Living with Employee Surveillance While Working From Home: A Sentiment Analysis on Tattleware and Bossware

Hai Jing(Jane) Tu
tuhai@iu.edu

March 10, 2022

word count= 2,019

1 Introduction and Research Background

Employee surveillance used to be mainly monitoring behavioral problems such as theft and fraud (Pierce, Snow, and McAfee, 2015). About forty years ago, many American employers were found to abuse employee rights by “bugging restrooms or placing informers among workers to find out who’s drinking or smoking dope on the job” (Crittenden,1980, June 20,p. D5). Such monitoring actions were often resisted by employees and questioned for violating employee rights.

In recent years, employee surveillance has become more pervasive with the development of information technology. Employee surveillance software, often referred to as ”tattleware” or ”bossware” allow employers to have deep access to the computer screens of employees (Klosowski, 2021, December 10). Other than logging work time, these programs can be aggressive by taking screenshots of employees’ screens, tracking locations and activities that include both working and entertainment, checking content such as emails and messages, and even detecting voice tones and facial expressions (Klosowski).

Not surprisingly, the use of employee surveillance software has increased significantly during the Covid-19 pandemic due to the big surge in numbers of work-from-home employees. Based on a survey by Gartner’s Brian Kropp, about 30% of companies in the survey now use surveillance programs to track remote workers, compared with only 10% of companies doing so before the pandemic (Klosowski). In the mean time, discussion that involves ”tattleware” and ”bossware” started to emerge more frequently on Twitter. Resistant to surveillance, many complain about the violation of privacy and question the necessity in such aggressive surveillance.

As employee surveillance programs such as tattleware and bossware negotiate their way into employees’ home offices, living rooms, and kitchens, it’s important to understand the logic behind employee surveillance and its role in today’s working environment. According to Westin (2003), “at the socio-cultural level, . . . privacy is determined by the individual’s power and social status” (p. 432). Privacy is not an equal right for people with different background and performing different roles in the society. Most surveillance programs are top-down by design. The structure of surveillance program decides that it serves the perspective of administrators and take employees as targets of surveillance. Therefore today's digital surveillance is no more democratic than the old system back in the 1980s that tie individuals to their work desk.

The world went through a deep change during the pandemic along with the rise and fall of infection numbers. Work from home as a temporary strategy in the beginning of the pandemic turned out to last longer than was expected. With employee surveillance becoming more prevalent, today's employees face the challenge of surveillance that's no less aggressive compared with those 40 years ago. It is necessary to find out what the sentiment toward employee surveillance tools such as "tattleware" or "bossware" evolves in the past couple years.

This research conducts sentiment analysis to find out the reactions to employee surveillance across different time periods: before the pandemic (Dec. 2019-Feb. 2020)), early pandemic(Apr. 2020-August. 2020), early middle of the pandemic (Oct. 2020-Jun.2021)), late middle of the pandemic (Jul, 2021-Dec. 2021), and towards the end of the pandemic (Jan. 2022 - Feb. 2022)). These findings will be combined with the social logic behind digital surveillance to explain the diminishing of private sphere.

This research also studies the difference in sentiments based on the popularity of tweets. Counts in retweets, reply, and likes are observed as indicators of popularity, and the changes across time periods in these counts will be identified and interpreted. In addition, sentiment scores between popular and not-so-popular tweets will be studied.

2 Research Questions

RQ 1. What are the differences in sentiment polarity and intensity on the subject of employee surveillance before, during, and towards the end of the COVID-19 pandemic?

RQ 2. What are the changes in tweets popularity on the subject of employee surveillance before, during, and towards the end of the COVID-19 pandemic?

RQ 3. What are the differences in sentiment polarity and intensity on the subject of employee surveillance between popular tweets and less popular tweets?

3 Data

The data are collected using Twitter Application Programming Interface (API). By applying for a developer account, any researcher can access twitter data using API. For this particular project, the PI is granted academic access to Twitter API with an elevated tweet caps of 10 million per month and a capacity of full-archive search. A tweet search query is constructed to include tweets that mentioned ”tattleware” or ”bossware” or ”employee surveillance” during the five different time periods mentioned above. The query also excludes retweets.

