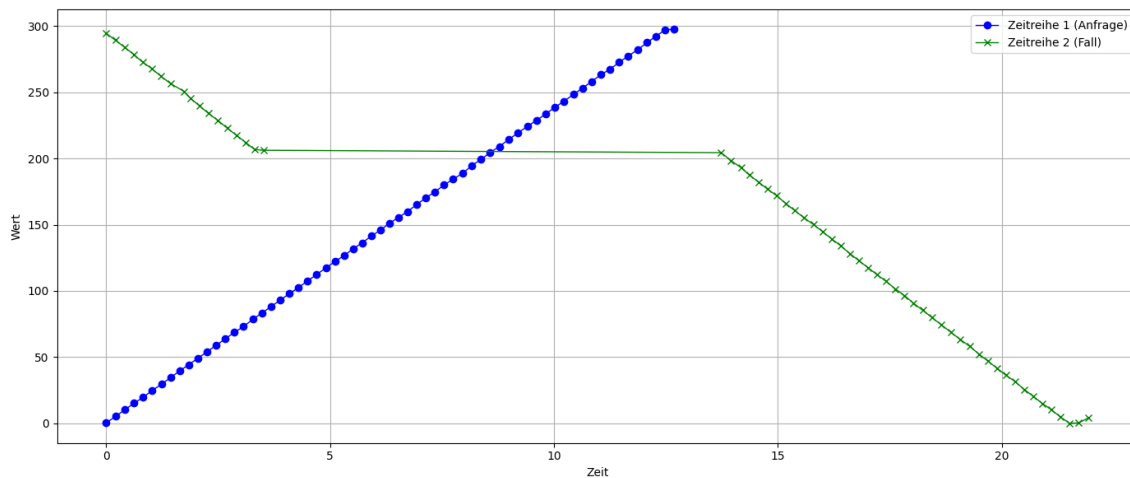


Evaluation von Erklärtexten

Die Grafik zeigt zwei Zeitreihen aus ProCAKE – W11 (Anfrage) und W397 (Fall). Der nachfolgende Text wurde von ChatGPT generiert und beschreibt die Ähnlichkeit der Zeitreihen mithilfe des Dynamic Time Warping (DTW)-Verfahrens. Ziel dieser Evaluation ist es, die Qualität der generierten Erklärung einzuschätzen. Bewertet werden dabei die fachliche Korrektheit, die inhaltliche Vollständigkeit sowie die Relevanz der Informationen in Bezug auf den konkreten Fall. Darüber hinaus werden die Verständlichkeit für fachfremde Nutzer:innen, logischer Aufbau und der sprachliche Lesefluss berücksichtigt.

Bitte lesen Sie die Erklärung sorgfältig durch und beantworten Sie anschließend die Bewertungsfragen anhand der bereitgestellten Skala. Ergänzend können Sie offene Anmerkungen zu möglichen Fehlern, Verständlichkeitsproblemen oder Verbesserungsvorschlägen machen.



LLM Output:

Im vorliegenden Vergleich werden zwei Zeitreihen im Detail untersucht: Die Anfrage-Zeitreihe umfasst 63 Messpunkte und wurde über einen Zeitraum von etwa 12,7 Sekunden aufgezeichnet. Die Fall-Zeitreihe besitzt insgesamt 59 Punkte und erstreckt sich über mehr als 21 Sekunden. **In der grafischen Darstellung wird deutlich, dass beide Reihen fast durchgehend mit ähnlichen Zeitabständen gemessen wurden.**

Der Vergleich erfolgt mithilfe des Dynamic Time Warping (DTW)-Verfahrens. Dieser Algorithmus ermöglicht es, Unterschiede in der Länge oder dem zeitlichen Ablauf der Reihen auszugleichen und so selbst bei Verschiebungen oder verzerrten Abläufen eine optimale Zuordnung der jeweiligen Messwerte zu erzielen. Bei DTW wird zunächst eine Matrix erstellt, die sämtliche möglichen Kombinationen von Messpunkten aus beiden Zeitreihen erfasst. Für jedes Paar wird ein sogenannter lokaler Ähnlichkeitswert berechnet. Hierbei fließt hauptsächlich ein, wie ähnlich sich die beiden Werte an einer bestimmten Stelle sind; außerdem wird berücksichtigt, wie groß der zeitliche Abstand zwischen den

gematchten Punkten ist. Ein perfekter Wert- oder Zeitunterschied ergibt die höchste Ähnlichkeit, größere Abstände verringern sie schrittweise, bis bei zu großem Unterschied keinerlei Übereinstimmung mehr erkannt wird.

