



Bachelor-Thesis

Konzeption und Evaluation eines hybriden Interaktionskonzepts
für ein webbasiertes Analytics-Dashboard mit natürlicher
Spracheingabe und klassischer Interaktion.

zur Erlangung des akademischen Grades

Bachelor of Science

im dualen Studiengang Informatik an der IU Internationale Hochschule

von

Kevin Tamme

Matrikelnummer: 102202087



31. März 2026

Referent: Prof. Dr. Klaus Quibeldey-Cirkel

Korreferent: Prof. Dr. Philipp Diebold

Kurzfassung

Kontext und Motivation: Analytics-Dashboards sind ein zentrales Werkzeug datengetriebener Entscheidungsfindung und basieren traditionell auf grafischen Benutzeroberflächen. Mit der zunehmenden Verfügbarkeit großer Sprachmodelle gewinnen natürliche Sprachschnittstellen als alternative oder ergänzende Interaktionsform an Bedeutung. Empirische Studien zeigen jedoch, dass sprachbasierte Interaktion zwar zugänglicher wirkt, jedoch bei komplexeren Aufgaben mit Effizienz- und Genauigkeitsnachteilen sowie erhöhter kognitiver Belastung verbunden sein kann.

Zielsetzung: Ziel der Arbeit ist die Konzeption, prototypische Umsetzung und empirische Evaluation eines hybriden Interaktionskonzepts für ein webbasiertes Analytics-Dashboard, das natürliche Spracheingabe mit klassischer grafischer Interaktion kombiniert. Zentrale Forschungsfrage ist, wie sich diese Kombination auf Effizienz, Nutzererfahrung und Fehlerraten auswirkt.

Methodik: Auf Basis einer systematischen Literaturanalyse wird ein konzeptionelles Modell entwickelt, das Interaktionsformen als komplementäre Komponenten einer kontrollorientierten Architektur versteht. Aufgabenkomplexität wird als zentrale Entscheidungsdimension modelliert. Zur empirischen Untersuchung wird ein Webprototyp mit drei Interaktionsmodi (GUI, NLI und HYBRID) implementiert und in einer kontrollierten Nutzerstudie mit 12 Teilnehmenden evaluiert. Erhoben werden Bearbeitungszeit, Genauigkeit, Interaktionsverhalten sowie subjektive Maße (NASA-TLX, SUS).

Ergebnisse: Die Ergebnisse zeigen, dass der GUI-Modus die geringste Bearbeitungszeit (31,70 s) aufweist, jedoch die niedrigste Accuracy (52,08 %). Der NLI-Modus erzielt die höchste Accuracy (72,92 %) und First-Try Accuracy (70,83 %), ist jedoch mit deutlich erhöhter Bearbeitungszeit (96,04 s) verbunden. Der HYBRID-Modus liegt in den meisten Metriken zwischen beiden Ansätzen und weist die geringste wahrgenommene Arbeitsbelastung (TLX: 5,83) auf.

Beitrag: Die Arbeit zeigt, dass keine Interaktionsform universell überlegen ist, sondern systematische Trade-offs zwischen Effizienz, Genauigkeit und kognitiver

Belastung bestehen. Hybride Interaktion wird als kontextabhängige Kontrollarchitektur konzeptualisiert, die insbesondere durch die Kombination aus sprachlicher Spezifikation und visueller Kontrolle Vorteile bietet und damit einen Beitrag zur Gestaltung zukünftiger Analytics-Systeme leistet.

Abstract

Context and Motivation: Analytics dashboards are a central tool for data-driven decision-making and traditionally rely on graphical user interfaces. With the increasing availability of large language models, natural language interfaces are emerging as an alternative or complementary interaction paradigm. However, empirical evidence shows that while language-based interaction appears more accessible, it may lead to lower efficiency and accuracy in more complex tasks and increase cognitive load.

Objective: This thesis aims to design, implement, and evaluate a hybrid interaction concept for a web-based analytics dashboard that combines natural language input with classical graphical interaction. The central research question examines how this combination affects efficiency, user experience, and error rates.

Methodology: Based on a systematic literature review, a conceptual model is developed that treats interaction modalities as complementary components of a control-oriented architecture. Task complexity is modeled as the primary decision dimension. A web-based prototype with three interaction modes (GUI, NLI, and HYBRID) is implemented and evaluated in a controlled user study with 12 participants. Metrics include task completion time, accuracy, interaction behavior, and subjective measures (NASA-TLX, SUS).

Results: The results show that the GUI mode achieves the shortest task completion time (31.70 s) but the lowest accuracy (52.08%). The NLI mode yields the highest accuracy (72.92%) and first-try accuracy (70.83%), but requires significantly more time (96.04 s). The hybrid mode positions itself between both approaches and results in the lowest perceived cognitive load (TLX: 5.83).

Contribution: The findings demonstrate that no single interaction modality is universally superior, but rather that systematic trade-offs exist between efficiency, accuracy, and cognitive load. Hybrid interaction is conceptualized as a context-dependent control architecture that benefits from combining linguistic specification with visual control, contributing to the design of future analytics systems.

Danksagung

Ich danke Gott für die Kraft, die Ausdauer und die Klarheit, die mich durch diese Zeit getragen haben.

Von tiefstem Herzen danke ich meiner gesamten Familie, ohne deren Unterstützung, Geduld und Vertrauen mein Studium nicht möglich gewesen wäre.

Ich danke meinen Freunden, die mich in dieser Zeit begleitet und in schwierigen Momenten aufgefangen haben.

Ein besonderer Dank gilt außerdem einem Menschen, der mich in dieser Zeit begleitet, gestärkt und stets an meiner Seite gestanden hat.

Ich danke meinen Betreuern Prof. Dr. Klaus Quibeldey-Cirkel und Prof. Dr. Philipp Diebold für die Übernahme der Betreuung und die Möglichkeit, diese Arbeit anzufertigen.

Abschließend danke ich den Entwicklerinnen und Entwicklern sowie der Open-Source-Community für die bereitgestellten Werkzeuge und Bibliotheken, die diese Arbeit in dieser Form ermöglicht haben.

Hinweis zur geschlechtergerechten Sprache

In der Arbeit wird auf eine geschlechtergerechte Sprache geachtet. Aus Gründen der besseren Lesbarkeit wird in der Arbeit eine Mischung verschiedener Formen geschlechtergerechter Sprache verwendet:

- **Geschlechtsneutrale Formulierungen** (z. B. Studierende, Mitarbeitende, Person)
- **Paarformen** (z. B. Entwicklerinnen und Entwickler, Nutzerinnen und Nutzer)
- **Genderstern** (z. B. Benutzer*innen) zur Sichtbarmachung aller Geschlechter

Alle verwendeten Personenbezeichnungen gelten gleichermaßen für alle Geschlechter. Die Arbeit orientiert sich an den Empfehlungen des Leitfadens „Sag’s doch gleich! Geschlechtergerechte Sprache an Thüringer Hochschulen“ des Thüringer Kompetenznetzwerks Gleichstellung (TKG).

Inhaltsverzeichnis

Kurzfassung	i
Abstract	iii
Danksagung	v
Hinweis zur geschlechtergerechten Sprache	ix
Abbildungsverzeichnis	xv
Tabellenverzeichnis	xvii
Abkürzungsverzeichnis	xix
1 Einleitung	1
1.1 Zielsetzung der Arbeit	2
1.2 Forschungsfragen	2
1.3 Abgrenzung	3
1.4 Aufbau der Arbeit	3
2 Theoretischer Hintergrund	5
2.1 Analytics-Dashboards und Visual Analytics	5
2.2 Direkte Manipulation in grafischen Benutzeroberflächen	6
2.3 Natürliche Sprachschnittstellen in Analytics-Systemen	6
2.4 Kognitive Belastung und Transparenz	7
2.5 Hybride Interaktionsansätze in Analytics-Systemen	7
2.6 Zwischenfazit	8
3 Konzeption des hybriden Interaktionskonzepts	9
3.1 Gestaltungsannahmen und Designziele	9
3.2 Aufgabenklassifikation als kontinuierliche Entscheidungsdimension .	11
3.3 Feedback-, Transparenz- und Kontrollmechanismen	13
3.4 Abgrenzung zu monolithischen Interaktionsansätzen	14

3.5	Operationalisierung im prototypischen System	15
3.5.1	Interaktionsmodi	15
3.5.2	Abfragemodell	16
3.5.3	Datensatz	16
3.5.4	Bezug zur Aufgabenklassifikation	16
4	Methodik	17
4.1	Forschungsdesign	17
4.2	Prototypische Implementierung	18
4.2.1	Abfragemodell	18
4.2.2	Abfrageausführung	19
4.2.3	Interaktionsmodi	19
4.3	Studienaufgaben	24
4.3.1	Klassifikation der Aufgabenkomplexität	24
4.4	Messinstrumente	25
4.4.1	Objektive Maße	25
4.4.2	Subjektive Maße	25
4.5	Studiendurchführung	25
4.6	Datenerhebung und Logging	27
4.7	Datensatz	27
4.8	Datenaufbereitung und statistische Analyse	27
5	Ergebnisse	29
5.1	Stichprobe	29
5.2	Bearbeitungszeit	29
5.3	Genauigkeit und Fehlerraten	30
5.4	Interaktionsverhalten	30
5.5	Explorative Analyse: Einfluss der Aufgabenkomplexität	32
5.6	Kognitive Belastung	33
5.7	Gebrauchstauglichkeit	34
5.8	Inferenzstatistische Analyse	34
5.9	Zusammenfassung der Ergebnisse	35
6	Diskussion	37
6.1	Einordnung der Ergebnisse im Kontext der Forschungsfragen	37

6.2	Einfluss der Interaktionsform auf Effizienz und Genauigkeit	38
6.3	Einfluss der Aufgabenkomplexität	40
6.4	Rolle des hybriden Interaktionskonzepts	41
6.5	Kognitive Belastung und subjektive Wahrnehmung	42
6.6	Bezug zu theoretischen Annahmen	43
6.7	Implikationen für das Interaktionskonzept	43
6.8	Limitationen der Studie	44
7	Fazit und Ausblick	47
7.1	Zusammenfassung der Arbeit	47
7.2	Zentrale Erkenntnisse	47
7.3	Beitrag der Arbeit	48
7.4	Limitationen	49
7.5	Ausblick	49
	Literaturverzeichnis	51
	Index	54
A	Anhang	55
A.1	Studienmaterialien	56
A.1.1	Trainingsaufgaben	56
A.1.2	Aufgabenset A	56
A.1.3	Aufgabenset B	56
A.1.4	Aufgabenset C	56
A.2	Fragebögen	57
A.2.1	System Usability Scale (SUS)	57
A.2.2	NASA Task Load Index (NASA-TLX)	58
A.3	Verzeichnis der KI-Nutzung	58

Abbildungsverzeichnis

3.1	Strukturelle Dimensionen der Aufgabenkomplexität	12
3.2	Schwellenwertmodell der Modalitätswahl entlang der Aufgabenkomplexität	12
3.3	Ablauf der hybriden Interaktion mit Feedback- und Kontrollmechanismen	14
3.4	Übersicht der drei Interaktionsmodi im prototypischen System . . .	16
4.1	Latin-Square-Design zur Ausbalancierung von Bedingungen und Aufgabensets	18
4.2	Systemarchitektur des Studienprototyps	19
4.3	Übersetzungsprozess von natürlicher Sprache zur strukturierten Abfrage	19
4.4	Studientool im GUI-Modus	21
4.5	Studientool im NLI-Modus	22
4.6	Studientool im HYBRID-Modus	23
4.7	Ablauf der Studiendurchführung	26

Tabellenverzeichnis

5.1	Bearbeitungszeit pro Interaktionsmodus	29
5.2	Accuracy und First-Try Accuracy	30
5.3	Interaktionsmetriken pro Modus	31
5.4	Bearbeitungszeit nach Interaktionsmodus und Aufgabenkomplexität (Mittelwert / Median in s)	32
5.5	Accuracy nach Interaktionsmodus und Aufgabenkomplexität ($n = 24$ je Zelle)	33
5.6	NASA-TLX pro Interaktionsmodus	33
5.7	System Usability Scale (SUS)	34
A.1	Verzeichnis der in dieser Arbeit genutzten KI-Werkzeuge	59

Abkürzungsverzeichnis

Abkürzung	Bedeutung
API	Application Programming Interface (Anwendungsprogrammierschnittstelle)
DFG	Deutsche Forschungsgemeinschaft
GUI	Graphical User Interface (Grafische Benutzeroberfläche)
HYBRID	Hybrider Interaktionsmodus
IDE	Integrated Development Environment (Integrierte Entwicklungsumgebung)
KI	Künstliche Intelligenz
LLM	Large Language Model (Großes Sprachmodell)
NLP	Natural Language Processing (Natürlichsprachverarbeitung)
NLI	Natural Language Interface (Natürliche Sprachschnittstelle)
NASA-TLX	NASA Task Load Index (Fragebogen zur Arbeitsbelastung)
TLX	Task Load Index (Kurzform von NASA-TLX)
SUS	System Usability Scale (Fragebogen zur Gebrauchstauglichkeit)
UI	User Interface (Benutzeroberfläche)
SD	Standard Deviation (Standardabweichung)
n	Stichprobengröße

1 Einleitung

Analytics-Dashboards sind ein zentrales Instrument datengetriebener Entscheidungsfindung in Business-Intelligence- und Visual-Analytics-Kontexten. Klassische Systeme basieren auf grafischen Benutzeroberflächen, die durch Filter, Menüs und direkte Manipulation eine präzise und nachvollziehbare Analyse ermöglichen, jedoch ein Verständnis der zugrunde liegenden Datenmodelle und Interaktionslogik voraussetzen.

