

BACHELORARBEIT

Wahrnehmung KI-generierter Software:

Auswirkungen der Offenlegung von KI-Genese auf das Nutzervertrauen

vorgelegt am 23. April 2026
Marcus Constantin Möller

Erstprüferin: Prof. Dr. Larissa Putzar
Zweitprüferin: Franziska Heil

**HOCHSCHULE FÜR ANGEWANDTE
WISSENSCHAFTEN HAMBURG**

Fakultät Informatik und Digitale Gesellschaft
Finkenau 35
22081 Hamburg

Zusammenfassung

Der jüngste Aufstieg generativer künstlicher Intelligenz hat den Paradigmenwechsel von manuellem Coding zu Vibecoding initiiert, bei dem Programmcode durch natürliche Sprache mit sogenannten Prompts erzeugt wird. Da diese Systeme auf Large Language Models basieren, weisen sie inhärente Blackbox-Eigenschaften und eine Neigung zu Halluzinationen auf, die durch ihren oft professionellen Klang plausibel wirken. Der derart generierte Code weist jedoch oft fehlerhafte Logik und Sicherheitslücken auf. Dies birgt die Gefahr, dass Nutzer der professionell aussehenden Benutzeroberfläche einer so erzeugten Anwendung ein unreflektiertes Übervertrauen entgegenbringen, ohne die potenziell fehlerhafte Berechnungslogik zu hinterfragen. Die vorliegende Bachelorarbeit untersucht daher, wie die Offenlegung der KI-Genese das Nutzervertrauen beeinflusst. Anhand eines vollständig mit Figma Make generierten Photovoltaik-Amortisationsrechners wurde eine qualitative Nutzerstudie ($N = 8$) mit zwei Untersuchungsgruppen (aufgeklärt über KI-Erzeugung vs. unaufgeklärt) durchgeführt. Die Datenerhebung erfolgte mittels Thinking-Aloud-Protokollen, semi-strukturierten Interviews und einem adaptierten Trust-in-Automation-Fragebogen. Die thematische Analyse der qualitativen Ergebnisse zeigt, dass eine ansprechende Benutzeroberfläche ein hohes initiales emotionales Vertrauen generieren kann. Die Offenlegung der KI-Genese kann jedoch als entscheidender Störfaktor wirken. Sie durchbricht das blinde Vertrauen und induziert ein angemessenes Misstrauen. Die Arbeit schließt mit der Erkenntnis, dass Transparenz über den generativen Ursprung von Software essenziell ist, um in Hochrisikokontexten eine angemessene Systemnutzung zu gewährleisten.

Abstract

The recent rise of generative artificial intelligence has initiated a paradigm shift known as vibecoding, in which program code is generated via natural language inputs, or prompts. Because these systems are based on Large Language Models, they exhibit inherent black-box characteristics and a tendency toward hallucinations, which appear plausible due to their often professional-sounding output. However, code generated in this manner frequently contains flawed logic and security vulnerabilities. This creates the risk that users might place unreflective overtrust in the professional-looking user interface of such an application without questioning the potentially flawed underlying computational logic. Therefore, this bachelor's thesis investigates how disclosing the AI genesis of a software application influences user trust. Using a photovoltaic amortization calculator entirely generated with Figma Make, a qualitative user study ($N = 8$) was conducted with two experimental groups (informed about AI-Origin vs. uninformed). Data collection was performed using think-aloud protocols, semi-structured interviews, and an adapted Trust in Automation questionnaire. The thematic analysis of the qualitative results demonstrates that an appealing user interface can generate a high level of initial affect-based trust. However, the disclosure of the AI genesis can act as a crucial disruptive factor. It shatters blind trust and induces warranted distrust. The thesis concludes that transparency regarding the generative origin of software is essential to ensure appropriate reliance in high-risk contexts.

Inhaltsverzeichnis

Abbildungsverzeichnis	III
Tabellenverzeichnis	IV
Codeblockverzeichnis	V
Abkürzungsverzeichnis	VI
Gender-Hinweis	VII
1 Einleitung	1
1.1 Problemstellung und Motivation	1
1.2 Zielsetzung und Forschungsfrage	3
1.3 Aufbau	4
2 Theoretische Grundlagen	5
2.1 Generative KI und Large Language Models	5
2.1.1 Foundation Model	5
2.1.2 Transformer und NLP	6
2.1.3 Begrenztes Sprachverständnis (Stochastischer Papagei)	7
2.1.4 Blackbox-Problem	7
2.1.5 Vibecoding und Agenten	9
2.1.6 Figma Make	10
2.2 Nutzervertrauen	11
2.2.1 Nutzervertrauen in Automation	12
2.2.2 Nutzervertrauen in Mensch-Computer-Interaktion (HCI)	15
2.2.3 Nutzervertrauen in KI	18
3 Implementierung und Architekturanalyse des PVA-Rechners	20
3.1 PV-Amortisationsrechner im Überblick	20
3.2 Architekturanalyse	24
4 Studiendesign und Methodik	29
4.1 Zielsetzung und Aufbau	29

4.2	Ablauf der Erhebung	31
4.3	Stichprobe	32
4.4	Erhebungsinstrumente	33
4.4.1	Quantitative Instrumente	33
4.4.2	Qualitative Instrumente	34
4.5	Datenauswertungsstrategie: Thematische Analyse nach Braun und Clarke	35
5	Ergebnisse	37
5.1	Quantitative Vertrauensmessung	37
5.1.1	Trust in Automation Fragebogen	37
5.1.2	Eigene Single Items	39
5.2	Qualitative Ergebnisse: Thematische Analyse	40
5.2.1	Theme 1: Process - Die KI als undurchsichtige Blackbox und Stochastischer Papagei	40
5.2.2	Theme 2: Performance - Spannungsfeld zwischen Usability und Automation Bias	43
5.2.3	Theme 3: Purpose - Wahrgenommene Intention und Professionalität	47
5.2.4	Theme 4: Appropriate Reliance und Warranted Distrust - Die Ambivalenz der KI-Offenlegung	50
6	Diskussion und Limitationen	54
6.1	Beantwortung der Forschungsfrage	54
6.1.1	Process	54
6.1.2	Performance	56
6.1.3	Purpose	58
6.2	Probleme und Limitationen der Studie	60
7	Fazit und Ausblick	62
	Literatur	64
	Anhang	67
	Studienergebnisse	67
	Instruktionsbogen	68
	Fragebogen	71
	Prompts	83

Abbildungsverzeichnis

1.1	Von Entwicklern und Designern berichtete Effizienzsteigerung durch KI der Jahre 2024 und 2025(l. Figma, 2025b)	1
1.2	Andrej Karpathys Definition von Vibecoding (Andrej Karpathy [@karpathy, 2025)	3
3.1	Startseite und Standorteingabe	22
3.2	Dachseiteneingabe	22
3.3	Ergebnisseite	23
3.4	Totes Objekt und Gott-Komponente (Screenshot aus dem Code) . .	25

Tabellenverzeichnis

5.1	Übersicht der Itemwerte sowie Mittelwerte und Standardabweichungen	38
5.2	Gruppenmittelwerte und Standardabweichungen	38
5.3	Zusammengefasste Übersicht der drei Items mit Einzelwerten sowie Gruppenmittelwerten und Standardabweichungen	39

Codeblockverzeichnis

3.1	Exemplarischer Codeausschnitt: Single Responsibility und Separations of Concern	25
3.2	Exemplarischer Ausschnitt aus dem Finanzierungs-Schritt für DRY .	26
3.3	Fail-Fast Verletzung und Magic Number	27

Abkürzungsverzeichnis

KI	Künstliche Intelligenz
HCI	Human Centered Interaction
UX	User Experience
UI	User Interface
IDE	Integrated Development Environment
LLM	Large Language Model
PV	Photovoltaik
PVA	Photovoltaik-Amortisation
PVA-Rechner	Photovoltaik-Amortisationsrechner
PV-Rechner	Photovoltaik-Amortisationsrechner
NLP	Natural Language Processing
NLU	Natural Language Understanding
TiA-Fragebogen	Trust in Automation-Fragebogen
FM	Foundation Model
ROI	Return On Investment
TA	Thematische Analyse
PPP-Modell	Performance-Purpose-Process-Modell
ABI-Modell	Ability-Benevolence-Integrity-Modell
KPI	Key Performance Indicator
DRY	Dont Repeat Yourself
SRP	Single Responsibility Principle
SOC	Separation of Concerns

Gender-Hinweis

Zur besseren Lesbarkeit dieser Bachelorarbeit wird bei Personenbezeichnungen das generische Maskulinum verwendet. Sämtliche Personenbezeichnungen beziehen sich, sofern nicht anders kenntlich gemacht, auf alle Geschlechter.

1 Einleitung

1.1 Problemstellung und Motivation

Seit der Veröffentlichung des generativen KI-Chatbots ChatGPT im November 2022 sind Chatbots sowohl im privaten als auch im beruflichen Kontext in vielen Branchen Teil des Alltags geworden. Laut McKinsey lag die Nutzung generativer KI-Systeme von Unternehmen in mindestens einem Businessbereich 2023 noch bei 33 Prozent und ist über 65 Prozent im Jahr 2024 auf 79 Prozent im Jahr 2025 gestiegen (McKinsey & Company, [2025](#)).

Dieser Trend zeigt sich besonders deutlich im Bereich der Software- und Design-Entwicklung. In einer Studie von *Figma*, einem primär für UX/UI-Prototyping bekannten Unternehmen, stimmten im Jahr 2024 71 Prozent der befragten Designer und Entwickler der Aussage zu, dass KI die Effizienz ihrer Arbeit erheblich steigert. Im Jahr 2025 stieg dieser Wert auf 78 Prozent an (I. Figma, [2025b](#)).

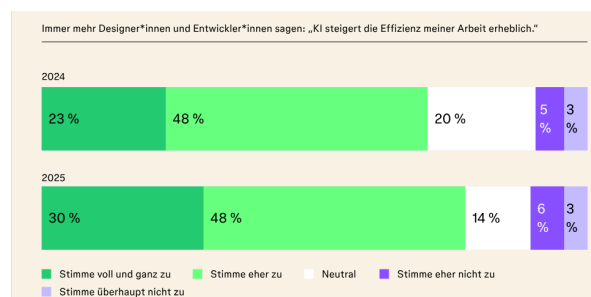


Abbildung 1.1: Von Entwicklern und Designern berichtete Effizienzsteigerung durch KI der Jahre 2024 und 2025(I. Figma, [2025b](#))

Mit Tools wie *GitHub Copilot* in integrierten Entwicklungsumgebungen (IDE) wie *Visual Studio Code* (Microsoft) oder *IntelliJ* (JetBrains), der KI-gestützten IDE *Cursor.AI* sowie *Figma Make* stehen mittlerweile umfangreiche Möglichkeiten zur teilweisen oder vollständigen generativen Erzeugung von Softwareanwendungen zur Verfügung. Diese Systeme reagieren auf textbasierte Eingaben (Prompts) und können daraus sowohl einzelne Code-Snippets als auch lauffähige Applikationen

generieren. Figma bewirbt seine Lösung Figma Make explizit damit, dass auch unerfahrene Akteure durch wenige Prompts komplexe Web-Apps erstellen können. Bemerkenswert ist hierbei, dass sich die generierten Artefakte nicht mehr nur auf statische Design-Mockups beschränken, sondern funktionale Business-Logik inkludieren (l. Figma, [2025a](#)). Diese zunehmende Integration generativer KI-Systeme in kreative und technische Arbeitsprozesse hat einen Paradigmenwechsel in der Softwareentwicklung initiiert.

