Comprehensive Evaluation of Emotional Intelligence in Large Language Models: A Dual-Focus on General Emotional Awareness and Personalization for Personality Structures

Background: This study, referencing Elyoseph et al., (2023a), and Hadar-Shoval et al., (2023) along with Elyoseph et al., (2023b), presents an in-depth analysis of the emotional intelligence of Large Language Models (LLMs) like ChatGPT-3.5 and ChatGPT-4. It uniquely distinguishes between LLMs' general emotional awareness capabilities and their personalized mentalizing abilities tailored to specific personality structures.

Methods: A multi-pronged approach evaluates key aspects of emotional intelligence. The Levels of Emotional Awareness Scale (LEAS) comprehensively examines ChatGPT-3.5's general emotional awareness benchmarked against human norms. Further personalized assessments probe ChatGPT-3.5's capacity to tailor responses to distinct personality constructs, scrutinizing its flexibility. Additionally, the widely recognized Reading the Mind in the Eyes Test (RMET) compares visual emotional recognition abilities in ChatGPT-4 versus Google Bard to gauge their proficiency in non-verbal cues. This dual verbal and visual testing toolkit provides rigorous insight into both general and specialized dimensions of emotional discernment across state-of-the-art LLMs.

Results: ChatGPT-3.5 significantly outperforms human norms in general emotional awareness, with a deviation exceeding 4 standard deviations above the mean. In personalizing responses for personality structures, it demonstrates sophisticated adaptability. ChatGPT-4's RMET performance aligns closely with human levels, scoring between 26-27, contrasting with Google Bard's lower proficiency.

Discussion: The studies collectively underscore the advanced emotional intelligence of LLMs. The contrast between ChatGPT-3.5's general emotional awareness and its targeted responses for personality structures highlights its flexibility and potential in nuanced human-AI interactions. Furthermore, ChatGPT-4 demonstrates proficiency in visual-emotional recognition on par with humans, as evidenced by its RMET performance. This is in contrast to Google Bard's lower scores, indicating there is still room for improvement. ChatGPT-4's well-rounded emotional capabilities across verbal and visual modalities underscore the rapid progress of LLMs toward human-level emotional aptitude. This research opens pathways for empathetic AI applications, particularly in mental health, necessitating further exploration into the ethical and practical implications of emotionally-aware AI in sensitive human-centric domains.

Referenze:
Elyoseph, Z., Hadar-Shoval, D., Asraf, K., & Lvovsky, M. (2023). ChatGPT outperforms humans in emotional awareness evaluations. Frontiers in Psychology, 14, 1199058.
Hadar-Shoval, D., Elyoseph, Z., & Lvovsky, M. (2023). The Plasticity of ChatGPT's Mentalizing Abilities: Personalization for Personality Structures. Frontiers in Psychiatry, 14, 1234397. Elyoseph, Z., Refoua, E., Asraf, K., Lvovsky, M., Shimoni, Y., & Hadar-Shoval, D. Can Large Language Models" Read Your Mind in Your Eyes"?.

Autore 1
Nome: Zohar
Cognome: Elyoseph
Affiliazione: The Max Stern Yezreel Valley College

Autore 2
Nome: Dorit
Cognome: Hadar-Shoval
Affiliazione: The Max Stern Yezreel Valley College