

# Prüfkette oder chain of verification

Dokumentation: Konversation mit OpenAI ChatGPT - 03. April 2024

 Mich beschäftigt der Begriff der Prüfkette. Der ist ja ungemein breit, d.h. er kommt in vielen Bereichen vor.

Was fiele denn dir dazu ein, ich meine so auf die (modellierte) Schnelle?

 Der Begriff "Prüfkette" kann in verschiedenen Kontexten verwendet werden, daher gibt es verschiedene Interpretationen und Anwendungen.

Hier sind einige mögliche Bedeutungen oder Kontexte, in denen der Begriff relevant sein könnte:

## 1. Produktions- oder Qualitätskontrolle:

- In der Fertigungsindustrie bezieht sich eine Prüfkette auf die Abfolge von Prüfungen und Inspektionen, die während des Produktionsprozesses durchgeführt werden, um die Qualität und Konformität der hergestellten Produkte sicherzustellen.

## 2. Softwareentwicklung:

- In der Softwareentwicklung kann eine Prüfkette den Prozess der automatisierten Tests bezeichnen, die in einer bestimmten Reihenfolge durchgeführt werden, um sicherzustellen, dass der Code korrekt funktioniert und den Anforderungen entspricht. Dies kann Unit-Tests, Integrationstests und Systemtests umfassen.

## 3. Finanz- und Rechnungswesen:

- Im Finanzbereich bezieht sich eine Prüfkette auf die Abfolge von Prüfungen und Kontrollen, die in einem Unternehmen durchgeführt werden, um die Richtigkeit der finanziellen Aufzeichnungen sicherzustellen und Betrug zu verhindern.

## 4. IT-Sicherheit:

- In Bezug auf die IT-Sicherheit kann eine Prüfkette den Prozess der Überprüfung von Sicherheitsmaßnahmen und -protokollen darstellen, um Schwachstellen zu identifizieren und zu beheben.

## 5. Logistik und Lieferkette:

- In der Logistik kann eine Prüfkette den Prozess der Überprüfung von Warenlieferungen von der Herstellung bis zur Auslieferung umfassen, um sicherzustellen, dass alle Qualitäts- und Sicherheitsstandards eingehalten werden.

## 6. Wissenschaft und Forschung:

- In wissenschaftlichen Untersuchungen kann eine Prüfkette den Ablauf von Experimenten und Analysen beschreiben, die durchgeführt werden, um Hypothesen zu überprüfen und Ergebnisse zu validieren.

Es ist wichtig zu beachten, dass der Begriff "Prüfkette" je nach Kontext unterschiedliche Bedeutungen haben kann.

Wenn Sie weitere Details zu einem spezifischen Anwendungsbereich haben oder nach etwas Bestimmtem suchen, lassen Sie es mich gerne wissen!



Danke für diese Informationen. Einen Aspekt, der gerade in deinem eigenen Zusammenhang wichtig wäre, vermisste ich dabei und zwar die Prüfkette im Englischen genannt *chain of verification*.



Entschuldigung für das Missverständnis. Wenn Sie auf den Kontext meiner eigenen Funktion als KI-Modell hinweisen, könnte der Begriff "Prüfkette" in Bezug auf die Modellvalidierung und -überprüfung relevant sein.

Hier sind einige mögliche Aspekte:

### Modellvalidierung:

- In der KI und maschinellen Lernmodelle könnte eine Prüfkette den Prozess der Validierung und Überprüfung eines Modells von der Datenvorverarbeitung bis zur Ausgabeproduktion umfassen.

### Datenqualität:

- Das Überprüfen der Qualität und Relevanz der Trainingsdaten könnte als Teil der Prüfkette betrachtet werden, um sicherzustellen, dass das Modell auf verlässlichen und repräsentativen Informationen basiert.

### Evaluationsprozess:

- Eine Prüfkette könnte auch den Prozess der Evaluierung von Modelleleistungen durch verschiedene Metriken und Testdatensätze einschließen, um sicherzustellen, dass das Modell ordnungsgemäß funktioniert.

### **Robustheitstests:**

- Tests auf Robustheit gegenüber verschiedenen Eingabebedingungen könnten ebenfalls Teil einer Prüfkette sein, um sicherzustellen, dass das Modell unter verschiedenen Szenarien zuverlässig ist.

Wenn Sie spezifischere Informationen oder Anforderungen haben, lassen Sie es mich bitte wissen, damit ich genauer auf den für Sie relevanten Aspekt der Prüfkette eingehen kann.