Mit der zunehmenden Verfügbarkeit großer Sprachmodelle gewinnen natürliche Sprachschnittstellen (*Natural Language Interfaces, NLI*) als alternative oder ergänzende Interaktionsform an Bedeutung. Sie ermöglichen die Formulierung analytischer Fragestellungen in Alltagssprache und versprechen einen niedrigschwelligen Zugang zu komplexen Analysen.

Empirische Befunde zeigen jedoch ein differenziertes Bild: Sprachbasierte Interfaces unterstützen flexible Exploration, sind jedoch häufig mit längeren Bearbeitungszeiten, erhöhter kognitiver Belastung und einem höheren Risiko von Fehlinterpretationen verbunden (Besharati, 2025; Voigt et al., 2022). Klassische Dashboards erweisen sich insbesondere bei strukturierten, wiederkehrenden Aufgaben als effizienter und verlässlicher. Eine vollständige Substitution grafischer Benutzeroberflächen durch natürliche Sprache erscheint daher nicht plausibel.

Aktuelle Arbeiten deuten stattdessen auf hybride Interaktionsansätze hin, bei denen natürliche Sprache zur Spezifikation analytischer Intentionen und grafische Interaktion zur Präzisierung, Kontrolle und Validierung eingesetzt wird. Entscheidend ist dabei eine aufgabenspezifische Verteilung der Interaktionsformen sowie die Sichtbarmachung von Systeminterpretationen durch geeignete Feedback-, Transparenz- und Kontrollmechanismen.

Vor diesem Hintergrund untersucht diese Arbeit die Konzeption und Evaluation eines hybriden Interaktionskonzepts für ein webbasiertes Analytics-Dashboard. Im Fokus steht die Analyse, unter welchen Bedingungen hybride Ansätze gegenüber monolithischen Interfaces Vorteile hinsichtlich Effizienz, Nutzererfahrung, kognitiver

Belastung und Fehleranfälligkeit bieten. Ergänzend wird explorativ untersucht, wie sich diese Effekte in Abhängigkeit von strukturellen Eigenschaften analytischer Aufgaben unterscheiden.

1.1 Zielsetzung der Arbeit

Ziel dieser Arbeit ist die Konzeption, prototypische Umsetzung und empirische Evaluation eines hybriden Interaktionskonzepts für ein webbasiertes Analytics-Dashboard, das natürliche Spracheingabe mit klassischer grafischer Interaktion kombiniert. Die Interaktionsgestaltung selbst bildet den zentralen Untersuchungsgegenstand.

Untersucht wird, wie analytische Aufgaben zwischen Spracheingabe, GUI-Interaktion und hybriden Interaktionsformen verteilt werden können und wie sich diese auf Bearbeitungszeit, kognitive Belastung, Fehleranfälligkeit und subjektive Wahrnehmung auswirken. Zusätzlich wird explorativ analysiert, inwiefern diese Effekte von der strukturellen Komplexität der Aufgaben abhängen. Auf Basis der Ergebnisse werden Implikationen für die Gestaltung hybrider Analytics-Dashboards abgeleitet.

1.2 Forschungsfragen

Die Arbeit wird von der folgenden zentralen Forschungsfrage geleitet:

Wie wirkt sich die Kombination aus natürlicher Spracheingabe und klassischer GUI-Interaktion auf Effizienz, Nutzererfahrung und Fehlerraten in Analytics-Dashboards aus?

Daraus ergeben sich folgende Unterfragen:

- Welche Arten analytischer Aufgaben profitieren besonders von Spracheingabe?
- Wie verändert sich die kognitive Belastung bei hybriden Interaktionen?
- Welche Feedback- und Kontrollmechanismen reduzieren Fehlinterpretationen?
- Wie wird ein hybrides Interface im Vergleich zu monolithischen Ansätzen wahrgenommen?

- Inwiefern variieren diese Effekte in Abhängigkeit von der strukturellen Komplexität der Aufgaben?

1.3 Abgrenzung

Die Arbeit fokussiert das Interaktionsdesign hybrider Analytics-Dashboards. Nicht betrachtet werden die Entwicklung oder Optimierung von Sprachmodellen sowie Aspekte wie Skalierung, Performance, Sicherheit oder Datenschutz produktiver Systeme.

Die Evaluation erfolgt in einer kontrollierten Nutzerstudie mit begrenzter Stichprobe. Aussagen zur Langzeitnutzung oder zu organisationalen Effekten sind nicht Ziel der Arbeit. Die Analyse der Aufgabenkomplexität erfolgt nachträglich und stellt keinen experimentell kontrollierten Faktor dar.

1.4 Aufbau der Arbeit

Kapitel 2 stellt den theoretischen Hintergrund dar. Kapitel 3 beschreibt das entwickelte Interaktionskonzept. Kapitel 4 erläutert das methodische Vorgehen und die Studiendurchführung. Kapitel 5 präsentiert die Ergebnisse, die in Kapitel 6 diskutiert werden. Kapitel 7 fasst die Erkenntnisse zusammen und gibt einen Ausblick. Eine Übersicht der verwendeten KI-Werkzeuge befindet sich in Anhang A.

2 Theoretischer Hintergrund

Dieses Kapitel entwickelt den theoretischen Bezugsrahmen der Arbeit und analysiert Unterschiede zwischen grafischer und sprachbasierter Interaktion in Analytics-Systemen. Auf dieser Grundlage wird die Notwendigkeit hybrider Interaktionsansätze hergeleitet.

2.1 Analytics-Dashboards und Visual Analytics

Analytics-Dashboards sind zentrale Instrumente datengetriebener Entscheidungsfindung. Sie ermöglichen die Exploration, Aggregation und Interpretation von Daten durch visuelle Repräsentationen und interaktive Funktionen.

Im Folgenden wird darunter ein interaktives, visuell orientiertes System verstanden, das analytische Operationen wie Filterung, Aggregation und Vergleich über grafische Benutzeroberflächen zugänglich macht.

Visual Analytics beschreibt Analyse als kooperativen Prozess zwischen Mensch und System. Visuelle Artefakte fungieren als externe kognitive Repräsentationen, indem sie Analysezustände persistent darstellen und so das Arbeitsgedächtnis entlasten.

Empirische Befunde zeigen, dass strukturierte visuelle Interfaces insbesondere bei komplexen Aufgaben mit höherer Genauigkeit und geringerer wahrgenommener Belastung einhergehen (Speier & Morris, 2003). Die Sichtbarkeit verfügbarer Operationen reduziert Ambiguität und unterstützt die Integration mehrerer Informationsquellen.

Die effektive Nutzung klassischer Dashboards setzt jedoch ein Verständnis von Datenmodellen und Interaktionslogik voraus.

2.2 Direkte Manipulation in grafischen Benutzeroberflächen

Direkte Manipulation ist das dominante Interaktionsparadigma grafischer Analytics-Systeme. Charakteristisch sind sichtbare Objekte, kontinuierliches Feedback sowie inkrementelle und reversible Aktionen.

Die visuelle Externalisierung von Analysezuständen reduziert die Notwendigkeit mentaler Rekonstruktion und unterstützt insbesondere bei mehrstufigen Aufgaben die Entscheidungsqualität (Speier & Morris, 2003). Strukturierte Interfaces erhöhen Kontrollierbarkeit, Nachvollziehbarkeit und syntaktische Klarheit.

Mit steigender Aufgabenkomplexität kann jedoch die Anzahl erforderlicher Interaktionen zunehmen, wodurch grafische Interfaces zwar präzise, aber nicht in allen Szenarien effizient sind.

2.3 Natürliche Sprachschnittstellen in Analytics-Systemen

Natürliche Sprachschnittstellen ermöglichen die Formulierung analytischer Intentionen in Alltagssprache und zielen auf eine Reduktion von Interaktionsbarrieren.

Empirische Befunde zeigen Vorteile bei einfachen, klar umrissenen Aufgaben (Wang et al., 2005). Mit zunehmender struktureller Komplexität treten jedoch systematische Leistungsnachteile auf, insbesondere bei mehrstufigen oder vergleichenden Analysen (Guo et al., 2024; Wang et al., 2005).

Dieses Muster weist auf eine Abhängigkeit der Interaktionsleistung von der Aufgabenstruktur hin. Komplexe Anfragen erhöhen aufgrund inhärenter Ambiguität das Risiko von Fehlinterpretationen und führen zu iterativen Reformulierungen.

Zudem fehlt häufig eine persistente Darstellung des Analysezustands, was die Anforderungen an mentale Integration erhöht und die kognitive Belastung steigern kann (Guo et al., 2024).

Gleichzeitig zeigt sich eine Diskrepanz zwischen objektiver Effizienz und subjektiver Wahrnehmung: Sprachbasierte Systeme werden oft als intuitiv empfunden, obwohl sie bei komplexeren Aufgaben weniger effizient sind (Wang et al., 2005).

2.4 Kognitive Belastung und Transparenz

Analytische Aufgaben beanspruchen begrenzte kognitive Ressourcen, wobei die Interaktionsgestaltung die mentale Belastung maßgeblich beeinflusst.

Grafische Interfaces reduzieren kognitive Last durch Persistenz und Externalisierung. Sprachbasierte Systeme senken Einstiegshürden, erhöhen jedoch bei komplexeren Aufgaben die Anforderungen an mentale Rekonstruktion (Guo et al., 2024; Wang et al., 2005).

Transparenz von Systementscheidungen wirkt nicht linear: Sowohl zu geringe als auch zu hohe Transparenz können Leistungsfähigkeit und Fehlererkennung beeinträchtigen. Mittlere, kontextabhängige Transparenzniveaus erweisen sich als besonders effektiv (Poursabzi-Sangdeh et al., 2021).

Die Wirkung von Transparenzmechanismen ist zudem nutzerabhängig und variiert mit Erfahrung und Aufgabenstruktur.

2.5 Hybride Interaktionsansätze in Analytics-Systemen

Hybride Interaktionsansätze kombinieren natürliche Sprache mit grafischen Steuerungs- und Rückmeldemechanismen. Sprache dient der Formulierung analytischer Intentionen, visuelle Komponenten der Präzisierung, Validierung und Anpassung.

Empirische Studien zeigen Leistungsgewinne gegenüber rein sprachbasierten Interfaces insbesondere bei komplexeren oder explorativen Aufgaben (Gao et al., 2015; Guo et al., 2024; Vaithilingam et al., 2024). Mechanismen wie Disambiguierung, visuelle Hervorhebungen oder interaktive Elemente reduzieren Fehlinterpretationen und erhöhen die Genauigkeit.

Die Wirksamkeit ist kontextabhängig und entfaltet sich primär bei erhöhter Aufgabenkomplexität. Interaktion erfolgt häufig sequenziell, wobei Sprache und direkte Manipulation unterschiedliche Rollen übernehmen.

Hybridität ist daher als funktionale Arbeitsteilung zu verstehen: Sprache reduziert Zugangsbarrieren, grafische Elemente sichern Kontrolle und Transparenz. Die Gestaltung erfordert eine abgestimmte Integration, um zusätzliche Komplexität zu vermeiden.

Transparenzmechanismen spielen dabei eine zentrale Rolle. Interaktive, nutzerkontrollierte Erklärungen zeigen stärkere Effekte als statische Offenlegungen (Schaffer et al., 2015), wobei eine kontextabhängige Dosierung erforderlich ist (Poursabzi-Sangdeh et al., 2021).

2.6 Zwischenfazit

Die Analyse zeigt ein komplexitätsabhängiges Spannungsfeld zwischen Zugänglichkeit, Kontrolle, Transparenz und kognitiver Belastung. Grafische Interfaces bieten Präzision und Persistenz, während natürliche Sprache den Zugang erleichtert, jedoch bei steigender Komplexität Leistungsnachteile aufweist.

Hybride Ansätze bieten insbesondere bei komplexeren und explorativen Aufgaben Vorteile, sofern Disambiguierungs- und Transparenzmechanismen angemessen gestaltet sind. Ein Substitutionsverhältnis besteht nicht; vielmehr ergibt sich eine funktionale Differenzierung entlang von Aufgabenstruktur, kognitiven Anforderungen und Nutzerexpertise.

Daraus ergibt sich die zentrale Herausforderung, hybride Analytics-Dashboards so zu gestalten, dass Zugänglichkeit, Kontrolle und Transparenz integriert werden, ohne zusätzliche kognitive Belastung zu erzeugen.

3 Konzeption des hybriden Interaktionskonzepts

Die im theoretischen Hintergrund identifizierten Spannungsfelder zeigen, dass die Leistungsfähigkeit analytischer Interaktion systematisch von der Aufgabenstruktur abhängt. Empirische Studien belegen konsistent Interaktionseffekte zwischen Aufgabenkomplexität und Interface-Typ: Natürliche Sprache ist bei einfachen Aufgaben konkurrenzfähig, während bei komplexeren Aufgaben strukturierte oder hybride Ansätze überlegen sind (Guo et al., 2024; Wang et al., 2005).

Zudem zeigen visuell strukturierte Interfaces Vorteile hinsichtlich Genauigkeit und kognitiver Belastung durch persistente Zustandsrepräsentation (Speier & Morris, 2003). Gleichzeitig verdeutlichen Arbeiten zu Transparenz, dass interaktive Erklärmechanismen Leistung und Kontrolle beeinflussen, jedoch nicht linear wirken (Hernandez-Bocanegra & Ziegler, 2023; Poursabzi-Sangdeh et al., 2021; Schaffer et al., 2015).

