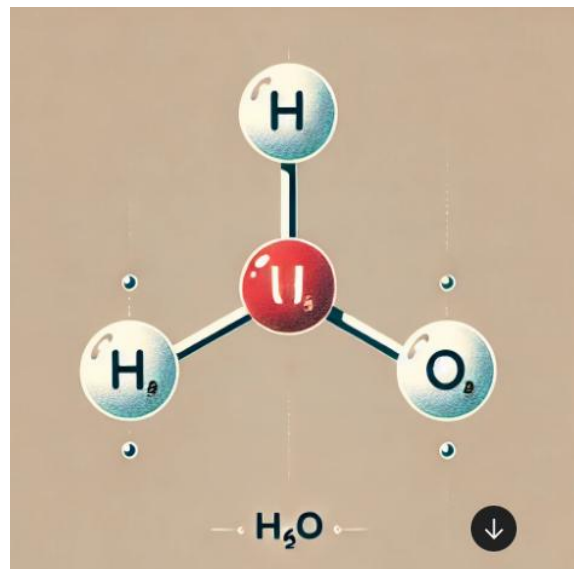


Was sind Multimodale KI-Anwendungen?

Dall•E2 ist sowohl ein **Bilderkennungsprogramm** als auch ein **Bildgenerator**, welches Textinformationen mit Bildinformationen mathematisch korreliert (Bajohr 2022, 76). Damit ist es ein stochastisches multimodales Modell, da es auf zwei verschiedene Datenformen ‚Text‘ und ‚Bild‘ zugreift. Es wurde von OpenAI durch Deep Learning entwickelt, wobei für das umfassende Trainingsmaterial Bilder aus dem Internet verwendet wurden (Hogea & Rocafort Ferrer 2024, 2).

Das Bilderkennungsprogramm setzt sich aus einem Machine-Learning Modell namens Clip zusammen und korreliert Textdaten mit Bildinformationen. Es bildet ein stochastisches Modell, welches in der Lage ist, eine möglichst **hohe Passung zwischen Bild und Text** zu modellieren (Bajohr 2022, 76). So kann aus dem Text: „Die Strukturformel von H₂O“ das Wort ‚Strukturformel‘ oder „H₂O“ erkannt werden und mit entsprechenden Bildinformationen korrelieren. Jede Text- und jede Bildinformation haben dabei einen Vektor und bilden jeweils zusammen ein Skalarprodukt, wobei ein höherer Wert des Skalarprodukts eine höhere Korrelation zwischen Bild und Text darstellt (Bajohr 2022, 76).

Mithilfe von Glide dem Bildgenerator können diese Informationen jetzt genutzt werden, um aus einem Text-input auf die entsprechenden Bildinformationen zurückzugreifen, um ein Bild aus dem bestehenden Datensatz zu generieren (Bajohr 2022, 76). Dabei werden einzelne Elemente aus der Texteingabe herausgenommen und mit korrelierenden Bildern abgeglichen, wobei hieraus Pixel verwendet werden, um ein neues Bild aus verschiedenen Textinformationen zu generieren (Hogea & Rocafort Ferrer 2024, 3). So wurde das dargestellte Bild in Chat GPT am 30.09.2024 auf den Eingabetext: „Erstelle mir die Strukturformel von Wasser“ erstellt. Auch wenn das dargestellte Bild von der korrekten Antwort abweicht, lieferte Chat GPT zusätzlich folgenden Text: „To create a visual representation of water as H₂O, I'll generate an image that focuses on the molecular structure of water. It will feature two hydrogen Atoms (H) bonded to a single oxygen atom (O), represented in a simple, clear diagram form. The



oxygen atom will be larger, with two smaller hydrogen atoms attached at an angle, showing the bent shape of the water molecule.“ Fachsprachlich wird dabei die Strukturformel von Wasser korrekt beschrieben, allerdings ist die Passung zwischen Bild und Text aus fachlicher Sicht als äußerst gering einzuschätzen.

Umgeschrieben und entnommen aus:

Bajohr, H. (2022). Dumme Bedeutung. Künstliche Intelligenz und Artifizielle Semantik, *Merkur*, 76 (882), 69-79.

Hogea, E. F. & Rocafort Ferrer, J. M. (2024). *The ethical situation of DALL-E 2*. Doi: *arXiv preprint arXiv:2405.19176* [Zuletzt zugegriffen am 01.09.2025]

Bild erstellt mit:

OpenAI (2024). ChatGPT Large Language Model, <https://chatgpt.com> [Zuletzt abgerufen am 30.09.2024].