Human Factors, Business Management and Society, Vol. 215, 2026, 264–274
https://doi.org/10.54941/ahfe1007607

AI Support for Improving Managerial Decision-Making Processes Regarding Workforce Planning

Ferhat Kutlu

Management Information Systems department of the Faculty of Economics, Administrative and Social Sciences in Fenerbahçe University, 34758, Ataşehir-İstanbul/Türkiye

ABSTRACT

The accelerating integration of AI into the labor market is transforming the foundational dynamics of modern work, particularly through the rise of remote work and the expansion of the gig economy. Freelance platforms increasingly rely on algorithmic management systems that automate task allocation, performance monitoring, and disciplinary decisions. While these systems promise substantial productivity gains, they also introduce new challenges, including heightened information asymmetry and reduced transparency for workers who often lack mechanisms to contest automated outcomes. This disconnect has significant implications for worker well-being, especially as critical human–resource functions become fully automated. This study proposes a methodological framework to leverage AI — specifically Natural Language Understanding — to support more transparent and equitable managerial decision making without compromising platform efficiency. Unlike conventional HR literature that assumes human involvement in conflict resolution, freelancers frequently encounter automated decisions that overlook nuanced indicators of distress. By incorporating computational techniques capable of detecting linguistic signals associated with emotional strain, this research addresses a critical gap in understanding how algorithmic systems can be designed to better account for human well-being. The study develops an empirically grounded infrastructure for processing unstructured worker generated text, integrating sentiment analysis through the "Valence Aware Dictionary and Sentiment Reasoner" model and discourse interpretation via a discourse relation classifier trained on the Penn Discourse Treebank. Forming a departure point to enable automated sensitivity analysis of worker feedback by discovering a strong indicator out of discourse relation type categories is the sole artefact of the research that could lead towards the proof of concept about automated sensitivity analysis of worker feedback to sustain a scalable pathway toward human centered algorithmic management.

Keywords: Natural language understanding, Artificial intelligence, Natural language model, Sentiment analysis

INTRODUCTION

The integration of Artificial Intelligence (AI) into the labor market is fundamentally changing the two key features of modern work as location independence and contract flexibility which could be renamed as remote work and gig economy. The rapidly expanding freelance economy has given rise to

Received February 6, 2026; Revised April 5, 2026; Accepted April 22, 2026; Available online July 20, 2026

algorithmic management, which uses software algorithms to assign tasks, monitor performance, and discipline employees. As noted in (Keegan and Meijerink, 2025), platform-based gig work often traps individuals in a precarious "gray zone" where their daily tasks and overall employment status are governed by opaque, automated systems. While this automation yields significant macroeconomic productivity and efficiency gains, it strips away traditional human oversight, replacing it with a "black box" environment where employees lack transparency. Recent scholarship emphasizes that AI's pervasive presence fundamentally alters the psychological contract between workers and platforms, diminishing perceptions of autonomy, dignity, and job security (Moghayedi et al., 2024; Tomprou and Lee, 2022). When human managers are replaced by algorithms, workers frequently experience dehumanization, feeling reduced to mere data points while navigating heightened technostress and isolation (Dang and Liu, 2025). This dynamic has catalyzed a novel form of workplace stress defined in recent digital discourse as "algorithmic anxiety" which goes beyond a simple fear of job displacement; it is a complex syndrome encompassing concerns about professional identity, shattered trust, and the erosion of human value in an automated future (Bankins et al., 2024; Soulami et al., 2024). Gig workers facing automated threats often exhibit various coping behaviors, where reactive approaches can lead to a vicious cycle of emotional exhaustion and negative reappraisal. Despite the escalating prevalence of algorithmic anxiety, current Human Resources (HR) literature remains inadequately equipped to address these digital-first challenges. Traditional HR paradigms operate under the assumption that human-to-human interaction is inherently present in conflict resolution and performance appraisal. Consequently, when gig workers encounter automated decisions (e.g., unfair task allocation or disciplinary actions), there is no clear path to objection. Because algorithms are not natively programmed to detect implicit human signals of distress, they routinely overlook the linguistic and behavioral indicators of anxiety that a human manager might recognize. To sustain a scalable pathway toward human-centered algorithmic management, there is an urgent need to translate unstructured worker feedback into actionable data. Addressing this requires not only acknowledging the psychological toll of automated management but developing standardized, empirically grounded metrics to measure and track worker distress over time. Natural Language Understanding (NLU) presents a viable methodology to bridge this gap by applying computational techniques capable of interpreting discourse and sentiment to systematically detect linguistic signals of emotional strain such as algorithmic anxiety index. Integrating a model for basic sentiment detection with an advanced discourse classifier can enable automated sensitivity analysis of employee feedback. So, an NLU-driven multi-model approach could be built to provide yet another facility of conducting automatic employee feedback analysis to understand sensitivity and identify areas for improvement.