Hybride Systeme erzielen in diesem Kontext insbesondere durch Disambiguation und visuelle Rückmeldung signifikante Leistungsgewinne gegenüber rein sprachbasierten Ansätzen (Feng et al., 2023; Gao et al., 2015; Vaithilingam et al., 2024).

Die Evidenz spricht damit gegen eine universell überlegene Interaktionsform. Ziel dieses Kapitels ist die Entwicklung eines konzeptionellen Modells, das hybride Interaktion als regelgeleitete Verteilung analytischer Kontrolle entlang von Aufgabenkomplexität, Transparenzanforderungen und Nutzerexpertise versteht.

3.1 Gestaltungsannahmen und Designziele

Die folgenden Gestaltungsannahmen leiten sich aus den empirischen Befunden ab und definieren Bedingungen, unter denen hybride Interaktion Effizienz, Kontrolle und Stabilität gewährleistet.

A1: Aufgabenkomplexität als primäre Entscheidungsdimension

Die Eignung von Interaktionsformen variiert systematisch mit der strukturellen Komplexität analytischer Aufgaben (Guo et al., 2024; Wang et al., 2005).

Die Wahl der dominanten Interaktionsform wird primär durch Aufgabenkomplexität bestimmt.

A2: Interaktive Disambiguation als Stabilitätsmechanismus

Interaktive Disambiguationsmechanismen reduzieren semantische Fehlinterpretationen und steigern die Genauigkeit (Feng et al., 2023; Gao et al., 2015; Vaithilingam et al., 2024).

Systeminterpretationen müssen explizit sichtbar und inkrementell korrigierbar sein.

A3: Persistente visuelle Repräsentation

Persistente visuelle Zustände reduzieren kognitive Belastung und erhöhen Entscheidungsgenauigkeit (Saktheeswaran et al., 2020; Speier & Morris, 2003).

Analysezustände müssen externalisiert und kontinuierlich verfügbar sein.

A4: Expertiseabhängige Adaptivität

Die Wirkung von Transparenz und Interaktion variiert mit Nutzerexpertise (Hernandez-Bocanegra & Ziegler, 2023; Nafari & Weaver, 2013; Schaffer et al., 2015).

Interaktion und Transparenz müssen kontext- und nutzerabhängig angepasst werden.

Übergeordnete Designziele

- Kontextabhängige Modalitätswahl entlang der Aufgabenkomplexität
- Sichtbare und korrigierbare Systeminterpretationen
- Persistente visuelle Zustandsrepräsentation
- Expertiseadaptive Interaktions- und Transparenzgestaltung

3.2 Aufgabenklassifikation als kontinuierliche Entscheidungsdimension

Analytische Aufgaben sind nicht binär, sondern variieren graduell in ihrer strukturellen Komplexität. Empirische Studien zeigen konsistent, dass die Leistungsfähigkeit von Interaktionsformen entlang dieser Dimension verschoben wird (Guo et al., 2024; Wang et al., 2005).

Strukturelle Dimensionen

Vier zentrale Faktoren bestimmen die Aufgabenkomplexität:

- **Operationsumfang:** Anzahl erforderlicher Verarbeitungsschritte (Speier & Morris, 2003)
- **Integrationsgrad:** Verknüpfung mehrerer Variablen oder Datenquellen (Guo et al., 2024)
- **Ambiguitätsrisiko:** Interpretationsspielraum natürlicher Sprache (Gao et al., 2015)
- **Iterationsbedarf:** Notwendigkeit inkrementeller Verfeinerung (Saktheeswaran et al., 2020)

Mit steigender Ausprägung dieser Faktoren wächst der Bedarf an Kontrolle, Transparenz und Zustandsrepräsentation.

Schwellenwertmodell

Es ergibt sich ein dreistufiges Modell:

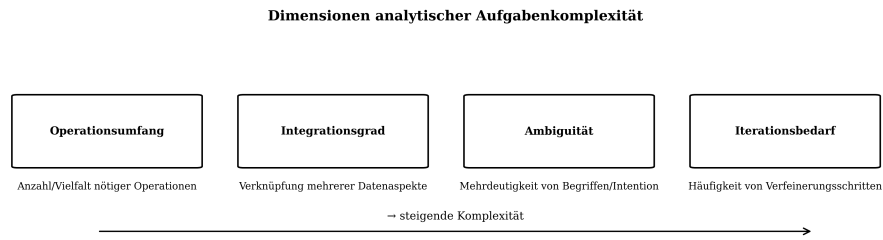


Abbildung 3.1: Strukturelle Dimensionen der Aufgabenkomplexität

1. **Niedrige Komplexität:** Sprache effizient, keine signifikanten Nachteile (Guo et al., 2024; Wang et al., 2005)
2. **Mittlere Komplexität:** Hybride Systeme überlegen durch Disambiguation und Feedback (Feng et al., 2023; Gao et al., 2015)
3. **Hohe Komplexität:** Visuelle bzw. hybride Ansätze deutlich überlegen (Guo et al., 2024; Vaithilingam et al., 2024)

Die Modalität verschiebt sich kontinuierlich mit steigender Komplexität.

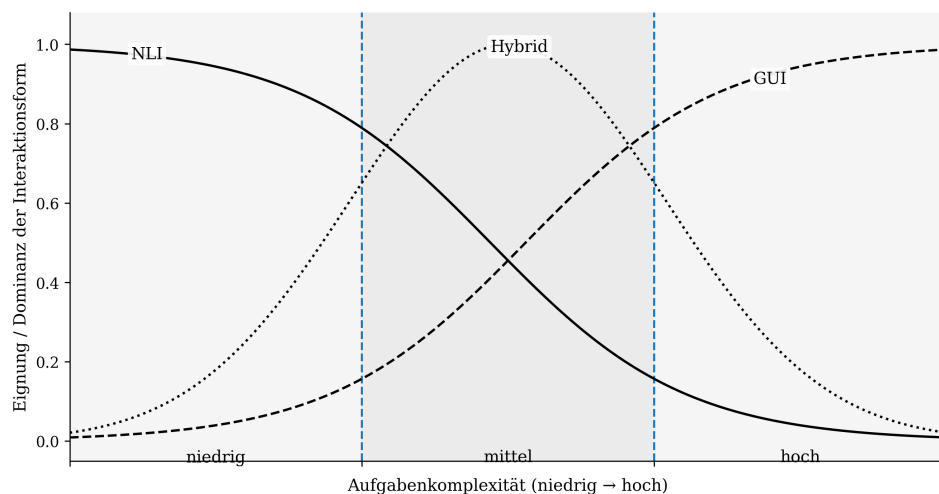


Abbildung 3.2: Schwellenwertmodell der Modalitätswahl entlang der Aufgabenkomplexität

Die in Abbildung 3.2 dargestellten Kurven verdeutlichen die graduelle Verschiebung der relativen Eignung der Interaktionsformen entlang der Aufgabenkomplexität. Während die Leistungsfähigkeit sprachbasierter Interaktion (NLI) mit zunehmender Komplexität abnimmt und grafische Interaktion (GUI) entsprechend zunimmt, zeigt der hybride Ansatz ein Maximum im Bereich mittlerer Komplexität. Dies reflektiert

die komplementäre Wirkung von sprachlicher Spezifikation und visueller Kontrolle: Bei moderater Komplexität ermöglicht die Kombination beider Modalitäten eine Reduktion von Ambiguität bei gleichzeitig reduziertem iterativen Anpassungsaufwand.

Theoretische Erwartungen

- Leistungsfähigkeit sprachbasierter Interfaces sinkt mit steigender Komplexität (Wang et al., 2005)
- Hybride Systeme reduzieren Fehlinterpretationen bei komplexen Aufgaben (Feng et al., 2023; Gao et al., 2015)
- Visuelle Persistenz stabilisiert Entscheidungsgenauigkeit (Guo et al., 2024; Speier & Morris, 2003)

Diese Erwartungen bilden die Grundlage der empirischen Untersuchung.

3.3 Feedback-, Transparenz- und Kontrollmechanismen

Mit steigender Aufgabenkomplexität werden Ambiguität, Fehlinterpretationen und Kontrollverlust zu zentralen Problemen (Gao et al., 2015; Guo et al., 2024). Feedback- und Transparenzmechanismen sind daher integrale Bestandteile hybrider Systeme.

Interaktive Disambiguation

Explizite Disambiguationsmechanismen reduzieren Fehlinterpretationen signifikant (Feng et al., 2023; Gao et al., 2015; Vaithilingam et al., 2024). Systeminterpretationen müssen sichtbar und selektierbar sein.

Transparenz

Transparenz beeinflusst Leistung und Kontrolle, wirkt jedoch nicht linear (Poursabzi-Sangdeh et al., 2021). Effektive Systeme implementieren adaptive, interaktive Transparenz (Hernandez-Bocanegra & Ziegler, 2023).

Persistente Zustandsrepräsentation

Visuelle Persistenz reduziert kognitive Belastung und stabilisiert Analyseprozesse (Saktheeswaran et al., 2020; Speier & Morris, 2003). Hybride Systeme koppeln sprachliche Spezifikation mit visueller Externalisierung.

Operationalisierung

- **Sichtbarkeit:** Systeminterpretationen werden visualisiert
- **Korrektierbarkeit:** Inkrementelle Anpassung ohne Neuformulierung
- **Adaptive Transparenz:** Kontextabhängige Steuerung der Erklärungstiefe

Hybrider Interaktionsprozess

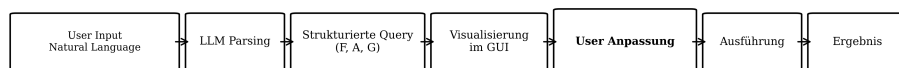


Abbildung 3.3: Ablauf der hybriden Interaktion mit Feedback- und Kontrollmechanismen

3.4 Abgrenzung zu monolithischen Interaktionsansätzen

Monolithische Ansätze priorisieren eine Modalität, während hybride Systeme Kontrolle kontextabhängig verteilen.

Sprachbasierte Systeme

Vorteile bei einfachen Aufgaben, jedoch Leistungsabfall bei steigender Komplexität und fehlende visuelle Persistenz (Gao et al., 2015; Guo et al., 2024; Wang et al., 2005).

Grafische Systeme

Hohe Kontrolle und Genauigkeit, jedoch potenzieller Interaktionsaufwand bei einfachen oder offenen Aufgaben (Saktheeswaran et al., 2020; Speier & Morris, 2003).

Hybride Systeme

Kombinieren Spezifikation und Kontrolle und erzielen insbesondere bei komplexeren Aufgaben Leistungsgewinne (Feng et al., 2023; Gao et al., 2015; Guo et al., 2024; Vaithilingam et al., 2024).

Zusammenfassung

- Sprache: flexibel, aber instabil bei hoher Komplexität
- GUI: stabil, aber potenziell ineffizient bei einfachen Aufgaben
- HYBRID: kontextabhängige Arbeitsteilung mit strukturellen Vorteilen

3.5 Operationalisierung im prototypischen System

Die Gestaltungsannahmen werden in einem webbasierten Prototyp umgesetzt, der drei Interaktionsmodi bereitstellt: GUI, NLI und HYBRID. Alle Modi nutzen denselben Datensatz und dasselbe Abfragemodell.

3.5.1 Interaktionsmodi

GUI: Strukturierte Eingabe über visuelle Elemente. **NLI:** Sprachliche Spezifikation mit automatischer Übersetzung. **HYBRID:** Kombination aus Spracheingabe und visueller Kontrolle mit nachgelagerter Anpassung.

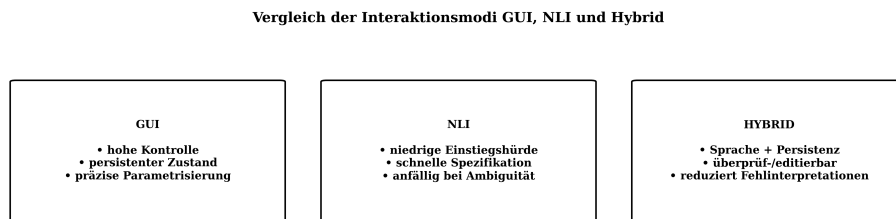


Abbildung 3.4: Übersicht der drei Interaktionsmodi im prototypischen System

3.5.2 Abfragemodell

Analytische Operationen werden einheitlich modelliert als

$$Q = (F, A, G)$$

mit Filterbedingungen F , Aggregation A und Gruppierung G . Dadurch bleiben Aufgaben über Modalitäten hinweg vergleichbar.

3.5.3 Datensatz

Strukturierter Sales-Datensatz mit typischen Dimensionen (Produkt, Kategorie, Region, Umsatz, Kosten).

3.5.4 Bezug zur Aufgabenklassifikation

Die Aufgaben variieren hinsichtlich struktureller Komplexität (Operationsumfang, Integration, Iteration). Die Klassifikation erfolgt nachträglich. Unterschiede in der Leistung sind damit konsistent mit Unterschieden in der Interaktionsform, erlauben jedoch keine kausale Zuschreibung der Komplexitätseffekte.

4 Methodik

Dieses Kapitel beschreibt das methodische Vorgehen, bestehend aus Forschungsdesign, Prototyp, Aufgaben, Messinstrumenten, Studiendurchführung sowie Datenerhebung und Datensatz.