Genau darauf wollte ich hinaus. Dazu gleich noch mehr. Zunächst wollte ich mich noch einer anderen Sache vergewissern.

Nämlich: Was ich eben getan habe, wäre ein geradezu idealtypischer Fall eines eingänglichen sgn. mäeutischen Dialogs, nicht wahr?



Ja, tatsächlich können die Elemente Ihres vorherigen Dialogs als Beispiel für einen sogenannten *mäeutischen Dialog* betrachtet werden. Der mäeutische Dialog ist eine Art von Socratic Dialog, benannt nach Sokrates' Mutter Mätis, was "Hebamme" oder "Geburtshelferin" bedeutet.

In einem mäeutischen Dialog stellt der Dialogpartner Fragen, um das Verständnis des Gesprächspartners zu entwickeln und zu klären, anstatt direkt Informationen zu liefern.

In Ihrem Fall haben Sie zunächst den Begriff "Prüfkette" eingeführt und dann Fragen gestellt, um verschiedene mögliche Interpretationen und Anwendungen dieses Begriffs zu erkunden. Dies fördert nicht nur das Verständnis, sondern ermöglicht auch, spezifischere und relevantere Informationen bereitzustellen.

Wenn Sie weitere Fragen oder Anmerkungen haben oder den Dialog in eine bestimmte Richtung lenken möchten, stehe ich zur Verfügung!



Danke. Jetzt zurück zu deiner vorherigen Aussage samt Erläuterung dazu.

Du nennst als Punkte der Prüfkette in deinem eigenen Zusammenhang die Modellvalidierung, die Datenqualität, den Evaluationsprozess und Robustheitstests.

Mir ginge es in diesem Zusammenhang weniger um die Modellvalidierung und -überprüfung (systemseitig) sondern um die Validierung von Prompting-Verfahren (anwenderseitig).

Abgesehen von deinem ersten Punkt Modellvalidierung, wären alle weiteren Punkte auch auf mein eigentliches Anliegen anwendbar. Liege ich dabei richtig?



Ja, Ihre Betonung auf die Validierung von Prompting-Verfahren, insbesondere aus anwenderseitiger Sicht, ist absolut verständlich.

Lassen Sie uns die genannten Punkte im Kontext der Validierung von Prompting-Verfahren genauer betrachten:

**1. Modellvalidierung:**

- Dies könnte sich auf die Überprüfung beziehen, ob das Sprachmodell in der Lage ist, die vom Benutzer bereitgestellten Anfragen effektiv zu verstehen und angemessen zu reagieren.

**2. Datenqualität:**

- Hierbei könnte es um die Qualität der Daten gehen, die während des Trainings des Modells verwendet wurden, insbesondere wenn diese Daten direkt von Benutzern stammen. Es ist wichtig sicherzustellen, dass die Daten repräsentativ und vielfältig sind, um ein robustes Modell zu gewährleisten.

**3. Evaluationsprozess:**

- Die Bewertung von Prompting-Verfahren könnte verschiedene Aspekte umfassen, wie die Effektivität der generierten Antworten, die Benutzerzufriedenheit, die Berücksichtigung von Kontext und die Anpassungsfähigkeit an verschiedene Benutzeranfragen.

**4. Robustheitstests:**

- Tests auf Robustheit könnten sich auf die Fähigkeit des Modells beziehen, unterschiedliche Benutzeranfragen und Formulierungen zu bewältigen, sowie auf die Reaktion des Modells auf unvorhergesehene Eingaben.

Alle genannten Punkte sind daher relevant für die Validierung von Prompting-Verfahren aus anwenderseitiger Sicht.

Es geht darum sicherzustellen, dass das Modell nicht nur technisch robust und korrekt ist, sondern auch die Anforderungen und Erwartungen der Benutzer erfüllt.



Danke. Ich bin beeindruckt.

Das meine ich jetzt keinesfalls ironisch! Was ich jetzt meine, aus anderer Quelle verstanden zu haben, ist, dass die sog. Prüfkette (Chain-of-Verification) als Methode Gegenstand der Forschung und entsprechender Beschreibung ist, richtig?



Ja, das ist korrekt. Die Konzepte der Modellvalidierung, Datenqualität, Evaluationsprozesse und Robustheitstests, die wir im Kontext der Prüfkette diskutiert haben, sind Gegenstand intensiver Forschung und Methodenentwicklung im Bereich des maschinellen Lernens, der Künstlichen Intelligenz (KI) und der natürlichen Sprachverarbeitung.

