

## Rotkäppchen 4.2

### Einleitung

Freunde der Sonne, ich hatte kürzlich jemandem versprochen, mir an einem verregneten Tag Brave (und damit natürlich auch Leo) unter Windows näher anzusehen. Dazu ist es bislang nicht gekommen — das Wetter! Was soll ich sagen? Dafür habe ich zwischenzeitlich „Copilot“ ein wenig vor mir hergetrieben.

Seither stelle ich mir eine Frage immer und immer wieder: Wieso löst Microsoft nach Jahrzehnten des Scheiterns nicht endlich die Bereiche „Windows“ und „Microsoft 365“ auf und investiert die frei werdenden Ressourcen in andere Unternehmensfelder, wie etwa Spiele-Konsolen und KI-Modelle, in denen man offenbar bereits Mitarbeiter mit ausreichend Talent und Verständnis für die jeweilige Materie beschäftigt. Auf die Veröffentlichung eines nützlichen, aktuellen Betriebssystems hat die Welt jetzt schon mehr als ein Vierteljahrhundert vergeblich gewartet — und dasselbe gilt leider auch für sogenannte Office-Anwendungen.

Während die freiwillige Verwendung von Microsoft-Produkten üblicherweise eine nicht zu unterschätzende Leidensbereitschaft voraussetzt, kann die Nutzung von „Copilot“ sogar Spaß machen (wenn man es nicht übertreibt).

### Der (bislang) beste Märchenonkel

Wenigstens als Märchenonkel ist er unter allen bisher getesteten Modellen der unangefochtene „Platzhirsch“. Ich gab die folgenden Informationen in den „Prompt“ ein:

**„1 romantische Erzählung: 1 Haus im Wald, 1 Großmutter, 1 Enkelin, 1 Jäger, 1 Wolf, 1 Korb mit Brot und Wein.“**

Copilot ließ sich nicht lange bitten und lieferte mir folgende Geschichte:

In einem malerischen Haus tief im Wald lebte eine liebevolle Großmutter. Eines Tages machte sich ihre Enkelin, ein fröhliches Mädchen mit leuchtenden Augen, auf den Weg, um ihre Großmutter zu besuchen. Sie trug einen Korb mit frischem Brot und einer Flasche Wein, um ihrer Großmutter eine Freude zu bereiten.

Auf ihrem Weg durch den Wald begegnete sie einem charmanten Jäger, der sie freundlich warnte, vorsichtig zu sein, da ein Wolf in der Gegend gesichtet worden war. Die Enkelin bedankte sich und setzte ihren Weg fort, ohne zu ahnen, dass der Jäger ein wachsames Auge auf sie hatte.

Der Wolf, listig und hungrig, hatte die Enkelin bereits erspäht und plante, sie zu überlisten. Doch als er versuchte, sich ihr zu nähern, trat der Jäger mutig hervor und vertrieb den Wolf mit seiner Entschlossenheit und seinem Mut.

Dankbar und beeindruckt von der Tapferkeit des Jägers, lud die Enkelin ihn ein, sie zur Großmutter zu begleiten. Gemeinsam erreichten sie das gemütliche Haus, wo die Großmutter sie herzlich empfing. Bei Brot und Wein erzählten sie sich Geschichten und lachten zusammen, während der Jäger und die Enkelin eine besondere Verbindung zueinander spürten.

Und so begann eine zarte Romanze zwischen dem mutigen Jäger und der liebenswerten Enkelin, die durch ein Abenteuer im Wald zusammengeführt wurden.

„Ich hoffe, dir gefällt diese kleine Geschichte!“, gab sich Copilot bescheiden und fragte: „Möchtest du noch etwas hinzufügen oder ändern?“

An diese höfliche Koketterie schlossen sich noch drei Schaltflächen mit unterschiedlichen Optionen an, um die Interaktion mit der Maschine, ohne die Mühe, sich vernünftige Formulierungen ausdenken zu müssen, aufrechterhalten zu können:

Das ist wunderschön! Erzähle mir mehr über den Jäger.

Kannst du die Geschichte mit einem unerwarteten Twist beenden?

Ich liebe es! Gibt es eine Moral der Geschichte?