Dieser neuartige Ansatz wird in der aktuellen Debatte unter dem Begriff Vibecoding subsumiert, welcher im Februar 2025 maßgeblich durch den ehemaligen OpenAI-Softwareentwickler Andrej Karpathy geprägt wurde:

„I’m building a project or webapp, but it’s not really coding – I just see stuff, say stuff, run stuff, and copy paste stuff, and it mostly works“
(Andrej Karpathy [[@karpathy](#), [2025](#)])

Bei diesem Paradigma formulieren Entwickler die funktionalen Anforderungen in natürlicher Sprache, während generative KI-Modelle den zugrunde liegenden Programmcode oder sogar komplette Anwendungen autonom generieren und implementieren (Theobald & Brandt, [2025](#)).

Dieses neue Entwicklungsparadigma führt jedoch auch zu neuen Herausforderungen und Problemen. Die Qualität des generierten Codes variiert stark in Abhängigkeit von der Prompt-Fähigkeit und Domänenexpertise des Anwenders (Theobald & Brandt, [2025](#)). Darüber hinaus unterliegt der Generierungsprozess einer Intransparenz. Da die Anwendungen auf Large Language Models (LLMs) basieren, ist für Endanwender oft nicht nachvollziehbar, auf welcher Datenbasis die generierten Programmteile und Berechnungslogiken beruhen. Diese Intransparenz wird in der Forschung als Blackbox-Problem bezeichnet. Nutzer tätigen Eingaben und erhalten Ausgaben, ohne den Prozess zu verstehen oder auf seine fachliche Validität hin überprüfen zu können (Zednik, [2019](#); Guidotti et al., [2018](#)).

Aus dieser Intransparenz resultiert eine relevante Herausforderung für die Mensch-Computer-Interaktion und die Vertrauensbildung der Nutzer. Da Endanwender die Leistungsfähigkeit und Korrektheit der durch Vibecoding generierten Berechnungslogiken nicht validieren können, besteht die Gefahr, dass sie sich bei der Beurteilung der Systemkompetenz unreflektiert auf die visuelle Professionalität der ebenfalls durch KI erzeugten Benutzeroberfläche verlassen (Schlicker & Langer, [2021](#)). Dies kann zu einem Übervertrauen in fehlerhafte Systemausgaben führen, insbesondere in Hochrisiko-Bereichen wie privaten Finanz- und Investitionsentscheidungen. Vor diesem Hintergrund stellt sich die drängende Frage, wie dieses mögliche Vertrauen in durch Vibecoding KI-generierte Applikationen

kalibriert werden kann, um eine angemessene Nutzung zu gewährleisten (Lee & See, 2004).



Abbildung 1.2: Andrej Karpathys Definition von Vibecoding (Andrej Karpathy [@karpathy, 2025])

1.2 Zielsetzung und Forschungsfrage

Basierend auf der skizzierten Problemstellung verfolgt die vorliegende Studie das Ziel, zu untersuchen, welchen Einfluss die Offenlegung, dass es sich um eine nahezu vollständig durch Vibecoding erzeugte Anwendung handelt, auf das Nutzervertrauen und die daraus resultierende Bewertung der Systemergebnisse hat. Um dieses Ziel zu erreichen, wird folgende zentrale Forschungsfrage (FF) formuliert:

Wie beeinflusst die Offenlegung der KI-Genese einer Softwareanwendung die subjektive Wahrnehmung und das Vertrauen der Nutzer in das System?

Zur Beantwortung dieser Frage dient ein Photovoltaik-Amortisationsrechner (PVA-Rechner), der die Anwender bei der Entscheidung unterstützen soll, ob sich ein Angebot für eine PV-Anlage wirtschaftlich lohnt. Die Amortisationszeit markiert hierbei den Zeitpunkt, an dem die Anschaffungskosten der Investition durch die Einsparungen aus der Stromerzeugung gedeckt sind. Bemerkenswert für das Forschungsdesign ist, dass dieser PVA-Rechner vollständig über das KI-Tool Figma Make generiert wurde. Die Untersuchung erfolgt im Rahmen einer Nutzerstudie mit insgesamt acht Probanden ($N = 8$), die in zwei Untersuchungsgruppen

(aufgeklärt über KI-Genese vs. unaufgeklärt) aufgeteilt werden. Die Teilnehmer erhalten ein standardisiertes Testszenario zur Bedienung des PV-Rechners. Zur Erfassung der Vertrauenswerte füllen die Probanden im Anschluss einen adaptierten Trust in Automation (TiA)-Fragebogen (Jian et al., 2000) aus. Der methodische Schwerpunkt liegt jedoch auf der qualitativen Auswertung. Über die gesamte Dauer der Interaktion verbalisieren die Probanden ihre Gedanken anhand der Thinking-Aloud-Methode (Ericsson & Simon, 1984) und begründen ihre Antworten im Fragebogen. Die so gewonnenen Transkripte werden abschließend durch eine Thematische Analyse nach Braun und Clarke (2006) ausgewertet.

1.3 Aufbau

Die vorliegende Arbeit gliedert sich in sieben aufeinander aufbauende Kapitel. Im Anschluss an diese Einleitung legt Kapitel 2 das theoretische Fundament, indem es die technischen Grundlagen generativer KI und des Vibecodings beleuchtet sowie den aktuellen Forschungsstand zum Nutzervertrauen in der Mensch-Computer-Interaktion (HCI) aufarbeitet. Daraufhin beschreibt Kapitel 3 die konkrete Implementierung und Architekturanalyse des generierten PVA-Rechners. Die Kapitel 4 und 5 widmen sich dem empirischen Teil der Arbeit. Hier werden das detaillierte Studiendesign und die Methodik dargelegt sowie die quantitativen und qualitativen Ergebnisse der Nutzerstudie systematisch ausgewertet. In Kapitel 6 werden diese Erkenntnisse im Kontext der bestehenden Forschung kritisch diskutiert. Die Arbeit schließt in Kapitel 7 mit einem Fazit sowie einem Ausblick auf zukünftige Forschungsbedarfe.

2 Theoretische Grundlagen

2.1 Generative KI und Large Language Models

Generative KI-Chatbots wie ChatGPT, Claude oder Gemini basieren auf sogenannten Large Language Models, kurz LLMs, die eine Unterkategorie der Foundation Models darstellen. Diese LLMs verwenden die von Vaswani et al. (2017) eingeführte Transformer-Architektur und sind probabilistische Werkzeuge zum Generieren von Text. Auch Figma Make nutzt mit Claude 3.7 Sonnet ein LLM als zugrunde liegende Modellkomponente. Es handelt sich um ein Werkzeug zur generativen Erstellung von Nutzeroberflächen, UI/UX-Prototypen und Web Apps mit bereits rudimentär vorhandener Anwendungslogik, sodass die gewünschten Grundfunktionen der Anwendung bereits genutzt und präsentiert werden können (Figma, 2025).

2.1.1 Foundation Model

Bei einem sogenannten Foundation Model (FM) handelt es sich um ein neues Paradigma der KI-Systeme. Diese Modelle sind mithilfe von Machine Learning vortrainierte tiefe neuronale Netze, die meist mittels Self-Supervised Learning auf extrem großen Datenmengen basieren. Sie sind sowohl in der Lage, die Aufgaben zu lösen, für die sie trainiert wurden, als auch die riesige Datenbasis zu nutzen, um Aufgaben zu bewältigen, für die das Modell nicht explizit trainiert wurde. Diese Eigenschaft wird als Emergenz bezeichnet. Emergenz macht FMs sehr flexibel und für ein breites Spektrum verschiedener Anwendungsaufgaben (Downstream-Tasks) adaptierbar. Aufgrund der Emergenz basieren viele Anwendungen auf wenigen großen FMs, wie z.B. *GPT*, *BERT* und einigen anderen. Diese Homogenisierung führt dazu, dass auf einem FM basierende Anwendungen sowohl die Fähigkeiten als auch die Stärken und Schwächen des zentralen Modells mit übernehmen (Bommasani et al., 2021).

Zu den bekanntesten FMs gehören LLMs, also Modelle, deren primärer Ein- und Ausgabekanal natürliche Sprache ist. So können mit riesigen Textmengen trainierte LLMs unter anderem dafür genutzt werden Texte zusammenzufassen, Programmiercode zu generieren, oder Texte zu übersetzen. Diese Downstream-Tasks werden oft mithilfe eines Chatbots, als Schnittstelle für natürliche Spracheingaben angewiesen. Es gibt auch andere Arten von FMs für verschiedene Ziele und Anwendungen. Bekannte KI-basierte Anwendungen, die einem FM zugrunde liegen, sind die Text-zu-Bild-Modelle *DALL-E* oder *Stable Diffusion*, welche mit umfangreichen Bilddatensätzen trainiert wurden, sowie das Text-zu-Video-Modell *Sora*, dem ebenfalls eine große Menge an Videodaten zugrunde liegt. Darauf aufbauend gibt es multimodale FMs, die sowohl ein LLM für Texte nutzen als auch auf Bild und Videomodelle Zugriff haben, um Aufgaben zu lösen, die sowohl Sprache als auch andere Medien erfordern.

2.1.2 Transformer und NLP

Damit das alles möglich ist, waren mehrere zentrale Entwicklungen erforderlich. Da das Trainieren von Neuronalen Netzen extrem rechenaufwendig ist, wird starke Hardware benötigt, insbesondere starke GPUs, die in den letzten Jahren um ein vielfaches leistungsfähiger geworden sind (Bommasani et al., 2021). Zudem sind große Datenmengen an z.B. Text, Bildern und Videos etc. erforderlich. Ein weiterer zentraler Baustein der FMs ist außerdem die sogenannte Transformer-Architektur, deren Kernprinzip es ist, Textsequenzen parallel zu verarbeiten und die Worte im Kontext zu betrachten, anstatt wie in älteren Ansätzen Wort für Wort nacheinander zu verarbeiten. Diese Architektur eignet sich besonders gut für Aufgaben des Natural Language Processing (NLP). Alte Modelle konnten keine langen Wortsequenzen parallel im Kontext betrachten durch Transformer ist es möglich, die gesamte Textsequenz mittels Self-Attention und Positionskodierung parallel zu verarbeiten und so mit genug Daten zu trainieren, um ein auf Sprache und Worten basierendes FM zu erstellen. Self-Attention berechnet für jedes Wort im Text, wie relevant alle anderen Wörter im selben füreinander sind. So ist es möglich, dass das Modell Abhängigkeiten innerhalb einer Textsequenz erfassen kann, auch wenn die Wörter weit voneinander entfernt stehen. Aufgrund der parallelen Verarbeitung der Textbestandteile enthalten diese von sich aus noch keine Informationen über die Wortreihenfolge. Um die Positionierung innerhalb des Textes dennoch zu berücksichtigen, wird jedem Wort eine Positionsinformation zugewiesen, die sogenannte Positionskodierung, welche es dem Modell deutlich einfacher macht, Kontexte innerhalb des Textes zuzuordnen (Vaswani et al., 2017).

