Introducing the Persuasive AI Transparency (PAIT) Model: The Effects of Explainability and Interpretability in Advertising

The technological advancements and increased accessibility and implementation of AI in advertising and marketing have fueled the debate around AI transparency. Given the introduction of new transparency concepts, as well as new objects of transparency in persuasion beyond the persuasive intent of the message, it is time to revisit the concept of transparency in the context of Persuasive AI. The current work aims to 1) introduce the Persuasive AI Transparency model, 2) provide conceptual clarity around AI transparency, including explainability and interpretability, 3) introduce the FORMAT typology that further explicates extrinsic explainability, and 4) provides a research agenda for future work on this area. The work will guide future research to systematically unravel and build our theoretical understanding of the implications of persuasive AI transparency. Furthermore, the model has implications for advertisers/tech industry that implement AI disclosures and guide legislative frameworks on transparency, as well as implications for consumer literacy.

Keywords: Artificial Intelligence (AI), Transparency, Explainability, Interpretability, Persuasion