

Infoblatt: Bildgeneration mittels Diffusionsverfahren

Bildgenerative Künstliche Intelligenz, wie Midjourney, DALL-E oder Stable Diffusion, nutzt das Diffusionsverfahren für künstliche neuronale Netzwerke, um Bilder zu generieren.

- Dabei lernt die KI, aus den Trainingsdaten relevante Merkmale abzuleiten. Vereinfacht gesagt, erfasst die KI die unterschiedlichen Anordnungen von Pixeln auf digitalen Bildern und ordnet diese immer komplexeren Konzepten zu.
- So werden aus Mustern, Texturen, Ecken und Kanten komplexe Objekte wie ein Hund.
- In einem letzten Schritt werden dabei den Konzepten Kennzeichnungen zugeordnet, sprich die Wörter oder Kategorien, in denen wir Bilder beschreiben.
- So können zunächst Bilder erkannt und klassifiziert werden.
- Auf dieser Basis kann in umgekehrter Richtung aus randomisierten Pixeln – auch Rauschen genannt – schrittweise ein Bild generiert werden.

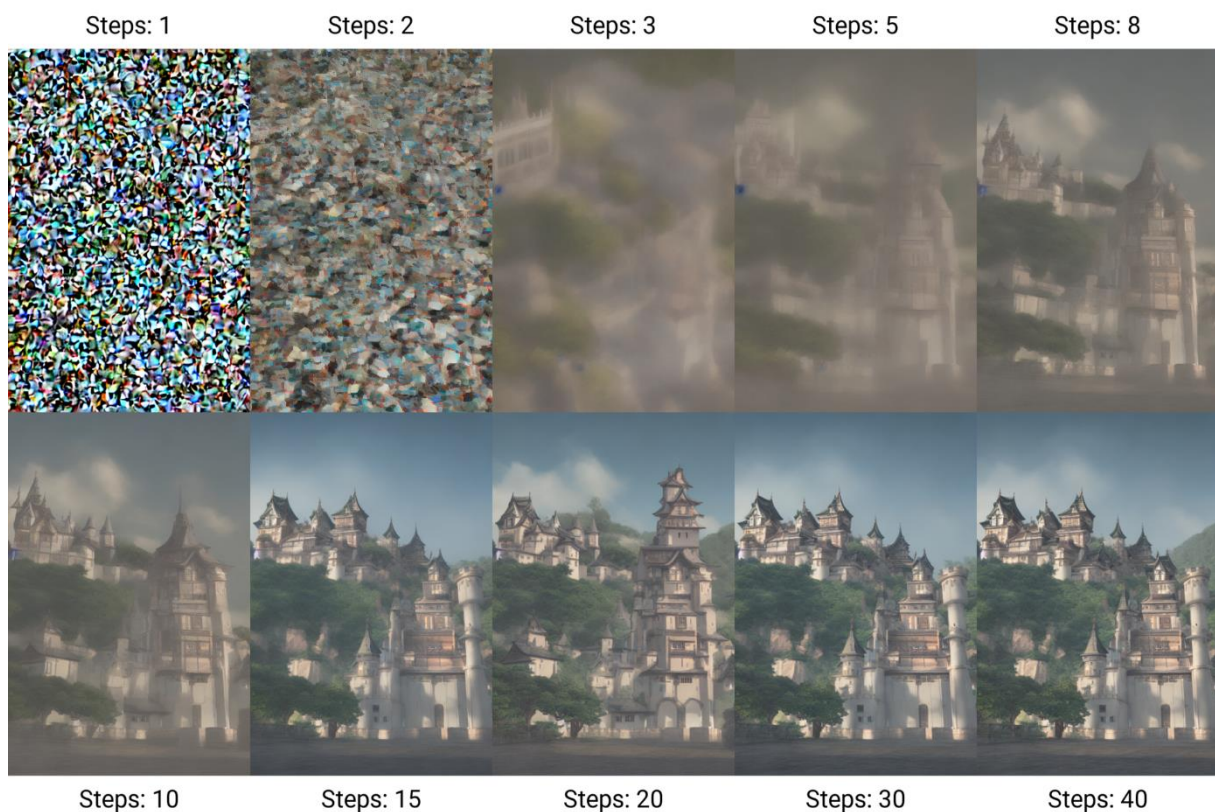


Abbildung: Grafische Veranschaulichung des Diffusionsverfahrens zur Bildgenerierung, online unter: https://de.wikipedia.org/wiki/Stable_Diffusion#/media/Datei:X-Y_plot_of_algorithmically-generated_AI_art_of_European-style_castle_in_Japan_demonstrating_DDIM_diffusion_steps.png

Ein bedeutendes Problem bei bildgenerativen KI-Systemen ist die Reproduktion von Verzerrungen, auch Bias genannt, die in den Trainingsdaten vorhanden sind. Diese können sich auf unterschiedliche Weise manifestieren, etwa in der überproportionalen Darstellung bestimmter Ethnien, Geschlechter oder Altersgruppen. Grund hierfür ist die Zusammensetzung des Trainingsdatensatzes: Wenn ein Bias in den Trainingsdaten vorhanden ist, wird das KI-Modell diese Muster lernen und in seinen Erzeugnissen reproduzieren. Daher gilt es, entweder die Trainingsdaten so auszuwählen, dass sie keinen Bias enthalten – was aufgrund der großen Menge an Daten, die benötigt werden, schwierig ist – oder zu versuchen, vorhandenen Verzerrungen während des Fine-Tunings entgegenzuwirken.