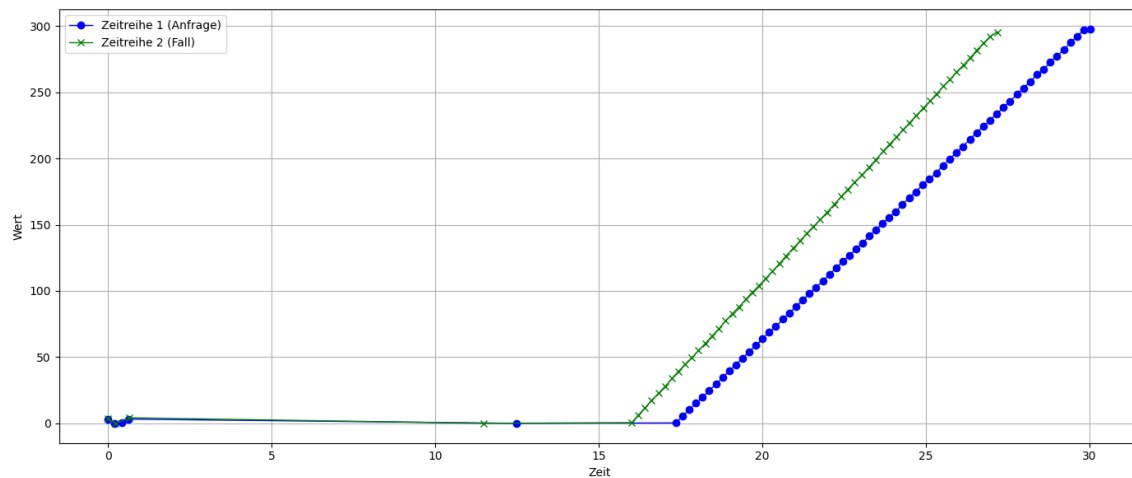


Evaluation von Erklärtexten

Die Grafik zeigt zwei Zeitreihen aus ProCAKE – W07 (Anfrage) und W391 (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 Folgenden werden zwei Zeitreihen verglichen: Die erste, die Anfrage-Zeitreihe, besteht aus 68 Messpunkten, die sich über rund 30 Sekunden erstrecken. Die zweite, die Fall-Zeitreihe, umfasst 59 Messwerte innerhalb von etwa 27 Sekunden. Die Messwerte sind bei beiden Reihen größtenteils regelmäßig verteilt, mit durchschnittlich wenigen Zehntelsekunden Abstand, abgesehen von einer gut sichtbaren längeren Lücke zu Beginn beider Reihen – in der Anfrage-Zeitreihe zwischen 0,6 und 12,5 Sekunden, in der Fall-Zeitreihe zwischen 0,6 und 11,5 Sekunden, jeweils mit nur einem einzelnen Wert in dieser Zeitspanne.

Für den Vergleich wurde der Dynamic Time Warping (DTW) Algorithmus eingesetzt. Diese Methode ermöglicht es, auch Zeitreihen unterschiedlicher Länge oder mit unregelmäßiger Taktung bestmöglich aufeinander abzustimmen. Dabei wird jedem denkbaren Punktpaar aus beiden Reihen ein lokaler Ähnlichkeitswert zugewiesen, der sich zur Hälfte aus dem zeitlichen Abstand, zur anderen Hälfte aus dem Unterschied der Messwerte ermittelt –

beide Anteile werden linear berücksichtigt, das heißt: je näher Zeit und Wert beieinander liegen, desto höher die Ähnlichkeit. Übersteigt der Zeitunterschied 100 Sekunden oder der Wertunterschied 200 Einheiten, ist die Ähnlichkeit gleich null.

Alle diese lokalen Ähnlichkeiten werden zu einer Matrix zusammengefasst. Aus dieser Matrix sucht DTW einen optimalen Pfad, mit dem Ziel, die Gesamtsumme aller zugeordneten Paarähnlichkeiten zu maximieren. Diagonale Schritte kennzeichnen die direkte Zuordnung eines Zeitpunkts aus beiden Reihen; sie werden doppelt gewichtet. **Horizontale und Vertikale Schritte entstehen, wenn zu einem Messpunkt aus der einen Zeitreihe zwei benachbarte Punkte aus der anderen zugeordnet werden – dies dient zum Ausgleich von unterschiedlichen Messfrequenzen oder Längen der Reihen.** Der tatsächliche DTW-Pfad zeigt im vorliegenden Fall für den überwiegenden Teil einen diagonalen Verlauf, was auf eine sehr enge Zuordnung zwischen den Punkten hinweist. Lediglich am Anfang sowie an einigen Stellen im weiteren Verlauf gibt es kurze Abschnitte mit horizontalen Schritten. Diese entstehen vor allem deshalb, weil die Anfrage-Zeitreihe ein paar Messpunkte mehr und einige zeitliche Überhänge aufweist.

Entscheidend für den berechneten DTW-Score ist, dass bei den meisten Zuordnungen die Messwerte sehr ähnlich sind, jedoch ein konstanter Zeitversatz von rund zwei Sekunden zwischen den beiden Reihen besteht – das ist sowohl im Zuordnungspfad als auch im Diagramm gut erkennbar. Obwohl die Positionen übereinstimmen, fällt dieser Zeitunterschied bei jedem einzelnen Paar ins Gewicht, da sowohl Zeit- als auch Wertähnlichkeit gleich stark zur lokalen Ähnlichkeit beitragen. Die lokalen Werte bleiben daher überall leicht unter dem Maximalwert, da die Zeitdifferenz zu einem konstanten Abzug führt – und zwar nicht nur punktuell, sondern im gesamten Verlauf des steilen Anstiegsabschnittes.

Aus der Summe dieser lokalen Ähnlichkeiten errechnet sich ein DTW-Score von 0,987. Der Wert zeigt eine fast perfekte Übereinstimmung beider Zeitreihen an – einzig der kleine, aber systematische zeitliche Abstand senkt die Gesamtsumme minimal.

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

1 = stimme überhaupt nicht zu 5 = stimme voll zu

1. 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.

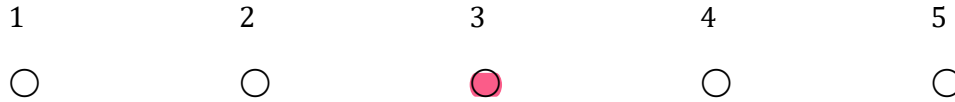
1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

7. Die Formulierungen sind flüssig und gut lesbar.

1	2	3	4	5
---	---	---	---	---



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



Offene Fragen:

- Gibt es fachliche Fehler oder kritische Auslassungen?

- Welche Aspekte der Erklärung könnten für Laien missverständlich sein?

Schwer zu verstehen woher die Zahlen 100 und 200 kommen; Dabei wird jedem denkbaren Punktpaar aus beiden Reihen ein lokaler Ähnlichkeitswert zugewiesen, der sich zur Hälfte aus dem zeitlichen Abstand, zur anderen Hälfte aus dem Unterschied der Messwerte ermittelt – beide Anteile werden linear berücksichtigt, das heißt: je näher Zeit und Wert beieinander liegen, desto höher die Ähnlichkeit. Übersteigt der Zeitunterschied 100 Sekunden oder der Wertunterschied 200 Einheiten, ist die Ähnlichkeit gleich null.

Fachbegriffe sollten erklärt werden

Die Begriffe vertikal, horizontal, diagonal sind nur schwer vorstellbar

- Welche konkreten Verbesserungen schlagen Sie inhaltlich oder didaktisch vor?