The data collection is done through `Snscrape`, a Python package. `Snscrape` is more advanced than `Tweepy` because the latter only allows query for recent post in the past 7 days. `Snscrape`, however, allows historical search that generates either `json` or `txt` file for analysis using pandas. Six attempts of data scraping were conducted to collect sample tweets from the different time periods.

After all the results are merged into one dataset for analysis, duplicates are dropped by checking identical tweets from the same twitter id. By dropping duplicates, the sample size drops from 6,611 to 5,450. Table 1 shows the size of sample from each time period before and during the pandemic.

Time	Tweet Count	VADER Compound Mean	STD
pre-pandemic (Dec. 2019-Feb. 2020)	500	-0.16	0.47
early-pandemic (Apr. 2020-August. 2020)	651	0.00	0.25
middle-pandemic(Oct. 2020-Jun.2021)	1439	-0.15	0.40
middle-pandemic(Jul, 2021-Dec. 2021)	2360	0.03	0.27
late-pandemic(Jan. 2022 - Feb. 2022)	500	-0.03	0.47

Table 1: Tweet Data Across Different Time Periods

4 Method

This study takes the dictionary-based approach and uses Valence Aware Dictionary for sEntiment Reasoning (`VADER`) to analyze Twitter data on employee surveillance. According to Hutto and Gilbert (2014), `VADER` measures not only polarity but also intensity of lexicons in documents. In addition, `VADER` excels in sentiment analysis for being sensitive to sentiment expressions in microblog-like contexts (Hutto and Gilbert). Twitter data will be analyzed using sample VADER code in Python for sentiment polarity and intensity during different time periods before and during the Covid-19 pandemic. A compound value that indicates the polarity and intensity of the tweet is shown in Table 1.

In addition to calculating sentiment scores, this paper also measures popularity of tweets using variables such as "reply count", "like count", and "retweet count." The change of these counts across different time periods will be analyzed, and the different in sentiment values between popular and not-so-popular tweets will be studied.

5 Results

5.1 RQ1

To find out the differences in sentiment polarity on the subject of employee surveillance before, during, and towards the end of the COVID-19 pandemic, a comparison was conducted using the means of `VADER` compound values (Figure 1). It shows that the sentiment towards employee surveillance was largely negative with a strong intensity before the pandemic. During the pandemic, the polarity start to revert. Another obvious negative polarity can be seen for the time period from October 2020 to June 2021.

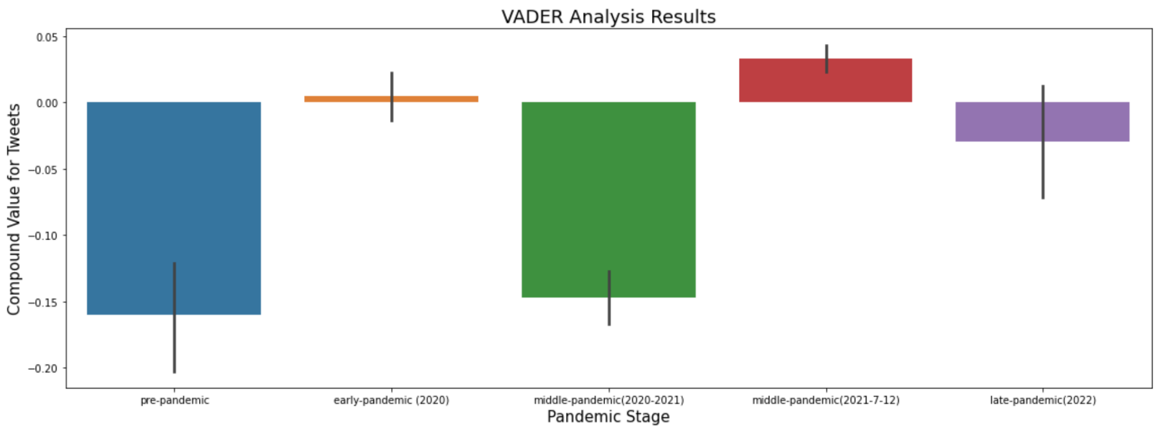

Figure 1: VADER Results on Polarity and Intensity of Tweets on Employee Surveillance

As shown in figure 1, the intensity of the tweets is shrinking as time develops toward the end of the pandemic. The intensity on the topic was rather strong before the pandemic. The second strongest intensity with a negative sentiment is during the pandemic in late 2020 to middle 2021. The score of intensity from the other three periods are rather mild.

