The PBSDS: A Dataset for the Detection of Pseudoprofound Bullshit

Evan D. DeFrancesco[1], Carlo Strapparava[1,2]

[1]Università degli Studi di Trento, Italy
[2]Fondazione Bruno Kessler

Abstract
This paper introduces the PBSDS, a dataset of tweets containing pseudoprofound bullshit—statements designed to appear profound but lacking substantive meaning. The PBSDS serves as a resource for studying pseudoprofound bullshit, exploring potential linguistic factors in perceiving bullshit. The dataset's creation and experiments with classifiers show promising results, despite limitations such as selection bias and subjective annotation.

Keywords
pseudoprofound bullshit, stylistic analysis, pragmatics

1. Introduction

"Bullshit" refers to communication that is designed to impress but is constructed without concern for truth [1]. Bullshit differs from lying in that the liar deliberately manipulates and subverts truth (usually with the intent to deceive), while the bullshitter is simply unconcerned with what is true and what is false. A liar needs to know the truth value of a proposition; the bullshitter simply does not care.

Although bullshit comes in different forms, in this project, we focused specifically on what is referred to as "pseudoprofound bullshit," which is designed to convey some sort of potentially profound meaning but is actually semantically vacuous [2], e.g., "Hidden meaning transforms unparalleled abstract beauty." Table 1 reports further examples of pseudoprofound bullshit and non-pseudoprofound bullshit sentences from our dataset.

The goal of this project is to construct a dataset of tweets that contain pseudoprofound bullshit in English (the PBSDS).[1] Operating under the assumption that bullshit is similar to spam email, we hypothesize that it should be possible to detect pseudoprofound bullshit using relatively simple classification algorithms.

Pseudopf BS?	*Sentence*
yes	The unpredictable is a reflection of humble excellence.
no	You must be good to yourself if you are ever going to be any good for others.
yes	The law of attraction is always responding to your thoughts. You are attracting in every moment of your life.
yes	Evolution is an ingredient of subjective excellence.
yes	Our consciousness is a reflection of the door of balance.
no	A garden is a zoo for plants.
no	Scientists are simply adults who retained and nurtured their native curiosity from childhood.

Table 1
Examples of pseudoprofound bullshit and non-pseudoprofound bullshit from the PBSDS.

2. Related work and motivation

Pennycook et al. [2] first explored the psychological nature of pseudoprofound bullshit, establishing an index of bullshit receptivity. They found that a tendency to judge pseudoprofound bullshit statements as profound was correlated with relevant variables such as an intuitive cognitive style and belief in the supernatural. They also found that detecting bullshit was not simply a matter of skepticism but rather of discerning deceptive vagueness in impressive-sounding claims. Walker et al. [3] established a link between illusory pattern perception and the propensity to rate pseudo-profound bullshit statements as profound. Later research by Pennycook and Rand [4] has found that low pseudoprofound bullshit receptivity correlates positively with perceptions of fake news accuracy and negatively with the ability to distinguish fake and real news. Littrell and Fugelsang [5] extended this understanding by exploring individuals' susceptibility to misleading information and its association with re-

CLiC-it 2023: 9th Italian Conference on Computational Linguistics, Nov 30 — Dec 02, 2023, Venice, Italy
evan.defrancesco@studenti.unitn.it (E. D. DeFrancesco); strappa@fbk.eu (C. Strapparava)

[1]The dataset is freely available upon request from the first author.

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

duced engagement in reflective thinking. They found that both highly receptive and highly resistant individuals exhibited limited awareness of their detection abilities for pseudo-profound bullshit. Turpin et al. [6] investigated the influence of different types of titles on the perceived profoundness of abstract art, revealing that pseudo-profound bullshit titles specifically enhanced the perceived profundity of the artwork. Nilsson et al. [7] found an association between pseudoprofound bullshit receptivity and social conservatism and economic progressivism. Relatedly, Evans et al. [8] examined scientific bullshit receptivity, which demonstrated positive correlations with pseudo-profound bullshit receptivity, belief in science, conservative political beliefs, and faith in intuition. They found that scientific literacy moderated the relationship between the two types of bullshit receptivity. These studies collectively shed light on the nature of pseudo-profound bullshit, its reception, and the underlying cognitive mechanisms. However, the development of a dedicated dataset of pseudoprofound bullshit can further facilitate comprehensive investigation and understanding of this phenomenon, contributing to future research endeavors.

Such a dataset could provide researchers with a standardized and reliable resource to study and analyze the phenomenon of pseudoprofound bullshit systematically. It would allow for the exploration of various linguistic, cognitive, and contextual factors that contribute to the perception of profoundness in nonsensical statements. Additionally, an annotated dataset could serve as a benchmark for developing and evaluating computational models and algorithms aimed at detecting and combating pseudoprofound bullshit. It would enable the training and testing of automated systems to recognize and classify instances of pseudoprofound bullshit accurately. This could be instrumental in building tools and technologies to enhance critical thinking, identify deceptive information, and improve media literacy.

