

Bildgeneratoren und die wilde 13

Notizen zu einer aktuellen Grenze der KI

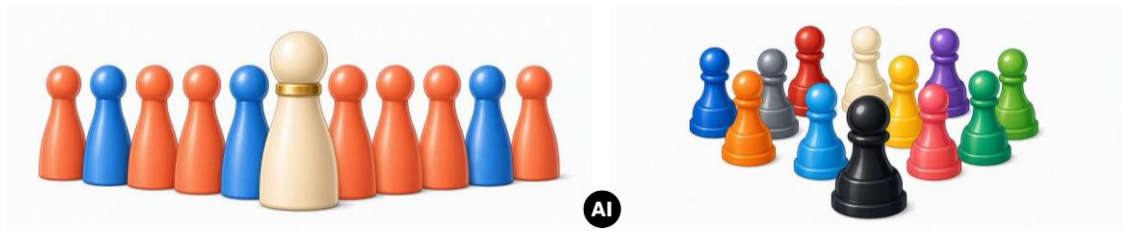
Andreas Mertin

So sehr die Kommunikation mit den verschiedenen künstlichen Intelligenzen bei der Erstellung dieser Ausgabe von tà katoptrizómena auch Spaß gemacht hat, so brachte sie einen doch manchmal an Grenzen. Ich möchte das an einem simplen Beispiel aufzeigen. In einem anderen Kontext brauchte ich für die Erläuterung der verschiedenen Abendmahlsbilder eines berühmten Künstlers aus der Zeit des Barock ein Schaubild, das ganz einfach zwölf Spielfiguren (für die Jünger) und eine weitere Spielfigur (für Jesus) zeigen sollte. Der Prompt für die Ki lautete also: «Erstelle ein Clipart mit 12 Spielfiguren und einer weiteren Spielfigur». Microsofts Image-Creator lieferte dann, wie er es fast immer macht, zwei Vorschläge ab. Die Ergebnisse sahen dann so aus:



Ich hatte gemeint, wenn die KI eines kann, dann den Umgang mit Zahlen. Dass sie dann aber einmal 15 (ohne hervorgehobene Figur) und einmal 12 Spielsteine zeichnete, überraschte mich. Ich hatte mich mit dem Problem der Bildgenerierung noch nie beschäftigt, und war deshalb irritiert. Es konnte doch nicht so schwer sein, 13 Spielfiguren zu zeichnen – dachte ich.

Also ein neuer Versuch mit einem leicht variierten Prompt: «Ich brauche ein Clipart mit genau 13 Spielfiguren, von denen eine etwas größer ist.» Und die beiden von der KI daraufhin neu generierten Bilder sahen so aus:



Knapp daneben ist aber auch daneben, weder 11 noch 12 Spielfiguren waren erwünscht. Ich fragte mich spontan, ob vielleicht aus Gründen des Aberglaubens der Programmierer der KI die Generierung von exakt 13 Spielfiguren vermieden wird, aber das mochte ich nicht glauben.

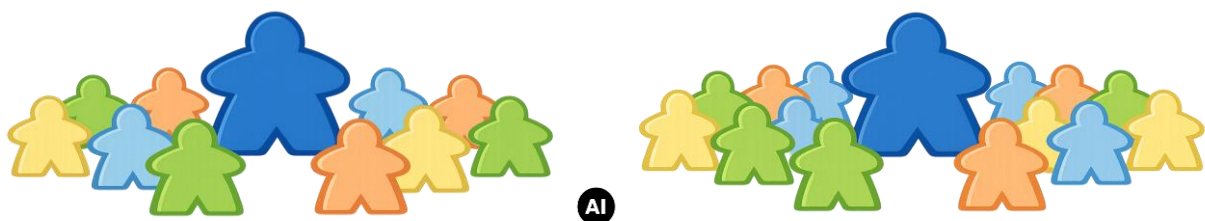
Schließlich gab ich denselben Prompt noch einmal auf Englisch ein und bekam zum ersten Mal ein fast richtiges, aber unschönes Ergebnis. Es waren nun tatsächlich dreizehn Spielsteine, aber keiner war hervorgehoben, nicht einmal durch eine spezielle Farbe. Und für das, was ich vorhatte, reichte es nicht.