5.2 RQ2

Table 2 summarizes the findings on popularity of tweets across different time periods before and during the pandemic. Retweet counts peaked at middle-pandemic periods (Oct. 2020- Jun. 2021 and Jul. 2021-Dec. 2021). In opposite to retweet counts, reply counts and Like counts show a valley the middle of the pandemic and peaks at the first and last time periods.

Time	Tweet Count	ReTweet Mean	Reply Mean	Like Mean
pre-pandemic (Dec. 2019-Feb. 2020)	500	2.66	0.99	6.87
early-pandemic (Apr. 2020-August. 2020)	651	6.29	0.25	3.91
middle-pandemic(Oct. 2020-Jun.2021)	1439	15.36	0.46	4.70
middle-pandemic(Jul. 2021-Dec. 2021)	2360	11.10	0.17	1.84
late-pandemic(Jan. 2022 - Feb. 2022)	500	0.94	0.58	5.74

Table 2: Tweet Popularity Across Different Time Periods

The following three figures (Figure 2) show the changes in retweet, reply, and likes of tweets on the subject of employee surveillance before, during, and towards the end of the COVID-19 pandemic.

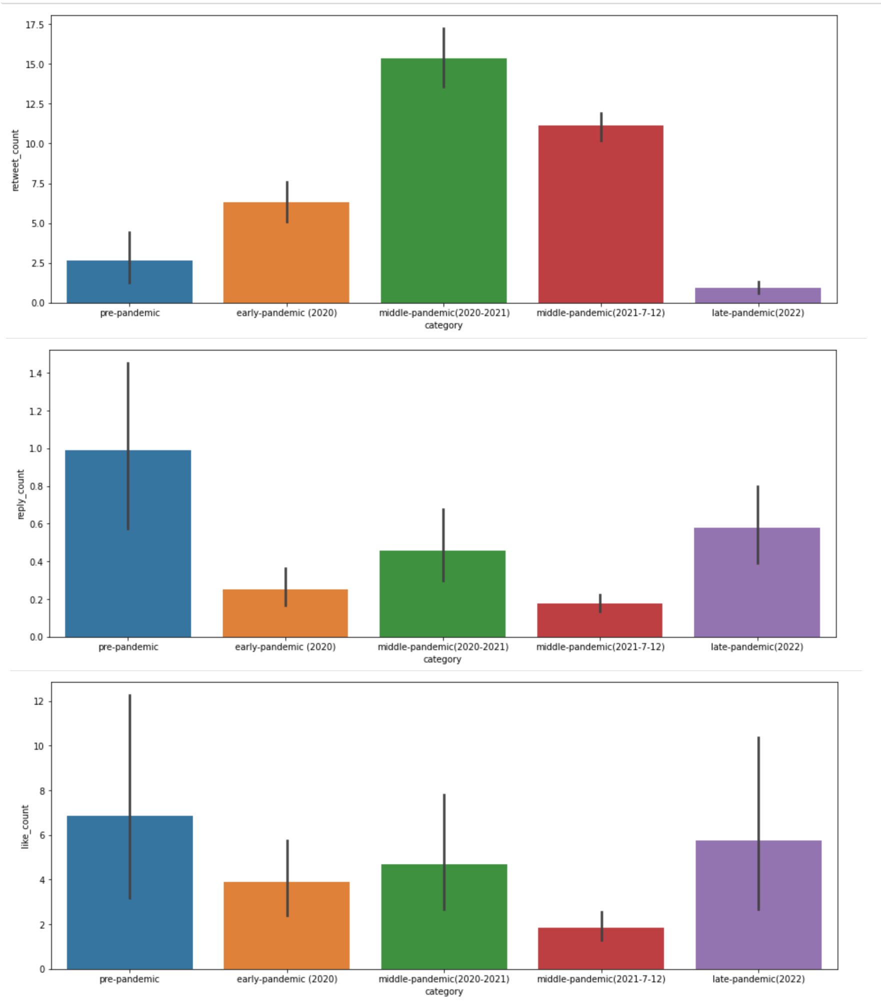

Figure 2: Popularity of Tweets Measured by Retweet$_Count, Reply_Count, andLike_Count$

