Barriers to Work-Integrated Learning. In *International Conference on Human-Computer Interaction*. Springer, 69–85.

[54] Kristina Gligorić, Ashton Anderson, and Robert West. 2019. Causal effects of brevity on style and success in social media. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–23.

[55] Crystal Gong, Koustuv Saha, and Stevie Chancellor. 2021. " The Smartest Decision for My Future": Social Media Reveals Challenges and Stress During Post-College Life Transition. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–29.

[56] Steven M Graham, Julie Y Huang, Margaret S Clark, and Vicki S Helgeson. 2008. The positives of negative emotions: Willingness to express negative emotions promotes relationships. *Personality and Social Psychology Bulletin* 34, 3 (2008), 394–406.

[57] Esther R Greenglass, Ronald J Burke, and Roman Konarski. 1997. The impact of social support on the development of burnout in teachers: Examination of a model. *Work & Stress* 11, 3 (1997), 267–278.

[58] 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.

[59] Zhihua Guo, Nengneng Ding, Minyu Zhai, Zhenwen Zhang, and Zepeng Li. 2023. Leveraging domain knowledge to improve depression detection on Chinese social media. *IEEE Transactions on Computational Social Systems* 10, 4 (2023), 1528–1536.

[60] John A Hartigan and Manchek A Wong. 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)* 28, 1 (1979), 100–108.

[61] Khan Md Hasib, Md Rafiqul Islam, Shadman Sakib, Md Ali Akbar, Imran Razzak, and Mohammad Shafiul Alam. 2023. Depression detection from social networks data based on machine learning and deep learning techniques: An interrogative survey. *IEEE Transactions on Computational Social Systems* 10, 4 (2023), 1568–1586.

[62] Tianran Hu, Haoyuan Xiao, Jiebo Luo, and Thuy-vy Thi Nguyen. 2016. What the language you tweet says about your occupation. In *Tenth International AAAI Conference on Web and Social Media*.

[63] Julie Hui, Nefer Ra Barber, Wendy Casey, Suzanne Cleage, Danny C Dolley, Frances Worthy, Kentaro Toyama, and Tawanna R Dillahunt. 2020. Community collectives: Low-tech social support for digitally-engaged entrepreneurship. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–15.

[64] Julie S Hui, Matthew W Easterday, and Elizabeth M Gerber. 2019. Distributed apprenticeship in online communities. *Human–Computer Interaction* 34, 4 (2019), 328–378.

[65] Loukas Ilias and Dimitris Askounis. 2023. Multitask learning for recognizing stress and depression in social media. *Online Social Networks and Media* 37 (2023), 100270.

[66] Loukas Ilias, Spiros Mouzakitis, and Dimitris Askounis. 2023. Calibration of transformer-based models for identifying stress and depression in social media. *IEEE Transactions on Computational Social Systems* 11, 2 (2023), 1979–1990.

[67] Kerry Joyce, Roman Pabayo, Julia A Critchley, and Clare Bambra. 2010. Flexible working conditions and their effects on employee health and wellbeing. *Cochrane database of systematic reviews* 2 (2010).

[68] Bridget Juniper. 2011. Defining employee wellbeing. *Occupational Health & Wellbeing* 63, 10 (2011), 25.

[69] Harmanpreet Kaur, Daniel McDuff, Alex C Williams, Jaime Teevan, and Shamsi T Iqbal. 2022. "I Didn't Know I Looked Angry": Characterizing Observed Emotion and Reported Affect at Work. In *CHI Conference on Human Factors in Computing Systems*. 1–18.

[70] Anna Kawakami, Shreya Chowdhary, Shamsi T Iqbal, Q Vera Liao, Alexandra Olteanu, Jina Suh, and Koustuv Saha. 2023. Sensing wellbeing in the workplace, why and for whom? envisioning impacts with organizational stakeholders. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–33.

