Non-Archival Abstract

A Window to the Mind: Analyzing Online Communication to Improve Child and Adolescent Mental Health Care

Jakob Prange[1] **Yu-Yin Hsu**[2] **Jan Pfister**[3]
Andreas Hotho[3] **Kerstin Paschke**[1]

[1]German Center for Addiction Research in Childhood and Adolescence (DZSKJ), University Medical Center Hamburg-Eppendorf (UKE)
[2]The Hong Kong Polytechnic University
[3]Data Science Chair & Center for Artificial Intelligence and Data Science (CAIDAS), Julius-Maximilians-Universität Würzburg
Correspondence: j.prange@uke.de

Much discourse related to youth's mental health crises takes place online. Forums like Reddit offer information access, community support, and customizable degrees of anonymity. The coping strategies discussed there also include unhealthy ones like excessive digital media use and substance use, which carry substantial mental and physical health risks themselves. In this ongoing work, we analyze language patterns in online communication about mental health and substance use and develop methods for early risk detection, in order to assist medical and socialwork professionals in delivering the best public education, prevention, and health care possible.

A dataset of algospeak and mental-health-related online communication. To limit the propagation and glorification of dangerous content such as suicide or substance abuse, online platforms employ content moderation systems. However, rather than removing harmful content entirely, this approach has instead accelerated the development of coded *algospeak*. In English, evasion strategies include phonetic respelling (e.g. "sewerslide" for *suicide*) and neologisms (e.g. "unalive" for *kill/die*). In German, productive compounding may be used to embed substance terms within novel lexical items like "Hustensaft-party" ('cough syrup party', referring to drinks containing codeine, an opiate). Across languages morphemes, words, or entire sentences can be substituted with emojis, e.g. the leaf 🍃, referring to cannabis, or the snowflake ❄️, referring to cocaine. Algospeak often not only succeeds in hiding stigmatized conversation topics from platform-internal moderation. It also increases the chance of at-risk children and adolescents falling through the gaps of public mental health support systems that have not yet adopted to the digital spaces in which they find comfort.

To bridge this support gap, we construct a multilingual map of algospeak patterns through systematic corpus collection and annotation from English and German sources. Starting from existing domain-specific English-language Reddit data sets annotated with substance use and mental health categories (Yang et al., 2024; Parapar et al., 2023), we experiment with automatic translation into German. Translation quality is monitored by humans and algospeak-related translation errors are recorded. In parallel, we collect a German Reddit corpus or adapt an existing one (Blombach et al., 2020), annotate it from scratch for evasion mechanism type, clinical risk, and annotator confidence. We then compare annotation reliability under cross-linguistic and coded-language conditions.

Risk detection. Our pilot experiments prior to the algospeak resource construction, following the CLEF eRisk methodology for gambling addiction risk detection (Parapar et al., 2023), show that both lexical (n-gram-based) and LLM-based classifiers are promising, achieving ROC-AUC values above 95%. Often few posts are enough to identify risky tendencies early. At the same time, these experiments reveal the importance of removing noise from the training data, and focusing on generalization to potentially sparse signals beyond the lexical level. Automatic risk detection systems are expected to profit substantially from algospeak pattern recognition.

References

Andreas Blombach, Natalie Dykes, Philipp Heinrich, Besim Kabashi, and Thomas Proisl. 2020. A corpus of German Reddit exchanges (GeRedE). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6310–6316, Marseille, France. European Language Resources Association.

Javier Parapar, Patricia Martín-Rodilla, David E. Losada, and Fabio Crestani. 2023. Overview of erisk 2023: Early risk prediction on the internet. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, pages 294–315, Cham. Springer Nature Switzerland.

Chenghao Yang, Tuhin Chakrabarty, Karli Hochstatter, Melissa Slavin, Nabila El-Bassel, and Smaranda Muresan. 2024. Identifying self-disclosures of use, misuse and addiction in community-based social media posts. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 2507–2521, Mexico City, Mexico. Association for Computational Linguistics.