Accordingly, within the work, an empirically grounded infrastructure is designed for processing unstructured worker-generated text in two phases. The first stage of the computational pipeline involves filtering unstructured text to identify explicit expressions of dissatisfaction or emotional strain. This phase relies on the detection of only negative comments via the Value-Aware Dictionary and Sentiment Reasoner (VADER) (Hutto and Gilbert, 2014) model. Once the comments are filtered, the negative feedback goes through an advanced classifier, created by fine-tuning the Bidirectional Encoder Representations from Transformers (BERT)[1] (Devlin et al., 2019) model. To enable sophisticated discourse interpretation, this model is re-trained specifically on the Penn Discourse Treebank Version 3.0 (PDTB) (Prasad et al., 2019). The results of this research and development form a solid departure point to enable automated sensitivity analysis of worker feedback by discovering strong indicators among Discourse Relation (DR) type categories.

METHODOLOGY

Development, experimentation, and analysis phases are conducted as almost parallel throughout the study to explore all possible means of automated sensitivity analysis performed by using Natural Language Processing (NLP) techniques to improve the computer's NLU capability in means of sensing the sign of anomaly, i.e., very high intention to leave the position. The methodology tested in the study has formed an empirically designed infrastructure capable of processing unstructured text data (e.g., worker forum posts, reviews, or survey) through (i) the model that exploits the VADER to specifically measure the emotional intensity of sentences and (ii) the BERT Model fine-tuned by a training through a supervised learning classification approach where the DRs have DR Type labels, given by expert PDTB annotators. The purpose of the former model is to process the filter input to focus on negative reviews shortly while the latter model is to do fact finding for the enhancement of precision.

VADER[2] Sentiment Analysis is a lexicon and rule-based tool that is specifically attuned to sentiments expressed in social media. It works well on texts from other domains. It is fully open source under the [MIT License][3]. Working for the analysis of word polarity to assign a sentiment score to each word, based on its emotional value, VADER uses a compound score, which is a normalized value between -1 and +1, representing the overall sentiment. Then sentiment categorization is done according to the thresholds given in Table 1 such that continuous sentiment scores were successfully categorized into 'Positive' (score > 0.05), 'Neutral' (-0.05 ≤ score ≤ 0.05), and 'Negative' (score < –0.05) for modeling purposes.

Table 1: Label mapping structure of VADER.

Sentiment Score	Sentiment Category
> 0.05	Positive sentiment (+)
BTWN -0.05 and 0.05	Neutral sentiment (0)
< –0.05	Negative sentiment (–)

PDTB is the third release in the Penn Discourse Treebank project[4], the goal of which is to annotate the Wall Street Journal (WSJ) articles with DRs which are grounded in an identifiable set of explicit words or phrases (discourse connectives)

[1] *https://huggingface.co/google-bert/bert-base-uncased*

[2] *VADER-Sentiment-Analysis: https://github.com/cjhutto/vaderSentiment*

[3] *https://pypi.org/project/vaderSentiment/*

[4] *https://catalog.ldc.upenn.edu/LDC2019T05*

or simply in the adjacency of two sentences. It has been used by many researchers in the natural language processing community and more recently, by researchers in psycholinguistics. Categorization of different types of DRs is done according to the PENN Discourse Treebank 3.0 Annotation Manual as explained in Table 2.