2.1.3 Begrenztes Sprachverständnis (Stochastischer Papagei)

LLMs sind probabilistische Textgeneratoren. Sie berechnen, welches das wahrscheinlichste nächste Wort (Token) ist, haben aber kein Verständnis für die semantische Bedeutung des Textes. Sie sind in der Lage, Natural Language Processing zu betreiben, nicht Natural Language Understanding (NLU). Bender et al. (2021) prägen den Begriff des Stochastic Parrot (Stochastischer Papagei), um auszudrücken, dass LLMs Sprache nur flüssig nachahmen, ohne kontextuelles Wissen oder die Fähigkeit, die Richtigkeit ihrer eigenen Trainingsdaten zu bewerten. Die Datenbasis ist unkuratiert oder nur automatisiert gefiltert, wodurch keine Garantie für faktische Korrektheit besteht, stattdessen können Vorurteile, Falschaussagen oder toxische Inhalte reproduziert werden. Dadurch, dass die Ausgaben sprachlich kohärent sind, besteht die Gefahr, dass Nutzer sie als kompetent und objektiv wahrnehmen und nicht mehr hinterfragen. Der Chatbot auf Basis eines LLM wird als kompetentes, allwissendes Gegenüber wahrgenommen, obwohl Halluzinationen (plausibel klingende, aber faktisch falsche oder erfundene Inhalte) oder Bias enthalten sein können. Dieses Problem wird allgemein als *Automation Bias* bezeichnet (Bender et al., 2021). Die Probleme, Fehler, Falschaussagen oder Bias in den Trainingsdaten des LLM können auf alle darauf basierenden Anwendungen übertragen werden. Wenn also in Claude 3.7 Sonnet Falschinformationen über Photovoltaik und dazugehörige Formeln oder Berechnungen enthalten sind, können diese von Figma Make, das auf diesem Modell basiert, geerbt werden. Auch im Kontext KI-generierter Systeme oder Systemteile, wie dem PVA-Rechner, ist dies eine wichtige Erkenntnis, insbesondere, wenn Nutzer nicht wissen, dass die Anwendungen und die im Hintergrund befindlichen Berechnungen auf der Basis eines LLMs entstanden sind. Die Berechnungen, Formeln und Daten werden möglicherweise als vollständig korrekt wahrgenommen und nicht mehr hinterfragt, da die Benutzeroberfläche professionell wirkt und die Erklärungen einzelner Bausteine plausibel klingen. Dies kann ohne Wissen über die Herkunft des Produkts zu Entscheidungen für oder gegen eine Investition führen und empfindliche finanzielle Fehlentscheidungen verursachen (Bommasani et al., 2021).

2.1.4 Blackbox-Problem

Trotz ihrer Leistungsfähigkeit besteht in der Forschung noch kein vollständiges Verständnis darüber, wie FMs im Detail funktionieren, was genau ihre Fähigkeiten sind, wo sie scheitern und wo ihre Grenzen liegen (Bommasani et al., 2021). Ein zentrales Problem ist dabei nicht nur das fehlende Verständnis der LLMs für die

Richtigkeit der Trainingsdaten (Bender et al., 2021) und ihrer eigenen Ausgaben, sondern auch die fehlende Transparenz des Generierungsprozesses. Es bleibt oft unklar, warum sich ein LLM bei einer Eingabe A für eine spezifische Antwort B entscheidet (Guidotti et al., 2018). Auf dieselbe Anfrage folgt nicht immer dieselbe Antwort (Bommasani et al., 2021), und es ist nicht nachvollziehbar, warum sich das Modell gerade jetzt für diese Ausgabe entschieden hat (Zednik, 2019; Guidotti et al., 2018).

Diese Thematik wird auch als Blackbox-Problem oder Opazität bezeichnet und thematisiert jene Intransparenz, wie eine Ausgabe entstanden ist. Nicht nur Nutzer haben Schwierigkeiten nachzuvollziehen, was während des Entscheidungsprozesses eines Modells passiert, oft auch die Entwickler selbst. FMs agieren nicht deterministisch und geben aufgrund der stochastischen Generierung von Token bei identischen Anfragen unterschiedliche Antworten. Diese fehlende Reproduzierbarkeit von Ausgaben schränkt deren Überprüfbarkeit ein, was zudem die klare Zuschreibung von Verantwortung erschwert, da weder Nutzer noch Entwickler eindeutig beurteilen können, auf welcher Grundlage eine Empfehlung oder Berechnung zustande kam (Zednik, 2019). Viele FMs haben emergente, unvorhergesehene Eigenschaften und Fähigkeiten, also Eigenschaften, für die das Modell nicht explizit trainiert wurde, und sie tun dies auf unvorhersehbare Weise. Das zeigt einerseits, wie leistungsfähig FMs sein können, andererseits hebt es hervor, dass jedes Modell Risiken mit sich bringt, die schwer einschätzbar oder vorhersehbar sind, insbesondere da bei immer größer werdenden Modellgrößen immer mehr emergente Eigenschaften auftreten können (Bommasani et al., 2021). Das ist vor Allem bei kritischen Entscheidungsprozessen von Bedeutung. Egal, ob es sich z.B. um wichtige unternehmerische oder medizinische Entscheidungen handelt oder um eine kostenintensive Privatinvestition, wie eine PV-Anlage. Wenn nicht nachvollziehbar ist, wie ein System eine Empfehlung oder Berechnung durchgeführt hat, dann ist es kritisch, eine Entscheidung auf Basis einer solchen Berechnung zu fällen (Bommasani et al., 2021).

Die Eigenschaften des Blackbox-Problems beeinflussen, wie Nutzer Vertrauen in KI-basierte Anwendungen aufbauen. Fehlt die Nachvollziehbarkeit der Funktionsweise und Entscheidungslogik, können Ergebnisse nicht rational überprüft werden. Nutzer sind daher gezwungen, ihre Bewertung auf subjektive Vertrauensannahmen zu stützen, etwa in Bezug auf die Kompetenz oder Professionalität des Systems.

2.1.5 Vibecoding und Agenten

Trotz der zuvor beschriebenen Problematiken wie Halluzinationen, Intransparenz und mangelnder Determiniertheit der Ausgaben stellt die Generierung von Programmiercode heutzutage einen der relevantesten Downstream-Tasks generativer KI dar. LLMs sind nicht nur in der Lage, Code zu analysieren und auf syntaktische Fehler zu überprüfen, sondern auch eigenständig komplexe Codesequenzen in diversen Programmiersprachen zu generieren. Infolgedessen haben viele Softwareentwickler, Ingenieure und Architekten LLM-basierte Modelle wie Claude, ChatGPT oder Gemini als festen Bestandteil in ihren Arbeitsalltag integriert. Diese stark von generativer KI gestützte Form der Softwareentwicklung wird in der aktuellen Debatte mehrheitlich als Vibecoding bezeichnet. Wie bereits in der Einleitung anhand der Definition von (Andrej Karpathy [@karpathy, 2025]) dargelegt, verschiebt sich der Fokus bei diesem Paradigma von der manuellen Code-Eingabe hin zur Beschreibung der gewünschten Funktionalität in natürlicher Sprache. Das LLM generiert den Programmcode autonom, während die Entwickler lediglich per Prompt iterieren, ohne sich primär mit Syntax oder manuellem Programmieren aufzuhalten (Theobald & Brandt, 2025). In der Praxis wird Vibecoding zunehmend in Zusammenarbeit mit KI-Agenten durchgeführt, die direkt in integrierte Entwicklungsumgebungen eingebunden sind. Ein KI-Agent ist ein autonomes System, das eigenständig spezifische und komplexe Aufgaben lösen kann. In der Softwareentwicklung unterstützen diese auf LLMs basierenden Agenten beim Planen, Schreiben, Orchestrieren, Reviewen und Korrigieren von Programmieraufgaben. Im Gegensatz zu herkömmlichen KI-Chatbots verfügen Agenten über ein kontextuelles Gedächtnis und sind in der Lage, die aktuelle Codebasis ganzheitlich zu betrachten, Code eigenständig auszuführen sowie externe Tools zu nutzen (Fraunhofer-IPA, n.d.). Die zentralen Vorteile des Vibecodings liegen in der drastisch gesenkten Einstiegshürde, die es auch Personen ohne tiefgreifende Programmierkenntnisse ermöglicht, Anwendungen zu erzeugen. Ideen und Prototypen lassen sich in kürzester Zeit umsetzen, ohne dass sich der Anwender im Detail mit Programmiercode, Schnittstellen oder Syntax auseinandersetzen muss. Die Möglichkeit zur schnellen, iterativen Weiterentwicklung der Ergebnisse erweist sich insbesondere in frühen Entwicklungsphasen als äußerst wertvoll, um die technische Machbarkeit zu überprüfen (Theobald & Brandt, 2025).

Dennoch bringt Vibecoding substantielle Risiken mit sich, die auf fundamentalen Problemen von Foundation Models basieren, wie der Problematik des stochastischen Papageien oder dem Blackbox-Charakter. Ähnlich wie bei generierter Sprache kann auch Programmcode halluziniert werden, was zu plausibel wirkenden, aber zum Teil funktionslosen oder nicht funktionierenden Codesequenzen

führt. Folglich entsteht beim Vibecoding oft Code mit geringer Wartbarkeit und Erweiterbarkeit (Theobald & Brandt, 2025). Aufgrund der hohen sprachlichen Qualität der Ausgaben erliegen Nutzer zudem leicht dem Automation Bias, sie vertrauen der KI ohne die Logik zu durchdringen.

Da die Modelle meist auf öffentlichen Datensätzen trainiert wurden, reproduzieren sie zudem mitunter bekannte Sicherheitslücken. Die Homogenisierung vieler KI-Agenten auf Basis derselben wenigen Modelle führt dazu, dass generierte Programme identische Schwachstellen erben und weitaus anfälliger für Angriffe werden. Der Verzicht auf systematische Tests und Dokumentation verringert die Nachvollziehbarkeit beim Vibecoding zusätzlich. Dennoch bietet der Ansatz eine mächtige Grundlage, um unkritische Funktionen oder Prototypen zeiteffizient zu erzeugen, vorausgesetzt, der generierte Code wird kritisch validiert und angepasst (Theobald & Brandt, 2025).

2.1.6 Figma Make

Das Paradigma des Vibecodings findet inzwischen auch in kommerziellen LLM-basierten Systemen Anwendung. Ein prominentes Beispiel hierfür ist das im Mai 2025 veröffentlichte Prompt-to-Code-Tool Figma Make (Figma, 2025). Das zugrundeliegende Softwareunternehmen Figma hat sich in den vergangenen Jahren als eine führende kollaborative Plattform für UI/UX-Design etabliert. Ursprünglich ermöglichte die Webanwendung lediglich die Erstellung statischer Design-Mockups und Prototypen mittels eines visuellen Editors. Dies erlaubte es Designern, App- und Web-Konzepte für Stakeholder aufzubereiten oder Usability-Tests durchzuführen, ohne dass bereits die finale Code-Logik implementiert werden musste. Mit Figma Make erweitert das Unternehmen sein Portfolio um eine generative Anwendung, die maßgeblich auf dem Sprachmodell Claude 3.7 Sonnet von *Anthropic* basiert. Der Nutzer formuliert seine Anforderungen in natürlicher Sprache. Zusätzlich zur reinen Texteingabe können bestehende Mock-ups, Screenshots oder Design-Referenzen als Kontext in den Prompt integriert werden (Figma, 2025). Das Tool transformiert diese Eingaben autonom in einen nutzbaren, mit grundlegender Anwendungslogik versehenen Prototyp. Durch iteratives Prompting kann dieser initiale Entwurf sukzessive verbessert und präzisiert werden, sodass im Ergebnis lauffähige Web-Applikationen entstehen. Der generierte Code wird in einem integrierten Editor zur Verfügung gestellt und lässt sich nahtlos an Versionsverwaltungssysteme wie *GitHub* anbinden. Der für die vorliegende empirische Untersuchung entwickelte Prototyp eines PVA-Rechners ist vollständig durch Prompting mittels Figma Make entstanden.