[71] Marios Kerasiotis, Loukas Ilias, and Dimitris Askounis. 2024. Depression detection in social media posts using transformer-based models and auxiliary features. *Social Network Analysis and Mining* 14, 1 (2024), 196.

[72] Os Keyes. 2018. The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proceedings of the ACM on human-computer interaction* 2, CSCW (2018), 1–22.

[73] Adnan Q Khan and Steven F Lehrer. 2013. The impact of social networks on labour market outcomes: New evidence from cape breton. *Canadian Public Policy* 39, Supplement 1 (2013), S1–S24.

[74] Faisal Khan, Amran Md Rasli, Rosman Md Yusoff, Tariq Ahmed, A Rehman, and Muhammad Muddassir Khan. 2014. Job rotation, job performance, organizational commitment: An empirical study on bank employees. *Journal of Management info* 3, 1 (2014), 33–46.

[75] Chelsea Kim and Hao-Chuan Wang. 2022. From Receivers to Givers: Understanding Practice of Reciprocity in an Online Support Community. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–17.

[76] Jaeseung Kim, Julia R Henly, Lonnie M Golden, and Susan J Lambert. 2020. Workplace flexibility and worker well-being by gender. *Journal of marriage and family* 82, 3 (2020), 892–910.

[77] Johanna Kluge, Julian Hildebrandt, and Martina Ziefle. 2019. The Golden Age of Silver Workers?. In *International Conference on Human-Computer Interaction*. Springer, 520–532.

[78] Mouna Knani, Pierre-Sébastien Fournier, and Caroline Biron. 2018. Psychosocial risks, burnout and intention to quit following the introduction of new software at work. *Work* 60, 1 (2018), 95–104.

[79] Tanja Koch, Charlene Gerber, and Jeremias J De Klerk. 2018. The impact of social media on recruitment: Are you LinkedIn? (2018).

[80] Ellen Ernst Kossek, Thomas Kalliath, and Parveen Kalliath. 2012. Achieving employee wellbeing in a changing work environment: An expert commentary on current scholarship. *International Journal of Manpower* (2012).

[81] Ellen Ernst Kossek, Monique Valcour, and Pamela Lirio. 2014. Organizational strategies for promoting work–life balance and wellbeing. *Work and wellbeing* 3 (2014), 295–319.

[82] Yubo Kou and Colin M Gray. 2018. " What do you recommend a complete beginner like me to practice?" Professional Self-Disclosure in an Online Community. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–24.

[83] Yubo Kou and Colin M Gray. 2018. Towards professionalization in an online community of emerging occupation: Discourses among UX practitioners. In *Proceedings of the 2018 ACM Conference on Supporting Groupwork*. 322–334.

[84] Yubo Kou, Colin M Gray, Austin L Toombs, and Robin S Adams. 2018. Understanding social roles in an online community of volatile practice: A study of user experience practitioners on reddit. *ACM Transactions on Social Computing* 1, 4 (2018), 1–22.

[85] Christian Krekel, George Ward, and Jan-Emmanuel De Neve. 2019. Employee wellbeing, productivity, and firm performance. *Saïd Business School WP* 4 (2019).

[86] Colin Latchem. 2014. Informal learning and non-formal education for development. (2014).

[87] Kaela Leyretana and Jose Eos Trinidad. 2022. Predicting or preventing lifelong learning? The role of employment, time, cost, and prior achievement. *Journal of Adult and Continuing Education* 28, 2 (2022), 658–673.

[88] Tianshi Li, Elizabeth Louie, Laura Dabbish, and Jason I Hong. 2021. How Developers Talk About Personal Data and What It Means for User Privacy: A Case Study of a Developer Forum on Reddit. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–28.

[89] Paul Pu Liang, Chiyu Wu, Louis-Philippe Morency, and Ruslan Salakhutdinov. 2021. Towards understanding and mitigating social biases in language models. In *International Conference on Machine Learning*. PMLR, 6565–6576.

