THE JOHN PAUL II CATHOLIC UNIVERSITY OF LUBLIN
FACULTY OF SOCIAL SCIENCES
INSTITUTE OF PSYCHOLOGY
EMOTION AND MOTIVATION PSYCHOLOGY DEPARTMENT
Al. Racławickie 14, 20-950 Lublin, POLAND

Lublin, 5 September 2025

dr hab. Agata Błachnio, prof. KUL

Institute of Psychology

The John Paul II Catholic University of Lublin

Review of the doctoral dissertation of Hubert Wojciech Plisiecki, MSc, entitled *Words, Vectors, and Feelings: Advancing Psychological Emotion Research Through Natural Language Processing*, prepared under the supervision of Dr hab. Artur Pokropek and with auxiliary supervision by Dr Grzegorz Pochwatko

The doctoral dissertation submitted for my review by MSc Hubert Wojciech Plisiecki comprises four articles. Three of them have been published in peer-reviewed journals with impact factors ranging from 2.9 to 3.9 and Polish ministerial scores between 70 and 140, while one single-authored article has been posted on arXiv. The author's contribution to all texts was leading and essential. The articles were published between 2023 and 2025, and in each case Hubert Plisiecki is the first author. The supervisor is a co-author of one article; the auxiliary supervisor is not a co-author of any. One of the studies was funded by an NCN SONATA BIS grant led by the supervisor. The submitted articles form a coherent research program that fills a gap in the psychology of emotion literature and introduces new NLP (Natural Language Processing) methods tailored to the needs of psychology, addressing domain-specific concerns

such as ecological validity, construct bias, and measurement fidelity. This supports the conclusion that the presented line of research has been published in reputable international journals and that the doctoral candidate made a substantial contribution to its development.

The dissertation addresses a highly timely and innovative topic, the application of Natural Language Processing (NLP) and Machine Learning (ML) methods in the psychology of emotion, and has a distinctly interdisciplinary character. It is worth emphasizing that the author effectively links computational methodology with psychological questions. In the studies, the author employs advanced ML techniques (word embeddings, transformer models, and graph neural networks (GNNs)). I assess the work very positively, both in terms of the research concept, its execution, and the presentation of results. The set of studies forms a coherent whole: each addresses a specific question, and together they constitute a logical program. The findings are consistent with and well explained in light of the psychological literature; they can be regarded as a contribution to knowledge in emotion psychology and also broaden the perspective on tools for studying emotions.

A number of strengths can be identified in the proposed studies, which underpin the positive conclusion of this review. It should be emphasized that the driving force behind the research goals was the limitation of prior psychological work using NLP, which often borrowed off-the-shelf tools from computer science and overlooked psychological nuance. The author proposes NLP methods tailored to the needs of psychology. In doing so, Hubert Plisiecki fills a gap by introducing methodological solutions specifically designed for psychological research.

The doctoral thesis is written very clearly. The research aims are presented on the basis of a solid literature review that covers both classic tools (LIWC, GI) and contemporary methods (Word2Vec, BERT, LLM), which highlights the continuity of developments in the psychology of language and situates the empirical results in a coherent context. There is no doubt that key strengths of the work include its interdisciplinarity and the proposal of computer science tools tailored to psychological questions, with due regard for the field's specificity and methodological rigor.

Particularly valuable is the reconstruction of the main dimensions of affect in textual data: the first two components correspond to valence and arousal, while the dominance dimension appears only partially, something the author clearly discusses, together with the interpretational limitations of the remaining components.

Next, the author proposes a model that can assess the emotional load of new words with very high agreement. This helps address the shortage of normative data and provides a simple tool ("stimuli-descent") for selecting pairs of similar words that differ only on the target dimension, which makes experimental planning easier.

Deserving particular attention is a rarely seen, multi-layered audit of political bias. The study demonstrates the presence of biases in emotion-inference models used for sentiment analysis (SA). These biases stem from annotators' personal attitudes, which the model "inherits." The author audits a Polish SA model, showing systematic differences in valence predictions depending on political affiliation, and then shows that removing mentions of those politicians reduces the bias (though does not completely eliminate). On this basis, the author recommends

caution when using such models in the social sciences, critical scrutiny of SA results, and consideration of lexicon-based solutions as a more ideologically neutral alternative. Such procedures strengthen the reliability of measurement and inference in psychological research.

