Discursive Representations of Loneliness in Human-Authored and AI-Generated Texts: A Lexical Multidimensional Analysis

This study investigates discursive representations of loneliness in human-authored and AI-generated texts using lexical multidimensional analysis (LMDA; Berber Sardinha & Fitzsimmons-Doolan, 2025). LMDA is an extension of the original MD framework introduced by Biber in the 1980s (Biber, 1988). Instead of using grammatical features as in the original formulation, LMDA uses lexical features only, and the resulting dimensions are interpreted as capturing the underlying discourses around a particular issue as represented in a corpus. In this case, we are interested in the discourses around loneliness; more specifically in how loneliness is represented in self-disclosure posts in specialized subreddits. For our understanding of loneliness, we rely on previous work that conceptualizes loneliness as a socially mediated and relational phenomenon (Ågren, 2017). Most prior work on AI and emotion expression focuses on empathetic responding rather than self-expression of an emotional state (Welivita & Pu, 2023). Typical tasks involve an AI replying empathetically to a user who expresses sadness, distress, or loneliness; the AI responses are analyzed usually via human ratings. In this project, however, we take a variation-based corpus perspective to the expression of loneliness by both human beings and AI through the linguistic analysis of the latent discourses expressed when human beings and machines talk about loneliness. Instead of relying on human ratings, we rely on the dimensions revealed by LMDA to detect how human beings and AI construct the expression of loneliness. To this end, three subcorpora are compiled: a corpus of human-authored Reddit posts, and two AI-generated corpora produced under different prompting conditions: one guided by summaries of the human posts and one generated from a generic prompt. To conduct the LMDA, we first tag each corpus for POS and extract content words, whose counts are used for keyword extraction by comparing each subcorpus against the other subcorpora. The resulting keywords are tabulated, and these counts are entered in a factor analysis. The resulting factors are interpreted as dimensions around loneliness. In the presentation, we will introduce and illustrate the dimensions, and will interpret and discuss the variation between the human and AI discourses on loneliness.

LONG ABSTRACT FOR YOUR REFERENCE ONLY; KEEP IT FOR LATER BUT DON'T SUBMIT THIS

Discursive Representations of Loneliness in Human-Authored and AI-Generated Texts: A Lexical Multidimensional Analysis

This study investigates discursive representations of loneliness in human-authored and AI-generated texts using lexical multidimensional analysis (LMDA; Berber Sardinha & Fitzsimmons-Doolan, 2025). LMDA is an extension of the original MD framework introduced by Biber in the 1980s (Biber, 1988). Instead of using grammatical features as in the original formulation, LMDA uses lexical features only, and the resulting dimensions are interpreted as capturing the underlying discourses around a particular issue as represented in a corpus. In this case, we are interested in the discourses around loneliness; more specifically in how loneliness is represented in self-disclosure posts in specialized subreddits. For our understanding of loneliness, we rely on previous work that conceptualizes loneliness as a socially mediated and relational phenomenon (Ågren, 2017; Cover, 2016; Imrie, 2018; Oakley, 2020; Barreto et al., 2024), which contrasts with classical psychological accounts that frame it primarily as an individual affective state (Weiss, 1973; Perlman & Peplau, 1982). We are also interested in how AI constructs discourses around loneliness. Previous research has shown that AI can take on human personas, thereby displaying particular human characteristics such as feelings and emotions (Jiang et al., 2023). This goes beyond the implementation of AI as intelligent machines: AI now becomes a feeling machine. We want to know to what extent AI can generate texts that express loneliness. If so, then the AI can be seen as being closer to a human being. Most prior work on AI and emotion expression focuses on empathetic responding rather than self-expression of an emotional state (Welivita & Pu, 2023). Typical tasks involve an AI replying empathetically to a user who expresses sadness, distress, or loneliness. In those studies, loneliness is treated as input context, not as the *voice* or *stance* the model must inhabit. Consequently, the output is evaluated for appropriateness, supportiveness, or perceived empathy, usually via human ratings. In this project, however, we take a corpus perspective to the expression of loneliness by both human beings and AI through the linguistic analysis of the latent discourses expressed when human beings and machines talk about loneliness. Instead of relying on human ratings, we rely on the dimensions revealed by LMDA to detect how human beings and AI construct the expression of loneliness. To do that, we collect three subcorpora: a corpus of Reddit posts (the human corpus), and two AI-generated corpora based on two prompting conditions – one where the AI was presented with a prompt that included a summary of each human post and asked to generate a Reddit post based on that summary, and another where the AI was only told to generate a Reddit post on loneliness. To conduct the LMDA, we first tag each corpus for POS and extract content words, whose counts are used for keyword extraction by comparing each subcorpus against the other subcorpora. The resulting keywords are tabulated, and these counts are entered in a factor analysis. The resulting factors are interpreted as dimensions around loneliness. In the presentation, we will introduce and illustrate the dimensions, and will interpret and discuss the similarities and differences between the human and AI discourses on loneliness.

Keywords: Loneliness; Lexical Multidimensional Analysis; Corpus Linguistics; AI-generated discourse; Register variation.

Berber Sardinha, T., & Fitzsimmons-Doolan, S. (2025). *Lexical Multidimensional Analysis: Identifying Discourses and Ideologies*. Cambridge University Press.

Biber, D. (1988). *Variation Across Speech and Writing*. Cambridge University Press.

Jiang, H., Zhang, X., Cao, X., Kabbara, J., & Roy, D. (2023). PersonaLLM: Investigating the Ability of GPT-3.5 to Express Personality Traits and Gender Differences. (arXiv:2305.02547). Retrieved May, from https://arxiv.org/abs/2305.02547

Welivita, A., & Pu, P. (2023). *Empathetic Response Generation for Distress Support* Proceedings of the SIGDIAL 2023 Conference, Prague, Czech Republic. https://aclanthology.org/2023.sigdial-1.59/

https://submit.jotform.com/260077964021051