[90] Jasmin Lui Ping Loi and Kenneth S Shultz. 2007. Why older adults seek employment: Differing motivations among subgroups. *Journal of Applied Gerontology* 26, 3 (2007), 274–289.

[91] Judith Lorber. 1993. Believing is seeing: Biology as ideology. *Gender & society* 7, 4 (1993), 568–581.

[92] Alex Jiahong Lu, Anna Gilhool, Joey Chiao-Yin Hsiao, and Tawanna R Dillahunt. 2022. Emotional Labor in Everyday Resilience: Class-Based Experiences of Navigating Unemployment Amid the COVID-19 Pandemic in the US. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–27.

[93] Alex Jiahong Lu and Xuecong Xu. 2020. " Learning for the Rise of China": Exploring Uses and Gratifications of State-Owned Online Platform. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–25.

[94] Susan Lund, James Manyika, Liz Hilton Segel, André Dua, Bryan Hancock, Scott Rutherford, and Brent Macon. 2019. *The future of work in America: People and places, today and tomorrow*. McKinsey Global Institute.

[95] Jalehsadat Mahdavimoghaddam. 2022. The Utility of Social Media in Understanding the Future of Work. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing*. 219–222.

[96] Jalehsadat Mahdavimoghaddam, Ayush Bahuguna, and Ebrahim Bagheri. 2022. Exploring the Utility of Social Content for Understanding Future In-Demand Skills. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–35.

[97] Jalehsadat Mahdavimoghaddam, Niranjan Krishnaswamy, and Ebrahim Bagheri. 2021. On the Congruence Between Online Social Content and Future IT Skill Demand. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–27.

[98] Dawid Majcherek, Arkadiusz Michał Kowalski, and Małgorzata Stefania Lewandowska. 2022. Lifestyle, Demographic and Socio-Economic Determinants of Mental Health Disorders of Employees in the European Countries. *International Journal of Environmental Research and Public Health* 19, 19 (2022), 11913.

[99] Jennifer Marlow and Laura Dabbish. 2014. From rookie to all-star: professional development in a graphic design social networking site. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 922–933.

[100] Ggaliwango Marvin, Majwega Jackson, and Md Golam Rabiul Alam. 2021. A machine learning approach for employee retention prediction. In *2021 IEEE Region 10 Symposium (TENSYMP)*. IEEE, 1–8.

[101] Tamir Mendel and Eran Toch. 2022. Meerkat: A Social Community Support Application for Older Adults. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–4.

[102] Sanjana Mendu, Anna Baglione, Sonia Baee, Congyu Wu, Brandon Ng, Adi Shaked, Gerald Clore, Mehdi Boukhechba, and Laura Barnes. 2020. A framework for understanding the relationship between social media discourse and mental health. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–23.

[103] Phyllis Moen, Erin L Kelly, Wen Fan, Shi-Rong Lee, David Almeida, Ellen Ernst Kossek, and Orfeu M Buxton. 2016. Does a flexibility/support organizational initiative improve high-tech employees' well-being? Evidence from the work, family, and health network. *American Sociological Review* 81, 1 (2016), 134–164.

[104] Phoebe V Moore. 2019. OSH and the future of work: benefits and risks of artificial intelligence tools in workplaces. In *International Conference on Human-Computer Interaction*. Springer, 292–315.

[105] Robson Morgan and Kelsey J O'Connor. 2022. Labor market policy and subjective well-being during the great recession. *Journal of Happiness Studies* 23, 2 (2022), 391–422.

[106] Antonio A Morgan-Lopez, Annice E Kim, Robert F Chew, and Paul Ruddle. 2017. Predicting age groups of Twitter users based on language and metadata features. *PloS one* 12, 8 (2017), e0183537.

[107] Naomi Nagata and Tomofumi Uetake. 2018. An e-Learning System Using Gamification to Support Preliminary Learning for Job Hunting. In *International Conference on Learning and Collaboration Technologies*. Springer, 173–184.

