Development, validation, and implementation of technology-based tools to study and address substance use and related conditions

Authors: *Tingting Liu, Salvatore Giorgi, Andrew Schwartz, Lyle Ungar, Brenda Curtis*

Incompletion of the treatment for substance use disorder (SUD) reduces the effectiveness of the treatment and increases the chances of relapse and negatively affects individuals socially, psychologically, and physically. Therefore, high-accuracy predictions of SUD treatment dropout and retention are needed. Using the social media language of patients, we developed and validated the predictive capability of an AI-based digital phenotype for SUD treatment outcomes, and further identified the linguistic features that have psychosocial relevance to SUD treatment outcomes.

Methods: Two studies were conducted using patients' Facebook posts two years before attending SUD outpatient treatment. A Deep Learning based AI model, Bidirectional Encoder Representations from Transformers (BERT) were applied and compared against the Addiction Severity Index (ASI v6) in Study 1 ($N = 269$). The Linguistic Inquiry and Word Count (LIWC) lexicon, Latent Dirichlet Analysis (LDA) topics, and AI-models of religiosity, affect, and temporal orientation were applied in Study 2 ($N = 206$; 55,415 posts).

Results: We found that these AI-based predictions outperformed the predictions based on a standard psychometric assessment scale taken at intake, the ASI (AUC of 0.725 versus AUC of 0.658; single-tailed permutation test $p < 0.001$). Linguistic analysis showed that patients who dropped out before 90 days discussed more diverse topics, focused on the past, and used more articles; while patients who stayed in the stayed in treatment talked more about religion, positive emotions, family, and the present, and used more first-person singular pronouns.

Discussion: The current findings demonstrates that social media language has strong validity as a risk assessment tool, either independently or combined with the standard psychometric assessment, to predict SUD treatment outcomes, and can be used as additional screener for personalizing treatment plans upon SUD treatment entry.

References: Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
Cacciola, J. S., Alterman, A. I., Habing, B., & McLellan, A. T. (2011). Recent status scores for version 6 of the Addiction Severity Index (ASI-6). *Addiction*, *106*(9), 1588-1602.
Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, *3*, 993-1022.