5.3 RQ3

The attention received by social media posts are known for it's long tail(Kordumova, van Gemert, and Snoek,2016). The Twitter data collected for this research also proves the long tail effect, with the majority of tweets receiving little attention and only a small amount of them receiving the most attention (Figure 3)

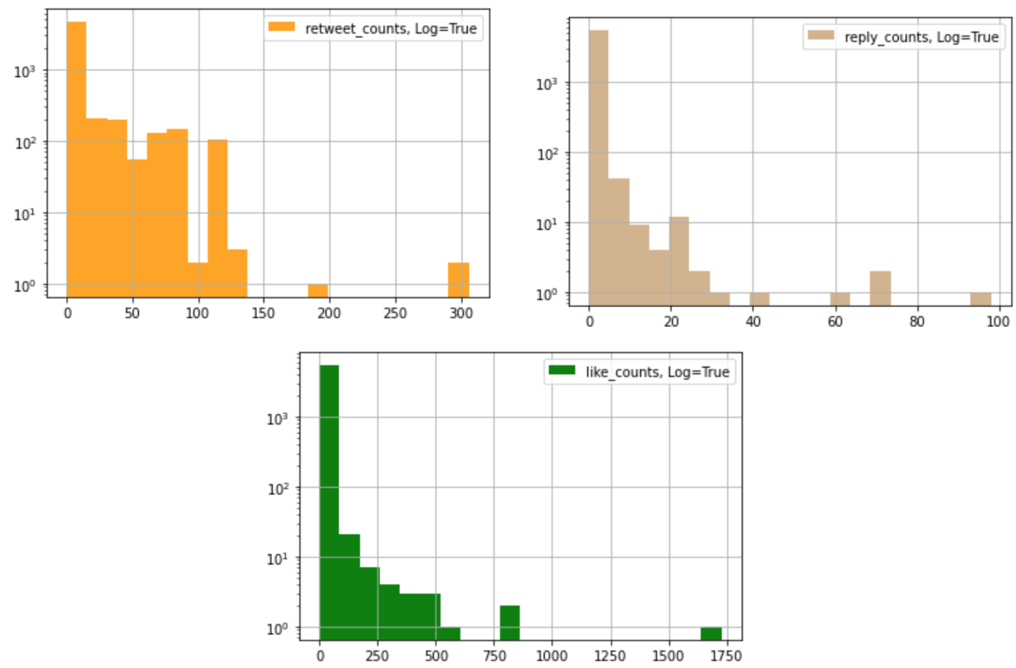

Figure 3: Histograms of Tweet Popularity Measured by Retweet$_{C}ount, Reply_{C}ount, andLike_{C}ount$

To find out the sentiment difference between popular and not-so-popular tweet, a new variable of "popularity" is created. For this research, the condition for a post to be considered popular is to have more then 20 replies, or more than 50 retweets, or more than 250 likes. The following (Figure 4) is the output of comparison between popular and not-so-popular tweets and their compound values.

.

	count	mean	std	min	25%	50%	75%	max
popularity								
popular	405.00	0.02	0.18	-0.90	0.00	0.00	0.00	0.69
not_so_popular	5045.00	-0.05	0.37	-0.97	-0.30	0.00	0.00	0.96

Figure 4: A Comparison of VADER Compound Values Based on Popularity of Tweets

Another comparison of the compound value distribution is done in histogram (Figure 5).

The compound value is between -1 and +1. It shows the values of the not-so-popular tweets are more spreaded out than the popular tweets in their compound values.