[108] James Ness et al. 2020. Technology in education. In *International Conference on Human-Computer Interaction*. Springer, 574–585.

[109] Matthew L Newman, Carla J Groom, Lori D Handelman, and James W Pennebaker. 2008. Gender differences in language use: An analysis of 14,000 text samples. *Discourse processes* 45, 3 (2008), 211–236.

[110] Karen O'Connor, Su Golder, Davy Weissenbacher, Ari Klein, and Graciela Gonzalez-Hernandez. 2022. Scoping Review of Methods and Annotated Datasets Used to Predict Gender and Age of Twitter Users. *medRxiv* (2022), 2022–12.

[111] 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.

[112] Jan-Peter Ostberg, Daniel Graziotin, Stefan Wagner, and Birgit Derntl. 2020. A methodology for psycho-biological assessment of stress in software engineering. *PeerJ Computer Science* 6 (2020), e286.

[113] Albert Park, Mike Conway, and Annie T Chen. 2018. Examining thematic similarity, difference, and membership in three online mental health communities from Reddit: a text mining and visualization approach. *Computers in human behavior* 78 (2018), 98–112.

[114] James W Pennebaker and Cindy K Chung. 2007. Expressive writing, emotional upheavals, and health. (2007).

[115] James W Pennebaker, Matthias R Mehl, and Kate G Niederhoffer. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology* 54, 1 (2003), 547–577.

[116] Oleksandra Poquet and Maarten De Laat. 2021. Developing capabilities: Lifelong learning in the age of AI. *British Journal of Educational Technology* 52, 4 (2021), 1695–1708.

[117] Claudette Pretorius, Darragh McCashin, Naoise Kavanagh, and David Coyle. 2020. Searching for mental health: a mixed-methods study of young people's online help-seeking. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[118] Jakkula V Rao and K Chandraiah. 2012. Occupational stress, mental health and coping among information technology professionals. *Indian journal of occupational and environmental medicine* 16, 1 (2012), 22.

[119] Signe Ravn, Ashley Barnwell, and Barbara Barbosa Neves. 2020. What is "publicly available data"? Exploring blurred public–private boundaries and ethical practices through a case study on Instagram. *Journal of empirical research on human research ethics* 15, 1-2 (2020), 40–45.

[120] Navid Rekabsaz, Robert West, James Henderson, and Allan Hanbury. 2021. Measuring Societal Biases from Text Corpora with Smoothed First-Order Co-occurrence. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 549–560.

[121] Elizabeth A Rider, Eman Ansari, Pamela H Varrin, and Joshua Sparrow. 2021. Mental health and wellbeing of children and adolescents during the covid-19 pandemic. *bmj* 374 (2021).

[122] Julian Risch and Ralf Krestel. 2020. Top comment or flop comment? predicting and explaining user engagement in online news discussions. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 14. 579–589.

[123] Pablo Robles-Granda, Suwen Lin, Xian Wu, Gonzalo J Martinez, Stephen M Mattingly, Edward Moskal, Aaron Striegel, Nitesh V Chawla, Sidney D'Mello, Julie Gregg, et al. 2021. Jointly predicting job performance, personality, cognitive ability, affect, and well-being. *IEEE Computational Intelligence Magazine* 16, 2 (2021), 46–61.

[124] Kat Roemmich, Florian Schaub, and Nazanin Andalibi. 2023. Emotion AI at work: Implications for workplace surveillance, emotional labor, and emotional privacy. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.

[125] Stuart Rose, Dave Engel, Nick Cramer, and Wendy Cowley. 2010. Automatic keyword extraction from individual documents. *Text mining: applications and theory* (2010), 1–20.

[126] Talma Rosenthal and Ariela Alter. 2012. Occupational stress and hypertension. *Journal of the American Society of Hypertension* 6, 1 (2012), 2–22.

[127] Veronica Rossano, Rosa Lanzilotti, and Teresa Roselli. 2020. A Simulation Game to Acquire Skills on Industry 4.0. In *International Conference on Human-Computer Interaction*. Springer, 730–738.