Table 2: DR Type categories*

DR Type	Definition & Example
Explicit Relation	The relations that are explicitly signaled through lexico-syntactic elements. They may be realized across or within sentences and involve coordinating conjunctions (and, so, but) subordinating conjunctions (because, after, while), adverbials (however, additionally, consequently), and prepositional phrases (in summary, on the contrary). *I cannot recall any disorder in currency markets* since {reason, succession} **the 1974 guidelines were adopted**.
Implicit Relation	Even though the absence of a connective supporting a DR, it is possible to sense the meaning of the inferred relation. *The parishioners of St. Michael and All Angels stop to chat at the church door, as members here always have*. (Implicit=while) {synchronous} **In the tower, five men and women pull rhythmically on ropes attached to the same five bells that first sounded here in 1614**.
AltLex and AltLexC Alternative Lexicalization	IF an Implicit DR is inferred to hold between or within sentences but the insertion of an implicit connective in the relation is perceived redundant, THEN the relation is referred to as being alternatively lexicalized. *Congressional Democrats and the Bush administration agreed on a compromise minimum-wage* . **opening the way for the first wage-floor boost in more than nine years**. {result} IF AltLex DR is inferred to hold between or within sentences but they can also be allowed to be found on their own, separate from other AltLex tokens, THEN the relation is labelled as AltLexC – as a new convention brought by PDTB V 3.0. *The fit is so good*, **we see this as a time of opportunity**, he said. {result} *Note: The alternative lexicalization is underlined. The senses of the connectives are given below the text.*
EntRel Entity Relation	There is not any DR can be inferred between adjacent sentences and adjacent sentences form entity-based coherence with the same entity being realized in both sentences either directly or indirectly. *Mr. Baker heads the Kentucky Association of Science Educators and Skeptics*. **Like Hollywood's Ghostbusters, Kentucky's stand ready to roll when haunts get out of hand.**
Hypophora	The DR where a question is asked and a meaningful response is provided. *What does this mean for trade policy*? **Too early to tell, but a trade deficit that is significantly smaller than we imagined does suggest a review of our trade posture**.
NoRel	Annotated only between adjacent sentences within a paragraph that are not linked to each other by a DR. *There is a small Renoir on the wall*. **Zurich-based magazine Bilanz lists Mr. Rey as having a fortune of about 1.5 billion Swiss francs**.

(Prasad et al., The PENN Discourse Treebank 3.0 Annotation Manual, 2019)
Note: Each example consists of a pipe-delimited record, and the token it corresponds to. The first argument is indicated in yellow, and the second argument is in blue.

EXPERIMENT

The experiment pipeline, shown in Figure 1, starts with "Online Employee Reviews Dataset"[5] (Huang et al., 2022) which contains 1,563 reviews in English and all are accepted as efficient for empirical study of the multi-model text analysis system. At the first step the experiment dataset is fed into VADER model, and each review received a sentiment value as shown in the descriptive statistics of the sentiment scores are given in Table 3, which reveals the negative abundance in the dataset. At the next step, the dataset is labeled according to the threshold values, given in Table 1, and the statistics of sentiment categories are retrieved as shown in Table 4. This showed that the VADER output cannot be regarded as ground truth since it is detected by human reading that at least two thirds of the reviews are full of workplace or organizational complaints. But the pipeline stuck to the model with the purpose of realizing automation and fact-finding study.