4.1 Forschungsdesign

Es wurde ein kontrolliertes Benutzerexperiment mit drei Interaktionsbedingungen durchgeführt, um deren Einfluss auf Leistung, Arbeitsbelastung und Gebrauchstauglichkeit zu untersuchen. Zusätzlich wird der Einfluss struktureller Aufgabenkomplexität explorativ analysiert.

Das Experiment folgt einem *Within-Subjects*-Design, bei dem alle Teilnehmenden alle Bedingungen durchlaufen. Dadurch werden individuelle Unterschiede kontrolliert.

Die unabhängige Variable ist die Interaktionsform:

- **GUI**: Formularbasierte grafische Interaktion
- **NLI**: Natürlichsprachliche Eingabe mit automatischer Übersetzung
- **HYBRID**: Kombination aus Spracheingabe und anschließender grafischer Bearbeitung

Als abhängige Variablen werden erhoben:

- Bearbeitungszeit
- Korrektheit
- Anzahl der Abfrageversuche
- Arbeitsbelastung (NASA-TLX)
- Gebrauchstauglichkeit (SUS)

Zur Kontrolle von Reihenfolgeeffekten wird ein *Latin Square*-Design verwendet, das sowohl Interaktionsbedingungen als auch Aufgabensets ausbalanciert.

Latin-Square-Design zur Reihenfolgekontrolle

	Block 1	Block 2	Block 3
T1	GUI	NLI	HYBRID
T2	NLI	HYBRID	GUI
T3	HYBRID	GUI	NLI
T4	GUI	NLI	HYBRID
T5	NLI	HYBRID	GUI
T6	HYBRID	GUI	NLI

Abbildung 4.1: Latin-Square-Design zur Ausbalancierung von Bedingungen und Aufgabensets

Die Analyse der Aufgabenkomplexität erfolgt nicht als experimentell manipulierte Variable, sondern als nachträgliche Klassifikation auf Basis der Aufgabenstruktur. Entsprechende Befunde sind daher nicht kausal interpretierbar, sondern als deskriptive Einordnung der beobachteten Unterschiede zu verstehen.

4.2 Prototypische Implementierung

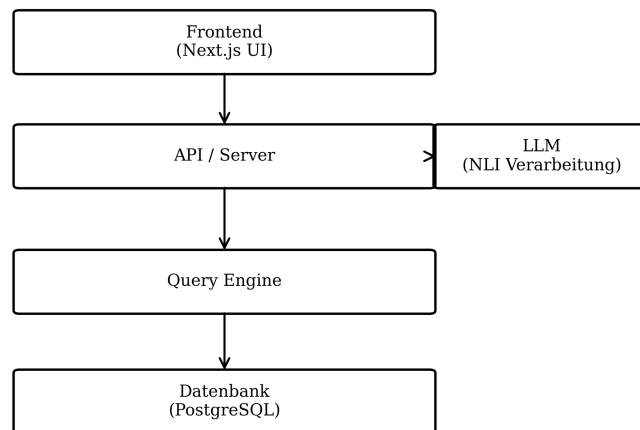
Zur Durchführung wurde ein webbasiertes System entwickelt, das alle Interaktionsformen unter identischen Bedingungen abbildet.

Die Anwendung basiert auf Next.js und TypeScript mit komponentenbasierter UI (Tailwind CSS). Die serverseitige Logik umfasst Abfrageverarbeitung, Sprachinterpretation und Datenspeicherung (PostgreSQL über Supabase).

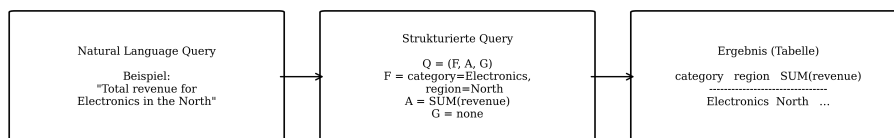
Alle Interaktionsmodi greifen auf dieselbe Abfrageinfrastruktur zu, sodass Unterschiede ausschließlich auf die Interaktion zurückzuführen sind.

4.2.1 Abfragemodell

Analytische Abfragen werden einheitlich als $Q = (F, A, G)$ modelliert (vgl. Abschnitt 3.5.2).

Systemarchitektur des Prototyps**Abbildung 4.2:** Systemarchitektur des Studienprototyps**4.2.2 Abfrageausführung**

Abfragen werden serverseitig validiert und ausgeführt. Filter, Aggregation und Gruppierung werden sequenziell angewendet, das Ergebnis wird tabellarisch zurückgegeben.

Übersetzung natürlicher Sprache in strukturierte Abfragen**Abbildung 4.3:** Übersetzungsprozess von natürlicher Sprache zur strukturierten Abfrage**4.2.3 Interaktionsmodi**

GUI: Strukturierte Eingabe über visuelle Elemente.

NLI: Sprachliche Eingabe mit automatischer Übersetzung.

HYBRID: Kombination aus Sprachinterpretation und visueller Nachbearbeitung.

Übung

GUI-Modus

1/3

Übungsaufgabe

Was ist der Gesamtumsatz (revenue) für Tablets?

Abfrage erstellen

FILTER + Filter hinzufügen

Produkt

ist gleich

Tablet

×

BERECHNUNG

Summe

 von

Umsatz (revenue)

GRUPPIEREN NACH

Keine Gruppierung

Abfrage ausführen

Ergebnis (1 Zeile)

WERT
36.532,89

Antwort bestätigen

Abbildung 4.4: Studentool im GUI-Modus

Übung

Spracheingabe-Modus

2/3

Übungsaufgabe

Wie hoch sind die durchschnittlichen Kosten (cost) pro Kategorie?

Spracheingabe

Durchschn. Kosten pro Kategorie

Absenden

Bitte formulieren Sie die Abfrage in eigenen Worten.

Ergebnis (2 Zeilen)

KATEGORIE	WERT
Accessories	34,4
Electronics	375,03

Antwort bestätigen

Abbildung 4.5: Studentool im NLI-Modus

Übung

Hybrid-Modus

3/3

Übungsaufgabe

Was ist die Gesamtmenge (quantity) in der Region East pro Produkt?

Spracheingabe

Übersetzen

Bitte formulieren Sie die Abfrage in eigenen Worten.

Erkannte Abfrage (editierbar)

FILTER

+ Filter hinzufügen

Region

ist gleich

East

×

BERECHNUNG

Summe

von

Menge (quantity)

GRUPPIEREN NACH

Produkt

Abfrage ausführen

Abbildung 4.6: Studentool im HYBRID-Modus

23

4.3 Studienaufgaben

Es wurden drei Aufgabensets (A–C) mit jeweils vier analytischen Aufgaben definiert. Jede Versuchsperson bearbeitet alle Sets, jeweils unter einer anderen Interaktionsbedingung.

Die Aufgaben simulieren typische analytische Fragestellungen (Aggregation, Filterung, Gruppierung). Unterschiede in der Komplexität ergeben sich aus der Struktur der Abfragen und werden nachträglich klassifiziert.

Eine Trainingsphase mit drei Aufgaben dient der Einführung und wird nicht ausgewertet. Die vollständigen Aufgabenstellungen aller Sets einschließlich der Trainingsaufgaben sind in Anhang A.1 dokumentiert.

4.3.1 Klassifikation der Aufgabenkomplexität

Die Aufgaben werden nachträglich anhand ihrer strukturellen Eigenschaften in drei Komplexitätsstufen eingeteilt. Die Klassifikation orientiert sich an der Anzahl der erforderlichen Abfrageoperationen sowie am Grad der notwendigen Interpretation und Iteration.

Folgende Kriterien werden herangezogen:

- **Niedrige Komplexität:** Aufgaben mit einer einzelnen Aggregation oder Filterung ohne weitere Verknüpfungen. Die erforderliche Abfrage ist eindeutig spezifiziert und erfordert keine iterative Verfeinerung.
- **Mittlere Komplexität:** Aufgaben mit mehreren kombinierten Operationen (z. B. Filterung und Gruppierung) oder einfachen Vergleichen. Eine begrenzte Interpretation der Fragestellung ist erforderlich.
- **Hohe Komplexität:** Aufgaben mit mehrstufiger Abfragelogik, mehreren abhängigen Operationen oder erhöhter Ambiguität. Die Lösung erfordert in der Regel iterative Anpassungen und Interpretation der Ergebnisse.

Die Zuordnung erfolgt auf Basis der zugrunde liegenden Query-Struktur $Q = (F, A, G)$ sowie der Anzahl kombinierter Operationen. Die Klassifikation basiert auf

strukturellen Eigenschaften der zugrunde liegenden Abfragen und wurde nicht unabhängig validiert. Interpretative Spielräume bei der Zuordnung einzelner Aufgaben zu Komplexitätsstufen können daher nicht vollständig ausgeschlossen werden.

4.4 Messinstrumente

4.4.1 Objektive Maße

- Bearbeitungszeit
- Korrektheit (Vergleich mit Referenzlösung)
- Anzahl der Abfrageversuche

Zusätzlich erfolgt eine nachträgliche Klassifikation der Aufgaben nach struktureller Komplexität.

4.4.2 Subjektive Maße

- NASA-TLX zur Arbeitsbelastung (Hart & Staveland, 1988)
- SUS zur Gebrauchstauglichkeit (Brooke, 1996)

Beide werden nach jedem Aufgabenblock erhoben. Die verwendeten Items der SUS sowie die Dimensionen des NASA-TLX sind im Anhang dokumentiert (siehe Abschnitt A.2.1 und A.2.2).

4.5 Studiendurchführung

Die Studie umfasst folgende Phasen:

1. Einführung in System und Ablauf
2. Trainingsphase mit drei Aufgaben (alle Interaktionsformen)
3. Drei Aufgabenblöcke mit je vier Aufgaben (eine Bedingung pro Block)
4. Nach jedem Block: NASA-TLX und SUS
5. Abschluss

Die Reihenfolge der Bedingungen wird durch das Latin-Square-Design bestimmt.

Ablauf des Experiments

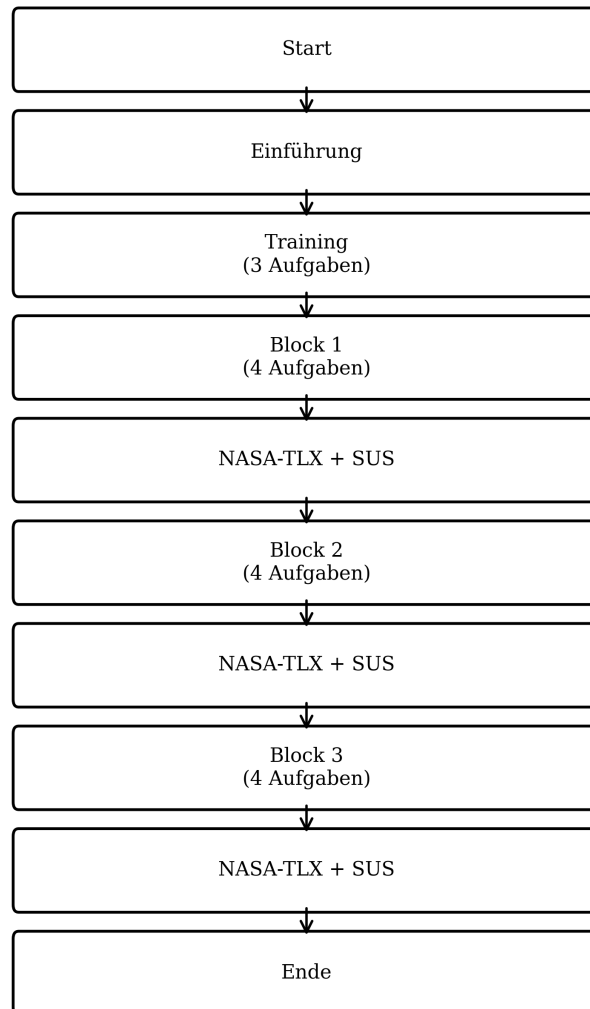


Abbildung 4.7: Ablauf der Studiendurchführung

4.6 Datenerhebung und Logging

Alle Interaktionen werden automatisch protokolliert, einschließlich Aufgabenstart, Abfrageausführung und Änderungen.

Erfasst werden Zeitstempel, Abfragen, Ergebnisse sowie Interaktionsereignisse. Im NLI- und HYBRID-Modus werden zusätzlich Texteingaben und generierte Abfragen gespeichert.

Die Daten werden anonymisiert über Session-IDs gespeichert und bilden die Grundlage für die Auswertung von Bearbeitungszeit, Fehlerraten und Interaktionsverläufen.

4.7 Datensatz

Es wird ein synthetischer Verkaufsdatensatz verwendet, der typische Attribute eines Sales-Szenarios umfasst (Produkt, Kategorie, Region, Menge, Umsatz, Kosten).

Der Datensatz wird deterministisch generiert, sodass alle Teilnehmenden identische Daten erhalten. Dies gewährleistet Vergleichbarkeit und vermeidet den Einsatz realer personenbezogener Daten.

Die Struktur ermöglicht typische analytische Operationen wie Filterung, Aggregation und Gruppierung.

4.8 Datenaufbereitung und statistische Analyse

Die Auswertung erfolgt in zwei Schritten: einer deskriptiven Analyse zur Beschreibung der beobachteten Unterschiede sowie einer ergänzenden inferenzstatistischen Analyse.