[128] Stephanie Rude, Eva-Maria Gortner, and James Pennebaker. 2004. Language use of depressed and depression-vulnerable college students. *Cognition & Emotion* 18, 8 (2004), 1121–1133.

[129] 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.

[130] Koustuv Saha and Shamsi T Iqbal. 2023. Focus Time for Wellbeing and Work Engagement of Information Workers. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–11.

[131] Koustuv Saha, Sang Chan Kim, Manikanta D Reddy, Albert J Carter, Eva Sharma, Oliver L Haimson, and Munmun De Choudhury. 2019. The language of LGBTQ+ minority stress experiences on social media. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–22.

[132] Koustuv Saha, Manikanta D Reddy, Stephen Mattingly, Edward Moskal, Anusha Sirigiri, and Munmun De Choudhury. 2019. Libra: On linkedin based role ambiguity and its relationship with wellbeing and job performance. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–30.

[133] Koustuv Saha, Jordyn Seybolt, Stephen M Mattingly, Talayeh Aledavood, Chaitanya Konjeti, Gonzalo J Martinez, Ted Grover, Gloria Mark, and Munmun De Choudhury. 2021. What life events are disclosed on social media, how, when, and by whom?. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–22.

[134] Koustuv Saha and Amit Sharma. 2020. Causal factors of effective psychosocial outcomes in online mental health communities. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 14. 590–601.

[135] Koustuv Saha, John Torous, Eric D Caine, Munmun De Choudhury, et al. 2020. Psychosocial effects of the COVID-19 pandemic: large-scale quasi-experimental study on social media. *Journal of medical internet research* 22, 11 (2020), e22600.

[136] Koustuv Saha, Asra Yousuf, Ryan L Boyd, James W Pennebaker, and Munmun De Choudhury. 2022. Social media discussions predict mental health consultations on college campuses. *Scientific reports* 12, 1 (2022), 1–11.

[137] Koustuv Saha, Asra Yousuf, Louis Hickman, Pranshu Gupta, Louis Tay, and Munmun De Choudhury. 2021. A social media study on demographic differences in perceived job satisfaction. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–29.