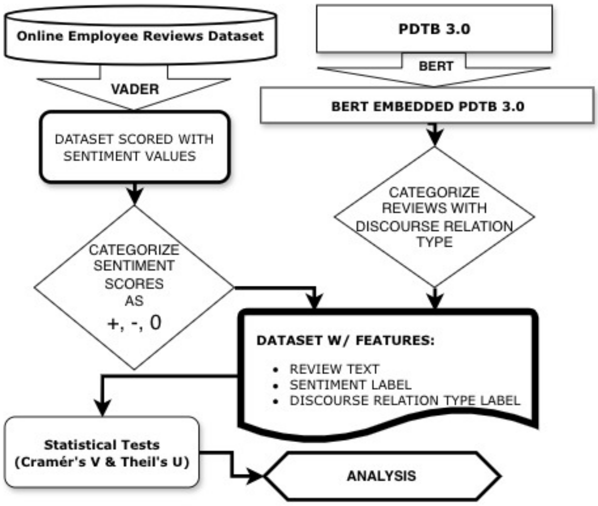

Figure 1: Experiment pipeline.

Table 3: The descriptive statistics for sentiment scores produced by VADER.

Parameter	Score
Mean	**–0.098155**
Standard Deviation	**0.430118**
Range	BTWN –0.974100 and 0.977300
25%	< –0.402300
50%	< 0.000000
75%	< 0.100950

[5] *Online Employee Reviews Dataset - Mendeley Data: https://data.mendeley.com/datasets/zw2shn7pv3/2*

Table 4: The statistics of sentiment categories designated by VADER.

Sentiment Category	Support	Ratio in the Dataset
Positive sentiment (+)	425	27.19
Neutral sentiment (0)	394	25.21
Negative sentiment (-)	**744**	**47.60**

As parallel to work on employee reviews dataset, BERT is fine-tuned by feeding PDTB, conforming to the same supervised learning for classification method explained in (Kutlu et al., 2024). So, a new BERT model for DR Type classification is obtained and the review text of the dataset is processed through DR Type classification with the new model. The statistics of DR Type categories are given in Table 5, and the abundance (%88.68) of AltLexC type is clearly seen. Except for NoRel, the other categories of DR Type are very less, but the test is done overall to verify the results.

Table 5: The statistics of DR type categories designated by BERT- PDTB model.

DR Type Category	Support	Ratio in the Dataset
AltLexC	**1386**	**88.68**
NoRel	163	10.43
Explicit	9	0.58
Implicit	3	0.19
EntRel	1	0.06
AltLex	1	0.06

A sample output from this stage is printed in Table 6, by choosing four of them as only from negative sentiment reviews. Three out of four are classified as AltLexC. The reviews are far from explicit type of discourse, and their sentences shoot arguments neither with connectives nor giving room for insertion of an implicit connective. They are also near to an atomic fashion as they could be allowed to be found on their own as separate from others. Considering the definition of AltLex and AltLexC given Table 2, AltLexC could be claimed to be fitting well to the negative reviews, forming a potential indicator to enhance anomaly detection in human input.

Table 6. Sample output of the experiment.

Sample Employee Review	*Sentiment Value*	Sentiment Category	DR Type
"Upper management does not trust its employees to make decisions', None: ["every decision has to be run up the ladder because they don't trust their employees to make the right ones. This includes minor ones that require VP approval causing unneeded delays. Don't look to advance in your career unless you have a 4-year degree", 'they will not promote you to any kind of management role unless you have that magic piece of paper."	*−0.4023*	**Negative**	**AltLexC**

(Continued)

Table 6: Continued.