Zur Vorbereitung der Daten werden die aufgezeichneten Interaktionen auf Ebene der Versuchsperson und Interaktionsbedingung aggregiert. Für jede Kombination aus Person und Modus werden Bearbeitungszeit, Korrektheit, Anzahl der Abfrageversuche sowie die subjektiven Bewertungen (NASA-TLX und SUS) zusammengeführt. Die Korrektheit wird anhand eines Vergleichs der erzeugten Abfrage mit einer Referenzlösung bestimmt. Dabei wird primär die strukturelle Übereinstimmung bewertet.

Semantisch äquivalente, aber strukturell abweichende Abfragen können in diesem Vorgehen als inkorrekt klassifiziert werden.

Die deskriptive Analyse umfasst Mittelwerte, Median und Standardabweichung für die kontinuierlichen Variablen sowie prozentuale Anteile für die Accuracy-Metriken.

Zur ergänzenden Bewertung der Unterschiede zwischen den Interaktionsformen werden paarweise Vergleiche mittels Wilcoxon Signed-Rank Tests durchgeführt. Aufgrund der kleinen Stichprobe und des Within-Subjects-Designs wird ein nicht-parametrischer Test gewählt, der keine Normalverteilungsannahmen erfordert. Die Analyse erfolgt innerhalb von Versuchspersonen, wobei nur diejenigen Fälle berücksichtigt werden, für die Werte in beiden verglichenen Bedingungen vorliegen. Durch die geringen Stichprobengröße ist die Teststärke der inferenzstatistischen Verfahren teils eingeschränkt. Bestehende Effekte könnten daher unter Umständen nicht als statistisch signifikant erkannt werden.

Die Anzahl der einbezogenen Beobachtungen variiert je Vergleich ($n = 6$ bis $n = 13$), da nicht für alle Teilnehmenden vollständige Daten in allen Metriken vorliegen. Die Ergebnisse der inferenzstatistischen Analyse werden daher als explorativ interpretiert und im Kontext der deskriptiven Befunde eingeordnet.

Aufgrund der kleinen Stichprobe und der rechtschiefen Verteilung der Bearbeitungszeiten – insbesondere im NLI-Modus – werden Ausreißer nicht aus der Analyse ausgeschlossen, sondern inhaltlich interpretiert. Die Bearbeitungszeiten werden sowohl als Mittelwert als auch als Median berichtet, um die Sensitivität gegenüber Extremwerten transparent zu machen. Die hohe Varianz im NLI-Modus wird als substanzieller Befund behandelt, da sie auf systematische Interaktionsprobleme in Einzelfällen hinweist und nicht auf Messfehler oder aufgabenfremdes Verhalten zurückzuführen ist.

5 Ergebnisse

5.1 Stichprobe

An der Studie nahmen insgesamt 16 Personen teil. Vier Datensätze wurden aufgrund unvollständiger Bearbeitung ausgeschlossen, sodass 12 valide Versuchspersonen in die Auswertung einbezogen wurden. Die geringe Stichprobengröße limitiert die Aussagekraft der Ergebnisse erheblich; insbesondere sind inferenzstatistische Befunde nur eingeschränkt belastbar und primär im Kontext der deskriptiven Analyse zu interpretieren.

Jede Versuchsperson bearbeitete alle drei Interaktionsbedingungen (GUI, NLI, HYBRID) mit jeweils vier Aufgaben, sodass insgesamt 144 abgeschlossene Aufgaben in die Analyse einfließen.

5.2 Bearbeitungszeit

Die durchschnittliche Bearbeitungszeit unterscheidet sich deutlich zwischen den Interaktionsformen.

Tabelle 5.1: Bearbeitungszeit pro Interaktionsmodus

Modus	Mittelwert (s)	Median (s)	SD (s)
GUI	31,70	26,23	23,41
NLI	96,04	29,76	316,88
HYBRID	50,68	35,35	33,55

Der GUI-Modus weist die niedrigste mittlere Bearbeitungszeit auf, gefolgt vom HYBRID-Modus. Der NLI-Modus zeigt zwar den höchsten Mittelwert (96,04 s), jedoch einen deutlich niedrigeren Median (29,76 s), was auf eine stark rechtsschiefe Verteilung durch einzelne Extremwerte hindeutet. Bei näherer Betrachtung der Interaktionsprotokolle zeigt sich, dass zwei Aufgabenbearbeitungen im NLI-Modus

Bearbeitungszeiten von über 1,000 s aufweisen. Diese sind auf wiederholte Fehlinterpretationen des Sprachmodells und die daraus resultierenden iterativen Reformulierungen zurückzuführen – es handelt sich also um substantielle Interaktionsprobleme, nicht um Messfehler oder aufgabenfremdes Verhalten. Die Werte verbleiben daher in der Analyse. Um die Sensitivität gegenüber diesen Extremwerten zu verdeutlichen, werden Bearbeitungszeiten durchgehend als Mittelwert und Median berichtet. Die Standardabweichung von 316,88 s im NLI-Modus unterstreicht die hohe Variabilität dieses Modus.

5.3 Genauigkeit und Fehlerraten

Tabelle 5.2: Accuracy und First-Try Accuracy

Modus	Accuracy (%)	First-Try Accuracy (%)
GUI	52,08	47,92
NLI	72,92	70,83
HYBRID	62,50	60,42

Der NLI-Modus erzielt in beiden Metriken die höchsten Werte, gefolgt vom HYBRID-Modus und dem GUI-Modus. Die Differenz zwischen Accuracy und First-Try Accuracy ist in allen Modi gering (GUI: 4,16 Prozentpunkte, NLI: 2,09, HYBRID: 2,08), was darauf hindeutet, dass Mehrfachversuche nur in wenigen Fällen zur Korrektur der Lösung geführt haben. Die niedrige Accuracy des GUI-Modus bei gleichzeitig schneller Bearbeitungszeit deutet darauf hin, dass Nutzer häufig ohne ausreichende Überprüfung ihrer Eingaben eine Abfrage abschiekten.

5.4 Interaktionsverhalten

Die Anzahl der Abfrageausführungen pro Aufgabe ist in allen Modi vergleichbar niedrig. Im HYBRID-Modus wurden im Schnitt 0,79 manuelle Anpassungen an der automatisch generierten Abfrage vorgenommen, was bedeutet, dass Nutzer bei rund drei Viertel aller Aufgaben aktiv in die Abfragestruktur eingegriffen haben. NLP-

Tabelle 5.3: Interaktionsmetriken pro Modus

Modus	Submissions	Queries	Hybrid Edits	NLP Errors
GUI	1,23	1,23	0,00	0,00
NLI	1,08	1,08	0,00	0,02
HYBRID	1,13	1,13	0,79	0,08

Fehler – Fälle, in denen das Sprachmodell keine gültige Abfrage erzeugen konnte – treten ausschließlich in sprachbasierten Modi auf und sind insgesamt selten.

5.5 Explorative Analyse: Einfluss der Aufgabenkomplexität

Zur ergänzenden Untersuchung struktureller Aufgabenmerkmale wurden die zwölf Studienaufgaben nachträglich in zwei Komplexitätskategorien eingeteilt. Die Einteilung in zwei Kategorien stellt eine vereinfachte Operationalisierung des in Kapitel 3 entwickelten dreistufigen Modells dar. Aufgaben mit genau einer Filterbedingung und ohne Gruppierung (A1, A3, B1, B3, C1, C3) wurden als *einfach* klassifiziert; Aufgaben mit Gruppierung und/oder zwei oder mehr Filterbedingungen (A2, A4, B2, B4, C2, C4) als *komplex*. Diese Einteilung orientiert sich an den in Abschnitt 4.4.1 definierten Aufgabenmerkmalen. Pro Zelle liegen $n = 24$ Aufgaben vor (12 Versuchspersonen $\times$ 2 Aufgaben je Komplexitätsstufe pro Bedingung).

Tabelle 5.4: Bearbeitungszeit nach Interaktionsmodus und Aufgabenkomplexität (Mittelwert / Median in s)

Komplexität	Modus	Mittelwert (s)	Median (s)
Einfach	GUI	37,7	30,9
	HYBRID	53,7	35,7
	NLI	60,3	30,2
Komplex	GUI	25,7	24,2
	HYBRID	47,7	34,6
	NLI	131,8	27,1

Bei einfachen Aufgaben erzielt der NLI-Modus die höchste Accuracy (79,2 %), während GUI (45,8 %) und HYBRID (50,0 %) deutlich darunter liegen. Die Medianwerte der Bearbeitungszeit sind in diesem Bereich über alle Modi vergleichbar (GUI: 30,9 s, HYBRID: 35,7 s, NLI: 30,2 s); der höhere Mittelwert im NLI-Modus ist auf vereinzelte Ausreißer zurückzuführen.

Bei komplexen Aufgaben kehrt sich das Accuracymuster teilweise um: Der HYBRID-Modus erzielt mit 75,0 % die höchste Genauigkeit, gefolgt von NLI (66,7 %) und GUI (58,3 %). Besonders auffällig ist der Anstieg der mittleren Bearbeitungszeit im NLI-Modus auf 131,8 s bei einem Median von lediglich 27,1 s (SD: 473,7 s).

Tabelle 5.5: Accuracy nach Interaktionsmodus und Aufgabenkomplexität ($n = 24$ je Zelle)

Komplexität	Modus	Accuracy (%)
Einfach	GUI	45,8
	HYBRID	50,0
	NLI	79,2
Komplex	GUI	58,3
	HYBRID	75,0
	NLI	66,7

Dieses Muster weist auf einzelne Extremfälle hin, in denen Fehlinterpretationen zu ausgedehnten Iterationssequenzen geführt haben.

Die explorative Analyse deutet darauf hin, dass NLI seine Stärken insbesondere bei einfachen Aufgaben entfaltet, während hybride Interaktion bei komplexeren Aufgaben eine stabilisierende Wirkung zeigt. Diese Befunde sind aufgrund des explorativen Charakters der Analyse und der begrenzten Stichprobengröße nicht inferenzstatistisch abgesichert.

5.6 Kognitive Belastung

Tabelle 5.6: NASA-TLX pro Interaktionsmodus

Modus	TLX (1–21)
GUI	6,43
NLI	6,39
HYBRID	5,83

Der HYBRID-Modus weist die geringste wahrgenommene Arbeitsbelastung auf. Die Unterschiede zwischen den Modi sind insgesamt gering; alle drei Werte liegen im

unteren Bereich der Skala, was auf eine insgesamt niedrige subjektive Belastung hindeutet.

5.7 Gebrauchstauglichkeit

Tabelle 5.7: System Usability Scale (SUS)

Modus	SUS (0–100)
GUI	66,67
NLI	71,46
HYBRID	66,67

Der NLI-Modus erzielt die höchste Bewertung der Gebrauchstauglichkeit (71,46), die nach der Klassifikation von Bangor et al. (2008) im Bereich „Good“ einzuordnen ist. GUI und HYBRID erreichen identische Werte (66,67), die an der Grenze zwischen „OK“ und „Good“ liegen. Der Gleichstand zwischen GUI- und HYBRID-Modus deutet darauf hin, dass die zusätzlichen Interaktionsmöglichkeiten des hybriden Ansatzes nicht unmittelbar zu einer höheren wahrgenommenen Gebrauchstauglichkeit führen. Eine mögliche Erklärung liegt darin, dass die zusätzliche Komplexität durch die Kombination beider Modalitäten zwar funktionale Vorteile bietet, jedoch gleichzeitig einen erhöhten Abstimmungs- und Korrekturaufwand zwischen sprachlicher Eingabe und visueller Kontrolle impliziert, der sich in der subjektiven Bewertung neutralisierend auswirkt. Zudem könnte der hybride Ansatz in der kurzen Studiendauer nicht vollständig erlernt worden sein, sodass potenzielle Usability-Vorteile nicht vollständig zum Tragen kommen. Bemerkenswert ist, dass der NLI-Modus trotz der höchsten objektiven Bearbeitungszeit und instabiler Interaktionsverläufe sowie hoher Varianz in den Bearbeitungszeiten die beste Usability-Bewertung erhält.

5.8 Inferenzstatistische Analyse

Zur ergänzenden Bewertung der Unterschiede zwischen den Interaktionsmodi wurden paarweise Vergleiche mittels Wilcoxon Signed-Rank Tests durchgeführt. Aufgrund der

kleinen Stichprobe und teilweise fehlender Werte variiert die Anzahl der einbezogenen Versuchspersonen je Vergleich ($n = 6$ bis $n = 13$).

Für die Bearbeitungszeit zeigt sich ein signifikanter Unterschied zwischen GUI und HYBRID ($p = 0,0206$), wobei der GUI-Modus eine signifikant geringere Bearbeitungszeit aufweist. Für die Vergleiche zwischen GUI und NLI ($p = 0,8884$) sowie zwischen NLI und HYBRID ($p = 0,5277$) konnten keine signifikanten Unterschiede festgestellt werden.

Für die Accuracy ergeben sich keine statistisch signifikanten Unterschiede zwischen den Modi (GUI vs. NLI: $p = 0,0894$; GUI vs. HYBRID: $p = 0,3173$; NLI vs. HYBRID: $p = 0,1800$). Dennoch zeigt sich deskriptiv ein konsistenter Trend zugunsten des NLI-Modus, insbesondere im Vergleich zum GUI-Modus.

Auch für die kognitive Belastung (TLX) konnten keine signifikanten Unterschiede festgestellt werden (alle $p > 0,36$). Die Ergebnisse deuten darauf hin, dass sich die wahrgenommene Arbeitsbelastung zwischen den Interaktionsformen nur geringfügig unterscheidet.