[138] Selim Sametoglu, Dirk Pelt, Lyle H Ungar, Meike Bartels, et al. 2022. The Value of Social Media Language for the Assessment of Wellbeing: A Systematic Review and Meta-Analysis. (2022).
[139] Leisa D Sargent and Deborah J Terry. 2000. The moderating role of social support in Karasek's job strain model. *Work & Stress* 14, 3 (2000), 245–261.
[140] Nina Savela, Max Pellert, Rita Latikka, Jenna Bergdahl, David Garcia, and Atte Oksanen. 2025. Affective, cognitive, and contextual cues in Reddit posts on artificial intelligence. *Journal of Computational Social Science* 8, 1 (2025), 6.
[141] Iina Savolainen, Reetta Oksa, Nina Savela, Magdalena Celuch, and Atte Oksanen. 2021. Covid-19 anxiety—a longitudinal survey study of psychological and situational risks among Finnish workers. *International Journal of Environmental Research and Public Health* 18, 2 (2021), 794.
[142] Ramit Sawhney, Harshit Joshi, Alicia Nobles, and Rajiv Ratn Shah. 2021. Tweet classification to assist human moderation for suicide prevention. In *Proc Int AAAI Conf Weblogs Soc Media*, Vol. 15. 609–620.
[143] Morgan Klaus Scheuerman, Jacob M Paul, and Jed R Brubaker. 2019. How computers see gender: An evaluation of gender classification in commercial facial analysis services. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–33.
[144] Morgan Klaus Scheuerman, Jacob M. Paul, and Jed R. Brubaker. 2019. How Computers See Gender: An Evaluation of Gender Classification in Commercial Facial Analysis Services. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 144 (nov 2019), 33 pages. doi:10.1145/3359246
[145] Morgan Klaus Scheuerman, Kandrea Wade, Caitlin Lustig, and Jed R Brubaker. 2020. How we've taught algorithms to see identity: Constructing race and gender in image databases for facial analysis. *Proceedings of the ACM on Human-computer Interaction* 4, CSCW1 (2020), 1–35.
[146] Julia Churchill Schoellkopf. 2012. Gender: An infinite and evolving theory. (2012).
[147] Eva Sharma and Munmun De Choudhury. 2018. Mental health support and its relationship to linguistic accommodation in online communities. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.
[148] Judy Hanwen Shen and Frank Rudzicz. 2017. Detecting anxiety through reddit. In *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality*. 58–65.
[149] Emily Sheng, Kai-Wei Chang, Premkumar Natarajan, and Nanyun Peng. 2019. The woman worked as a babysitter: On biases in language generation. *arXiv preprint arXiv:1909.01326* (2019).
[150] Sheng-Pao Shih, James J Jiang, Gary Klein, and Eric Wang. 2013. Job burnout of the information technology worker: Work exhaustion, depersonalization, and personal accomplishment. *Information & Management* 50, 7 (2013), 582–589.
[151] Donghoon Shin, Subeen Park, Esther Hehsun Kim, Soomin Kim, Jinwook Seo, and Hwajung Hong. 2022. Exploring the Effects of AI-assisted Emotional Support Processes in Online Mental Health Community. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7.
[152] Glorian Sorensen, Jack T Dennerlein, Susan E Peters, Erika L Sabbath, Erin L Kelly, and Gregory R Wagner. 2021. The future of research on work, safety, health and wellbeing: A guiding conceptual framework. *Social Science & Medicine* 269 (2021), 113593.
[153] Wyke Stommel and Lynn de Rijk. 2021. Ethical approval: none sought. How discourse analysts report ethical issues around publicly available online data. *Research Ethics* 17, 3 (2021), 275–297.
[154] Abdul-Razak Suleman, Bernard Bekuni Boawei Bingab, Kwame Owusu Boakye, and Richard Sam-Mensah. 2022. Job rotation practices and employees performance: do job satisfaction and organizational commitment matter? *SEISENSE Business Review* 2, 1 (2022), 13–27.
[155] Vedant Das Swain, Koustuv Saha, Gregory D Abowd, and Munmun De Choudhury. 2020. Social media and ubiquitous technologies for remote worker wellbeing and productivity in a post-pandemic world. In *2020 IEEE Second International Conference on Cognitive Machine Intelligence (CogMI)*. IEEE, 121–130.
[156] Monideepa Tarafdar, Qiang Tu, Bhanu S Ragu-Nathan, and TS Ragu-Nathan. 2007. The impact of technostress on role stress and productivity. *Journal of management information systems* 24, 1 (2007), 301–328.
[157] Louis Tay, Vincent Ng, Lauren Kuykendall, and Ed Diener. 2014. Demographic factors and worker well-being: An empirical review using representative data from the United States and across the world. *The role of demographics in occupational stress and well being* (2014).
[158] David R Thomas. 2006. A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation* 27, 2 (2006), 237–246.
[159] Sigal Tifferet and Iris Vilnai-Yavetz. 2018. Self-presentation in LinkedIn portraits: common features, gender, and occupational differences. *Computers in Human Behavior* 80 (2018), 33–48.
[160] Maria Tomprou, Laura Dabbish, Robert E Kraut, and Fannie Liu. 2019. Career mentoring in online communities: Seeking and receiving advice from an online community. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.
[161] Isabelle van der Vegt and Bennett Kleinberg. 2020. Women worry about family, men about the economy: Gender differences in emotional responses to COVID-19. In *International Conference on Social Informatics*. Springer, 397–409.
[162] Jessica Vitak, Katie Shilton, and Zahra Ashktorab. 2016. Beyond the Belmont principles: Ethical challenges, practices, and beliefs in the online data research community. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. 941–953.
[163] Svitlana Volkova, Kyungsik Han, and Courtney Corley. 2016. Using social media to measure student wellbeing: a large-scale study of emotional response in academic discourse. In *International Conference on Social Informatics*. Springer, 510–526.
[164] Fenglian Wang, Mingqing Hu, and Min Zhu. 2020. Threat or Opportunity–Analysis of the Impact of Artificial Intelligence on Future Employment. In *International Conference on Human-Computer Interaction*. Springer, 296–308.
[165] Liuping Wang, Dakuo Wang, Feng Tian, Zhenhui Peng, Xiangmin Fan, Zhan Zhang, Mo Yu, Xiaojuan Ma, and Hongan Wang. 2021. Cass: Towards building a social-support chatbot for online health community. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–31.
[166] Zijian Wang, Scott Hale, David Ifeoluwa Adelani, Przemyslaw Grabowicz, Timo Hartman, Fabian Flöck, and David Jurgens. 2019. Demographic inference and representative population estimates from multilingual social media data. In *The world wide web conference*. 2056–2067.
[167] Zijian Wang and Christopher Potts. 2019. Talkdown: A corpus for condescension detection in context. *arXiv preprint arXiv:1909.11272* (2019).
[168] Nancy M Wendlandt and Aaron B Rochlen. 2008. Addressing the college-to-work transition: Implications for university career counselors. *Journal of Career Development* 35, 2 (2008), 151–165.
[169] Sue Williamson, Alicia Pearce, Helen Dickinson, Vindhya Weeratunga, and Fiona Bucknall. 2021. Future of work literature review: Emerging trends and issues. (2021).
[170] Jamie Wilson, Cora Sargeant, Natalie Jago, and Sarah Wright. 2023. Co-creating a new theory of gender beyond the binary: A Delphi study in the United Kingdom. *Journal of LGBT Youth* (2023), 1–25.
[171] Sang Eun Woo, Andrew T Jebb, Louis Tay, and Scott Parrigon. 2018. Putting the "person" in the center: Review and synthesis of person-centered approaches and methods in organizational science. *Organizational Research Methods* 21, 4 (2018), 814–845.
[172] Jue Wu, Junyi Ma, Yasha Wang, and Jiangtao Wang. 2021. Understanding and Predicting the Burst of Burnout via Social Media. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–27.