In dieser Matrix prüft der Algorithmus für jeden Punkt, auf welchem Weg aus den vorherigen Schritten die höchste Summenähnlichkeit erreichbar ist. Ein diagonaler Schritt bedeutet dabei, dass jeweils ein Messpunkt aus beiden Reihen direkt zugeordnet wird, was als besonders aussagekräftig betrachtet und daher stärker gewichtet wird. Horizontale und vertikale Schritte gleichen hingegen Dehnungen oder Sprünge im Verlauf aus und zählen jeweils einfacher.

Der DTW-Pfad, also die tatsächlich ermittelte optimale Zuordnung der Messpunkte, spiegelt die Strukturunterschiede der beiden Zeitreihen deutlich wider. Zu Beginn bewegen sich beide Reihen noch weitgehend gemeinsam, jedoch zeigt sich direkt im weiteren Verlauf ein starkes Auseinanderdriften: **Praktisch alle Werte der ersten Zeitreihe im Bereich von etwa Messwert 12 bis etwa 36 werden ausschließlich demselben Messpunkt der zweiten Zeitreihe zugeordnet. Es folgt eine weitere Phase, in der jeweils ein Messwert aus der ersten Zeitreihe dem jeweils folgenden Messwert der zweiten Zeitreihe zugeordnet wird.** Der DTW-Pfad ist dadurch über weite Strecken nicht diagonal, sondern enthält lange horizontale und vertikale Abschnitte. Dies deutet auf fehlende Entsprechungen und großflächige Abweichungen zwischen den Zeitreihen hin.

In der Grafik ist auch klar zu erkennen, dass sich die Reihen an keinem Punkt wirklich begegnen – die Anfrage-Reihe steigt an, während die Fall-Reihe bereits abfällt.

Durch diese anhaltend hohen Wertdifferenzen erreicht die DTW-Analyse nur einen sehr niedrigen Gesamtähnlichkeitswert von 0,18 auf einer Skala von 0 bis 1. Das Ergebnis zeigt eindeutig, dass eine starke strukturelle Verschiedenheit zwischen den Zeitreihen besteht und kaum Gemeinsamkeiten sowohl im Werteverlauf als auch im zeitlichen Verlauf vorhanden sind. Entsprechend bestätigt die grafische Darstellung, dass die beiden Reihen zu keiner Phase ähnliche Werte oder ähnliche Kurvenformen aufweisen.

Bitte bewerten Sie die folgenden Aussagen auf einer Skala von 1 bis 5:

1 = stimme überhaupt nicht zu 5 = stimme voll zu

1. Die Informationen sind fachlich korrekt dargestellt. Die Erklärung ist fachlich korrekt - alle dargestellten Fakten, Begriffe, Interpretationen und DTW-spezifischen Konzepte sind richtig und frei von sachlichen Fehlern.

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

2. Sowohl die relevanten Zeitreihenmerkmale als auch die Funktionsweise des DTW-Algorithmus (Matrix, Pfad, lokale und globale Ähnlichkeit) werden vollständig beschrieben.

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

3. Die Erklärung beschränkt sich auf zentrale Aspekte, die für das Verständnis des konkreten DTW-Vergleichs wichtig sind, und vermeidet allgemeine oder irrelevante Informationen.

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

4. Die Erklärung bezieht sich konkret auf den dargestellten Fall und erklärt ihn kontextbezogen.

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

5. Die Erklärung ist für Laien angemessen einfach gehalten.

1	2	3	4	5
<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

6. Die Sprache der Erklärung ist kohärent und logisch aufgebaut.

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

7. Die Formulierungen sind flüssig und gut lesbar.

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

8. Die Erklärung ist präzise, vermeidet unnötige Wiederholungen und Abschweifungen.

1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

Offene Fragen:

- Gibt es fachliche Fehler oder kritische Auslassungen?

- Welche Aspekte der Erklärung könnten für Laien missverständlich sein?

Fachbegriffe ohne vereinfachte Umschreibung.

Die Erklärung der Matrix- und Pfadsuche ist zwar korrekt, könnte aber für Nicht-Informatiker schwer vorstellbar sein, weil visuelle oder anschauliche Beispiele fehlen.

Schwer verständlicher Satz: Praktisch alle Werte der ersten Zeitreihe im Bereich von etwa Messwert 12 bis etwa 36 werden ausschließlich demselben Messpunkt der zweiten Zeitreihe zugeordnet

Ebenfalls schwer vorzustellen: Der

DTW-Pfad ist dadurch über weite Strecken nicht diagonal, sondern enthält lange horizontale und vertikale Abschnitte. Dies deutet auf fehlende Entsprechungen und großflächige Abweichungen zwischen den Zeitreihen hin.

- Welche konkreten Verbesserungen schlagen Sie inhaltlich oder didaktisch vor?