

Hindernisse und Möglichkeiten von ChatGPT 3.5 bei der Erstellung eines interaktiven Dashboards und über dessen Usability

Sarah Bauhofer

Pädagogische Hochschule Weingarten

ABSTRACT

This project explored the creation of a dashboard using ChatGPT 3.5, documenting the challenges and opportunities that arise when working with large data sets and visualizations. The dashboard was evaluated for its usefulness through a usability test.

The test results show that the dashboard worked well overall. However, the build process reveals that ChatGPT 3.5 is unable to create dashboards specific to individual needs and has difficulty processing large data sets. Despite these limitations, ChatGPT 3.5 can be viewed as a supporting tool that can save time and money, especially in a professional context.

Keywords: Chatbot, Visualization, interactive Dashboard, ChatGPT, Usability

1 EINLEITUNG

Softwaregestützte Visualisierungssysteme veranschaulichen große Datensätze, um Nutzende bei der effektiven Ausführung von Aufgaben zu unterstützen (Munzner, 2015). Sie helfen auch dabei, zugrunde liegende Muster in umfangreichen Datenanalysen aufzudecken (Li et al., 2024).

Mit dem Aufkommen großer Sprachmodelle wie ChatGPT ergeben sich neue Möglichkeiten für die automatisierte Erstellung von Dashboards und die interaktive Datenanalyse. Kim et al. (2024) weisen auf die Notwendigkeit hin, die Herausforderungen und Möglichkeiten der Datenvisualisierungen mit ChatGPT zu untersuchen. Angesichts der steigenden Beliebtheit von ChatGPT ist es wichtig zu verstehen, wie gut dieses Tool in der Lage ist, mit Visualisierungen und großen Datensätzen zu arbeiten. Solche Erkenntnisse sind entscheidend für die zukünftige Entwicklung und Anwendung dieser Technologien in verschiedenen Domänen.

Bisherige Studien haben sich auf die allgemeinen Fähigkeiten von ChatGPT konzentriert, menschenähnliche Konversationen zu generieren und Aufgaben autonom auszuführen (Bukar et al., 2024; Spitzer, 2023; Deng & Lin, 2023; Gentil & Merle, 2023). Allerdings gibt es noch wenige Arbeiten, die sich speziell mit der Integration von ChatGPT in Visualisierungssysteme befassen.

Das Ziel dieses Projekts war es daher, zu erforschen, inwiefern ChatGPT 3.5 zur Erstellung von Visualisierungen und der Verarbeitung großer Datensätze eingesetzt werden kann. Hierfür wurde ein für den Datensatz domänenspezifisches Dashboard mit Hilfe von ChatGPT 3.5 erstellt, wobei die Möglichkeiten und Hindernisse dieses Prozesses dokumentiert wurden.

Ein weiterer Schwerpunkt war die Bewertung der Nutzbarkeit dieses Dashboards durch einen Usability-Test, um den tatsächlichen Nutzen für mögliche Personen dieser Domäne zu ermitteln. Um diese Ziele zu erreichen, wurde folgende Fragestellung entwickelt: *Inwiefern kann ChatGPT 3.5 zur Erstellung eines interaktiven Dashboards verwendet werden und wie nutzerfreundlich ist dieses Dashboard?*

2 UMSETZUNG

Die Erstellung eines Dashboards bietet eine praxisnahe Möglichkeit, die Interaktion von ChatGPT mit großen Datensätzen zu bewerten und die Nutzerfreundlichkeit der Visualisierungen zu untersuchen. Dashboards sind weit verbreitete Werkzeuge zur Visualisierung und Analyse großer Datenmengen, da sie eine intuitive Oberfläche bieten, die es Nutzenden ermöglicht, Daten effizient zu erfassen und zu interpretieren. Ein Dashboard kombiniert verschiedene Visualisierungselemente, wie Diagramme und Grafiken, um komplexe Datensätze verständlich darzustellen (Few, Stephen, 2006.; Jacobs & Hensel-Börner, 2020; Pappas & Whitman, 2011).

Die Interaktivität, mit der ein Dashboard erstellt werden kann, hilft zusätzlich dabei Komplexe Daten und große Mengen an Informationen zu vereinfachen (Munzner, 2015; Shneiderman, 2003). Aufgrund dieser Eigenschaften eignet sich ein Dashboard hervorragend, um die Fähigkeiten von ChatGPT 3.5 in Bezug auf Datenvisualisierung und -analyse zu testen.