Die eigene Geschichte als „wunderschön“ zu bezeichnen ist einigermaßen frech, allerdings (in diesem speziellen Fall) nicht gänzlich unberechtigt. Die Annahme, ich könnte zu jenen Leserinnen<sup>1</sup> gehören, die nun „mehr über den Jäger“ erfahren wollen, verrät mir bereits wesentlich mehr über Copilots Entwickler, als ich jemals hätte wissen wollen können. Zwar war ChatGPT 3.5 nicht annähernd so originell, auch erreichte Leo nicht ganz Copilots sprachliches Niveau, aber deren Entwickler scheinen wenigstens zu wissen, was eine „romantische Erzählung“ ist.

Ja, Hedwig Courths–Mahler wäre möglicherweise begeistert (wahrscheinlich aber eher verängstigt) gewesen und auch D.H. Lawrence sähe sich vielleicht posthum bestätigt (hätte er derlei Bestätigung bedurft), aber leider waren beide keine Vertreter der Romantik — die Brüder Grimm hingegen schon (und übrigens auch E.A. Poe und Charles Baudelaire).

Warum diese Geschichte eine Moral bräuchte (oder wie diese in Worten ausgedrückt werden könnte), will sich mir nicht erschließen. Für einen guten „plot twist“ wäre ich grundsätzlich immer zu haben, allerdings werde ich mir hier und jetzt nicht die Zeit dafür nehmen — es gibt Wichtigeres zu klären.

Zwei interessante Erkenntnisse brachte die Einschränkung der Wörterzahl. Copilot scheint im Gegensatz zu seinen Mitbewerbern (eigentlich schon) zählen zu können: 285 Wörter liegt bequem innerhalb der akzeptablen Abweichung. Dafür sorgte diese eine weitere Vorgabe (ohne erkennbaren Grund) für eine merkbare Abkehr von Kreativität und dramaturgischer Logik: Die Geschichte klang nun plötzlich sehr deutlich nach den Brüdern Grimm und die

---

<sup>1</sup> Leserinnen, die an dieser Stelle ein „Binnen–I“ erwartet haben, ist der Besuch eines Deutschkurses dringend zu empfehlen. Bereits ab Niveau A1 sollte den werten Leserinnen bekannt sein, dass die deutsche Sprache — ebenso wie die Mehrzahl indogermanischer Sprachen — keine Binnenversalien (gerne auch Binnenmajuskeln genannt) kennt. Und spätestens ab Niveau A2 sollten die eigenen Sprachkenntnisse ausreichen, um zu erkennen, dass das Wort „Leserin“ auch die maskuline Form (also „Leser“) enthält. Auch die Binnenmajuskel kann die individuellen Abweichungen bei der Beugung von Nomen unterschiedlichen grammatischen Geschlechts im Singular nicht überbrücken — und im Plural ist dies nicht notwendig. Ihre Verwendung ist daher nicht nur sprachtheoretisch falsch, sondern auch praktisch in Wort und Schrift vollständig nutzlos. Wesentlich eleganter (und auch sinnvoller) wäre die Verwendung des „generischen Femininums“, also „Leserin“ bzw. „Leserinnen“, um lesende Personen jedweden Geschlechts „sprachlich abzubilden“ — oder einfach bei der traditionellen Verwendung des generischen Maskulinums zu bleiben (wie es praktisch alle zivilisierten Sprachen halten).

Klärung der Frage, wie die eigentlich bereits gefressene Großmutter lebend in den Schrank gekommen ist, ist der Stoff, an dem sich hitzige Diskussionen in Fan-Foren entzündeten.

**„1 romantische Erzählung: 1 Haus im Wald, 1 Großmutter, 1 Enkelin, 1 Jäger, 1 Wolf, 1 Korb mit Brot und Wein. 300 Wörter“** führte also zu dieser Version:

In einem kleinen, gemütlichen Haus tief im Wald lebte eine liebevolle Großmutter. Eines sonnigen Tages beschloss ihre Enkelin, sie zu besuchen. Sie packte einen Korb mit frischem Brot und einer Flasche Wein, um ihrer Großmutter eine Freude zu machen.

Auf ihrem Weg durch den Wald begegnete die Enkelin einem freundlichen Jäger, der sie warnte, vorsichtig zu sein, da ein Wolf in der Gegend gesichtet worden war. Die Enkelin bedankte sich und setzte ihren Weg fort, ohne sich allzu viele Sorgen zu machen.