2.2 Nutzervertrauen

Da der durch Figma Make generierte Code im fertigen Produkt unsichtbar ist, bleibt dem Endnutzer die KI-Genese der Anwendung in aller Regel verborgen. Er nutzt das Tool unter der Prämisse, es sei ein klassisch programmiertes System. Die Offenlegung ist deshalb eine so interessante Variable, weil sie dieses Nichtwissen beendet und den Nutzer über die stochastische Herkunft des Rechners aufklärt. Mit diesem Wissen steigt das wahrgenommene Risiko einer Fehlfunktion (Schilke & Reimann, 2025). Die Bereitschaft, das Tool trotz dieses Wissens und der mangelnden Kontrollmöglichkeit zu nutzen, lässt sich am besten durch das Konstrukt des Nutzervertrauens beschreiben.

Eine der anerkanntesten Definitionen von Vertrauen, die häufig in der Forschung über Nutzervertrauen sowie Mensch-Computer-Interaktion Verwendung findet, stammt von Mayer et al. (1995) aus ihrer Arbeit *An Integrative Model of Organizational Trust* über Vertrauen in Organisationsstrukturen:

„the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party“ (Mayer et al., 1995, S.712).

Es werden drei Kernfaktoren von Vertrauen genannt: *Ability* (Fähigkeit/Kompetenz), *Benevolence* (Wohlwollen) und *Integrity* (Integrität). Diese werden folgendermaßen beschrieben:

Ability: „The group of skills, competencies, and characteristics that enable a party to have influence within some specific domain“ (Mayer et al., 1995, S.717).

Benevolence: „The extent to which a trustee is believed to want to do good to the trustor, aside from an egocentric profit motive“ (Mayer et al., 1995, S.718).

Integrity: „The extent to which the trustee adheres to a set of principles the trustor finds acceptable“ (Mayer et al., 1995, S.719).

Diese Definition wurde ursprünglich für Menschen in Organisationsstrukturen entwickelt, ist aber auch in HCI weit verbreitet und dient als Grundlage für unterschiedliche weiterführende oder spezifizierende Vertrauensdefinitionen, um die Beziehung und das Vertrauen in technische Systeme zu beschreiben. Da technischen Systemen jedoch menschliche Eigenschaften wie moralische Integrität oder

eigenständiges Wohlwollen fehlen, wurde diese Definition von Lee und See (2004) auf den Kontext der Automation adaptiert. Sie übersetzen die menschlichen Faktoren in technische Äquivalente, um die Nuancen der Mensch-Maschine-Beziehung präziser abzubilden.

2.2.1 Nutzervertrauen in Automation

Lee & See definieren Vertrauen wie folgt:

„Trust can be defined as the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability“ (Lee & See, 2004, S.51).

Diese Definition passt gut zu unserem Anwendungsfall. Der Nutzer, der daran interessiert ist, eine PV-Anlage auf seinem Dach zu installieren und Angebote vorliegen hat, übergibt Teile der Entscheidung dem Rechner. Er stellt dem Rechner nur die Parameter des Angebots und die Gegebenheiten seines Dachs zur Verfügung. Dieser berechnet auf Basis unterschiedlicher Algorithmen, zu welchem Zeitpunkt die Investition refinanziert sein würde.

Der Nutzer hat kaum Kontrolle darüber, wie und mit welchen Algorithmen dieser Rechner arbeitet. Er erhält lediglich ein Ergebnis und muss diesem glauben schenken oder selbst umfangreiche Recherchen betreiben, um das Ergebnis auf Korrektheit zu überprüfen.

Lee & Sees Arbeit *Trust in Automation: Designing for Appropriate Reliance* nimmt sich verschiedener vorheriger Arbeiten an, die unterschiedliche, aber sehr ähnliche Vertrauensdefinitionen und Teilaspekte behandeln. Diese werden diskutiert und auf automatisierte Prozesse und Systeme adaptiert. Sie führen angemessene Nutzung (Appropriate Reliance) als zentrales Designziel ein, bei dem weder Übervertrauen (Overtrust) herrscht, sodass Systeme sowie deren Fähigkeiten überschätzt werden, noch Misstrauen (Distrust), bei dem das Vertrauen in ein System geringer ist, als die tatsächlichen Fähigkeiten es hergeben. Das Gegenteil zur Appropriate Reliance ist die unangemessene Nutzung (Unappropriate Reliance). Das Ziel ist kalibriertes Vertrauen (Calibrated Trust), das den tatsächlichen Fähigkeiten eines Systems entspricht.

Lee & See entwickeln das PPP-Framework (Performance-Process-Purpose), das Mayers ABI-Dimensionen auf technische Systeme überträgt, um zu vermeiden, dass Nutzer sich durch Overtrust auf falsche Berechnungen verlassen und z.B.

eine finanzielle Fehlentscheidung bei der PV-Anlage treffen oder durch Distrust den korrekten Rechner ablehnen (Lee & See, 2004).

Ability wird zu *Performance* und folgendermaßen definiert:

„performance refers to the competency or expertise as demonstrated by its ability to achieve the operator’s goals [...] The operator will tend to trust automation that performs in a manner that reliably achieves his or her goals (Lee & See, 2004, S.59).“

Es beschreibt *was* die Automation tut, also die Leistungsfähigkeit eines technischen Systems das erwünschte Ziel zu erreichen. Nutzer tendieren dazu, Automation zu vertrauen, die zuverlässig das gewünschte Ziel erreicht.

Im Kontext unseres PVA-Rechners wäre *Performance* gegeben, wenn Nutzer dem System zutrauen, präzise und fachlich korrekte Amortisationsrechnungen durchzuführen.

Benevolence entspricht *Purpose* und wird wie folgt definiert:

„Purpose refers to the degree to which the automation is being used within the realm of the designer’s intent [...] The purpose basis of trust reflects the attribution of these characteristics to the automation. Frequently, whether or not this attribution takes place will depend on whether the designer’s intent has been communicated to the operator. If so, the operator will tend to trust automation to achieve the goals it was designed to achieve (Lee & See, 2004, S.59-60).“

Es beschreibt, *warum* eine Anwendung entwickelt wurde, also den Sinn und Zweck bzw. die Aufgabe, die eine Anwendung erfüllen soll.

Im Zusammenhang mit unserem PVA-Rechner wäre dieser Sinn das Berechnen des Amortisationszeitpunkts oder auch den Break-Even einer PV-Anlage. Also den Zeitpunkt, wann eine PV-Anlage mit bestimmten Parametern, unter gegebenen Voraussetzungen der Immobilie und zu einem festgelegten Preis ihre Kosten wieder eingenommen hat. Der Rechner soll in diesem Fall den Nutzer dabei unterstützen die bestmögliche finanzielle Entscheidung zu treffen. Dafür muss der Entwickler der diesen Rechner gebaut hat, *Benevolence* erfüllen und dem System einen *Purpose*, also einen Zweck induzieren, der den Interessen des Nutzers dient. Er muss tatsächlich daran interessiert sein, dass der Nutzer die beste Entscheidung treffen kann und keine Eigeninteressen verfolgen.

Integrity wird mit *Process* gleichgestellt und wie folgt beschrieben:

„Process is the degree to which the automation’s algorithms are appropriate for the situation and able to achieve the operator’s goals [...] The operator will tend to trust the automation if its algorithms can be understood and seem capable of achieving the operator’s goals in the current situation (Lee & See, 2004, S.59)“.

Prozessvertrauen ist gegeben, wenn die Berechnungen sowohl transparent als auch verlässlich sind, die Parameter faktisch geprüft wurden und keinen einseitigen Bias aufweisen. Des Weiteren müssen die Algorithmen nachvollziehbar und für den Nutzer verständlich sein. Wenn diese Bedingungen von Process erfüllt sind, tendieren Nutzer dazu, dem System zu vertrauen.

Diese Anwendung der Definitionen von Mayer et al. (1995) und der Weiterentwicklung durch Lee und See (2004) ist besonders passend für generative KI-Systeme wie Figma Make. Solche Systeme gelten als Black Box, da ihre internen Entscheidungsprozesse für Nutzer opak sind. Diese Intransparenz schränkt die Kontroll- und Überwachungsmöglichkeiten stark ein und erhöht die Vulnerabilität, insbesondere bei finanziellen Entscheidungen (Lee & See, 2004).

Die Dimension der Undurchsichtigkeit ist auch dann bereits gegeben, wenn ein solcher Rechner verwendet wird, ohne zu berücksichtigen, dass dieser durch KI erstellt wurde, da auch bei von Menschen konstruierten Anwendungen Fehler oder eine bewusste Täuschung nicht auszuschließen sind. Wenn die Konstruktion jedoch mehrheitlich durch generative KI entstanden ist, entstehen laut Bommasani et al. (2021) weitere Dimensionen der Unsicherheit durch die Abhängigkeit von vorgelagerten Modellen.

Es existiert nicht nur menschlicher Bias oder Fehler, sondern auch maschineller. Das Vertrauensmodell muss daher auf mehreren Ebenen angewendet werden:

1. Der Anbieter (Mensch/Firma): Diesem muss Kompetenz, Wohlwollen und Integrität zugesprochen werden (Mayer et al., 1995). Zudem muss darauf vertraut werden, dass dieser bei möglichen Halluzinationen und Fehlern der KI diese auch korrigiert.
2. Die Anwendung (Figma Make): Hier muss auf die korrekte Umsetzung der Algorithmen vertraut werden.
3. Das Foundation Model: Die zugrundeliegende KI, im Fall von Figma Make das Modell Claude 3.7 Sonnet (Figma, 2025), muss eine verlässliche Performance aufweisen. Außerdem ist es möglich, dass falsche Trainingsdaten einen Einfluss auf die generierte Anwendung haben und z.B. falsche Algorithmen oder Fakten halluziniert werden.

Dies ergibt drei Ebenen, in denen Fehler, Bias oder bewusste Täuschungen Einfluss auf das Endergebnis nehmen können.

2.2.2 Nutzervertrauen in Mensch-Computer-Interaktion (HCI)

Während klassische Modelle wie das von Mayer et al. (1995) den Fokus auf zwischenmenschliches Vertrauen legen und die Weiterentwicklung von Lee und See (2004) Mayers Modell für Automation im allgemeinen adaptiert hat, ist für den vorliegenden Untersuchungsgegenstand eine Präzisierung des Vertrauensnehmers notwendig.

Da der Nutzer bei der Verwendung des PVA-Rechners nicht mit einer realen Person kommuniziert, sondern direkt auf eine Softwareanwendung reagiert, verschiebt sich die Vertrauensfrage, insbesondere auch, weil das System vollständig durch Vibecoding und generative KI entstanden ist. Gemäß Gulati et al. (2024) ist es in solchen Szenarien essenziell, das Konzept des *Trust in Technical Artifacts* anzuwenden. Das bedeutet, dass nicht die Intentionen eines menschlichen Gegenübers bewertet werden, sondern die Eigenschaften des technischen Artefakts selbst im Zentrum der Vertrauensbildung stehen. Die Technologie ist hier nicht Vermittler, sondern das direkte Objekt des Vertrauens.