Die Erstellung des Dashboards orientierte sich an den Leitfragen „was“, „warum“ und „wie“ visualisiert werden sollte (Munzner, 2015). Die Auswahl der Visualisierungen erfolgte basierend auf den Eigenschaften des Datensatzes, dem spezifischen Anwendungsszenario sowie der Komplexität der Interpretierbarkeit. Klare und einfache Designs wurden aufgrund ihrer Vorteile hinsichtlich der Verständlichkeit bevorzugt (Jacobs & Hensel-Börner, 2020).

Ziel des Usability Tests war es auch, das Dashboard sowohl effektiv (Vorhandensein relevanter Informationen, die auch vom Empfangenden verstanden werden) als auch effizient (schnelle Informationsaufnahme ohne Wahrnehmungsprobleme) zu gestalten (Jacobs & Hensel-Börner, 2020). Die Interaktionen wurden auf Basis des Shneiderman's Mantra: Overview first, zoom and filter, details on demand (1996) erstellt. Die wesentlichen Komponenten des Dashboards waren daher: Filteroptionen, tabellarische Abbildung der Ergebnisse und detaillierte Ansicht zu den einzelnen Leistungen (siehe Abb. 1).

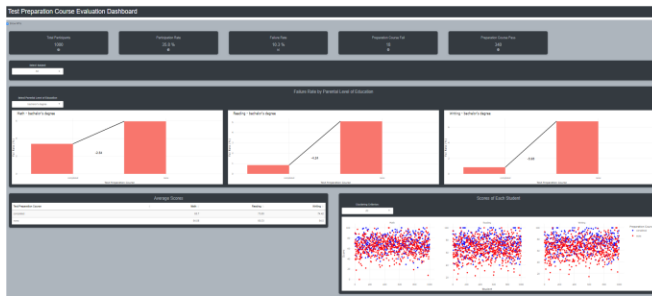


Abb. 1 Dashboard

ERSTELLUNG DES DASHBOARDS MIT CHATGPT UND R

Das Dashboard wurde in RStudio anhand eines ShinyApp-Codes erstellt. GitHub diente zur Dokumentation der Möglichkeiten und Hindernisse von ChatGPT 3.5 sowie zur Dokumentation diverser Codes. Die vollständige Dokumentation lässt sich hier einsehen: [GitHub](#)

Der verwendete Datensatz „Studentsperformance“ enthält Angaben zu 1000 Studierenden über erreichte Punkte in verschiedenen Tests (Mathe, Schreiben, Lesen). Es ist jedoch nicht ersichtlich, um welche Schülergruppe (Student, Grundschule, Weiterführende Schule) es sich handelt. Im Datensatz sind verschiedene demografische Einflussfaktoren enthalten, darunter Geschlecht, Ethnizität, Bildungsstand der Eltern, Art des Mittagessens und Teilnahme an einem Vorbereitungskurs für den Test.

Der Datensatz wurde bereinigt und an die spezifischen Szenarien angepasst. Faktoren wie Mittagessen und Ethnizität wurden bei der Erstellung des Dashboards nicht berücksichtigt, da sie sich nicht für das ausgewählte Szenario eigneten.

Durch die Übermittlung der Datenstruktur, der Forschungsfrage und Szenarien, war ChatGPT in der Lage, passende Visualisierungen vorzuschlagen. In Kombination mit Shiny, das Beispiele für die Codierung von Visualisierungen enthält, war es ChatGPT möglich, die Visualisierungen zu modifizieren.

Das Prompting orientierte sich dabei stets an einer deklarativen Visualisierungssprache (DVL), welche Datenzuordnung, Diagrammtypen, Stilrichtlinien, Interaktionsfunktionen und Layoutdesign beinhaltet (Li et al., 2024).

MÖGLICHKEITEN UND HINDERNISSE VON CHATGPT 3.5

Die Nutzung präziser Ausdrucksweise und Schritt-für-Schritt-Ansätze sind essenziell. Bei der Code-Generierung und der Anpassung von Interaktionen treten gelegentlich Schwierigkeiten auf, weshalb ein Verständnis für Coding-Fehler notwendig ist

ChatGPT 3.5 unterstützt jedoch effektiv bei der Ideenfindung und Gestaltung von Dashboards. Ein wesentlicher Nachteil ist die Unfähigkeit von ChatGPT 3.5, auf Templates oder externe Codes zuzugreifen sowie Kompatibilitätsprobleme mit aktuellen Softwareversionen von Shiny oder RStudio. ChatGPT 3.5 ermöglicht die Anpassung des Datensatzes, wie beispielsweise das Entfernen von Spalten, kann aber lediglich auf den vorhandenen Datensatz und dessen

Variablenbezeichnungen zugreifen und Berechnungen dafür erstellen. Eine Übersetzung des Datensatzes in eine andere Sprache ist nicht möglich, da die Werte dann nicht mehr den Bezeichnungen zugeordnet werden können.