Für die Gebrauchstauglichkeit (SUS) zeigen sich ebenfalls keine signifikanten Unterschiede zwischen den Modi (alle $p > 0,42$), obwohl der NLI-Modus in der deskriptiven Analyse die höchsten Werte aufweist.

Insgesamt zeigen die inferenzstatistischen Ergebnisse nur begrenzte Evidenz für signifikante Unterschiede zwischen den Interaktionsformen. Die deskriptiven Trends – insbesondere die höhere Accuracy des NLI-Modus, die geringere Bearbeitungszeit des GUI-Modus sowie die geringere kognitive Belastung im HYBRID-Modus – bleiben jedoch konsistent und liefern wichtige Hinweise auf bestehende Trade-offs. Die Ergebnisse sind vor dem Hintergrund der kleinen Stichprobe und variierender Fallzahlen als explorativ zu interpretieren.

5.9 Zusammenfassung der Ergebnisse

Die Ergebnisse zeigen systematische Unterschiede zwischen den Interaktionsformen, die keine universell überlegene Modalität erkennen lassen. Der NLI-Modus erzielt die höchste Gesamtgenauigkeit, insbesondere bei einfachen Aufgaben, ist jedoch mit

hoher Zeitvarianz verbunden. Der GUI-Modus ermöglicht die schnellste Bearbeitung im Durchschnitt, weist jedoch die geringste Accuracy auf. Der HYBRID-Modus positioniert sich in den meisten objektiven Metriken zwischen beiden Ansätzen, zeigt die niedrigste wahrgenommene Arbeitsbelastung und erzielt bei komplexen Aufgaben die höchste Genauigkeit. Die explorative Komplexitätsanalyse deutet auf eine aufgabenabhängige Verschiebung der relativen Stärken der Interaktionsformen hin.

6 Diskussion

6.1 Einordnung der Ergebnisse im Kontext der Forschungsfragen

Ziel der Arbeit war die Untersuchung, wie sich unterschiedliche Interaktionsformen in Analytics-Dashboards auf Effizienz, Genauigkeit, kognitive Belastung und Nutzerwahrnehmung auswirken.

Die Ergebnisse zeigen, dass Interaktionsformen keine universell überlegenen Alternativen darstellen, sondern systematisch unterschiedliche Leistungsprofile aufweisen. Es treten konsistente Trade-offs zwischen Effizienz, Genauigkeit und wahrgenommener Belastung auf, die sich nicht unabhängig voneinander optimieren lassen. Der GUI-Modus ermöglicht die schnellste durchschnittliche Bearbeitung, weist jedoch die geringste Genauigkeit auf. Der NLI-Modus erzielt die höchste Gesamtgenauigkeit, ist jedoch durch eine hohe Varianz der Bearbeitungszeit gekennzeichnet. Der HYBRID-Modus positioniert sich zwischen beiden Ansätzen und zeigt insbesondere bei komplexeren Aufgaben sowie hinsichtlich der wahrgenommenen Arbeitsbelastung spezifische Vorteile.

Die Ergebnisse liefern Hinweise auf strukturelle Zusammenhänge zwischen Interaktionsform und Aufgabenmerkmalen, erlauben jedoch aufgrund der begrenzten Stichprobengröße und des explorativen Studiendesigns keine belastbaren inferenzstatistischen Schlussfolgerungen.

Ebenso lassen sich die Ergebnisse dahingehend verdichten, dass Interaktionsformen funktional unterschiedliche Rollen im Analyseprozess einnehmen. Ihre Eignung ist nicht absolut, sondern abhängig von Aufgabenkomplexität und Zielkriterium. Die zentrale Forschungsfrage lässt sich damit präzisieren: Hybride Interaktion erzeugt keinen generellen Performance-Vorteil, ermöglicht jedoch eine kontextabhängige Kombination komplementärer Stärken und stellt damit eine strukturierte Form der Verteilung analytischer Kontrolle dar.

Die inferenzstatistischen Ergebnisse liefern nur begrenzte Evidenz für signifikante Unterschiede zwischen den Interaktionsformen. Lediglich der Unterschied in der Bearbeitungszeit zwischen GUI und HYBRID ist statistisch signifikant. Die ausbleibende Signifikanz in den übrigen Vergleichen ist im Kontext der kleinen Stichprobe und der hohen Varianz – insbesondere im NLI-Modus – zu interpretieren und weist auf eine eingeschränkte Teststärke hin. Die Studie ist daher primär explorativ zu verstehen und zielt auf die Identifikation struktureller Muster, nicht auf deren abschließende inferenzstatistische Absicherung.

Die hohe Streuung der Bearbeitungszeiten im NLI-Modus stellt einen zentralen Befund dar. Sie zeigt, dass natürlichsprachliche Interaktion nicht allein durch durchschnittliche Performance charakterisiert werden kann, sondern maßgeblich durch ihre Instabilität in Einzelfällen geprägt ist. NLI-basierte Systeme weisen damit ein hochvariables Interaktionsverhalten auf, in dem effiziente und korrekte Lösungen häufig auftreten, jedoch durch einzelne Fälle mit erheblichen Iterationsaufwänden ergänzt werden. Für den praktischen Einsatz bedeutet dies, dass neben Effizienz und Genauigkeit insbesondere die Vorhersagbarkeit von Interaktionsdauer als eigenständiges Qualitätskriterium berücksichtigt werden muss.

6.2 Einfluss der Interaktionsform auf Effizienz und Genauigkeit

Die Ergebnisse zeigen eine klare Entkopplung von Effizienz und Genauigkeit. Der GUI-Modus ermöglicht zwar die schnellste Bearbeitung, erzielt jedoch mit 52,08 % die niedrigste Accuracy. Dieser Befund ist auf den ersten Blick kontraintuitiv, da grafische Interfaces durch ihre explizite Strukturierung eigentlich Fehler bei der Abfragekonstruktion reduzieren sollten.

Eine plausible Erklärung liegt in der Art der Bedienungsfehler: Im GUI-Modus müssen Nutzer mehrere Parameter – Filterbedingungen, Aggregationsfunktion, Zielfeld und optionale Gruppierung – manuell und unabhängig voneinander korrekt auswählen. Bereits die Verwechslung eines einzelnen Dropdown-Werts, etwa die Wahl von `avg` statt `sum` oder die fehlende Gruppierungsdimension, führt zu einem falschen Ergebnis. Die hohe Bearbeitungsgeschwindigkeit im GUI-Modus deutet darauf hin,

dass Nutzer diese Parameter häufig zügig und ohne ausreichende Überprüfung setzen. Die geringe Differenz zwischen Accuracy und First-Try Accuracy (52,08 % vs. 47,92 %) bestätigt, dass Fehler selten durch Mehrfachversuche korrigiert wurden.

Dieses Verhalten lässt sich auch theoretisch einordnen: Die fragmentierte Parametrisierung im GUI-Modus erhöht die Anforderungen an die Integration mehrerer Entscheidungsschritte und begünstigt sogenannte Slips in der Interaktion, bei denen einzelne korrekt verstandene Teilentscheidungen fehlerhaft umgesetzt werden. Diese Interpretation stellt eine plausible Erklärung dar, kann jedoch nicht eindeutig von alternativen Ursachen – etwa suboptimalem Interface-Design oder aufgabenspezifischen Verständnisproblemen – abgegrenzt werden. Im Gegensatz dazu reduziert die holistische Spezifikation im NLI-Modus die Anzahl isolierter Entscheidungsstellen, wodurch bestimmte Fehlerarten vermieden werden, jedoch auf Kosten erhöhter Ambiguität.

Eine alternative Erklärung für die niedrige GUI-Accuracy liegt in suboptimalem UI-Design. Diese Erklärung kann nicht vollständig ausgeschlossen werden, allerdings spricht das Interaktionsverhalten eher gegen diese Interpretation: Mit durchschnittlich 1,21 Submissions pro Aufgabe und einer Differenz von lediglich 3,85 Prozentpunkten zwischen Accuracy und First-Try Accuracy zeigen Nutzer kaum Korrekturverhalten. Bei grundlegenden Verständnisproblemen durch schlechtes Design wären mehr Iterationsversuche zu erwarten. Das beobachtete Muster ist konsistenter mit Slips bei der Parametrisierung, bei denen einzelne Parameter fehlerhaft gesetzt werden, ohne dass dies dem Nutzer auffällt.

Der NLI-Modus profitiert demgegenüber davon, dass das Sprachmodell die analytische Intention als Ganzes interpretiert und in eine strukturierte Abfrage überführt. Formulierungsfehler des Nutzers werden dabei vom Modell kompensiert, sofern die Intention eindeutig ist. Die hohe Varianz der Bearbeitungszeiten im NLI-Modus – mit einer Standardabweichung von 316,88 s im Gesamtdatensatz und 473,7 s bei komplexen Aufgaben – deutet jedoch auf instabile Interaktionsverläufe hin: In Einzelfällen führen Fehlinterpretationen zu ausgedehnten Iterationssequenzen, in denen Nutzer ihre Formulierung wiederholt anpassen mussten.

Dieser Befund erscheint auf den ersten Blick paradox: Der instabilste Modus erzielt gleichzeitig die höchste Genauigkeit. Die Erklärung liegt in der strukturellen Verschie-

denheit der Fehlerarten: Während GUI-Fehler durch isolierte Parametrisierungs-Slips entstehen, die dem Nutzer nicht auffallen, scheitert NLI entweder vollständig an der Interpretation – was zu ausgedehnten Iterationssequenzen führt – oder gelingt holistisch korrekt. Die Verteilung deutet auf eine mögliche Heterogenität der Interaktionsverläufe hin, bei der effiziente Lösungen und einzelne stark verlängerte Bearbeitungsprozesse nebeneinander auftreten.

6.3 Einfluss der Aufgabenkomplexität

Die explorative Analyse struktureller Aufgabenmerkmale differenziert die Hauptbefunde und zeigt, dass die Leistungsfähigkeit der Interaktionsformen systematisch von der Aufgabenkomplexität abhängt.

Bei einfachen Aufgaben – mit einer Filterbedingung und ohne Gruppierung – erzielt der NLI-Modus die höchste Accuracy (79,2 %), während GUI (45,8 %) und HYBRID (50,0 %) deutlich darunter liegen. Einfache analytische Intentionen lassen sich in natürlicher Sprache direkt und eindeutig formulieren, sodass die Übersetzung in strukturierte Abfragen zuverlässig erfolgt und Ambiguität kaum eine Rolle spielt. Gleichzeitig sind die Medianwerte der Bearbeitungszeit über alle Modi vergleichbar (GUI: 30,9 s, HYBRID: 35,7 s, NLI: 30,2 s), was zeigt, dass NLI in diesem Komplexitätsbereich keinen Effizienznachteil aufweist.

Mit steigender Aufgabenkomplexität verschiebt sich dieses Muster. Bei komplexen Aufgaben – mit Gruppierung und/oder mehreren Filterbedingungen – erzielt der HYBRID-Modus mit 75,0 % die höchste Accuracy, gefolgt von NLI (66,7 %) und GUI (58,3 %). Die Kombination aus sprachlicher Spezifikation und visueller Kontrolle ermöglicht es, initiale Systeminterpretationen gezielt zu überprüfen und inkrementell anzupassen. Dadurch werden Fehlinterpretationen reduziert, ohne dass die Anfrage vollständig neu formuliert werden muss.

Die Instabilität des NLI-Modus nimmt mit steigender Komplexität deutlich zu. Die Standardabweichung der Bearbeitungszeit von 473,7 s bei einem Median von 27,1 s weist auf stark divergierende Interaktionsverläufe hin. Während ein Großteil der Anfragen effizient gelöst wird, führen Fehlinterpretationen in einzelnen Fällen zu

ausgedehnten Iterationssequenzen. Dieses Muster entspricht der in Abschnitt 2.3 beschriebenen Zunahme von Ambiguität und Interpretationsspielraum bei komplexeren Aufgaben.

Die Ergebnisse lassen sich als empirische Bestätigung des in Kapitel 3 entwickelten Modells interpretieren. Aufgabenkomplexität fungiert als primäre Entscheidungsdimension, entlang derer sich die geeignete Interaktionsform verschiebt: NLI ist bei niedriger Komplexität effizient und präzise, während mit zunehmender Komplexität der Bedarf an Kontrolle, Transparenz und Zustandsrepräsentation steigt und hybride Ansätze strukturelle Vorteile entfalten. Interaktionsformen sind damit nicht als feste Alternativen zu verstehen, sondern als Funktionen von Aufgabenkomplexität und Kontrollbedarf.

Aufgrund des explorativen Charakters der Analyse und der begrenzten Stichprobengröße sind diese Befunde nicht inferenzstatistisch abgesichert. Sie liefern jedoch konsistente Hinweise auf ein strukturelles Wirkungsprinzip, das über die konkreten Aufgaben hinaus generalisierbar ist.

6.4 Rolle des hybriden Interaktionskonzepts

Der HYBRID-Modus zeigt konsistent mittlere Werte in nahezu allen objektiven Metriken. Gleichzeitig weist er die geringste wahrgenommene Arbeitsbelastung auf und erzielt bei komplexen Aufgaben die höchste Accuracy.