Da das technische Artefakt für den Nutzer primär durch seine Benutzeroberfläche (User Interface/UI) erfahrbar wird, manifestiert sich die Vertrauensbildung maßgeblich im Design. Die UI ist somit das zentrale Kommunikationsmedium, über das Vertrauen beim Nutzer aufgebaut werden muss. Bei der erstmaligen Verwendung des KI-generierten PVA-Rechners ist es Nutzern nicht sofort möglich die technische Korrektheit der Berechnungen zu überprüfen. Um dieser Problematik zu begegnen, wird zwischen der tatsächlichen (Actual Trustworthiness) und der wahrgenommenen Vertrauenswürdigkeit (Perceived Trustworthiness) eines Systems (Schlicker & Langer, 2021) differenziert. Während die Actual Trustworthiness objektive Eigenschaften wie algorithmische Validität und fehlerfreie Funktionalität beschreibt, bleiben diese für den Anwender oft opak. Mangels direkter Überprüfbarkeit dieser objektiven Kriterien stützen sich Nutzende stattdessen auf sichtbare Hinweisreize (Cues) im User Interface, um die Vertrauenswürdigkeit subjektiv zu bewerten (Schlicker & Langer, 2021). Das Resultat dieses kognitiven Bewertungsprozesses bildet die Perceived Trustworthiness. Aus diesem Grund ist das initiale Vertrauen (Initial Trust) entscheidend für die Nutzungsbereitschaft.

Zieglmeier und Lehene (2021) legen in ihrer Arbeit *Designing Trustworthy User Interfaces* dar, dass das Design der Benutzeroberfläche maßgeblichen Einfluss auf Initial

Trust hat. Sie überführen die theoretischen Dimensionen vom PPP-Framework (Lee & See, 2004) in eine konkrete Taxonomie für das Interface Design. Dies macht es möglich die abstrakten Vertrauensaspekte direkt auf die Gestaltung des PVA-Rechners anzuwenden. Das PPP-Framework erhält eine konkrete, auf UX und UI angepasste Adaption mit Faktoren, die jede einzelne Dimension ausmachen. Außerdem unterscheiden sie zwei Varianten von Vertrauen: emotionales Vertrauen (Affect-Based Trust), das durch die Wahrnehmung der Ästhetik, sowie die subjektive Schönheit beeinflusst wird und sich auf die Gefühle der Nutzer bezieht. Kognitives Vertrauen (Cognitive Trust) basiert dagegen auf der Wahrnehmung der Systemzuverlässigkeit durch den Nutzer und steht im Zusammenhang mit logischem Denken. Allgemein scheint Affect-Based Trust einen größeren Einfluss auf das Gesamtvertrauen zu haben als das kognitive Vertrauen, wird jedoch auch durch das Niveau des kognitiven Vertrauens beeinflusst (Zieglmeier & Lehene, 2021).

1. Purpose in UI ist unterteilt in die Aspekte Benevolence (Wohlwollen), Perceived Security (wahrgenommene Sicherheit) und Credibility (Glaubwürdigkeit).

Die Purpose-Dimension hängt vom durch die Entwickler intendierten Nutzen des Systems ab. Ein professionelles Erscheinungsbild (Credibility) und eine hilfsbereite Ansprache (Benevolence) tragen dazu bei, dem Nutzer positive Absichten zu signalisieren (Zieglmeier & Lehene, 2021).

2. Process in UI ist aufgeteilt in die Aspekte Integrity (Integrität), ähnlich wie im ABI-Modell von Mayer et al. (1995), Predictability (Vorhersehbarkeit), Familiarity (Vertrautheit), Transparency (Transparenz), Communication (Kommunikation), Usability (Nutzbarkeit), Personalization (Personalisierung).

Die Process-Dimension beschreibt ähnlich wie bei Lee und See (2004), *wie* ein System funktioniert und wie passend das Systemdesign für den intendierten Zweck ist. Dieses *wie* kann natürlich nicht immer in allen Details für jeden Nutzer erklärt werden, sodass für die meisten Leute ein solcher PVA-Rechner einen Blackbox-Charakter hat. Dennoch kann diese Dimension anteilig abgebildet werden. Auch wenn hier nicht der Code offengelegt wird, ist es möglich die einzelnen Bestandteile und Faktoren der Rechnung zu erläutern. In unserem Fall kann das durch Erklärungen der einzelnen Parameter erreicht werden:

- Was bedeuten die Parameter?
- Welchen Zweck haben die Parameter
- Wie kann man herausfinden, welche Parameter das eigene Haus oder das Angebot hat, das dem Nutzer vorliegt?

Der Gefahr einer möglichen kognitiven Überlastung (Cognitive Load) bei der Erklärung von Systemfunktionen, insbesondere wenn der Nutzer noch kein Fachwissen zum Zweck des Rechners hat, wie bei dem PV-Rechner, kann durch schrittweise Offenlegung (Progressive Disclosure) entgegengewirkt werden (Springer & Whittaker, 2018). Die kontextspezifische Einblendung von Details optimiert die Usability und bedient somit direkt die Process-Dimension des Vertrauens. Diese offene und schrittweise Kommunikation gibt den Nutzern die Chance mehr vom Prozess zu verstehen, fördert Transparenz, Vorhersehbarkeit und verhindert Überforderung, weil der Nutzer weniger Recherche betreiben muss, um zu verstehen, was die wichtigsten Faktoren beim Erwerb einer PV-Anlage sind. Das fördert die Bereitschaft der Nutzer, dem System zu vertrauen. Außerdem ist ein konsistentes, vertrautes Design von Vorteil, das anderen Systemen ähnelt, die ähnliche Ziele verfolgen, um möglichst vielen Nutzern einen leichten Einstieg in das System zu ermöglichen. Des Weiteren sollte das System eine gute Usability aufweisen und die Nutzer nicht überfordern. Das steigert die Bereitschaft, das System zu nutzen (Zieglmeier & Lehene, 2021).

3. Performance in UI zeichnet sich durch die Faktoren Competence (Kompetenz), Reliability (Zuverlässigkeit) und Validity (Richtigkeit) aus.

Ähnlich wie bei Lee und See (2004) wird hier bewertet, wie gut oder kompetent das System den intendierten Zweck umsetzt. Also hauptsächlich die Richtigkeit der Ergebnisse. Es ist natürlich nicht immer möglich, die Performance im System darzustellen, da der Nutzer in unserem Anwendungsfall oft nicht das nötige Wissen besitzt, um die Validität des Ergebnisses vollumfänglich beurteilen zu können. Dies kann teilweise durch Interface Design adressiert werden, sowie durch klare, transparente Kommunikation der Ergebnisse, aber natürlich auch durch Systemstabilität und Systemgeschwindigkeit in der Ausführung. Ein professionelles, ästhetisches Auftreten der Webseite wird oft als Ersatzindikator empfunden. Es zählt auf Affect-Based Trust ein, welches in Ermangelung von Prüfmöglichkeiten Cognitive Trust stützt (Zieglmeier & Lehene, 2021). Sie nutzen für ihre Evaluation Fragen, die auf Basis von Likert Skalen beantwortet und evaluiert werden. In dieser Studie wird dies ähnlich durchgeführt.

Es wird sowohl ein Trust in Automation (TiA)-Fragebogen genutzt, um die Vertrauensaspekte zu erfassen, als auch Interviews durchgeführt, die während und nach der Nutzung des Systems qualitative Fragen stellen, welche im Nachhinein durch die thematische Analyse nach Braun und Clarke (2006) analysiert und evaluiert werden.

2.2.3 Nutzervertrauen in KI

All diese Indikatoren für das Nutzervertrauen in UI sind relevant, um innerhalb der Studie zu untersuchen, wie Nutzer auf die KI-generierte Anwendung reagieren. Insbesondere wichtig ist zu betrachten, dass untersucht wird, wie sich die Reaktion und das Vertrauen verändern, wenn der Nutzer weiß, dass es sich um eine KI-generierte Anwendung handelt.

Aktuelle Studien beschreiben hier ein Transparenz-Paradoxon. Während Transparenz in der klassischen HCI-Forschung meist als vertrauensbildend gilt, zeigt sich bei generativer KI ein gegenteiliger Effekt. Hinweise darauf, dass Inhalte KI-generiert wurden, führen zu Skepsis und geringerem Vertrauen (Buder et al., 2024). Dies gilt nicht nur für Marketing-Inhalte, sondern, wie Schilke und Reimann (2025) nachweisen, auch für arbeitsbezogene Aufgaben und Entscheidungshilfen.

Diese Erkenntnis erzwingt eine Neubewertung dessen, was als Appropriate Reliance (Lee & See, 2004) gelten kann. Das Vertrauensobjekt ist hierbei weder ein rein deterministisches Werkzeug noch ein menschliches Gegenüber. Es handelt sich vielmehr um ein hybrides Konstrukt: Ein deterministisches System, das jedoch auf Basis menschlicher Eingaben durch einen stochastischen Prozess entstanden ist. Für den Nutzer entstehen dadurch mehrere Ebenen der Opazität. Ohne den expliziten Hinweis auf die KI-Genese geht der Nutzer jedoch fälschlicherweise von einer einzigen, deterministischen Blackbox aus, also der Berechnungslogik im Hintergrund. Die Offenlegung korrigiert diese Fehlannahme, aktiviert aber gleichzeitig das beschriebene Transparenz-Paradoxon.

Visser et al. (2023) diskutieren für die Bewertung von Nutzervertrauen im Zeitalter von KI einen weiteren Aspekt, zusätzlich zur Appropriate Reliance. Sie betrachten Distrust nicht mehr als bloße Abwesenheit von Vertrauen, sondern als eigene Dimension. Distrust charakterisiert sich durch Wachsamkeit, kritische Reflexion und Skepsis. Idealerweise empfindet der Nutzer nicht nur ein angemessenes Vertrauen (Warranted Trust) gegenüber Anwendungen, sondern auch ein angemessenes Misstrauen (Warranted Distrust). Insbesondere wenn die durch KI-Tools erzeugten Nutzeroberflächen eine hohe Qualität aufweisen, die hohen Affect-Based Trust erzeugen (Zieglmeier & Lehene, 2021), ist Warranted Distrust ein wichtiger Faktor.

Diese hochwertigen Interfaces generieren durch ihre professionelle Gestaltung starke positive Cues, die zu einer hohen Perceived Trustworthiness bei den Nutzenden führen. Wenn diese subjektiv wahrgenommene Vertrauenswürdigkeit jedoch die Actual Trustworthiness der fehleranfälligen, halluzinierenden KI-Logik im Hintergrund deutlich übersteigt, kommt es zu einer Fehleinschätzung (Schlicker

& Langer, 2021). Dies verletzt das Ziel der Appropriate Reliance und führt zu Overtrust. Die Nutzer schließen von der professionellen Form der UI fälschlicherweise auf eine korrekte Funktionalität der Berechnungslogik (Bender et al., 2021).

Wenn eine Anwendung hauptsächlich durch Vibecoding oder sogar gänzlich durch einen KI-Agenten erzeugt wurde, kann der Nutzer im Nachhinein nicht überprüfen, wie verantwortungsbewusst die Entwickler mit dem durch KI erzeugten Code umgegangen sind. Dazu kommt das Risiko von Sicherheitslücken in solchem Code (Theobald & Brandt, 2025). Problematisch ist hierbei die Zielgruppe: Figma Make richtet sein Marketing explizit an Nutzer, die ohne komplexen Code interaktive Prototypen erstellen wollen (l. Figma, 2025a). Auch wenn es geboten ist, den Code und die Ergebnisse der Prototypen zu validieren, fehlt oft die Expertise dazu. Werden solche Anwendungen oder Web Apps öffentlich angeboten, verschiebt sich die Verantwortung auf die Endanwender der erzeugten Apps. Da diese das Tool gerade nutzen, weil sie selbst keine Experten für Photovoltaikanlagen sind, entsteht ein hohes Risiko für Unappropriate Reliance (Lee & See, 2004).