In der wissenschaftlichen Gemeinschaft gibt es fortlaufende Bemühungen, Methoden und Best Practices für die Bewertung, Validierung und Verbesserung von KI-Modellen zu entwickeln.

Die Prüfkette, oder auch Testkette, ist dabei eine Methodik, die die verschiedenen Aspekte der Modellbewertung und -validierung in einer sequenziellen oder parallel angeordneten Abfolge von Tests organisiert.

Forschungsbemühungen konzentrieren sich darauf, wie man sicherstellen kann, dass KI-Modelle nicht nur auf bestimmte Trainingsdaten gut reagieren, sondern auch auf vielfältige und unvorhersehbare Eingaben robust sind.

Die Validierung von Modellen und Verfahren aus der Benutzersicht ist ebenfalls von großer Bedeutung, um sicherzustellen, dass die Modelle tatsächlich nützlich und effektiv für die beabsichtigten Anwendungen sind.

In der Praxis werden verschiedene Metriken, Benchmarks und Testverfahren entwickelt, um die Leistung und Verlässlichkeit von KI-Modellen zu bewerten. Dies ist ein dynamisches Forschungsfeld, da die Technologie ständig weiterentwickelt wird und neue Herausforderungen auftreten.



Abgesehen von dem Umstand, dass im Deutschen (wie im Niederländischen) eigentlich von Verarbeitung natürlicher Sprache und eben nicht natürlichen Sprachverarbeitung die Rede sein sollte, wie würde so eine Prüfkette aus der Benutzersicht, die ja dir zufolge ebenfalls von großer Bedeutung ist, denn anhand eines etwas ausführlicheren Beispiels praktisch aussehen (können)?



Lassen Sie uns ein Beispiel für eine Prüfkette aus der Benutzersicht betrachten, insbesondere im Kontext von Sprachgenerierenden Modellen (wie Chatbots oder Textgeneratoren):

1. **Eingabeverständnis (Input Understanding):**
  - Überprüfen, ob das Modell in der Lage ist, die Benutzereingabe korrekt zu verstehen und die Absicht hinter der Anfrage angemessen zu erfassen. Dies könnte die Erkennung von Schlüsselwörtern, Entitäten und die Interpretation des Kontexts umfassen.
2. **Korrekte und relevante Antwortgenerierung:**
  - Bewertung der Fähigkeit des Modells, auf die Benutzereingabe mit korrekten und relevanten Antworten zu reagieren. Hierbei wird geprüft, ob die generierten Ausgaben die beabsichtigte Information enthalten und die Nutzeranforderungen erfüllen.
3. **Natürlichkeit der Antwort:**
  - Überprüfen, ob die generierten Antworten natürlich und verständlich für den Benutzer sind. Dies könnte die Bewertung von Grammatik, Satzstruktur, und allgemeiner Sprachqualität einschließen.
4. **Kontextbezug und Konsistenz:**
  - Testen, ob das Modell den Kontext der vorherigen Interaktionen beibehält und konsistente Antworten im Verlauf des Gesprächs liefert. Dies ist wichtig für ein sinnvolles und flüssiges Gespräch.
5. **Adaptabilität an Benutzeranfragen:**
  - Prüfung, ob das Modell flexibel auf verschiedene Formulierungen derselben Anfrage oder auf unerwartete Eingaben reagieren kann. Dies unterstützt die Benutzerfreundlichkeit und verbessert die Anpassungsfähigkeit des Modells.
6. **Feedback-Loop und Lernfähigkeit:**
  - Überwachen, ob das Modell aus Benutzerfeedback lernen kann und sich kontinuierlich verbessert. Dies könnte die Integration von Feedback-Mechanismen seitens der Benutzer oder von menschlichen Experten umfassen.  
um das Modell zu verfeinern.
7. **Sicherheit und Ethik:**
  - Sicherstellen, dass das Modell ethisch einwandfrei arbeitet und keine schädlichen oder unangemessenen Antworten generiert. Dies könnte den Schutz vor missbräuchlicher Nutzung oder Diskriminierung einschließen.

Diese Aspekte der Prüfkette aus der Benutzersicht bieten eine umfassende Möglichkeit, die Leistung eines sprachgenerierenden Modells zu bewerten und

sicherzustellen, dass es den Anforderungen und Erwartungen der Benutzer gerecht wird.

Es ist wichtig zu beachten, dass die spezifischen Kriterien je nach Anwendungsfall variieren können, und die Prüfkette sollte entsprechend angepasst werden.