An original contribution is the introduction of the SProp GNN architecture, an explainable sentiment analysis model that relies solely on syntactic structure and word-level emotional cues. Thanks to semantic blinding, the model is "blind" to information about specific words, which makes it more resistant to social biases. SProp GNN outperforms lexicon-based solutions (e.g., VADER, EmoAtlas) across two tasks and two languages, while approaching the accuracy of transformer-based models; it significantly reduces biases and enables tracing the propagation of emotional information, thus bridging the gap between the interpretability of lexicon-based approaches and the powerful yet opaque deep models.

I appreciate that the author does not stop at an enthusiastic presentation of NLP's potential, but consistently calibrates the conclusions against the methods' limitations. In the first article, he explicitly emphasizes the exploratory nature of the analyses and the low yet significant correlation with valence norms ($r = 0.31$). In the second article, he treats extrapolated norms as a heuristic rather than a substitute for empirical data, and also introduces the stimuli-descent algorithm for selecting semantically matched stimuli. In the third article, he presents a bias audit conducted on a Polish sentiment analysis model and recommends replication before deployment, whereas in the fourth he proposes a bias-safe architecture that limits the transfer of annotators' biases into model predictions. The author also explicitly discusses the noise and context dependence of textual data, the distinction between author and reader perspectives, the need for

rigorous validation and cross-linguistic replication, and the risks of inferring properties of individuals from public texts. Thanks to these measures, the overall narrative remains methodologically cautious, and the work has high practical value for experimental psychology, providing tools for stimulus selection as well as procedures for identifying and controlling biases.

It is also worth noting several limitations, which in no way detract from my overall positive assessment. First, in the author's summary it would be helpful to include a separate section devoted to limitations, directions for future research, and practical implications.

Second, the behavioral validation in the first article is primarily correlational (agreement with valence norms $r \approx 0.31$); beyond t-SNE visualizations and logistic regression on the GoEmotions corpus, no prospective human studies were conducted to demonstrate that the models' predictions translate into behaviors or ratings in new tasks (e.g., speeded emotion categorizations, intensity judgments, reaction times).

Third, the generalizability of the conclusions requires verification on other corpora and in different cultural contexts: replications are needed in other languages and beyond Reddit. Although Article 2 reports strong results in several languages and Article 4 covers two languages, the current conclusions rely largely on GoEmotions (English, Reddit) and on Polish political data, which limits their direct transfer to other platforms, registers, and languages. Because norms of emotional expression and lexical connotations vary across languages and dialects, further cross-domain and cross-cultural replications are necessary.

Finally, in the second article the author reports an average performance drop of about 11% for words that deviate substantially from the training distribution, indicating weaker model performance on rare or novel vocabulary. The work does not include tests of temporal stability (data drift) or broader robustness evaluations, which is a natural avenue for further validation.

In sum, the dissertation is mature, ambitious, and valuable. It combines a solid theoretical foundation with original methodological solutions and delivers findings of clear cognitive and applied significance. The work opens new avenues of research at the intersection of psychology and computer science. The presented series of publications forms a coherent whole, both logically and substantively. The dissertation is interdisciplinary and highly timely, which increases its significance and potential impact. The author has carried out a very ambitious project distinguished by high research quality. Hubert Plisiecki has shown that he can effectively plan and conduct studies grounded in the literature, analyze results, and, moreover, successfully publish them in reputable international journals. A review of available databases indicates that the presented article series is not the doctoral candidate's only scholarly output; within just three years he has published over a dozen papers in international venues. It is worth emphasizing that in the articles presented the candidate played a key role at every stage of the research process, from developing the concept, through conducting the studies and analyzing the results, to publishing them in high-quality scientific journals. This attests to his advanced research skills and professionalism. Although he is at the beginning of his academic career, his activity demonstrates strong potential and a clear commitment to its further development.

THE JOHN PAUL II CATHOLIC UNIVERSITY OF LUBLIN
FACULTY OF SOCIAL SCIENCES
INSTITUTE OF PSYCHOLOGY
EMOTION AND MOTIVATION PSYCHOLOGY DEPARTMENT
Al. Racławickie 14, 20-950 Lublin, POLAND

Taking into account all of the above, I find that the doctoral dissertation presented as a series of four articles meets the requirements set out in Article 187 of the Act of 20 July 2018 – Law on Higher Education and Science (Journal of Laws of 2021, items 478, 619, 1630). I therefore recommend that the Scientific Council of the Institute of Psychology of the Polish Academy of Sciences admit the candidate to the subsequent stages of the doctoral proceedings.