

8-10-2023

The Language of Privacy Policies: Uncovering Hidden Meaning with NLP

Khole Gwebu
University of New Hampshire, khole.gwebu@unh.edu

Jing Wang
University of New Hampshire, jing.wang@unh.edu

Lin Guo
Old Dominion University, lguo@odu.edu

Chuanyi Tang
Old Dominion University, ctang@odu.edu

Follow this and additional works at: https://aisel.aisnet.org/treos_amcis2023

Recommended Citation

Gwebu, Khole; Wang, Jing; Guo, Lin; and Tang, Chuanyi, "The Language of Privacy Policies: Uncovering Hidden Meaning with NLP" (2023). *AMCIS 2023 TREOs*. 5.
https://aisel.aisnet.org/treos_amcis2023/5

This material is brought to you by the TREO Papers at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2023 TREOs by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

The Language of Privacy Policies: Uncovering Hidden Meaning with NLP

TREO Talk Paper

Jing Wang

University of New Hampshire
jing.wang@unh.edu

Lin Guo

Old Dominion University
lin.guo@odu.edu

Kholekile Gwebu

University of New Hampshire
khole.gwebu@unh.edu

Chuanyi Tang

Old Dominion University
chuanyi.tang@odu.edu

Abstract

Privacy policies remain the single most important reference point for users to find out how their data is collected, stored, used, and shared (Jensen, & Potts, 2004). Companies thus have a moral obligation to ensure that both elements of the “Notice and Choice” framework are handled ethically. The Notice element is deemed ethical only if the privacy policy’s content (i.e., the privacy policy fully informs the key company practices in all four stages of the data consumption life cycle: collection, storage, usage, and sharing) as well as its presentation (i.e., the privacy policy is presented in a transparent, fair, and easy to comprehend manner) are handled ethically. Similarly, the Choice element is deemed only ethical if the privacy policy’s content (i.e., the privacy policy includes choices/controls for users to opt-in or opt out of a company’s data collection, usage, storage, and sharing practices) as well as its presentation (i.e., such choices are presented in a transparent, fair, and easy-to-comprehend manner) are handled ethically.

Currently, it remains unknown if privacy policies are handled consistently and ethically across companies. To fill this important void in the literature, this study focuses on both the content and presentation of companies’ privacy policies and employs Natural Language Processing (NLP) and machine learning techniques to examine the extent to which Fortune 500 companies adhere to the Notice and Choice framework.

This research utilizes several data analytics approaches, such as text classification, readability analysis, reading time estimation, and NLP to process and analyze a large corpus of data. It adds to what is known about privacy policies and data protection by shedding light on how privacy policies are written and how they are constructed. The study also highlights how clear and open privacy policies are, as well as how they could be improved to encourage informed consent.

References

Jensen, C., & Potts, C. (2004, April). Privacy policies as decision-making tools: an evaluation of online privacy notices. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems (pp. 471-478).