

Untersuchung KI-generierter Bilder zur Evaluation von Faktoren für die Verbesserung von Text-to-Image-Generation

Thema:

Untersuchung KI-generierter Bilder zur Evaluation von Faktoren für die Verbesserung von Text-to-Image-Generation

Art:

BA

BetreuerIn:

Michael Achmann

BearbeiterIn:

Julia Niedermaier

ErstgutachterIn:

Niels Henze

ZweitgutachterIn:

N.N.

Status:

abgeschlossen

Stichworte:

Text-to-Image-Generation, Generative Modelle, HCI, Bildwahrnehmung

angelegt:

2023-01-10

Antrittsvortrag:

2023-01-23

Abgabe:

2023-06-20

Hintergrund

Künstliche Intelligenzen haben mit Sprachassistenten wie Siri oder Alexa längst Einzug in den Alltag der Gesellschaft erhalten. Bereits Mitte des letzten Jahrhunderts wurde der Begriff künstliche Intelligenz von John McCarthy mit dem Ziel Maschinen zu entwickeln, die wie Menschen lernen und denken können erstmals in Verbindung gesetzt [1]. Neben großen Fortschritten in Bereichen wie Robotik und Natural Language Processing gibt es ebenso voranschreitende Prozesse im Bereich der generativen Modelle, insbesondere der Text-to-Image-Generation. Dabei werden auf Basis von Texteingaben in natürlicher Sprache Bilder erzeugt, die den eingegebenen Satz widerspiegeln. State-of-the-art Modelle wie Parti [2] oder Stable Diffusion [3] beeindrucken neben vielen anderen Systemen [4,5,6] bereits mit qualitativ hochwertigen Ergebnissen bei der Generierung von fotorealistischen Bildern. Jedoch stoßen die Systeme in verschiedenen Bereichen an ihre Grenzen: DALL-E 2 wird zum Beispiel für mangelndes relationales Verständnis [7] und nicht ausreichende Symbol-Entität-Zuordnung [8] kritisiert. Daraus folgend stellt sich die Frage, ob menschliche Betrachter trotz solcher Limitationen einen Unterschied zwischen generiertem und echtem Bild erkennen können und anhand welcher Kriterien sie eine Unterscheidung festmachen.

Zielsetzung der Arbeit

Ziel der Arbeit ist die Untersuchung der menschlichen Fähigkeit, zwischen KI-generierten und echten Bildern zu unterscheiden. Hierbei sollen Faktoren, die zur Unterscheidung zwischen echt und gefälscht beitragen, genauer betrachtet werden. Die Forschungsfragen werden wie folgt definiert: Können Menschen KI-generierte Bilder von echten Bildern unterscheiden? Wie sicher sind sie sich in ihrer Einschätzung? Welche Faktoren tragen zur Unterscheidung der Bilder bei? Welche Elemente lassen ein generiertes Bild „unecht“ erscheinen?

Konkrete Aufgaben

- Aufbereitung der Literatur zu Text-to-Image-Generation und Prompt Engineering
- Generierung eines Pools aus Bildern mittels eines KI-Systems
- Erstellung, Durchführung und Auswertung einer Benutzerstudie
- Erstellung eines Kriterienkatalogs

Erwartete Vorkenntnisse

- Grundkenntnisse im Bereich HCI und Bildwahrnehmung
- Kenntnisse im Bereich der Bildgenerierung und Prompt Engineering
- Design, Durchführung und Auswertung empirischer Studien

Weiterführende Quellen

[1] Wennker, P. (2020). Künstliche Intelligenz in der Praxis. Anwendung in Unternehmen und Branchen: KI.

[2] Yu, J., Xu, Y., Koh, J.Y., Luong, T., Baid, G., Wang, Z., Vasudevan, V., Ku, A., Yang, Y., Ayan, B.K., Hutchinson, B.C., Han, W., Parekh, Z., Li, X., Zhang, H., Baldrige, J., & Wu, Y. (2022). Scaling Autoregressive Models for Content-Rich Text-to-Image Generation. ArXiv, abs/2206.10789.

[3] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10674-10685.

[4] Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E.L., Ghasemipour, S.K., Ayan, B.K., Mahdavi, S.S., Lopes, R.G., Salimans, T., Ho, J., Fleet, D.J., & Norouzi, M. (2022). Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. ArXiv, abs/2205.11487.

[5] Oppenlaender, J. (2022). The Creativity of Text-to-Image Generation. 25th International Academic Mindtrek conference.

[6] Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical Text-Conditional Image Generation with CLIP Latents. ArXiv, abs/2204.06125.

[7] Marcus, G., Davis, E., & Aaronson, S. (2022). A very preliminary analysis of DALL-E 2. arXiv preprint arXiv:2204.13807.

[8] Rassin, R., Ravfogel, S., & Goldberg, Y. (2022). DALLE-2 is Seeing Double: Flaws in Word-to-Concept Mapping in Text2Image Models. arXiv preprint arXiv:2210.10606.

From:

<https://wiki.mi.ur.de/> - **MI Wiki**

Permanent link:

https://wiki.mi.ur.de/arbeiten/wip_ai_generated_realistic_photographs

Last update: **19.07.2023 08:53**