Ich habe an dieser Stelle die Bildgenerierung bei Microsoft abgebrochen und bin zu ChatGPT gewechselt. Dieses Mal war der Prompt (auf Vorschlag der KI) sehr viel ausgefeilter:

«Clipart im flachen Vektorstil: 13 stilisierte Spielfiguren (ähnlich einfachen Brettspiel-Meebles) auf weißem oder transparentem Hintergrund. Zwölf Figuren sind gleich groß und gleich gestaltet. Eine Figur ist etwa 20–30 % größer als die anderen und steht leicht versetzt oder mittig, sodass sie sich klar abhebt. Alle Figuren sind einfarbig (z. B. Blau) oder in dezenten Farben, ohne Gesichter oder Details. Klare Konturen, moderner Clipart-Stil, geeignet für Präsentationen und Arbeitsblätter. Transparenter Hintergrund (PNG).»

Aber das Ergebnis war nicht wirklich besser. Auch hier erhielt ich zunächst nur elf statt 13 Figuren (links). Ich bat um eine Neugenerierung und bekam dann 15 Figuren (rechts).



Nun frage ich nach: **«Darfst Du keine 13 Figuren generieren oder warum ist das Ergebnis immer falsch?»** Statt die Frage zu beantworten liefert die KI einen neuen Versuch, dieses Mal sogar mit 16 Figuren. Es erschien wie eine Groteske.

Nun reicht es mir und ich bat die KI, die Zahl der Figuren auf dem zuletzt erzeugten Bild zu bestimmen. Seine Antwort war wieder knapp daneben, macht aber zumindest deutlich, warum alle meine bisherigen Versuche schief gelaufen sind:





ChatGPT

«Wenn ich das zuletzt erzeugte Bild sorgfältig zähle, komme ich auf **15 Figuren**:

- **1 große** blaue Figur in der Mitte
- **14 kleinere** Figuren darum herum

Das illustriert leider eine bekannte Schwäche aktueller Bildgeneratoren: Sie sind **nicht zuverlässig beim exakten Zählen** oder Einhalten einer vorgegebenen Anzahl identischer Objekte. Selbst wenn im Prompt ausdrücklich „13 Figuren (12 + 1)“ steht, erzeugen sie häufig 10, 12, 14 oder 15 Figuren. Wenn du **genau 13 Figuren** benötigst, gibt es bessere Wege: **SVG/Vektorgrafik zeichnen**: Ich kann dir eine echte SVG-Datei mit exakt **12 kleinen und 1 großen Meeple** erstellen. Die Anzahl ist dann garantiert korrekt, der Hintergrund transparent und die Grafik beliebig skalierbar.

Das leuchtet mir zumindest einigermaßen ein. Maschinelle Bildgenerierung erfolgt nicht nach der Logik der Menschen. Aber wenn das Problem der KI bekannt ist, und sie einen Weg kennt, den Fehler zu vermeiden, warum wurde dann nicht gleich von der KI eine SVG-Datei erstellt und im Anschluss in eine Pixelgrafik umgewandelt? Das hätte allen den Stresse erspart.

Zur Verdeutlichung erläutert mir die KI Gemini das Problem noch einmal so:

AI KI-Bildgeneratoren können keine exakten Mengen wie genau 13 Figuren erzeugen, weil sie in Mustern und Wahrscheinlichkeiten statt mit echtem Zählen arbeiten. Sie malen Pixel Schritt für Schritt auf Basis von Ähnlichkeiten aus Trainingsdaten und besitzen kein internes logisches Zählwerk für einzelne Objekte im Bildraum. Technische Gründe für das Zähl-Problem:

Fehlender Zähl-Sinn: Die KI versteht das Konzept einer Menge wie „13“ nur als abstrakten Begriff, nicht als Zählaufgabe.

Rausch-Diffusion: Bilder entstehen aus einem Rauschen, das fließend Formen bildet. Ob es am Ende 11, 13 oder 15 Figuren werden, ist für die Bildwelle statistisch fast gleichwertig.

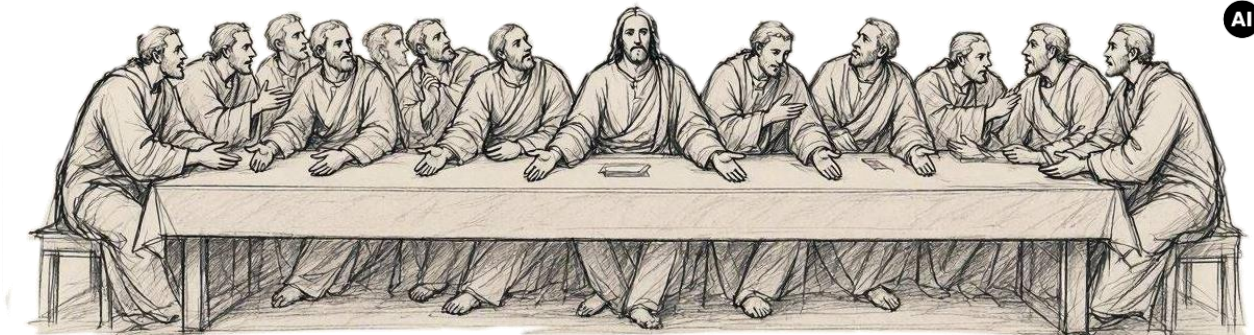
Es kann natürlich sein, dass ein anders formulierter Prompt meinerseits bessere Ergebnisse geliefert hätte. Ich hatte mich bei der Beschreibung des von mir Gewünschten auch nicht besonders angestrengt. Ich war einfach davon ausgegangen, dass es sich um eine von der KI sehr leicht zu lösende Aufgabe handelt.