ChatGPT 3.5 unterstützt bei verschiedenen Interaktionen, jedoch gibt es Schwierigkeiten bei der Anpassung von Interaktionen und Voreinstellungen, wie der Farbgebung von Balken in ShinyApp. Es bietet jedoch Hilfe bei der Auswahl von Farben und der Integration von Bildern und Symbolen.

Insgesamt bietet ChatGPT 3.5 verschiedene Möglichkeiten zur Erstellung eines Dashboards in RStudio, einschließlich des Zugriffs auf R-Datensätze. Geduld und Aufmerksamkeit sind bei der Arbeit mit ChatGPT 3.5 jedoch weiterhin erforderlich.

3 USABILITY TEST

Um zu testen, ob und inwiefern das Dashboard auch nutzbar ist, wurde ein Usability Test durchgeführt. Die ISO-Norm 9241/10 beschreibt Usability als gegeben, wenn der Benutzende seine Aufgaben effektiv, effizient und zufriedenstellend mithilfe eines technischen Objekts erledigen kann und wie ergonomisch die interaktiven Systeme gestaltet sind (Prümper, 1997).

Daraus ergeben sich 7 Grundprinzipien, Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Steuerbarkeit, Erwartungskonformität, Fehlerrobustheit, Individualisierbarkeit und Lernförderlichkeit (Prümper, 1997).

Im vorliegenden Projekt besitzt der Fragebogen 26 angepasste Items mit 6 der 7 Grundprinzipien. Die Bewertung der bipolaren Aussagen erfolgte auf einer siebenstufigen Ratingskala von sehr negativ ("---") bis sehr positiv ("+++") (Figl, 2009). Zudem wurde die Anonymität der Teilnehmenden sichergestellt.

DATENERHEBUNG UND -AUSWERTUNG

Die Daten wurden an demselben Computer erhoben ($N=10$). Vorerst wurden die Aufgaben (Szenarien) für das Dashboard bearbeitet, danach wurde der Fragebogen zur Usability des Dashboards ausgehändigt. Die durchschnittliche Bearbeitungszeit lag bei 20 Minuten. Die erhobenen Daten wurden nach dem mit dem Mittelwert berechnet (Prümper, 1997).

4 ERGEBNISSE

Nach Prümper (1997) ist eine Anwendung gut gestaltet, wenn Sie in den einzelnen Kategorien sowie in der Gesamtwertung einen Mittelwert von 1 aufweist. Den höchsten Mittelwert erreichte die Aufgabenangemessenheit ($M = .83$). Den niedrigsten Wert erreichte die Fehlertoleranz ($M = .43$). Insgesamt wurde das Dashboard von den Teilnehmenden jedoch mit einer hohen Nutzerfreundlichkeit bewertet.

5 DISKUSSION

Die Kollaboration mit RStudio und ShinyApp ist zuverlässig. Allerdings gibt es Hindernisse, gerade Fehlerbehandlungen,

die es zu beachten gilt. Allein durch die Arbeit mit ChatGPT wird vermutlich nie das gewünschte Ziel (Dashboard) erreicht werden. Die Ergebnisse des Usability Tests zeigen ebenfalls, dass Anpassungen nötig sind, um eine bessere Usability zu gewährleisten.

Die Ergebnisse des Usability Tests könnten durch die Möglichkeiten und Hindernisse durch ChatGPT 3.5 beeinflusst worden sein. Dennoch konnte sich das Dashboard als nützlich in dieser Domäne herausstellen. Die Stichprobengröße ist ebenfalls zu bemängeln. Hier könnten Optimierungen am Design und der Erhebung vorgenommen werden, die zu einer größeren Stichprobenreichweite führen kann.

Zukünftige Forschungen könnten an dieser Arbeit anknüpfen, um zu testen, ob und inwiefern ChatGPT in der Lage ist, solche Anpassungen vorzunehmen. Zudem wurde ChatGPT 3.5 aktualisiert auf ChatGPT 4. Hier gibt es nun die Möglichkeit, Dokumente und Bilder hochzuladen, was die Arbeit mit Visualisierungen oder großen Datensätzen verbessern könnte. Im Arbeitskontext könnten diese Ergebnisse interessant sein, da ChatGPT dabei unterstützen kann, große Datenmengen zu analysieren und diese aufzubereiten, was Zeit und Geld sparen könnte.