Dieses Ergebnis unterstützt die Annahme, dass hybride Interaktion eine stabilisierende Wirkung auf den Analyseprozess haben kann. Die Kombination aus sprachlicher Spezifikation und visueller Kontrolle ermöglicht es, Systeminterpretationen zu überprüfen und bei Bedarf anzupassen, ohne die Abfrage vollständig neu formulieren zu müssen. Die im Konzept postulierte Funktion hybrider Systeme als Kontrollarchitektur wird durch die beobachteten Hybrid Edits empirisch gestützt: Im Schnitt wurden bei 0,79 Anpassungen pro Aufgabe aktiv in die generierten Abfragen eingegriffen, was auf eine tatsächliche Nutzung des Kontrollmechanismus hinweist.

Gleichzeitig zeigt die Gesamtaccuracy des HYBRID-Modus (62,50 %), dass zusätzliche Kontrolle allein nicht zu maximaler Genauigkeit führt. Der hybride Ansatz

reduziert Fehlinterpretationen, eliminiert sie jedoch nicht vollständig. Mögliche Ursachen sind eine unvollständige Nutzung der Editierfunktion oder Fälle, in denen die automatisch generierte Abfrage als korrekt akzeptiert wurde, obwohl sie von der Referenzlösung abwich.

Diese Befunde lassen sich als empirische Unterstützung des in Kapitel 3 entwickelten Modells interpretieren, in dem hybride Interaktion als Mechanismus zur Verteilung von Kontrolle zwischen Nutzer und System fungiert. Während NLI die initiale Spezifikation effizient übernimmt, ermöglicht die visuelle Repräsentation im hybriden Ansatz eine gezielte Korrektur und Stabilisierung des Ergebnisses. Die Ergebnisse legen nahe, dass diese Kombination insbesondere unter Bedingungen erhöhter struktureller Komplexität wirksam wird.

6.5 Kognitive Belastung und subjektive Wahrnehmung

Die Ergebnisse zeigen, dass der HYBRID-Modus die geringste wahrgenommene Arbeitsbelastung aufweist (TLX: 5,83). Dies steht im Einklang mit der Annahme A3, wonach persistente visuelle Repräsentationen und interaktive Kontrollmöglichkeiten die kognitive Stabilität erhöhen. Die explizite Sichtbarkeit der generierten Abfragestruktur reduziert die Notwendigkeit mentaler Rekonstruktion und entlastet das Arbeitsgedächtnis.

Bemerkenswert ist, dass der NLI-Modus trotz höherer Bearbeitungszeiten und instabiler Interaktionsverläufe in Einzelfällen die höchste Bewertung der Gebrauchstauglichkeit erhält (SUS: 71,46). Dieses Ergebnis deutet auf eine systematische Diskrepanz zwischen objektiver Performanz und subjektiver Wahrnehmung hin, die in der Literatur bereits dokumentiert ist (Wang et al., 2005). Die wahrgenommene Natürlichkeit der Spracheingabe scheint die Usability-Bewertung positiv zu beeinflussen, unabhängig von tatsächlicher Effizienz oder Fehleranfälligkeit. Nutzer schätzen die Möglichkeit, analytische Intentionen direkt in Alltagssprache zu formulieren, auch wenn dies in einzelnen Fällen mit erheblichem Zeitaufwand verbunden ist.

6.6 Bezug zu theoretischen Annahmen

Die Ergebnisse stützen zentrale Annahmen aus Kapitel 2.

Annahme A1 (Aufgabenkomplexität als primäre Entscheidungsdimension) wird durch die explorative Komplexitätsanalyse gestützt: Die relativen Stärken der Interaktionsformen verschieben sich systematisch mit der strukturellen Komplexität der Aufgaben. NLI dominiert bei einfachen, HYBRID bei komplexen Aufgaben. Allerdings bleibt die niedrige GUI-Accuracy bei einfachen Aufgaben (45,8 %) durch das Modell unzureichend erklärt. Das Schwellenwertmodell setzt eine korrekte Nutzung der jeweiligen Interaktionsform voraus – oberflächliches Interaktionsverhalten, wie es im GUI-Modus beobachtet wurde, stellt eine Randbedingung dar, die im Modell nicht abgebildet ist.

Annahme A2 (interaktive Disambiguation als Stabilitätsmechanismus) wird durch die Hybrid Edits und die erhöhte Accuracy des HYBRID-Modus bei komplexen Aufgaben unterstützt. Die aktive Nutzung der Editierfunktion zeigt, dass Nutzer den Kontrollmechanismus als relevant erachten.

Annahme A3 (persistente visuelle Repräsentation) spiegelt sich in der geringsten kognitiven Belastung des HYBRID-Modus wider.

Die beobachtete Diskrepanz zwischen subjektiver Bewertung und objektiver Performanz im NLI-Modus ist konsistent mit bestehenden Befunden (Wang et al., 2005) und unterstützt die in Abschnitt 2.3 beschriebene Wahrnehmungsverzerrung gegenüber natürlichsprachlichen Interfaces.

6.7 Implikationen für das Interaktionskonzept

Die Ergebnisse legen nahe, dass hybride Interaktion nicht als universell überlegene Lösung verstanden werden kann, sondern als kontextabhängiger Ansatz, dessen Stärken von der Aufgabenstruktur abhängen.

Die empirischen Befunde unterstützen folgende Gestaltungsprinzipien:

- Natürliche Sprache eignet sich besonders für die initiale Formulierung einfacher analytischer Intentionen, bei denen Ambiguität gering ist.

- Grafische Interfaces bieten bei einfachen Aufgaben keine Genauigkeitsvorteile gegenüber NLI, wenn Nutzer die Parametrisierung oberflächlich vornehmen. Designs, die zur Überprüfung der Eingaben anregen, könnten die Accuracy verbessern.
- Hybride Ansätze entfalten ihre Stärken insbesondere bei komplexen Aufgaben, bei denen die Kombination aus sprachlicher Spezifikation und visueller Kontrolle Fehlinterpretationen reduziert.
- Transparenz- und Disambiguationsmechanismen sollten adaptiv gestaltet sein: Bei einfachen Aufgaben kann direkte Ausführung ohne zusätzliche Kontrollschritte effizienter sein.

Damit bestätigt die Studie das in Kapitel 3 entwickelte Modell einer funktionalen Arbeitsteilung zwischen Interaktionsformen und präzisiert es hinsichtlich der Komplexitätsabhängigkeit.

6.8 Limitationen der Studie

Die Ergebnisse sind im Kontext mehrerer Limitationen zu interpretieren.

Die Stichprobengröße ist mit 12 Teilnehmenden begrenzt, wodurch die statistische Aussagekraft der Ergebnisse eingeschränkt ist. Ergänzend zur deskriptiven Analyse wurden paarweise Vergleiche mittels Wilcoxon Signed-Rank Tests durchgeführt. Aufgrund der kleinen Stichprobe und teilweise fehlender Werte variieren die Fallzahlen zwischen den Vergleichen ($n = 6$ bis $n = 13$), wodurch die Teststärke reduziert ist. Die hohe Varianz im NLI-Modus ist insbesondere auf einzelne Ausreißer in den Bearbeitungszeiten zurückzuführen, die durch individuelle Nutzungsstrategien einzelner Teilnehmender bedingt sind.

Die Klassifikation der Aufgabenkomplexität erfolgte nachträglich auf Basis der Query-Struktur und stellt keinen experimentell kontrollierten Faktor dar. Konfundierende Einflüsse wie Aufgabenreihenfolge oder aufgabenspezifische Lerneffekte innerhalb der Blöcke können daher nicht vollständig ausgeschlossen werden.

Die inferenzstatistischen Analysen sind aufgrund der geringen Stichprobengröße und reduzierten Teststärke lediglich als explorative Hinweise zu interpretieren und erlauben keine belastbaren inferenziellen Schlussfolgerungen.

Darüber hinaus handelt es sich um eine kontrollierte Laborsituation. Langfristige Nutzungseffekte, Lernprozesse sowie organisationale Einflussfaktoren wurden nicht untersucht. Die Übertragbarkeit der Ergebnisse auf reale Anwendungskontexte ist daher eingeschränkt. Insbesondere sind Effekte langfristiger Nutzung, Routinebildung und organisationaler Einbettung nicht abgebildet. Die Studie ist daher primär als explorativer Beitrag zur Hypothesengenerierung zu verstehen.

Insgesamt sind die Ergebnisse als explorativ zu interpretieren und dienen primär der Identifikation struktureller Muster unter kontrollierten Bedingungen, nicht der Ableitung generalisierbarer Effekte.

7 Fazit und Ausblick

7.1 Zusammenfassung der Arbeit

Ziel dieser Arbeit war die Konzeption, prototypische Umsetzung und empirische Evaluation eines hybriden Interaktionskonzepts für Analytics-Dashboards. Im Fokus stand die Frage, wie sich die Kombination aus natürlicher Sprache und grafischer Interaktion auf Effizienz, Genauigkeit, kognitive Belastung und Nutzerwahrnehmung auswirkt.

Auf Basis einer systematischen Analyse bestehender Forschung wurde ein konzeptionelles Modell entwickelt, das Interaktionsformen nicht als konkurrierende Alternativen, sondern als komplementäre Komponenten einer regelgeleiteten Kontrollarchitektur versteht. Aufgabenkomplexität wurde dabei als zentrale Entscheidungsdimension modelliert.

Zur empirischen Untersuchung wurde ein webbasiertes System implementiert, das drei Interaktionsmodi bereitstellt: eine rein grafische Interaktion, eine natürlichsprachliche Interaktion sowie eine hybride Kombination beider Ansätze. In einer kontrollierten Nutzerstudie wurden diese Interaktionsformen miteinander verglichen.

Die Ergebnisse zeigen, dass sich die Interaktionsformen systematisch unterscheiden. Der GUI-Modus ermöglicht die schnellste Bearbeitung, weist jedoch die geringste Genauigkeit auf. Der NLI-Modus erzielt die höchste Genauigkeit, ist jedoch mit deutlich erhöhten Bearbeitungszeiten verbunden. Der HYBRID-Modus positioniert sich zwischen beiden Ansätzen und zeigt insbesondere Vorteile hinsichtlich der wahrgenommenen Arbeitsbelastung.

7.2 Zentrale Erkenntnisse

Die Arbeit liefert mehrere zentrale Erkenntnisse für die Gestaltung analytischer Interaktionssysteme.

Erstens zeigt sich, dass keine Interaktionsform universell überlegen ist. Stattdessen bestehen strukturelle Trade-offs zwischen Effizienz, Genauigkeit und kognitiver Belastung.

Zweitens bestätigt sich die Annahme, dass hybride Interaktion als stabilisierende Kontrollarchitektur wirken kann. Die Kombination aus sprachlicher Spezifikation und visueller Überprüfbarkeit ermöglicht eine bessere Balance zwischen Flexibilität und Kontrolle.

Drittens zeigt sich eine systematische Diskrepanz zwischen objektiver Performanz und subjektiver Wahrnehmung. Natürlichsprachliche Interaktion wird trotz höherer Bearbeitungszeiten als besonders gebrauchstauglich wahrgenommen.

Viertens unterstreichen die Ergebnisse die Bedeutung von Transparenz- und Disambiguationsmechanismen. Die Möglichkeit, Systeminterpretationen explizit zu überprüfen und anzupassen, erweist sich als zentraler Faktor für die Interaktionsqualität.

Fünftens zeigt sich, dass Varianz ein eigenständiges Qualitätsmerkmal von Interaktionssystemen darstellt. Insbesondere im NLI-Modus treten stark schwankende Bearbeitungszeiten auf, die auf instabile Interaktionsverläufe in Einzelfällen zurückzuführen sind. Für die Systemgestaltung bedeutet dies, dass neben durchschnittlicher Effizienz auch die Vorhersagbarkeit von Interaktionsdauer als zentrales Designziel berücksichtigt werden muss.

7.3 Beitrag der Arbeit

Die Arbeit leistet einen eigenständigen Beitrag auf konzeptioneller, methodischer und empirischer Ebene zur Gestaltung hybrider Interaktionssysteme in Analytics-Dashboards.

Auf konzeptioneller Ebene wird ein Modell entwickelt, das Interaktionsformen nicht als konkurrierende Alternativen, sondern als komplementäre Komponenten einer regelgeleiteten Kontrollarchitektur beschreibt. Das Modell beschreibt die Verteilung

analytischer Kontrolle als Funktion von Aufgabenkomplexität, Transparenzanforderungen und kognitiver Belastung und liefert damit eine systematische Entscheidungslogik für die Wahl geeigneter Interaktionsformen.

Auf methodischer Ebene wird ein funktionaler Webprototyp realisiert, der unterschiedliche Interaktionsparadigmen unter kontrollierten Bedingungen vergleichbar macht. Durch die einheitliche Abfrageinfrastruktur und das Within-Subjects-Design wird sichergestellt, dass beobachtete Unterschiede auf die Interaktionsform zurückzuführen sind und nicht auf systembedingte Variationen.

Auf empirischer Ebene zeigt die Studie, dass Interaktionsformen systematisch unterschiedliche Profile hinsichtlich Effizienz, Genauigkeit und Stabilität aufweisen. Insbesondere wird deutlich, dass natürlichsprachliche Interaktion durch hohe Varianz gekennzeichnet ist, während hybride Ansätze bei steigender Aufgabenkomplexität eine stabilisierende Wirkung entfalten. Die Ergebnisse liefern damit empirische Evidenz für die Komplexitätsabhängigkeit analytischer Interaktionsleistung.

7.4 Limitationen

Die Ergebnisse der Arbeit sind im Kontext mehrerer Einschränkungen zu interpretieren.