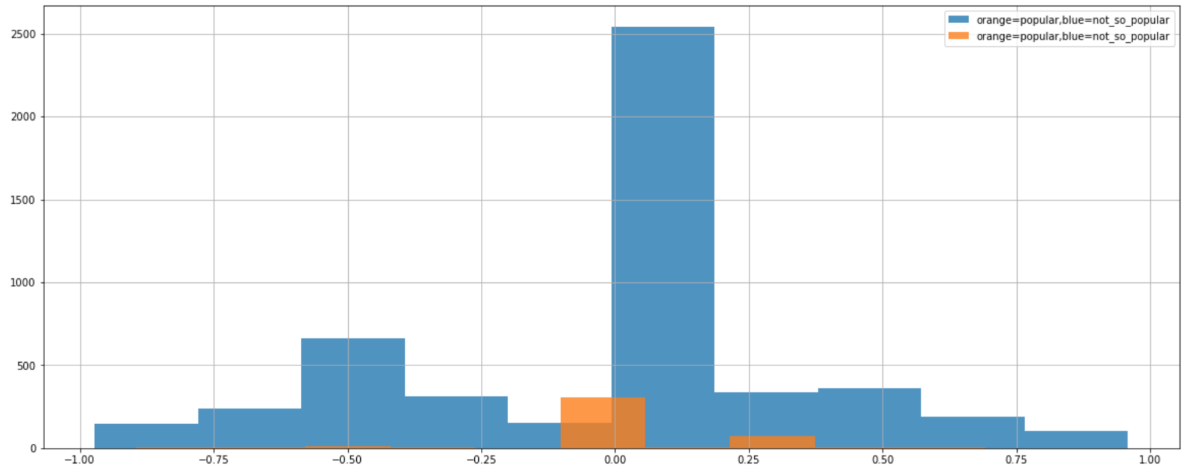

Figure 5: Histogram of Compound Tweets Based on Popularity

6 Discussion

6.1 Sentiment of Employee Surveillance on Social Media

This research answered all three research questions. First of all, the polarity and intensity of tweets sentiment on the subject of employee surveillance does not show a clear pattern based on the time frames. But it does show that the intensity is relatively weak in the past year, compared with the time periods of before pandemic and early-middle pandemic.

Secondly, this research studies the popularity of tweets on the subject of employee surveillance over different time periods. Retweet counts show that the early-middle pandemic period is when tweets on this subject are the most frequently retweeted. Reply counts and like counts show a very similar pattern, with more reply and like on each end of the time periods. Because most people choose to react to a tweet in only one way, not all three, it explains why retweet count and the other two counts are complementary to each other.

The last research question asks if the popularity of tweets make a difference in its sentiment score. The mean compound scores of popular tweets and not-so-popular are very small (less than 0.1). At the same time, it should be noted that popularity of social media posts follow the long tail distribution, and there is a significantly small number of popular tweets (N=405) compared with not-so-popular tweets(N=5045). A statistic comparison of the two groups would be too risky due to large difference in sample size.

6.2 Algorithm Accuracy and Limitations

The results of this research is based on the polarity and intensity compound values calculated in `VADER`. `VADER` is known for being ”quick and computationally economical without sacrificing accuracy” (Hutto and Gilbert,2014). The reported accuracy of `VADER` is on par with other machine learning models that reported accuracy levels at around 80% when conducting sentence-level sentiment analysis (Hutto and Gilbert).