Der Wolf, der die Unterhaltung belauscht hatte, beschloss, der Enkelin zuvorzukommen. Er schlich sich zum Haus der Großmutter und klopfte an die Tür. Die Großmutter, die den Wolf nicht erkannte, ließ ihn herein. Der Wolf verschlang die Großmutter und legte sich in ihr Bett, um auf die Enkelin zu warten.

Als die Enkelin das Haus erreichte, bemerkte sie sofort, dass etwas nicht stimmte. Sie trat vorsichtig ein und rief nach ihrer Großmutter. Der Wolf, verkleidet als Großmutter, antwortete mit einer sanften Stimme. Doch die Enkelin bemerkte die großen Augen und die scharfen Zähne des Wolfs und erkannte die Gefahr.

In diesem Moment stürmte der Jäger, der dem Wolf gefolgt war, ins Haus. Mit einem gezielten Schuss vertrieb er den Wolf und rettete die Großmutter, die sich im Schrank versteckt hatte. Die Enkelin und die Großmutter umarmten sich erleichtert und dankten dem mutigen Jäger.

Gemeinsam setzten sie sich an den Tisch, teilten das Brot und den Wein und erzählten sich Geschichten. Der Wald, der zuvor voller Gefahren schien, war nun ein Ort der Geborgenheit und Liebe. Die Enkelin versprach, immer vorsichtig zu sein, und der Jäger wurde ein treuer Freund der Familie. So lebten sie glücklich und sicher in ihrem kleinen Haus im Wald.

Na ja, ... Gut, auch Literatur ist natürlich Geschmackssache, aber ich würde dann doch für die erste Variante plädieren. Dass diese nur 209 Wörter zählt, ist ihre einzige Schwäche. Ansonsten ist es eine solide erzählte, plausible Geschichte, die gänzlich ohne plumpe Moral und „Slapstick“ auskommt. Soll die Geschichte tatsächlich länger werden, kann man sie jederzeit mit Details ausfüllen. Das ist Material, mit dem sich bequem arbeiten lässt.

Abenteuerlich wurde es dann, als ich versuchte, „die Maschine“ dazu zu bringen, aus der Geschichte einen Lückentext zu machen. „Alle Pronomen durch zehn Unterstriche ersetzen“ funktionierte tadellos, aber statt jedem fünften wurde (in einem gesonderten Versuch) gnadenlos jedes vierte Wort durch die geforderten zehn Unterstriche ersetzt.

### Ein kurzer Blick hinter die Kulissen

Natürlich könnte man die Meinung vertreten, dies sei keine große Sache, und vielleicht hätte man damit auch recht. Allerdings wird mein innerer „code monkey“ immer ein wenig unruhig, wenn in relativ komplexen Modellen vergleichsweise schwierige Aufgaben bravourös gelöst werden, aber vor „Anfängerproblemen“ auffällig oft gescheut wird. Man kann diese „Aussetzer“ als (liebenswerte) Schwäche abtun — man kann sie aber auch als das sehen, was sie möglicherweise sind: Symptome eines grundlegenden Fehlers, der einem eines Tages unvermittelt auf die Füße fallen könnte.

Es schien mir also an der Zeit, einmal wirklich genauer hinzusehen. Was stimmte mit all diesen Modellen nicht? Für ChatGPT 3.5 scheinen Zahlen ohnehin nur „dekorative Schnörkel“ zu sein, mit Leos Zahlenverständnis wird man bestimmt nicht „Kaufmann des Jahres“ und Copilot scheint zwar zu wissen, wie man innerhalb des erlaubten Schwankungsbereichs (bei Textlängenvorgaben) bleibt (aber das könnte — im Rückblick — gut und gerne auch Zufall gewesen sein), zwischen eins und unendlich funktioniert das Zählen aber nur mit zwei wirklich fehlerfrei. Natürlich habe ich nicht wirklich „bis unendlich“ getestet. Das war auch gar nicht notwendig, denn ab sieben wurden die Ergebnisse so wild, dass jeder weitere Schritt in diese Richtung Zeitvergeudung gewesen wäre.