Sample Employee Review	*Sentiment Value*	Sentiment Category	DR Type
"No work life balance whatsoever; - Daily firefighting, contradictory requests regarding what to focus on; - No concept, no plan, no communication; - Inefficient processes, that's saying if existing; - Company is going downhill."	*−0.786*	Negative	AltLexC
"The salary was low for the position. Benefits were standard. Company was a bit overstructured for my taste. Felt that a 50k position should be paid closer to 70k."	*−0.728*	Negative	AltLexC
"Low empowerment to employees of purchasing department …"	*−0.2732*	Negative	NoRel

ANALYSIS

To conduct a distribution analysis of the results the BoxPlot Distribution of DR Types vs Sentiment Scores is plotted in Figure 2, which shows that only the NoRel class has normal distribution while AltLexC and Explicit classes are below zero with their three quartiles as to favor negative sentiment.

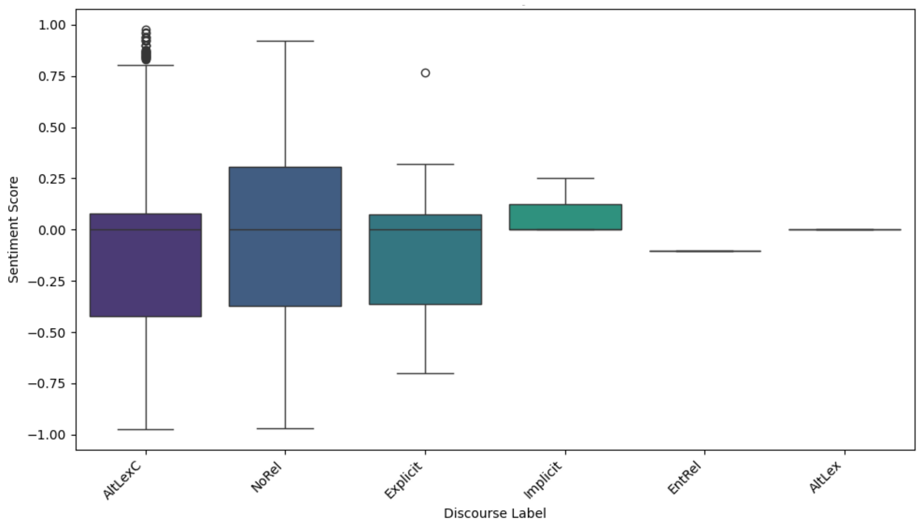

Figure 2: BoxPlot distribution of DR type categories vs sentiment scores.

The Mean and Standard Deviation of DR Type categories in means of Sentiment Scores are given in Table 7. AltLex and EntRel labels showed a mean sentiment of 0.000000 and -0.102700, respectively, both with NaN standard deviations, suggesting uniform or single sentiment scores within these categories. Implicit had the most positive average sentiment (0.083333) with a relatively low standard deviation (0.144338). AltLexC and Explicit displayed slightly negative average sentiments (-0.106605 and -0.041811) with standard deviations around 0.42. NoRel showed the highest variability in sentiment with a standard deviation of 0.466432 around a slightly negative mean of -0.033334.

Table 7: The mean and standard deviation of DR types vs sentiment scores.

DR Type Category	Mean	Standard Deviation
AltLexC	–0.106605	**0.425980**
NoRel	–0.033334	**0.466432**
Explicit	–0.041811	0.425022
Implicit	0.083333	0.144338
EntRel	–0.102700	NaN
AltLex	0.000001	NaN

Except for AltLexC and NoRel, the support numbers of all other DR Type categories are very less according to values given in Table 5. Thus, the proportion of each sentiment category in DR Type categories is given only for AltLexC and NoRel in Table 8, revealing the coverage of negative sentiment (48.63 + 39.88 = 88.51 %) is so wide.

Table 8: Proportion of each sentiment category in DR type categories.