3. Data

3.1. Scraping Twitter

We used snscrape[2], an easy-to-use Python package, to crawl the Twitter[3] profiles of six accounts and return the 2,000 most recent tweets from each account. The accounts were scraped on 8 August 2023. We selected accounts that, we hoped, would provide a mix of pseudoprofound bullshit, non-pseudoprofound bullshit, profound philosophy and generic statements. For the initial dataset, we chose accounts that were associated with alternative medicine, pseudoscience, new age spirituality, philosophy and scientific communication. In particular, we scraped the following accounts, from which we collected a total of 12,000 tweets:

- **@DeepakChopra**: Deepak Chopra is a new-age author and alternative medicine promoter. His writing has been described as "incoherent babbling strewn with scientific terms."[4]
- **@WisdomofChopra**: WisdomOfChopra is operated by a bot that produces tweets that are meant to replicate the tone and structure (but not necessarily the content) of Deepak Chopra. The tweets are generated by a simple algorithm: words and phrases are contained within four PHP arrays. The first array contains sentence subjects; the second array contains verb phrases; the third contains determiner phrases and adjectives; the fourth contains nouns. Words and phrases from each array are then combined to generate tweets.
- **@TheSecret**: The Secret's Twitter account is largely composed of messages that promote the pseudoscientific "law of attraction," which claims that positive thoughts attract positive experiences and negative thoughts attract negative experiences.
- **@realNDWalsche**: Neale Donald Walsch is an American new-age writer and speak whose work has appeared in a film version of *The Secret*. His own writing consists primarily of new-age spirituality texts.
- **@kate_manne**: Kate Manne is an associate professor of philosophy at Cornell University. Her research focuses on moral philosophy, metaethics, moral psychology, feminist philosophy and social philosophy. In 2019, Manne was named one of the world's top fifty thinkers.[5]
- **@neiltyson**: Neil deGrasse Tyson is an astrophysicist and science communicator.

We recognize that the decision to include artificially generated content from @WisdomofChopra may be seen as a controversial one. However, the distinction between human and artificial origins of the content was secondary for our purposes. What remained paramount was the essence of the content itself: its pseudoprofound nature.

3.2. Data cleaning

From the initial 12,000 tweets collected, we excluded: duplicate tweets; single-word tweets; tweets that were composed only of hashtags; tweets that were direct replies

[2]https://github.com/JustAnotherArchivist/snscrape
[3]As of 2023, called X.

[4]https://www.washingtonpost.com/news/answer-sheet/wp/2015/05/15/scientist-why-deepak-chopra-is-driving-me-crazy/
[5]https://www.prospectmagazine.co.uk/magazine/prospect-worlds-top-50-thinkers-2019

to other Twitter users; tweets that contained URLs; and tweets that contained emojis. We also removed the hashtag (#) and at-sign (@) from tweets. Finally, we decided to remove tweets that explicitly referenced a personal and individual deity (represented in the tweets as “God"), as we did not wish to cause any inadvertent offense by labelling religious beliefs as pseudoprofound bullshit. After data cleaning, we were left with 5,196 tweets, comprising the initial PBSDS.

3.3. Annotation

Two volunteer annotators provided judgments of whether a tweet constituted pseudoprofound bullshit. The annotators were both students in their mid-20s and were previously not familiar with the concept of pseudoprofound bullshit. The annotators were provided with a working definition of pseudoprofound bullshit (*i.e., statements that sound profound and meaningful but that are actually semantically vacuous; pseudoprofound bullshit may use grandiose terms to deceive people*) as well as several examples of sentences that constituted pseudoprofound bullshit and that did not constitute pseudoprofound bullshit. The working definition was left purposefully vague, given the general difficulty of defining pseudoprofound bullshit. After all, what one person may consider to be pseudoprofound, another person might consider to be actually profound. Annotators were instructed to label the tweet ‘1’ if they believed that it constituted pseudoprofound bullshit and ‘0’ if they did not. Perhaps reflecting the difficulty of arriving at a single sense of pseudoprofound bullshit, Cohen’s kappa was calculated at 0.52, indicating moderate inter-rater reliability [9]. The first author of this paper adjudicated disagreements between the two annotators’ judgments.

4. Dataset description

After annotation, the PBSDS contains 2756 tweets judged as pseudoprofound bullshit (53.04% of the total dataset) and 2440 tweets judged as non-pseudoprofound bullshit (46.96% of the total dataset). Although the two classes are reasonably well-balanced, pseudoprofound bullshit may be disproportionately represented in the dataset compared to its overall occurrence in natural language. However, this is not unexpected, given that the dataset was sourced primarily from Twitter accounts that were likely to include a large amount of pseudoprofound bullshit.

5. Experiments and Results

We trained six machine learning classifiers and compared the performance to test the validity of the dataset. The six classifiers selected for the task were the Support Vector Classifier (SVC), K-nearest Neighbors (KNN), Multinomial Naive Bayes (MNB), Decision Tree Classifier (DTC), Logistic Regression Classifier (LRC) and Random Forest Classifier (RFC). All models were implemented via the scikit-learn library [10].

The tweets were vectorized using tf-idf vectorization, and the data was split into a training set (85%) and a testing set (15%).