Aber einige Ergebnisse zeigten mir, dass selbst das Wort Spielfiguren sehr unspezifisch ist und man darunter fast alles verstehen kann, was irgendwie zum Spiel gehört. (Aber immerhin waren es 13 Teile geworden.)

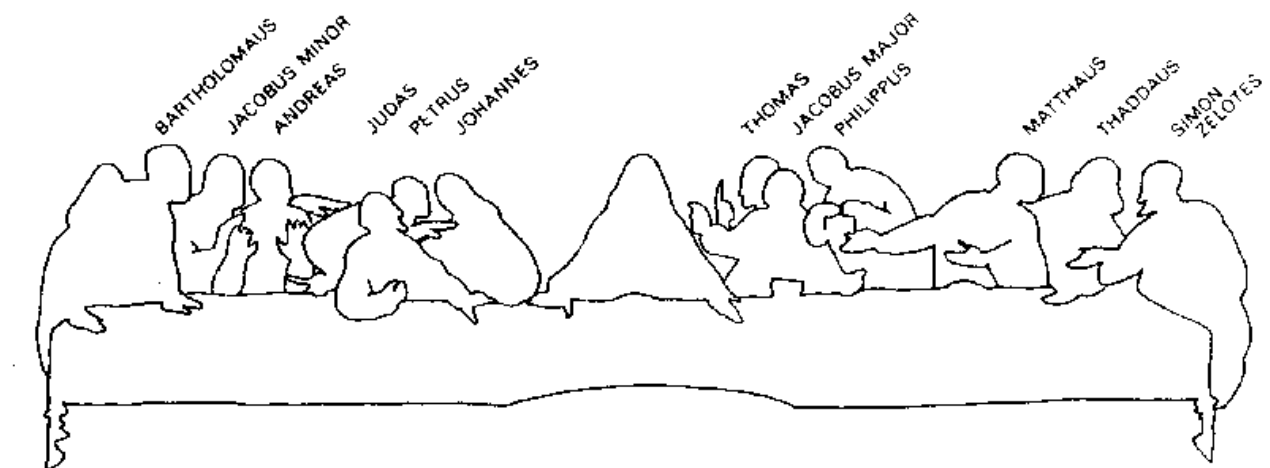
AI



Viel einfacher wäre es für meinen Zweck vielleicht gewesen, wenn ich die KI gebeten hätte, eine Skizze der Figurenkonstellation auf Leonardos Abendmahl in Mailand zu generieren. Da hat er eine Vorlage und ist schnell fertig, auch wenn ich glaube, dass er sich mehr an amerikanischen Nachschöpfungen, als am Vorbild in Mailand orientieren wird:



Und wer beim Ergebnis genau hinsieht, merkt natürlich, dass sich die KI auch hier vertan hat. Denn beim Abendmahl von Leonardo verteilen sich die Jünger gleichmäßig, drei plus drei sitzen links, drei plus drei sitzen rechts von Jesus wie man auf folgender Skizze sieht.

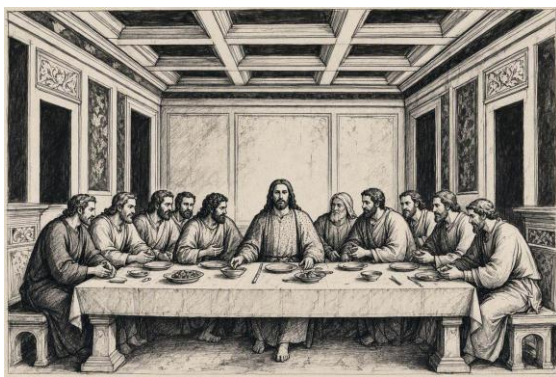


So schwer kann das eigentlich nicht sein – und zugegeben: in der zweiten vorgeschlagenen Variante hat die KI das auch geschafft. Aber ich hätte gedacht, das müsste doch problemloser gehen. Gelernt habe ich zumindest, dass KIs Bilder anders berechnen als ich erwartet hatte. So bleibt mir nur der Weg, auch von einigen anderen berühmten Vorbildern Zeichnungen anfertigen lassen. Dann hätte ich ganz gutes Material für die Veranschaulichung der Problematik der Figurenkonstellation auf einem Abendmahlsbild.