DR Type Category / Sentiment	Negative (%)	Neutral (%)	Positive (%)
AltLexC	**48.63**	25.32	26.05
NoRel	**39.88**	23.31	36.81

The correlation analysis of "DR Types vs Sentiment Categories" is applied to verify the output statistics mentioned in here. The correlation coefficients matrix, shown in Figure 4, is produced besides the conduct of additional statistical tests (Cramér's V and Theil's U). At first glance, it is concluded that the matrix reveals that DR Type categories and sentiment categories are weakly correlated and explanatory power of the BERT-PDTB model over the sentiment categories is not strong. That is, knowing the discourse label of a review gives you almost no information about whether the review is Positive, Neutral, or Negative. As seen in the correlation matrix, the correlation coefficients between specific labels and sentiment categories are extremely low, ranging mostly between –0.07 and +0.07. The highest positive correlation is only 0.074 (between NoRel and Positive), which is statistically trivial, implying that no linear relationship exists. Cramér's V (Overall Association) measures the strength of association between two categorical variables on a scale of 0 to 1 and it is calculated as 0.073, indicating that a weak association. Generally, a value below 0.1 is considered negligible. This also signs that the sentiment categories are largely independent of each other. Theil's U (Directional Explanatory Power) measures how much uncertainty in one variable is reduced by knowing the other, and it is calculated as 0.005 in this experiment. Label → Sentiment (U = 0.005): Knowing the label reduces the uncertainty about the classification of sentiment by only 0.5%.

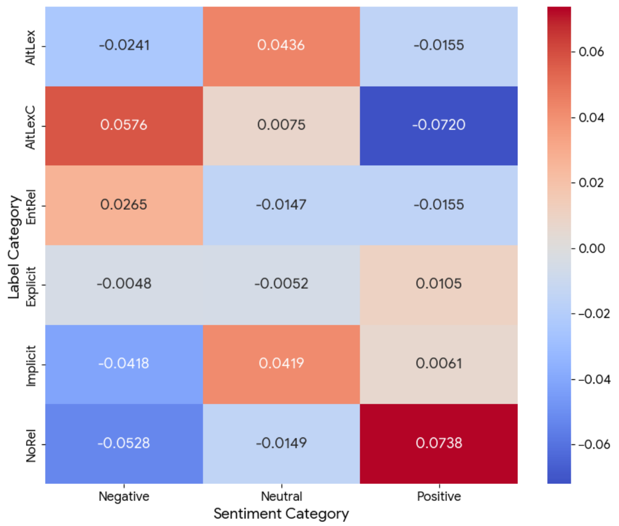

Figure 4: Correlation coefficients matrix between DR type vs sentiment categories.

Finally, a multinomial logistic regression model is built to interpret the relationship between sentiment and DR Type categories, ending with the coefficients given in Table 9. The coefficients represent the change in the log-odds of a sentiment category for a one-unit increase in the corresponding discourse label relative to a baseline, providing a quantified relationship between discourse labels and sentiment categories. Positive coefficients indicate an increased likelihood of that sentiment category given the discourse label, while negative coefficients indicate a decreased likelihood. Focusing on the main target of negative sentiments the AltLexC discourse label demonstrated the best positive coefficient (0.282) while Implicit received the smallest likelihood (-0.614). This formed yet another indicator for increased likelihood of negative sentiment associated with AltLexC.

Table 9: Logistic regression model coefficients.

DR Type Category / Sentiment	Negative	Neutral	Positive
AltLexC	**0.281951**	−0.197999	−0.083951
NoRel	0.061568	−0.295402	0.233834
Explicit	0.123633	−0.258471	0.134837
Implicit	**−0.613607**	0.515483	0.098124
EntRel	**0.428552**	−0.216191	−0.212362
AltLex	−0.265621	0.480617	−0.214997

CONCLUSION

The core question of this study was in two folds: (i) to conduct in depth investigation of the relationship between DR Type labels and sentiment categories, and (ii) to search for the best indicator associated with negative sentiment likelihood. Should there be viable fact finding, research and development to follow-up the proof-of-concept for automated sensitivity analysis of worker feedback to sustain a scalable pathway toward human centered algorithmic management and more accountable digital labor platforms.

