Background
Humor is one of the most creative and socially embedded forms of human communication, reflecting both emotional expression and cultural interaction. With the rapid development of generative artificial intelligence, the differences between human and AI humor have become an important issue in computational humor research. However, existing comparative studies on human and AI humor face two major limitations: first, most rely on small-scale corpora or subjective ratings, making it difficult to uncover systematic linguistic differences; second, even when linguistic features are considered, prior analyses are often impressionistic or confined to a single dimension, failing to explain the deeper mechanisms underlying human-AI humor differences. Without systematic research, we cannot determine how far AI still is from human humor, which may result in continued constraints on its ability to replicate and simulate human humor.

Aims
This study aims to adopt a quantitative linguistics approach, using a data-driven paradigm to systematically compare human- and AI-generated humor across multiple linguistic dimensions, including lexical features, syntactic structures, sentiment, semantic patterns, and prosodic features.

Samples
A large-scale parallel humor corpus was first constructed, consisting of 22,000 human-authored jokes (sourced from Reddit, short jokes, and puns) and 66,000 AI-generated jokes (produced by GPT-4.1-mini, Llama-3.3-70B, and DeepSeek-V3-Chat).

Methods
An automated program was developed to extract and quantify linguistic indices, and Elastic Net logistic regression was employed to identify the most discriminative linguistic features between human and AI humor. This approach overcomes the shortcomings of earlier studies with respect to corpus size and insufficient data-driven methods.

Results
The results show that AI-generated jokes rely more heavily on connectives and retrospective narration. By contrast, human-authored jokes exhibit greater lexical density, conditional subordination, emotional diversity, semantic incongruity, and thematic consistency, as well as more frequent use of nouns, interrogatives, social references, affective markers, and rhythmic devices. These differences highlight the significant gap between human and AI humor in terms of creativity, emotional depth, and prosodic design.

Conclusions
These findings reveal important linguistic differences between human and AI-generated humor, offering empirical insights that may support the theoretical study of humor mechanisms. The study hopes to contribute both to the theoretical validation of humor mechanisms and to the practical advancement of explainable and human-like humor generation in LLMs.