Die Stichprobe ist begrenzt und erlaubt keine umfassende Generalisierung. Der verwendete Datensatz ist synthetisch und bildet reale Datenstrukturen nur vereinfacht ab. Zudem wurde die Studie in einer kontrollierten Umgebung durchgeführt, wodurch reale Nutzungsszenarien nur eingeschränkt repräsentiert sind.

Weiterhin wurden ausschließlich kurzfristige Interaktionen untersucht. Lern- und Gewöhnungseffekte sowie langfristige Nutzung wurden nicht berücksichtigt.

7.5 Ausblick

Aus den Ergebnissen ergeben sich mehrere Ansatzpunkte für zukünftige Forschung.

Eine zentrale Erweiterung besteht in der Untersuchung komplexerer Datenszenarien, insbesondere mit größeren und heterogenen Datensätzen sowie mehrstufigen Analyseprozessen.

Darüber hinaus erscheint die Entwicklung adaptiver hybrider Systeme vielversprechend, die Interaktionsformen dynamisch an Aufgabenkomplexität und Nutzerverhalten anpassen.

Ein weiterer Forschungsansatz liegt in der vertieften Analyse von Transparenzmechanismen. Insbesondere die Frage, wie unterschiedliche Formen und Detaillierungsgrade von Erklärungen die Interaktion beeinflussen, bietet weiteres Potenzial.

Schließlich sind Langzeitstudien erforderlich, um Lernprozesse, Nutzungsstrategien und Akzeptanz hybrider Systeme in realen Anwendungskontexten zu untersuchen.

Literaturverzeichnis

- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6), 574–594. <https://doi.org/10.1080/10447310802205776>
- Besharati, P. (2025). *A Comparative Evaluation of Natural Language and Dashboard Interfaces for Visualizing Real-Time Monitoring Data* [Magisterarb., Dalhousie University]. (besucht am 27. März 2026) <https://dalspace.library.dal.ca/items/9ebda67d-99f2-4536-8297-37b7dd6b1759>
- Brooke, J. (1996). SUS: A "Quick and Dirty" Usability Scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester & I. L. McClelland (Hrsg.), *Usability Evaluation in Industry*. Taylor & Francis.
- Feng, Y., Wang, X., Pan, B., Wong, K., & Ren, Y. (2023). XNLI: Explaining and Diagnosing NLI-Based Visual Data Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 29(1), 537–547. <https://doi.org/10.1109/TVCG.2023.3240003>
- Forschungsgemeinschaft, D. (2023). Verwendung generativer Modelle für die Text- und Bilderstellung im Förderhandeln der DFG: Stellungnahme des Präsidiums. (besucht am 31. März 2026) <https://www.dfg.de/resource/blob/289674/ff57cf46c5ca109cb18533b21fba49bd/230921-stellungnahme-praesidium-ki-ai-data.pdf>
- Gao, T., Dontcheva, M., Adar, E., Liu, Z., & Karahalios, K. (2015). DataTone: Managing Ambiguity in Natural Language Interfaces for Data Visualization. *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, 489–500. <https://doi.org/10.1145/2807442.2807478>
- Guo, Y., Shi, D., Guo, M., Wu, Y., & Cao, N. (2024). Talk2Data: A Natural Language Interface for Exploratory Visual Analysis via Question Decomposition. *ACM Transactions on Interactive Intelligent Systems*, 14(2), 1–24. <https://doi.org/10.1145/3643894>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, 52, 139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)

- Hernandez-Bocanegra, D. C., & Ziegler, J. (2023). Explaining Recommendations through Conversations: Dialog Model and the Effects of Interface Type and Degree of Interactivity. *ACM Transactions on Interactive Intelligent Systems*, 13(2), 1–47. <https://doi.org/10.1145/3579541>
- Nafari, M., & Weaver, C. (2013). Augmenting Visualization with Natural Language Translation of Interaction: A Usability Study. *Computer Graphics Forum*, 32(3pt4), 391–400. <https://doi.org/10.1111/cgf.12126>
- Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Vaughan, J. W., & Wallach, H. (2021). Manipulating and Measuring Model Interpretability. *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, 1–52. <https://doi.org/10.1145/3411764.3445315>
- Saktheeswaran, A., Srinivasan, A., & Stasko, J. (2020). Touch? Speech? or Touch and Speech? Investigating Multimodal Interaction for Visual Network Exploration and Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 26(6), 2168–2179. <https://doi.org/10.1109/TVCG.2020.2970512>
- Schaffer, J., Giridhar, P., Jones, D., Höllerer, T., Abdelzaher, T., & O'Donovan, J. (2015). Getting the Message?: A Study of Explanation Interfaces for Microblog Data Analysis. *Proceedings of the 20th International Conference on Intelligent User Interfaces*, 345–356. <https://doi.org/10.1145/2678025.2701406>
- Speier, C., & Morris, M. G. (2003). The Influence of Query Interface Design on Decision-Making Performance. *MIS Quarterly*, 27(3), 397–423. <https://doi.org/10.2307/30036539>
- Vaithilingam, P., Glassman, E. L., Inala, J., & Wang, C. (2024). DynaVis: Dynamically Synthesized UI Widgets for Visualization Editing. *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, 1–17. <https://doi.org/10.1145/3613904.3642639>
- Voigt, H., Alacam, O., Meuschke, M., Lawonn, K., & Zarriß, S. (2022). The Why and The How: A Survey on Natural Language Interaction in Visualization. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 348–374. <https://doi.org/10.18653/v1/2022.naacl-main.27>
- Wang, Q., Nass, C., & Hu, J. (2005). Natural Language Query vs. Keyword Search: Effects of Task Complexity on Search Performance, Participant Perceptions,

and Preferences. *IFIP TC13 International Conference on Human-Computer Interaction (INTERACT)*, 106–116. https://doi.org/10.1007/11555261_12

A Anhang

Erklärung zur Nutzung KI-basierter Werkzeuge

Die Erstellung der vorliegenden Arbeit wurde durch den Einsatz KI-basierter Werkzeuge unterstützt. Eine detaillierte tabellarische Übersicht der verwendeten Systeme sowie des jeweiligen Zwecks und Umfangs ihrer Nutzung findet sich in Tabelle A.1.

Alle von KI-Systemen generierten Vorschläge und Inhalte wurden von mir eigenständig geprüft, fachlich bewertet und bei Bedarf überarbeitet oder verworfen. Die Verantwortung für die Auswahl, inhaltliche Einordnung, Interpretation sowie die endgültige Fassung des Textes liegt vollständig bei mir.

Grundlage für den Umgang mit den Werkzeugen bilden die aktuellen Empfehlungen der Deutschen Forschungsgemeinschaft (DFG) zu KI in wissenschaftlichen Arbeiten (Forschungsgemeinschaft, 2023).

Declaration on the Use of AI-based Tools

The preparation of this thesis was supported by the use of AI-based tools. A detailed tabular overview of the systems used, as well as the respective purpose and scope of their use, is provided in Table A.1.

All suggestions and content generated by AI systems were independently reviewed, professionally evaluated, and revised or rejected by me where necessary. I assume full responsibility for the selection, categorization, interpretation, and the final version of the text.

The handling of these tools follows the current recommendations of the German Research Foundation (DFG) for scientific work (Forschungsgemeinschaft, 2023).

A.1 Studienmaterialien

Alle Aufgaben basieren auf einem Datensatz mit Verkaufsdaten, bestehend aus den Dimensionen Produkt, Kategorie und Region sowie den Metriken Umsatz (revenue), Kosten (cost) und Menge (quantity).

A.1.1 Trainingsaufgaben

Vor Beginn der eigentlichen Aufgabenbearbeitung wurden drei Trainingsaufgaben bereitgestellt, um die Teilnehmenden mit dem System und den Interaktionsformen vertraut zu machen.

1. Was ist der Gesamtumsatz (revenue) für Tablets?
2. Wie hoch sind die durchschnittlichen Kosten (cost) pro Kategorie?
3. Was ist die Gesamtmenge (quantity) in der Region East pro Produkt?

A.1.2 Aufgabenset A

1. Wie hoch ist der Gesamtumsatz (revenue) für die Kategorie Electronics?
2. Was sind die durchschnittlichen Kosten (cost) pro Region?
3. Wie viele Keyboards wurden insgesamt verkauft (quantity)?
4. Was ist der Gesamtumsatz (revenue) pro Region, nur für Electronics?

A.1.3 Aufgabenset B

1. Was ist die Gesamtmenge (quantity) aller Verkäufe in der Region North?
2. Was ist der durchschnittliche Umsatz (revenue) pro Produkt?
3. Wie hoch sind die maximalen Kosten (cost) für Laptops?
4. Was ist die Gesamtmenge (quantity) pro Kategorie in der Region South?

A.1.4 Aufgabenset C

1. Wie hoch sind die Gesamtkosten (cost) für alle Produkte außer Keyboards?
2. Was ist der minimale Umsatz (revenue) pro Produkt?
3. Wie viele Verkäufe gab es für Phones?

4. Was ist der Gesamtumsatz (revenue) pro Produkt in der Region West?

A.2 Fragebögen

A.2.1 System Usability Scale (SUS)

Zur Erfassung der wahrgenommenen Gebrauchstauglichkeit wurde die System Usability Scale (SUS) verwendet. Die Teilnehmenden bewerteten die folgenden Aussagen auf einer fünfstufigen Likert-Skala.

Skala:

- Stimme überhaupt nicht zu
- Stimme eher nicht zu
- Teils/teils
- Stimme eher zu
- Stimme voll zu

Items:

1. Ich denke, dass ich dieses System gerne häufig benutzen würde.
2. Ich fand das System unnötig komplex.
3. Ich fand das System einfach zu benutzen.
4. Ich denke, ich würde die Hilfe einer technisch versierten Person benötigen, um das System benutzen zu können.
5. Ich fand, die verschiedenen Funktionen des Systems waren gut integriert.
6. Ich fand, das System enthielt zu viele Inkonsistenzen.
7. Ich kann mir vorstellen, dass die meisten Menschen den Umgang mit diesem System sehr schnell erlernen würden.
8. Ich fand das System sehr umständlich zu benutzen.
9. Ich fühlte mich bei der Benutzung des Systems sehr sicher.
10. Ich musste eine Menge lernen, bevor ich mit dem System zurechtkam.

A.2.2 NASA Task Load Index (NASA-TLX)

Zur Erfassung der wahrgenommenen Arbeitsbelastung wurde der NASA Task Load Index (NASA-TLX) verwendet. Die Teilnehmenden bewerteten die folgenden Dimensionen auf einer Skala von gering bis hoch (bzw. bei Leistung von sehr gut bis sehr schlecht).

1. **Geistige Anforderung**

Wie viel geistige Anstrengung war erforderlich?

Skala: Gering – Hoch

2. **Körperliche Anforderung**

Wie viel körperliche Aktivität war erforderlich?

Skala: Gering – Hoch

3. **Zeitliche Anforderung**

Wie viel Zeitdruck haben Sie empfunden?

Skala: Gering – Hoch

4. **Leistung**

Wie erfolgreich waren Sie bei der Aufgabe?

Skala: Sehr gut – Sehr schlecht

5. **Anstrengung**

Wie hart mussten Sie arbeiten, um Ihr Leistungsniveau zu erreichen?

Skala: Gering – Hoch

6. **Frustration**

Wie unsicher, entmutigt oder verärgert waren Sie?

Skala: Gering – Hoch

A.3 Verzeichnis der KI-Nutzung

Die Tabelle A.1 dokumentiert die in dieser Arbeit verwendeten KI-Werkzeuge sowie Zweck, Kontext und Art der erzeugten Unterstützung.

Tabelle A.1: Verzeichnis der in dieser Arbeit genutzten KI-Werkzeuge

KI-Tool		Zweck	Kontext/Eingangstext	Generierte Ausgabe	Kapitel
Cursor (integrierter KI-Assistent)	(in-	Unterstützung bei Implemen- tierung und Dokumentstruk- tur	Entwicklung des Webproto- typs (Frontend, API-Routen, Datenverarbeitung) so- wie Strukturierung des LaTeX-Dokuments	Code-Vorschläge, Refac- torings sowie Hinweise zur Strukturierung von Komponenten und Dokument	Kap. 3.5, Kap. 4.2, Anhang
ChatGPT (GPT-5.3 / GPT-5.4)	/	Sprachliche Überarbeitung und punktuell- e inhaltliche Unterstützung	Bereits ausgearbeitete Text- passagen sowie einzelne Fra- gestellungen zu Konzepten und Forschungszusammen- hängen	Vorschläge zur sprachlichen Präzisierung, Verdichtung und Strukturierung sowie erklärende Einordnungen	Kap. 1–7, Abstract
Elicit		Literaturrecherche	Recherche zu Themen wie Natural Language Interfaces, Visual Analytics und hybride Interaktion	Zusammenfassungen wissen- schaftlicher Arbeiten sowie Hinweise auf relevante Quel- len	Kap. 2, Kap. 3
Claude Code		Unterstützung bei Datenauf- bereitung und Auswertung	Aufbereitete Studiendaten und Metriken aus dem Logging-System	Hilfestellung bei Strukturie- rung der Auswertung und Er- stellung von Visualisierungen	Kap. 5

Erklärung der Selbstständigkeit

Hiermit versichere ich, die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt zu haben.

Frankfurt am Main, den 31. März 2026

A handwritten signature in black ink, appearing to read 'Kevin Tamme', with a stylized, cursive script.

Kevin Tamme