[173] Chuang-Wen You, Chien Wen Yuan, Nanyi Bi, Min-Wei Hung, Po-Chun Huang, and Hao-Chuan Wang. 2021. Go Gig or Go Home: Enabling Social Sensing to Share Personal Data with Intimate Partner for the Health and Wellbeing of Long-Hour workers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.

[174] Aijia Yuan, Edlin Garcia Colato, Bernice Pescosolido, Hyunju Song, and Sagar Samtani. 2025. Improving workplace well-being in modern organizations: A review of large language model-based mental health chatbots. *ACM Transactions on Management Information Systems* 16, 1 (2025), 1–26.

[175] Hannes Zacher. 2013. Older job seekers' job search intensity: The interplay of proactive personality, age and occupational future time perspective. *Ageing & Society* 33, 7 (2013), 1139–1166.

[176] Ben Zefeng Zhang, Tianxiao Liu, Shanley Corvite, Nazanin Andalibi, and Oliver L Haimson. 2022. Separate Online Networks During Life Transitions: Support, Identity, and Challenges in Social Media and Online Communities. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–30.

[177] Jiawei Zhou, Koustuv Saha, Irene Michelle Lopez Carron, Dong Whi Yoo, Catherine R Deeter, Munmun De Choudhury, and Rosa I Arriaga. 2022. Veteran Critical Theory as a Lens to Understand Veterans' Needs and Support on Social Media. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–28.

[178] Chen Zhu, Hengshu Zhu, Hui Xiong, Pengliang Ding, and Fang Xie. 2016. Recruitment market trend analysis with sequential latent variable models. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 383–392.