The first practical implication is the promising results about the convergence of negative sentiment category on the DR Types of AltLexC and NoRel. Especially the majority (88.68%) of negative reviews had hit AltLexC type of DRs, making it abundant over its kinds in the experiment dataset. In addition, the proportion of negative samples in the AltLexC type DRs is 48.63% – the highest score in the resulting statistics. The findings were examined by calculating several secondary parameters, along with a correlation analysis that weakened the effect of previous findings. Finally, the classification results tested through multinomial logistic regression analysis that ended up with a potential sign of yet another model which could be built upon AltLexC, Implicit and EntRel types of DRs.

From a general perspective, a departure point to enable automated sensitivity analysis of worker feedback by discovering a strong indicator out of DR Type categories of PDTB is the sole artefact of the research that could lead towards the proof of concept. Further analysis could involve exploring the predictive power of the multinomial logistic regression model which could potentially incorporate other linguistic features to enhance sentiment prediction.

ACKNOWLEDGMENT

The author would like to acknowledge The Applied Human Factors and Ergonomics International (AHFE) Conference Comity and Fenerbahçe University Rectorate.

REFERENCES

Bankins S., Hu X., Yuan Y., 2024. Artificial intelligence, workers, and future of work skills. Current Opinion in Psychology. Volume 58, 101828, ISSN 2352-250X, https://doi.org/10.1016/j.copsyc.2024.101828

Dang J., Liu L., 2024. Extended artificial intelligence aversion: People deny humanness to artificial intelligence users. *Journal of Personality and Social Psychology*. Advance online publication. https://doi.org/10.1037/pspi0000480

Devlin J., Chang M.W., Lee K., and Toutanova K., 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the ACL: Human Language Technologies, *NAACL-HLT 2019, V*ol. 1 (Long and Short Papers), June 2–7, Minneapolis, MN, USA, ACL, pp. 4171–4186.

Huang L., Abrahams A., Sithipolvanichgul J., Gruss R., Ractham P., 2015. Online Employee Reviews Dataset", Mendeley Data, Version 2, Published:15 Aug 2022, https://doi.org/10.17632/zw2shn7pv3.2

Hutto C.J., Gilbert E.E., 2014. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI.

Keegan R., Meijerink J., 2025. Algorithmic Management in Organizations? From Edge Case to Center Stage. Annual Review Organizational Psychology and Organizational Behavior. 12:395-422. https://doi.org/10.1146/annurev-orgpsych-110622-070928

Kutlu F., Zeyrek D., Kurfalı M., 2024. Toward a shallow discourse parser for Turkish. Cambridge University Natural Language Engineering Journal. Vol. 30(5):1009-1034, https://doi.org/10.1017/S1351324923000359

Moghayedi A., Mahachi J., Lediga R., Mosiea T., Phalafala E., 2024. Revolutionizing affordable housing in Africa: A comprehensive technical and sustainability study of 3D-printing technology. Sustainable Cities and Society, Volume 105, 2024, 105329, ISSN 2210-6707, https://doi.org/10.1016/j.scs.2024.105329

Prasad R., Webber B., Lee A., and Joshi A., 2019. Penn Discourse Treebank Version 3.0. LDC Catalog No.: LDC2019T05, ISBN: 1-58563-877-3, ISLRN: 977-491-842-427-0, https://doi.org/10.35111/qebf-gk47, March 15, 2019.

Prasad R., Webber B., Lee A., and Joshi A., 2019. The PENN Discourse Treebank 3.0 Annotation Manual. Journal Philadelphia, University of Pennsylvania, Volume35.

Soulami M., Benchekroun S., Galiulina A., 2024. Exploring how AI adoption in the workplace affects employees: A bibliometric and systematic review. Front Artif Intell. https://doi.org/10.3389/frai.2024.1473872

Tomprou M., Lee M. K. (2022). Employment relationships in algorithmic management: A psychological contract perspective. *Computers in Human Behavior*, 126, Article 106997. https://doi.org/10.1016/j.chb.2021.106997