C34

Maissa Trabilsy MA, BA Advisor(s): Antonio Forte Ph.D.,M.D.

Co-author(s): -

Utilizing Generative Text to Image AI Models to Explore Race, Sex, and Age in Aesthetic Plastic Surgery

It is unclear how representative and inclusive of various patient populations text to image AI models are. Therefore, this project explores the diversity of race, gender, and age in the images generated by AI models: DALL-E3, Midjourney, and Adobe Firefly, in response to prompts focused on popular aesthetic procedures: liposuction, blepharoplasty, and rhinoplasty.

Prompts were designed to ask each AI model to generate images of surgical outcomes for liposuction, blepharoplasty, and rhinoplasty for each gender, race and age combination: male vs. female, Caucasian or white, Black or African American, Latino or Hispanic, and age groups: 20-30 years, 31-45 years, and 46+ years. Each generated image was evaluated for representation of skin color by Fitzpatrick and Monk scales, and sex parity using a 4-item questionnaire. The Kruskal-Walis test was used for the overall comparison of continuous variables between the 3 models ($p<0.05$) and the Wilcoxon rank sum test for pairwise comparisons ($p<0.017$, after adjustment based on the Bonferroni method for multiple comparisons). The Fischer's east test was used for the overall comparison of categorical variables between the 3 models ($p<0.05$) as well as pairwise comparisons ($p<0.017$).

There was no significant difference between representation of light skin color (Fitzpatrick I-III & Monk 1-5) vs. dark skin color (Fitzpatrick IV-VI & Monk 6-10) between the 3 AI text to image generative models ($p=0.26$ & $p=0.31$). A significant difference was found generally between all 3 AI models ($p<0.0001$) as well as when comparing female vs male ($p=0.0009$) regarding the depiction of aging.

There appears to be inclusivity and fair representation of light skin colors and dark skin colors, but there is still room for improvement regarding the depiction of gender bias.