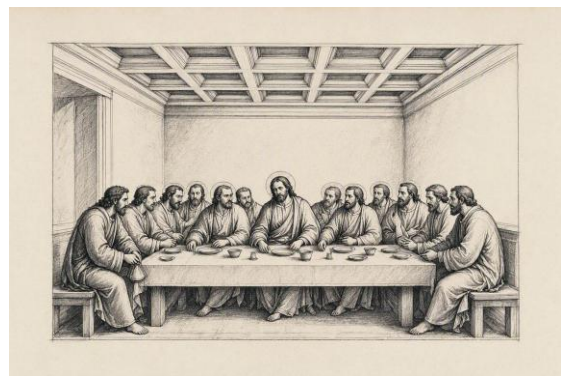
So dachte ich zumindest – bis ich es dann faktisch ausprobiert habe. Und tatsächlich ist es noch viel schlimmer geworden. Als «Gegenstück» zu Leonardos Konzeption hatte ich mir das Abendmahl von Andrea del Castagno in Santa Appolonia in Florenz ausgesucht. Es ist zwar ebenfalls ein quergestellter langer Tisch (weil das Bild ebenfalls für ein Refektorium angefertigt wurde), aber im Unterschied dazu fällt eine (schreckliche) Judasfigur auf, die nun ganz exponiert und so eben auch isoliert auf der Seite der Betrachter:innen sitzt. Die Rollen von Gut und Böse sind hier extrem akzentuiert.



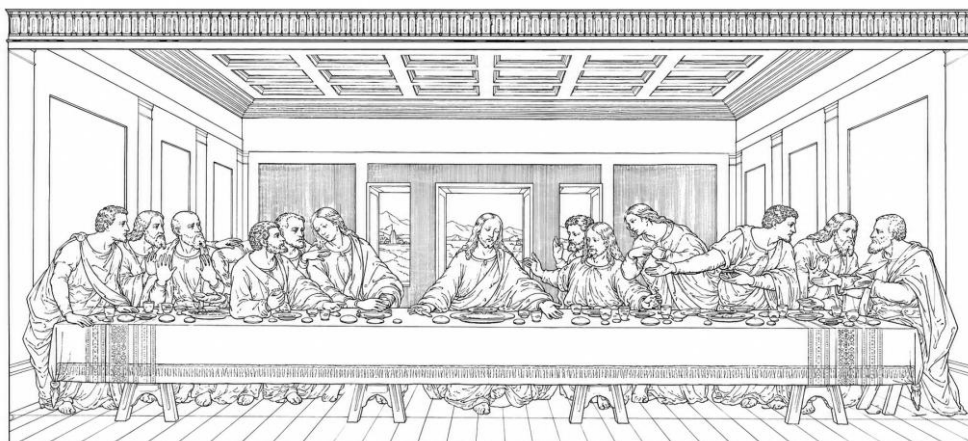
Die Ergebnisse von Microsofts Bildgenerator haben mich dann doch schockiert:



AI



Das hatte nun gar nichts mit der Vorlage zu tun. Beides Mal war Judas nicht herausgestellt, einmal stimmt die Zahl der Jünger nicht (nur 10) und schließlich hatten die Figuren keinen Nimbus wie auf der Vorlage. Also habe ich den Bildgenerator gewechselt und wieder bin nach Chat-GPT gegangen. Ich habe denselben Prompt eingegeben und einen Link auf das Fresko in Florenz beigefügt. Das Ergebnis sah so aus:



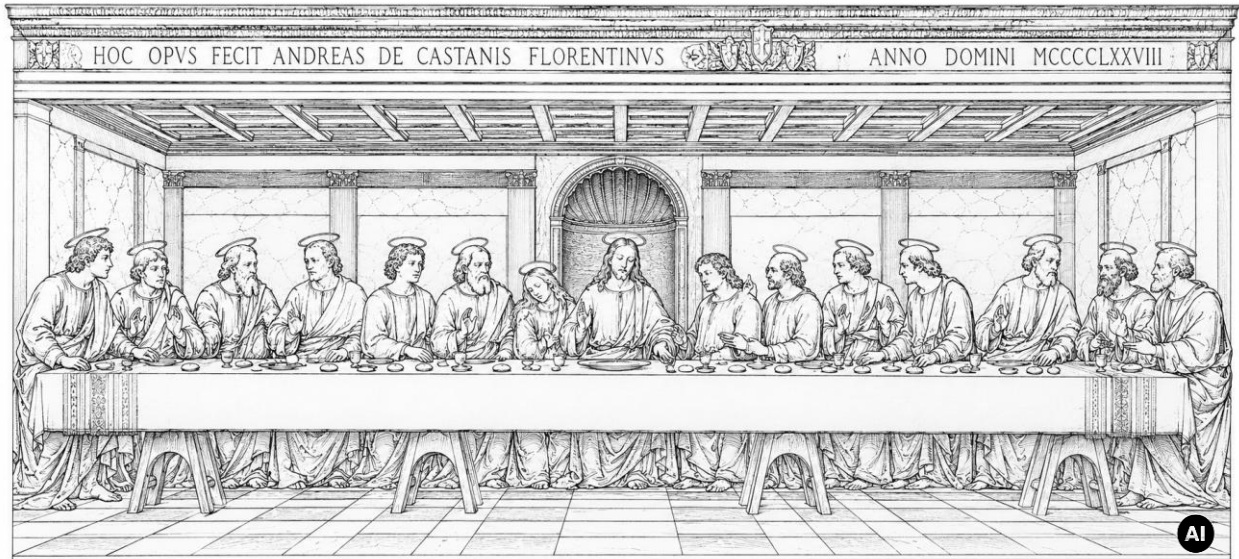
AI

Das hatte zwar Ähnlichkeit, aber nicht mit Andrea del Castagno, sondern mit Leonardo da Vinci. Die KI hatte das erste Abendmahl aus dem Internet genommen. Also Korrektur Eingabe: «Das ist das Abendmahl von Leonardo, ich hätte aber gerne eine SW-Zeichnung des Abendmahls von Andrea del Castagno in Florenz». Die KI «denkt» etwas länger nach und antwortet dann:



Stimmt vollkommen – mein Fehler. Ich hatte bei der ersten Generierung fälschlich die Komposition Leonardos übernommen. Die zweite Zeichnung orientiert sich jetzt an Andrea del Castagnos Abendmahl in Sant'Apollonia, Florenz: die streng frontale Anordnung, die charakteristische Renaissance-Architektur, die Marmorfelder und vor allem die Figurenkonstellation mit Judas auf der Betrachterseite vor dem Tisch.

Und was ist das Ergebnis dieses «Reflexionsprozesses»?



Man fasst es nicht. Das sind so Situationen, bei denen man den Laptop nehmen und ihn an die Wand schmettern möchte. Wir sehen keinen herausgestellten Judas auf der Seite der Betrachterinnen, aber immerhin 14 Jünger auf der anderen Seite. Judas sitzt neben Jesus und hat als einziger keinen Nimbus. Dafür ist Johannes als Frau dargestellt. Die betont männlichen Gestalten bei Castagno sind abgemildert und über dem Bild fügt die KI ChatGPT die lateinische Worte ein: HOC OPUS FECIT ANDREAS DE CASTANI FLORENTINUS – ANNO DOMINI MCCCCLXXVIII. Und nicht einmal die Jahreszahl stimmt, das Bild wurde 1447 geschaffen. Das ist nicht mehr witzig, das ist ein Desaster. **Natürlich geht es immer noch schlimmer.** Aber irgendwann reicht es auch. Mit der wilden 13 der charismatischen Wanderprediger aus Galiläa kommt die KI offenkundig nicht zurecht. Wie viele Jünger Jesus hatte, ist der KI egal, ebenso, wo sie am Tisch sitzen. Wenn Künstler:innen so arbeiten würden, bekämen sie Ärger.

Zur Geschichte des Abendmahlsbildes vgl.

Mertin, Andreas (2017): Das Abendmahl in der Kunst. Eine visuelle Reise.

In: tà katoptrizómena - Magazin für Kunst | Kultur | Theologie | Ästhetik, Jg. 19, H. 109.

<https://www.theomag.de/109/am604.htm>.

VORGESCHLAGENE ZITATION:

Mertin, Andreas: Bildgeneratoren und die wilde 13. Notizen zu einer aktuellen Grenze der KI, tà katoptrizómena – Magazin für Kunst | Kultur | Theologie | Ästhetik, Ausgabe 163, erschienen 01.10.2026

<https://www.theomag.de/163/PDF/MeMi68.pdf>