Ein Ansatz, um diesen potenziell entstehenden Overtrust zu minimieren, besteht darin, den Nutzern durch einen Hinweis mitzuteilen, dass die Anwendung größtenteils oder vollständig mithilfe von KI entwickelt wurde. Dies ermöglicht Warranted Distrust (Visser et al., 2023) und ist somit nicht zwangsläufig negativ. Aus Anwenderperspektive ist dieser Vertrauensverlust notwendig, um eine angemessene Risikoeinschätzung vorzunehmen.

3 Implementierung und Architekturanalyse des PVA-Rechners

3.1 PV-Amortisationsrechner im Überblick

Als Test-Anwendung für die Untersuchung dient ein funktionaler Prototyp eines Photovoltaik-Amortisationsrechners. Die Anwendung ist für beide Versuchsgruppen identisch. Der Rechner wurde durch das KI-Tool Figma Make, auf Basis von zunächst sieben iterativen Prompts (siehe Anhang: 7) erzeugt. Das von Figma Make genutzte Modell ist in diesem Fall Claude 3.7 Sonnet. Die verwendeten Eingabefelder und Kennzahlen wurden mithilfe anderer existierender PVA-Rechner, wie dem PVGIS-Rechner (PVGIS24, [n. d.](#)) oder dem Amortisationsrechner von [rechner-photovoltaik.de](#) ([n. d.](#)), recherchiert, identifiziert, ausgewählt und im Startprompt mitgegeben.

Außerdem wurde Figma Make angewiesen, generell benutzerfreundlich und explizit mit Progressive Disclosure der Eingabefelder zu arbeiten, um eine moderne, benutzerfreundliche Oberfläche zu erhalten, die den Cognitive Load gering hält (Springer & Whittaker, 2018). Zudem wurde die Möglichkeit zur Konfiguration mehrerer Dachseiten bewusst im Prompt als Anforderung formuliert. Zusätzlich wurde explizit gefordert, dass kurze, sofort sichtbare Infotexte und etwas längere, durch einen Klick auf einen Info-Button erreichbare Erklärungen eingebaut werden, um größeren Recherchebedarf zu vermeiden, Transparenz zu schaffen und erste Nachfragen zu den Kennzahlen direkt abzufangen. Dies adressiert die Process-Dimension des PPP-Modells von Lee und See (2004). Außerdem sollten auf der Ergebnisseite Erzeugungswerte nach Monat angezeigt werden.

In den ersten Nutzertests wurden die dunkle Farbgebung sowie die von dem KI-Tool designten Logos kritisiert. Auch wurde die fehlende Möglichkeit, den Berechnungsprozess im Verlauf neu zu starten, angemerkt. Dies wurde mit sechs weiteren Prompts angepasst. Die Logos wurden ausgetauscht, die Neustartfunktion während des Prozesses ermöglicht und ein helleres, freundlicheres sowie zu Photovoltaik optisch passendes Gesamtbild gewählt.

Da Figma Make auch von Anwendern ohne Programmier- oder Fachexpertise genutzt werden könnte, um vollwertige Anwendungen generieren zu lassen, wurden sowohl die UI als auch die Berechnungslogik sowie die Textinformationen komplett vom KI-Tool erstellt. Der komplette textliche und algorithmische Inhalt des Rechners ist nicht auf Validität und vollwertige Funktionalität überprüft worden, da das Ziel die Untersuchung des Nutzervertrauens im Kontext der KI-Genese ist, insbesondere wenn kein Experte die Qualität des Designs, des Inhalts und der Technik überprüft hat. Es wird folglich nicht die Actual Trustworthiness evaluiert, sondern ausschließlich die durch Cues im Interface subjektive Perceived Trustworthiness (Schlicker & Langer, 2021) der Anwendenden anhand der Dimensionen des PPP-Modells untersucht.

Die inhaltliche Struktur des Rechners orientiert sich an etablierten, auf dem Markt existierenden PVA-Anwendungen ([Link zum Rechner](#)). Die Dateneingabe erfolgt schrittweise über sieben aufeinanderfolgende Interaktionskarten:

1. **Standort:** Standort der Immobilie (siehe Abbildung 3.1)
2. **Photovoltaiksystem:** Konfiguration der Module im Kontext der Dachseiten des Hauses (siehe Abbildung 3.2)
3. **Nutzungsdauer:** Geplante Nutzungsdauer und Leistungsabnahme pro Jahr
4. **Kosten:** Investitionskosten- und Betriebskosten und mögliche Inflationsrate
5. **Einspeisevergütung:** Informationen über die Einspeisevergütung
6. **Stromverbrauch:** Energiebedarf und Verbrauchsdaten des Haushalts
7. **Finanzierung:** Finanzierungsinformationen, wie Kreditbetrag, Laufzeit und Zinssatz.

Nach Abschluss der Eingabe präsentiert das System eine detaillierte Ergebnisansicht. Diese gibt die Berechnungsergebnisse aus und visualisiert sie durch interaktive Diagramme (siehe Abbildung 3.3). Zu den ausgegebenen Metriken zählen unter anderem die jährliche Stromproduktion, der Eigenverbrauchsanteil, die Netzeinspeisung, der jährliche Ertrag, die Amortisationszeit, die Gesamtertritte über die Nutzungsdauer (ROI), der Gesamtgewinn sowie der Gesamtertrag. Die visuelle Aufbereitung umfasst Diagramme zur monatlichen Stromproduktion, dem Verhältnis von Eigenverbrauch zu Einspeisung sowie der Produktions- und Ertragsentwicklung über zehn Jahre.

PV-Amortisationsrechner
Wirtschaftlichkeitsanalyse für Photovoltaik-Anlagen

Schritt 1 von 7
14% abgeschlossen

Standort

Wo wird die Anlage installiert?

Der Standort Ihrer PV-Anlage ist entscheidend für die Menge an Sonneneinstrahlung, die Sie erhalten. Deutschland hat unterschiedliche Sonneneinstrahlungswerte je nach Region.

Standort (PLZ oder Stadt)

Südliche Regionen erhalten tendenziell mehr Sonnenstunden pro Jahr.

Zurück
Weiter

Abbildung 3.1: Startseite und Standorteingabe

PV-Amortisationsrechner
Wirtschaftlichkeitsanalyse für Photovoltaik-Anlagen

Schritt 2 von 7
Neu starten
29% abgeschlossen

Photovoltaiksystem

Technische Details Ihrer Anlage

Konfigurieren Sie alle Dachseiten, auf denen PV-Module installiert werden sollen. Jede Dachseite kann unterschiedliche Ausrichtungen, Neigungen und Modultypen haben.

Neue Dachseite hinzufügen

Name / Bezeichnung

Montageart

Aufdach-Montage ist die häufigste und kostengünstigste Variante.

Dachneigung (Grad)
Dachausrichtung

Optimal: 30-35° Süd ist ideal

Modultyp

Monokristalline Module sind Standard und bieten den besten Wirkungsgrad.

Installierte Leistung (kWp)
Systemverluste (%)

Pro kWp: ca. 5-7 m² Typisch: 10-20%

Abbrechen
Dachseite abschließen

Zurück
Weiter

Abbildung 3.2: Dachseiteneingabe

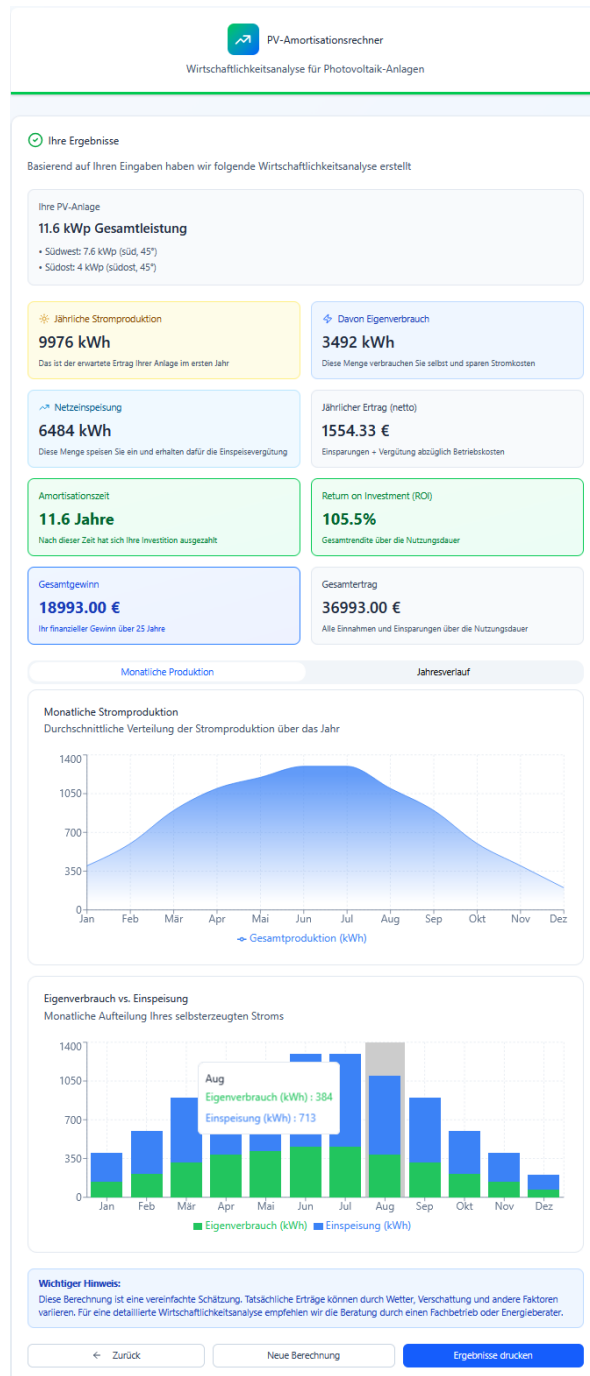


Abbildung 3.3: Ergebnissseite

3.2 Architekturanalyse

Im vorherigen Unterkapitel haben wir uns der UI der Anwendung gewidmet, nun kommen wir zur Codebasis. Wenngleich das Tool damit beworben wird, dass es zum Prototypenbau verwendet werden kann (I. Figma, 2025a), besteht in der Praxis das Risiko, dass Anwender die generierten Ergebnisse als Grundlage für produktive Systeme missverstehen oder sogar als vollwertige Anwendungen einsetzen, insbesondere wenn die visuelle Qualität über die technische Beschaffenheit hinwegtäuscht (Zieglmeier & Lehene, 2021). Gerade in diesem Spannungsfeld ist eine fundierte Auseinandersetzung mit der gelieferten Software-Qualität und Architektur unabdingbar.

Laut Theobald und Brandt (2025) vom Fraunhofer Institut ist KI-generierter Code oft unstrukturiert, redundant, schwer wartbar, enthält häufig Sicherheitslücken oder unsichere Implementierungen, die z.B. SQL-Injection oder Cross-Site-Scripting erlauben. Außerdem gibt es oft keine oder nur wenige Unit Tests.