REFERENCES

- Bukar, Umar Ali, Md Shohel Sayeed, Siti Fatimah Abdul Razak, Sumendra Yogarayan, Oluwatosin Ahmed Amodu, und Raja Azlina Raja Mahmood. „Text Analysis on Early Reactions to ChatGPT as a Tool for Academic Progress or Exploitation“. *SN Computer Science* 5, Nr. 4 (29. März 2024): 366. <https://doi.org/10.1007/s42979-024-02714-7>.
- Deng, Jianyang, und Yijia Lin. „The Benefits and Challenges of ChatGPT: An Overview“. *Frontiers in Computing and Intelligent Systems* 2, Nr. 2 (5. Januar 2023): 81–83. <https://doi.org/10.54097/fcis.v2i2.4465>.
- Few, S. (2006) *Information Dashboard Design: the Effective Visual Communication of Data*, pp. 40–46. O'Reilly Media, Inc, Sebastopol
- Figl, Kathrin. „ISONORM 9241/10 und Isometrics: Usability-Fragebögen im Vergleich“. In *Mensch & Computer 2009*, von Hartmut Wandke, Saskia Kain, und Doreen Struve, 143–52. Oldenbourg Verlag, 2009. <https://doi.org/10.1524/9783486598551.143>.
- Gentil, Marie-Hélène, und Catherine Merle. „The Use of Quality Tools, Concepts and Methods to Help the Transition to Industry 4.0: On the Way to Quality 4.0“, 2023.
- Jacobs, Luise, und Susanne Hensel-Börner. „Die Kraft effektiver Daten-Visualisierung – CLEAR(I): Ein Leitfaden zur wirkungsvollen Dashboard-Gestaltung“. In *Data-driven Marketing*, herausgegeben von Silvia Boßow-Thies, Christina Hofmann-Stölting, und Heike Jochims, 43–75. Wiesbaden: Springer Fachmedien Wiesbaden, 2020. https://doi.org/10.1007/978-3-658-29995-8_3.
- Kim, Nam Wook, Grace Myers, und Benjamin Bach. „How Good Is ChatGPT in Giving Advice on Your Visualization Design?“ *arXiv*, 30. Januar 2024. <http://arxiv.org/abs/2310.09617>.
- Kompendium multimediales Lernen. X.media.press. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. <https://doi.org/10.1007/978-3-540-37226-4>.
- Li, Shuaimin, Xuanang Chen, Yuanfeng Song, Yunze Song, und Chen Zhang. „Prompt4Vis: Prompting Large Language Models with Example Mining and Schema Filtering for Tabular Data Visualization“. *arXiv*, 29. Januar 2024. <http://arxiv.org/abs/2402.07909>
- Munzner, Tamara. „Visualization Analysis & Design“, 2015
- Pappas, L., Whitman, L. (2011). *Riding the Technology Wave: Effective Dashboard Data Visualization*. In: Smith, M.J., Salvendy, G. (eds) *Human Interface and the Management of Information. Interacting with Information. Human Interface 2011. Lecture Notes in Computer Science*, vol 6771. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21793-7_29
- Prümper, Jochen. „Der Benutzungsfragebogen ISONORM 9241/10: Ergebnisse zur Reliabilität und Validität“. In *Software-Ergonomie '97*, herausgegeben von Rüdiger Liskowsky, Boris M. Velichkovsky, und Wolfgang Wüschmann, 49:253–62. Berichte des German Chapter of the ACM. Wiesbaden: Vieweg+Teubner Verlag, 1997. https://doi.org/10.1007/978-3-322-86782-7_21.
- Prümper, Jochen, und Michael Anft. „Die Evaluation von Software auf Grundlage des Entwurfs zur internationalen Ergonomie-Norm ISO 9241 Teil 10 als Beitrag zur partizipativen Systemgestaltung - ein Fallbeispiel“, o. J.
- Shneiderman, B. (1996). *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations*. In: *Proc. of the IEEE Symposium on Visual Languages*, Washington, 1996 (pp. 336–343). IEEE Computer Society Press.
- Spitzer, Manfred. „ChatGPT: Nur ein weiterer Trend oder eine Revolution?“ *Nervenheilkunde* 42, Nr. 04 (April 2023): 192–99. <https://doi.org/10.1055/a-1948-8785>.