Typically, the first two steps of the sentiment analysis are subjectivity classification and sentence-level sentiment classification. The first determines whether the text is a subjective sentence or an objective sentence. The second step of sentiment analysis determines whether the sentence, if subjective, expresses a positive, negative, or neutral opinion. Unlike `LIWC`, `VADER` does not seek to determine if a sentence is objective or subjective, fact or opinion. Instead, it examines if a sentence expresses a positive, negative or neutral opinion. If the classification is neutral, then the sentence is more likely an objective statement. To check on the accuracy of this analysis, I examined the polarity values of five tweets, and the output is shown below (Figure 6)

```
Tweet1:  School custodian refuses to download phone app that monitors location, says it got her fired .
 VADER Polarity Scores= {'neg': 0.205, 'neu': 0.795, 'pos': 0.0, 'compound': -0.5574}


Tweet2:  As COVID-19’s spread has prompted an expansion of work-from-home policies across various industries, the use
of more-pervasive monitoring software, also known as “tattleware” or “bossware,” has increased.
 VADER Polarity Scores= {'neg': 0.0, 'neu': 0.925, 'pos': 0.075, 'compound': 0.2732}


Tweet3:  Employees Balking At “Tattleware” - Todays General Counsel https://t.co/yVpz900FgD #emplaw #hr #employees
 VADER Polarity Scores= {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}


Tweet4:  @business What a fucking disgraceful advert. Employee surveillance software?
What about judging people by their outputs and contribution, trusting them to do their best? And managing performance
if they fall short.
 Not watching them like factory chickens.
 VADER Polarity Scores= {'neg': 0.05, 'neu': 0.779, 'pos': 0.171, 'compound': 0.7311}


Tweet5:  @CIPD has warned business leaders that surveillance tech can seriously damage the trust between employees an
d employers, but what other ways can you remote manage employee performance? We explore this and much more in our Fut
ure of Work Part II report
 VADER Polarity Scores= {'neg': 0.113, 'neu': 0.838, 'pos': 0.049, 'compound': -0.2144}
```

Figure 6: Manual Test Using 5 Tweets

Based on the results above, the only obvious error is with `tweet4`, which is a strong negative statement itself, but is rated neutral with a positive compound. What is confusing is probably the inclusion of multiple sentences in this tweets and the use of question marks as negative statement. According to Liu and Zhang (2012), sentence-level classification is not suitable for compound and complex sentences. In addition, the lexicon-based approach basically use a list of opinion words and a set of rules to determine the orientations of opinions in a sentence, and one shortcoming is that these opinion words do not cover all types of expressions of opinions in human language (Liu and Zhang,2012).

It’s worth to note that `VADER` sentiment expressions are validated through a wisdom-of-the-crowd (WotC) approach (Hutto and Gilbert, 2014). For example, in Hutto and

Gilbert's research, they relied on human raters available through Amazon Turk to acquire valid point estimate for sentiment valence of about 9,000 lexical feature candidates that are context-free. The validity is guaranteed through screening, training, and data quality checking on crowd-sourced evaluations. The generalizable heuristics they identified include punctuation (exclamation point), capitalization, degree modifiers, contrastive conjunction, and the tri-gram preceding a sentiment-laden lexical feature. While these generalizable heuristics can catch the majority of the sentiment, they fail to interpret the question mark in tweet4. It shows the algorithm does not perform well in circumstances where euphemism and sarcasm are used. In the case of tweet4 or other similar cases, a corpus-based approach for sentiment analysis may be a better choice.

References

Crittenden, A. (1980, June 20). Experts find abuse of employee rights. *The New York Times*, pp.D1,D5.

Hutto, C. J., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, *8*(1), 216-225. Retrieved from `https://ojs.aaai.org/index.php/ICWSM/article/view/14550`

Klosowski, T. (2021, December 10). How your boss can use your remote-work tools to spy on you. *The New York Times*. Retrieved 2022-02-26, from `https://www.nytimes.com/wirecutter/blog/how-your-boss-can-spy-on-you/?searchResultPosition=1`

Kordumova, S., van Gemert, J., & Snoek, C. G. M. (2016). Exploring the long tail of social media tags. In Q. Tian, N. Sebe, G.-J. Qi, B. Huet, R. Hong, & X. Liu (Eds.), *Multimedia modeling* (pp. 51–62). Cham: Springer International Publishing.

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In C. C. Aggarwal & C. Zhai (Eds.), *Mining text data* (pp. 415–463). Boston, MA: Springer US. Retrieved from `https://doi.org/10.1007/978-1-4614-3223-4_13` doi: 10.1007/978-1-4614-3223-4_13

Pierce, L., Snow, D. C., & McAfee, A. (2015). Cleaning house: The impact of information technology monitoring on employee theft and productivity. *Management Science*, *61*(10), 2299–2319.

Westin, A. F. (2003). Social and political dimensions of privacy. *Journal of Social Issues*, *59*(2), 431–453.