Vor diesem Hintergrund werden wir einige architektonische Eigenschaften des PVA-Rechners analysieren und mit Vorwürfen gegenüber KI-generiertem Code vergleichen. Zunächst werden die technischen Grundlagen des Prototypen dargelegt. Die Anwendung basiert auf einer komponentenbasierten React-Architektur und ist in TypeScript implementiert. Für das Styling und die visuelle Gestaltung der Benutzeroberfläche kommt Tailwind CSS zum Einsatz. Es gibt eine *Attributions.md* Datei, in welcher darüber informiert wird, dass einige Komponenten und Fotos von <https://ui.shadcn.com/> und <https://unsplash.com> stammen. Es gibt einen UI-Ordner, in welchem diverse UI-Komponenten abgelegt sind, auf die das Programm zugreift. Nahezu die gesamte Applikationslogik ist in einer einzigen Hauptkomponente (*GuidedKpiCalculator.tsx*) organisiert. Neben dieser aktuellen Hauptkomponente enthält die Codebasis zudem eine obsoletere Version (*KPICalculator*), die aus einem früheren Iterationsschritt stammt, wie in der roten Umrandung in Abbildung 3.4 ersichtlich. Dieser tote Code wurde von der KI nicht bereinigt, was die Wartbarkeit der Codebasis belastet.

Bei der aktuellen Hauptkomponente handelt es sich um eine klassische Ausprägung des *God Object* Anti-Patterns mit über 1300 Zeilen, wie Abbildung 3.4 zeigt. Mit Ausnahme mancher React UI-Komponenten ist das komplette Programm hier organisiert. Die Komponente enthält mehrere Interfaces, Datentypen und eine riesige Export-Funktion. Sowohl Business-Logik, State-Management, die gesamte UI, die Navigationslogik und Interfaces sind enthalten. Dies führt zu einer hohen Kopplung bei gleichzeitig niedriger Kohäsion, eine architektonische Konstellation, die in der Softwaretechnik als besonders fehleranfällig gilt (Martin, 2003).

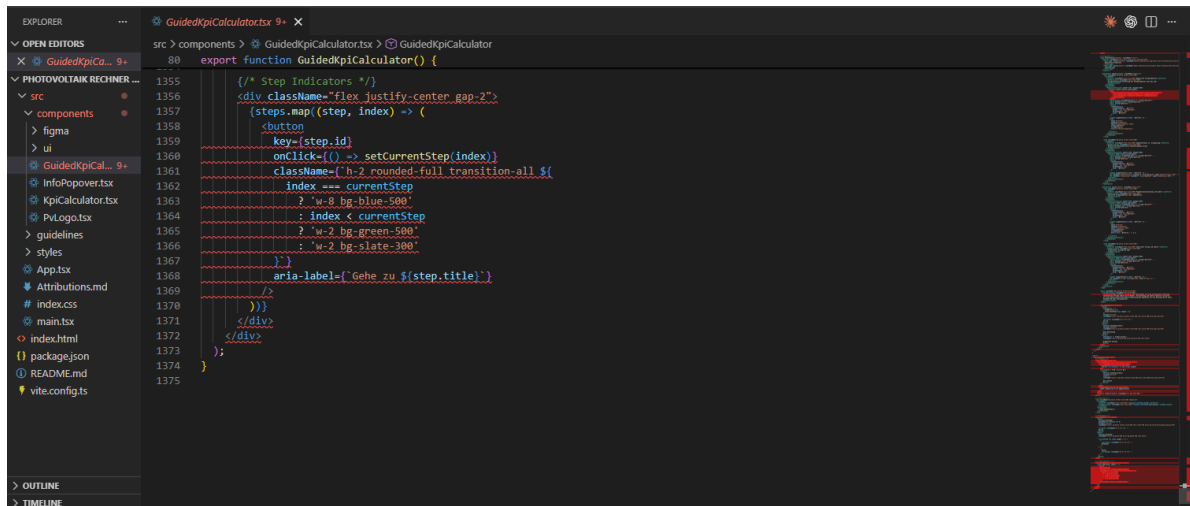


Abbildung 3.4: Totes Objekt und Gott-Komponente (Screenshot aus dem Code)

Das *Single Responsibility Prinzip*, welches besagt, dass jede Einheit, Klasse, Komponente oder Modul genau eine Verantwortlichkeit tragen soll (Martin, 2003), wird massiv verletzt, da sich die Komponente aus mehreren Gründen ändern könnte, z.B. die Darstellung und die Berechnungslogik, siehe Codeblock 3.1. Sowohl UI-Logik als auch Berechnungslogik finden in GuidedKpiCalculator.tsx statt. Dies erschwert sowohl die Verständlichkeit als auch die Wartbarkeit des Codes erheblich.

Zudem wird das Prinzip *Separation of Concerns* verletzt. Dieses appelliert daran, dass ein System unterschiedliche Belange, wie z.B. Business Logik, UI-Logik oder Datenstrukturen voneinander trennen sollte (Dijkstra, 1982). Dies ist ebenfalls kaum erfüllt, siehe Codeblock 3.1. Lediglich einige UI-Komponenten sind außerhalb organisiert. Die komplette Berechnungslogik findet in der Gott-Komponente statt, genauso wie die Interfaces, welche die Dachseiten und die KPIs zur Berechnung enthalten.

Listing 3.1: Exemplarischer Codeausschnitt: Single Responsibility und Separations of Concern

```

1 // UI-Logik (Rendering eines Formularschritts)
2 case 0:
3   return (
4     <Input
5       value={formData.standort}
6       onChange={e => handleInputChange('standort', e.target.value)}
7     />

```

```

8   );
9
10  // Berechnungslogik – direkt in derselben Komponente
11  const calculateResults = () => {
12    const jahresproduktion = formData.dachseiten.reduce((sum, dach) => {
13      const leistung = parseFloat(dach.installierteLeistung) || 0;
14      return sum + (leistung * 1000 * (1 - verluste / 100));
15    }, 0);
16
17    setResults({ ... });
18  };

```

Auch das Prinzip *Dont Repeat Yourself* - DRY wird kaum beachtet, welches besagt, dass jede Wissensseinheit innerhalb eines Systems eine einzige, eindeutige Repräsentation haben muss (Hunt & Thomas, 2011). So werden beispielsweise Eingabefelder redundant definiert, anstatt diese durch Bausteine wiederverwendbar zu machen. Dies erhöht die Fehleranfälligkeit und den Wartungs und Änderungsaufwand bei zukünftigen Änderungen. In Codeblock 3.2 zeigt sich eine mehrfache Wiederholung identischer UI-Strukturen innerhalb eines Schritts. Hier wäre eine Kapselung in wiederverwendbare Komponenten möglich, da sich dadurch Wartbarkeit und Konsistenz der Eingabefelder verbessert.

Listing 3.2: Exemplarischer Ausschnitt aus dem Finanzierungs-Schritt für DRY

```

1  case 6:
2    return (
3      <>
4        <Label htmlFor="fremdkapital">Kreditbetrag / Fremdkapital €(</Label>
5        <InfoPopover
6          title="Finanzierung einer PV-Anlage"
7          content="Bei Barzahlung geben Sie 0 ein."
8        />
9        <Input id="fremdkapital" type="number" value={formData.fremdkapital} />
10
11        <Label htmlFor="zinssatz">Zinssatz (</Label>
12        <InfoPopover
13          title="Zinssätze für PV-Kredite"
14          content="KfW-Kredite liegen oft bei etwa 2-4 %."
15        />
16        <Input id="zinssatz" type="number" step="0.1" value={formData.zinssatz} />
17      </>
18    );

```

Sichere, moderne und leicht zu wartende Software folgt dem sogenannten *Fail Fast-Prinzip*, um Fehlerzustände sofort sichtbar zu machen und die Ursachenforschung zu erleichtern. Dies erfordert, dass Datentypen und Logik durch Assertions und Typisierung direkt bei der Anwendung geprüft werden. Fehleingaben der Nutzer sollten durch Validierung und Error-Handler abgefangen werden (Shore, 2004). Auch dieses Prinzip wird nur unzureichend angewendet, siehe Codeblock 3.3. Der Code verletzt das Fail Fast-Prinzip, weil ungültige oder leere Eingaben nicht frühzeitig abgefangen werden. Statt eine Fehlermeldung auszugeben oder die Berechnung abzubrechen, werden problematische Werte durch `|| 0` bzw. `|| 14` stillschweigend ersetzt. Dadurch läuft die Berechnung weiter, obwohl die Eingabedaten potenziell unplausibel sind, und es kann ein fachlich falsches Ergebnis entstehen. Stattdessen sollte vor der Berechnung geprüft werden, ob die Eingaben gültige numerische Werte enthalten. Ist dies nicht der Fall, sollte die Anwendung den Fehler direkt im betroffenen Feld anzeigen und die Nutzer informieren. Dieses Anti-Pattern bezeichnet Shore (2004) als *Fail Slow*.

Eine weitere Konvention, die Figma Make bei der Genese kaum beachtet hat, sind die sogenannten *Magic Numbers*, also hartcodierte Zahlenwerte ohne erklärenden Kontext wie z.B. eine Konstante (Fowler, 2018). Im vorliegenden Code (Codeblock 3.3) treten diese in zwei Ausprägungen auf, welche beide beim Fallbackwert `|| 14` ersichtlich sind. Einerseits ist die 14 der Standardwert für die Systemverluste, (Codeblock 3.3) andererseits ein stillschweigender Fallback-Wert, wie auch `|| 0`, der ungültige Eingaben ohne Fehlermeldung ersetzt.

Listing 3.3: Fail-Fast Verletzung und Magic Number

```
1 const jahresproduktion = formData.dachseiten.reduce((sum, dach) => {  
2   const leistung = parseFloat(dach.installierteLeistung) || 0;  
3   const verluste = parseFloat(dach.systemverluste) || 14;  
4   return sum + (leistung * 1000 * (1 - verluste / 100));  
5 }, 0);
```

Zusammenfassend bestätigt die Code-Analyse die Eingangs zitierte Warnung von Theobald und Brandt (2025). Das generierte Artefakt weist typische Merkmale von Vibecoding auf. Das User Interface und die komplette äußere Erscheinung wirken professionell, während der Blick in den Code strukturelle Mängel offenbart. Der Code missachtet gängige Clean Code-Standards: Magic Numbers existieren und die Anwendung des Fail Fast-Prinzips wird vernachlässigt. Der Großteil der Web-App ist eine monolithische Gott-Komponente mit hoher Kopplung und niedriger Kohäsion. Essenzielle Prinzipien wie DRY oder die SOLID-Regeln finden durch die riesige Komponente kaum Verwendung, was die Wartbarkeit, Erweiterbarkeit und

Testbarkeit stark einschränkt. Eine Integration in bestehende produktive Systeme wäre ohne tiefgreifendes Refactoring kaum möglich. Kritisch ist diese Diskrepanz zwischen visueller Professionalität und technischer Fragilität im Hinblick auf das Nutzervertrauen. Da die technische Beschaffenheit für Anwender opak ist, erhöht sich das Risiko für Overtrust, da Nutzer durch die Qualität der Oberfläche Rückschlüsse auf die Berechnungslogik ziehen könnten (Zieglmeier & Lehene, 2021). Dies betrifft auch die fachliche Korrektheit der im Code verwendeten Konstanten und KPIs. Weil der Fokus auf der softwaretechnischen und nutzerzentrierten Analyse liegt, wurde die fachliche Korrektheit der Berechnungsformeln im Rahmen dieser Arbeit nicht explizit verifiziert. Die Code-Analyse zeigt jedoch, dass eine solche Verifikation durch die monolithische Struktur erschwert wird, was insbesondere für Laien, an die sich Figma Make richtet, noch schwieriger ist. Da Foundation Models zu Halluzinationen neigen (Bommasani et al., 2021), wie dem Erfinden falscher Berechnungsmethoden oder Fakten, muss in Frage gestellt werden, ob die von Figma Make generierten Ergebnisse der Realität entsprechen.

In den nächsten Kapiteln der Arbeit wird untersucht, wie sich dieses Spannungsfeld auf die Nutzer auswirkt. Dazu werden im Folgenden der Studienaufbau, die Analyse und die Auswertungsmethoden dargelegt.