In order to evaluate and compare the results of the six classifiers, we used the standard metrics in text classification: Precision (*P*), Recall (*R*), F-score (*F1*) and Accuracy (*Acc*). The results achieved with the six classifiers are reported in Table 2.

Classifier	*P*	*R*	*F1*	*Acc*
SVC	0.9307	0.7943	0.8571	0.8564
KNN	0.8406	0.8227	0.8315	0.8192
MNB	0.9008	0.8156	0.8561	0.8513
DTC	0.8719	0.8203	0.8453	0.8372
LRC	**0.9435**	0.7896	0.8597	0.8603
RFC	0.9309	**0.8274**	**0.8761**	**0.8731**

Table 2
Results obtained from the six classifiers, reported in terms of precision, recall, F-score and accuracy.

6. Limitations

The PBSDS has several limitations that could be addressed in future versions of the dataset. The dataset was collected from specific Twitter accounts presumed to contain pseudoprofound bullshit. This may have resulted in an overrepresentation of pseudoprofound content compared to its overall occurrence in natural language. The dataset thus may not fully capture the range and diversity of pseudoprofound bullshit found in other contexts. Relatedly, the PBSDS’s reliance on tweets from specific Twitter accounts limits its generalizability to other platforms or sources of pseudoprofound bullshit. The characteristics and patterns observed in the dataset may not be representative of pseudoprofound content found elsewhere. Future versions of the PBSDS could address this concern by diversifying the sources of data collection. This would involve not only expanding the range of Twitter accounts under examination but also branching out to other social media platforms, blogs, articles, printed publications and even, perhaps, spoken word content. By incorporating a broader spectrum of sources, the dataset would provide a more comprehensive and varied representation of pseudoprofound bullshit.

Additionally, defining and identifying pseudoprofound bullshit can be challenging and subjective. The annotation process relied on the judgments of two annotators, which may have introduced inherent biases and variations in interpretations. Although efforts were made to establish guidelines, the subjective nature of the task may have affected the consistency of annotations. While the inter-rater reliability between the annotators was measured to be moderate, there was still inherent subjectivity and disagreement in determining whether a tweet constituted pseudoprofound bullshit. The resolution of disagreements by a single adjudicator introduced another layer of subjectivity. Introducing a multi-rater system, in which multiple individuals assess the content's (pseudo)-profundity, could add layers of reliability and objectivity to the dataset.

Finally, the PBSDS comprises 5,196 tweets, which is relatively small in comparison to other text corpora. This limited size may restrict the scope and statistical power of analyses, potentially impacting the generalizability of findings derived from the dataset.

7. Conclusion

Despite its limitations, the PBSDS offers valuable insights into the phenomenon of pseudoprofound bullshit and its detection. The dataset provides a foundation for further research, enabling comprehensive investigations into linguistic patterns, cognitive biases, and societal implications associated with pseudoprofound bullshit. By better understanding and identifying pseudoprofound bullshit, researchers can develop tools and strategies to enhance critical thinking, combat deceptive communication, and promote media literacy in an increasingly complex information landscape.

Acknowledgments

We acknowledge the support of the PNRR project FAIR - Future AI Research (PE00000013), under the NRRP MUR program funded by the NextGenerationEU.

References

[1] H. G. Frankfurt, On bullshit, in: On Bullshit, Princeton University Press, 2005.

[2] G. Pennycook, J. Allan Cheyne, N. Barr, D. J. Koehler, J. A. Fugelsang, On the reception and detection of pseudo-profound bullshit, Judgment and Decision Making 10 (2015) 549–563. doi:10.1017/S1930297500006999.

[3] A. C. Walker, M. H. Turpin, J. A. Stolz, J. A. Fugelsang, D. J. Koehler, Finding meaning in the clouds: Illusory pattern perception predicts receptivity to pseudo-profound bullshit, Judgment and Decision Making 14 (2019) 109–119.

[4] G. Pennycook, D. G. Rand, Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking, Journal of Personality 88 (2020) 185–200.

[5] S. Littrell, J. A. Fugelsang, Bullshit blind spots: The roles of miscalibration and information processing in bullshit detection, Thinking & Reasoning (2023) 1–30.

[6] M. H. Turpin, A. C. Walker, M. Kara-Yakoubian, N. N. Gabert, J. A. Fugelsang, J. A. Stolz, Bullshit makes the art grow profounder, Judgment and Decision making 14 (2019) 658–670.

[7] A. Nilsson, A. Erlandsson, D. Västfjäll, The complex relation between receptivity to pseudo-profound bullshit and political ideology, Personality and Social Psychology Bulletin 45 (2019) 1440–1454.

[8] A. Evans, W. Sleegers, Ž. Mlakar, Individual differences in receptivity to scientific bullshit, Judgment and Decision Making 15 (2020) 401–412.

[9] J. R. Landis, G. G. Koch, The measurement of observer agreement for categorical data, Biometrics (1977) 159–174.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn: Machine learning in python, The Journal of Machine Learning Research 12 (2011) 2825–2830.