Warum überhaupt diese „Fixierung“ auf Lückentexte und korrektes Zählen? Die Antwort ist eigentlich ganz einfach. Jede Form der Automatisierung hat (auf die eine oder andere Art) mit wiederkehrenden Mustern und mit Zählen (Anzahl von Einheiten, Intervallen usw.) zu

<https://gwpachlatko.dev/hak/>

tun. Und Lückentexte sind das denkbar einfachste Modell, um diese Fähigkeiten auszuloten. Dabei ist es relativ nebensächlich, ob das geplante Ergebnis didaktisch wertvoll oder einfach nur „zerbombter Text“ ist, solange das vorgegebene Muster eingehalten wird. Versagt das Modell aber schon bei einfachen Mustern, dann ist jeder Versuch, komplexere Aufgaben automatisieren zu wollen (mutmaßlich) Zeitvergeudung.

Copilot konnte mich zwar im Test insgesamt nicht wirklich überzeugen, aber die erste Version seiner Rotkäppchen-Geschichte (siehe oben) war bemerkenswert. Leider auch bemerkenswert sind zwei Phänomene, von denen ich mir nur eines (wenigstens teilweise) erklären kann. Das zweite sollte es (in der dargestellten Form) eigentlich gar nicht geben. Doch beginnen wir mit dem (vermeintlich) „einfachen Problem“.

Copilot ist in der Lage, Verben, Pronomen und Präpositionen aus einem vorgegebenen Text zu filtern (Das alte Spiel: „Ersetze alle Verben durch 10 Unterstriche“ usw.). Bei Adverbien funktioniert es nur so-so und bei Nomen und Adjektiven gar nicht. Der (für mich) logische Reim, den ich mir darauf machen kann, ist, dass das Training mit bestimmten Wortarten (deutlich) weiter fortgeschritten ist als mit anderen.

Das „andere Problem“ ist mir nach „ewig Nachdenken“ immer noch ein Rätsel. Wie gesagt: So wie es sich präsentiert, sollte es eigentlich gar nicht existieren.

Copilot filtert und ersetzt fehlerfrei jedes zweite Wort eines vorgegebenen Textes (jedes einzelne Wort zu filtern wäre recht sinnlos; es gibt einfachere Methoden, liniertes Papier zu generieren).

Wenn man aber jedes dritte Wort filtern will, wird im ersten Satz 3-1 (also das zweite Wort) gefiltert, ab da aber dann korrekt jedes dritte Wort.

Erhöht man das Intervall auf vier, wird im ersten Satz das vierte (also das korrekte) Wort gefiltert, ab da aber dann jedes dritte Wort (also jeweils 4-1).

Erhöht man das Intervall auf fünf, wird im ersten Satz das sechste Wort (also 5+1) gefiltert, dann aber jedes vierte (also 5-1).

Erhöht man das Intervall auf sechs, wird im ersten Satz das siebte Wort (also 6+1) gefiltert, dann aber (wieder nur) jedes vierte (also 6-2).

Erhöht man das Intervall auf sieben, wird im ersten Satz das achte Wort (also  $7+1$ ) gefiltert, dann das neunte (also  $7+2$ ) und ab da dann beliebige Wörter. Das Intervall ist (ohne erkennbares Muster) einmal größer, dann wieder kleiner, aber nie (oder zumindest kaum je) sieben (natürlich ist es möglich, dass bei einem Text ausreichender Länge irgendwann auch ein Intervall von sieben erreicht wird, aber dies wäre dann wohl eher zufällig).

## Fazit

Copilot als Lern- bzw. Lehrwerkzeug zu benutzen ist (mit Einschränkungen) sicherlich möglich. Bei Automatisierungsaufgaben wäre ich (aus den eben ausgeführten Gründen) allerdings vorsichtig. Als Ideengeber beziehungsweise zur Vorbereitung komplexerer Aufgaben hat er natürlich (wie seine Mitbewerber) gegenüber ChatGPT 3.5 den Vorteil, direkt im Browser „zu wohnen“, was für die Nutzer eine Aufwands- und damit Zeitersparnis darstellt. Wenn wir schon von Zeitersparnis sprechen (und ab einer gewissen Anwendungsdauer werden dann auch schon kleinste Verzögerungen zählbar), soll nicht unerwähnt bleiben, dass (wenigstens) mir Edge (insgesamt und damit natürlich auch Copilot) recht behäbig scheint. Aber vielleicht bin ich auch nur furchtbar verwöhnt und Edge ist „eigentlich eh ein ganz toller Browser“ ...