4 Studiendesign und Methodik

4.1 Zielsetzung und Aufbau

Ziel der vorliegenden Studie ist es, die Wahrnehmung und das Nutzervertrauen der Studienteilnehmer gegenüber dem PVA-Rechner zu erfassen und zu untersuchen. Die Ergebnisse sollen Erkenntnisse darüber liefern, wie stark das Wissen über die KI-Genese einer Anwendung das Nutzervertrauen in die Funktionsweise und die Ergebnisse beeinflusst. Es soll herausgearbeitet werden, wie ein fehlendes KI-Label die kritische Betrachtung von Nutzern beeinflusst und somit zu Overtrust führen kann, oder ob die Information über die KI-Herkunft den Nutzer in Richtung Distrust prägt.

Die Studienteilnehmer werden in zwei Gruppen aufgeteilt. Gruppe 1 erhält vorab die Information, dass es sich bei dem PV-Rechner um eine vollständig KI-generierte Webanwendung handelt. Gruppe 2 wird diese Information vorenthalten. Ihnen wird lediglich mündlich mitgeteilt, dass im Rahmen der Bachelorarbeit der PV-Rechner evaluiert wird.

Die Gruppen bestehen jeweils aus vier Teilnehmenden. Aufgrund des qualitativen Charakters der Studie und mangelnder Ressourcen wurde die Gesamtstichprobengröße auf $N=8$ begrenzt. Die Interviews werden sowohl mit Personen vor Ort, als auch über Microsoft Teams standardisiert durchgeführt. Nach Einholung des Einverständnisses zur Aufnahme, Transkription und anonymisierten Weiterverarbeitung der Daten, wird eine Aufnahme gestartet.

Nach einer standardisierten Instruktion erhalten beide Gruppen eine Szenariobeschreibung mit Aufgaben. Diese umfasst die Aufgabenstellung, einen Instruktionsbogen (7) mit den spezifischen Daten für den PV-Rechner, um die Aufgaben zu lösen, sowie die Links zur Umfrage und zum PVA-Rechner. Die feste Vorgabe der Eingabedaten dient dazu, die kognitive Belastung der Probanden gezielt zu steuern und den Fokus auf die Interaktion mit dem Rechner zu lenken. Der Umfragebogen besteht aus einem Teil vor der Aufgabebearbeitung. Dort werden demografische

Daten erhoben und drei Fragen zur allgemeinen technischen Affinität gestellt, um die qualitativen Ergebnisse im Nachhinein kontextualisiert zu betrachten.

Die Teilnehmenden sollen während der Nutzung und des Lösens der Aufgabe nach dem Prinzip der Thinking-Aloud-Methode (Ericsson & Simon, 1984) ihre Gedankengänge und Entscheidungen laut mitteilen, um qualitativ begründete Aussagen zu erhalten. Der Versuchsleiter hält sich während der Aufgabe im Hintergrund, steht aber für technische Fragen und bei erheblichen Schwierigkeiten zur Verfügung.

Die Aufgabe enthält eine kleine Schwierigkeit, um die Aufmerksamkeit bei der Durchführung zu erhöhen, da moderat herausfordernde Aufgaben das Engagement steigern können (Sayali et al., 2023). Die Anwender müssen die Kilowattpeak-Leistung der Anlage anhand der Modulmenge multipliziert mit der Leistung in Watt pro Modul selbst berechnen. Dies wird mit dem eigenen Smartphone-Taschenrechner oder dem PC-eigenen Taschenrechner durchgeführt. Außerdem werden drei Fragen zu den Ergebnissen des Rechners gestellt, die im Fragebogen zu beantworten sind.

Im Anschluss an die Bearbeitung werden beide Gruppen dazu aufgefordert, einen standardisierten Trust in Automation-Fragebogen (Jian et al., 2000) auszufüllen, um eine quantitative Flankierung der Aussagen zu ermöglichen. Um weitere qualitative Aussagen zu erheben, werden die Probanden dazu aufgefordert Thinking-Aloud auch während der Beantwortung des Fragebogens anzuwenden. Die 7-stufige Likert-Skala des originalen TiA-Fragebogens wurde zu einer 5-stufigen Likert-Skala transformiert. Diese Modifikation stellt eine methodische Limitation hinsichtlich der Konstruktvalidität dar, wurde jedoch zugunsten der Einheitlichkeit der Erhebung und zur Minimierung der kognitiven Belastung bewusst in Kauf genommen.

Danach kommt es zu einem kurzen, leitfadenorientierten Interview, dessen Aufbau mithilfe des PPP-Modells von Lee und See (2004) entstanden ist. Dazu gibt es zwei Fragen, die sowohl begründet als auch im Rahmen einer Single-Item-Messung beantwortet werden. Es werden Fragen gestellt, die Bezug nehmen auf die drei Dimensionen Process, Performance und Purpose. So sollen unterschiedliche Aspekte des Nutzervertrauens qualitativ erfasst werden. Im Laufe des Interviews wird Gruppe 2 darauf hingewiesen, dass es sich um eine KI-generierte Anwendung handelt. Die Probanden werden mit einer Single-Item-Messung auf einer 5-stufigen Likert-Skala darum gebeten, folgendes Statement zu beantworten: *Der Einfluss von KI bei der Erstellung des PV-Rechners beeinflusst mein Vertrauen in das System positiv mit einer Antwortvarianz von Stimme überhaupt nicht zu bis stimme voll und ganz zu.* Gruppe 1 wird ebenfalls erneut auf die KI-Genese der Anwendung hingewiesen und aufgefordert, das gleiche Single-Item zu beantworten wie Gruppe 2.

Zum Abschluss beantworten die Teilnehmenden drei Fragen zu ihrer generellen KI-Affinität, losgelöst von der vorher bearbeiteten Aufgabe. Außerdem wird um eine kurze Einschätzung gebeten, wie vertrauenswürdig die Ergebnisse von Figma Make empfunden werden, um weitere Einordnungsmöglichkeiten zu erhalten.

4.2 Ablauf der Erhebung

1. **Begrüßung und Aufklärung:** Zu Beginn erfolgte die Begrüßung sowie die Aufklärung über den Zweck der Studie. Außerdem wurde das Einverständnis für Aufnahme, Transkription und Analyse der Daten, im Fragebogen eingeholt. Gruppe 1 wird über die KI-Genese des Rechners, mithilfe von Figma Make, aufgeklärt.
2. **Aufgabenübergabe und Instruktion:** Den Teilnehmenden wurde das Arbeitsblatt (7) übergeben. Zudem erhielten sie eine Einführung in die Aufgabenstellung und in die Thinking-Aloud-Methode (Ericsson & Simon, 1984). Sie wurden instruiert, während der gesamten Bearbeitung ihre Gedankengänge laut zu verbalisieren und im späteren Fragebogen Begründungen für ihre jeweiligen Entscheidungen zu formulieren.
3. **Vorumfrage:** Vor der Nutzung des Rechners wurden soziodemografische Daten sowie die allgemeine technische Affinität der Probanden erhoben.
4. **Interaktion:** Die Versuchspersonen interagierten mit dem PVA-Rechner, bearbeiteten die vorgegebene Aufgabe und beantworteten die drei inhaltlichen Fragen. Hierbei wandten sie kontinuierlich die Thinking-Aloud-Methode an.
5. **Anschlussumfrage und Interview:** Im Anschluss füllten die Teilnehmenden den standardisierten Trust in Automation-Fragebogen (7) aus (Jian et al., 2000). Die abgegebenen Bewertungen wurden von den Probanden gleichzeitig qualitativ begründet. Daraufhin wurden die vier qualitativen Anschlussfragen gestellt.
6. **Offenlegung oder Erinnerung an KI-Genese:** Gruppe 1 wurde noch einmal an die KI-Genese des Rechners erinnert und gebeten, den Einfluss dieses Wissens auf das eigene Vertrauen in den Rechner auf einer 5-stufigen Likert-Skala quantitativ einzuschätzen und qualitativ zu begründen. Gruppe 2 wurde in diesem Schritt darüber aufgeklärt, dass der Rechner mit dem KI-Tool Figma Make generiert wurde. Auch diese Gruppe wurde anschließend

gebeten, den Einfluss dieser neuen Information auf ihr Vertrauen auf einer 5-stufigen Likert-Skala einzuschätzen und zu begründen.

7. **Abschluss:** Abschließend beantworteten alle Teilnehmenden Fragen zu ihrer generellen KI-Affinität. Zudem bewerteten sie das Statement: *Mein Vertrauen in die Ergebnisse von Figma Make ist hoch* im spezifischen Kontext des PVA-Rechners sowie eigener möglicher Erfahrungen mit Figma Make. Danach erfolgte die Verabschiedung.

4.3 Stichprobe

Bedingt durch die explorative und qualitative Ausrichtung der Studie wurde eine Gelegenheitsstichprobe von $N = 8$ Teilnehmenden aus dem direkten und akademischen Umfeld der Versuchsleitung herangezogen. Die Versuchspersonen wurden paritätisch in zwei Gruppen unterteilt: Während Gruppe 1 explizit über die KI-Genese des Systems aufgeklärt wurde, fand bei Gruppe 2 keine initiale Offenlegung statt. Die ausschließlich männliche Stichprobe umfasste sechs Personen (75 %) zwischen 18 und 29 Jahren sowie zwei Personen (25 %) zwischen 30 und 39 Jahren. Hinsichtlich des Bildungsabschlusses gaben drei Teilnehmende einen Bachelor- und einer einen Masterabschluss an. Die übrigen Probanden verfügten über einen Abschluss als staatlich geprüfter Techniker (Maschinenbau), eine Berufsausbildung, das Abitur sowie eine Kombination aus Abitur und Berufsausbildung. Die Berufsfelder erstreckten sich primär über die Informatik (drei Personen), gefolgt von jeweils einer Person aus der Baubranche, dem E-Commerce, dem Marketing, der Touristik und der Unternehmensberatung. Fünf der Probanden gaben an, zur Miete in einer Wohnung zu leben, während drei weitere bei ihren Eltern wohnten. Bei der Informationssuche für kostenintensive technische Anschaffungen verließen sich die Teilnehmenden (Mehrfachnennungen waren möglich) vorwiegend auf Online-Vergleichsportale (5), gefolgt von Videos (4) und Herstellerwebseiten (4) sowie Empfehlungen von Freunden (3) und Online-Foren (3). KI-Chatbots (2), Social Media (1) und Fachzeitschriften (1) spielten eine untergeordnete Rolle, während Zeitungsartikel oder Online-Medien sowie der Fachhandel gar nicht genannt wurden. Die Auswertung der prä-experimentellen Befragung (gemessen auf 5-stufigen Likert-Skalen von 1 = *stimme überhaupt nicht zu* bis 5 = *stimme vollumfänglich zu*) zeigte ein moderates bis hohes grundsätzliches Interesse an der Anschaffung einer Photovoltaikanlage ($Mdn = 4,0$). Konkret positionierten sich hierbei drei Personen *neutral*, zwei stimmten *eher zu* und drei *vollumfänglich zu*. Hinsichtlich der allgemeinen Technikaaffinität zeigte sich ein homogenes, sehr

positives Bild: Alle acht Teilnehmenden (100 %) gaben an, neuen technologischen Entwicklungen generell positiv gegenüberzustehen ($Mdn = 5,0$). Auch die Einarbeitung in neue Software fiel der Mehrheit sehr leicht ($Mdn = 4,5$; vier *vollumfänglich*, drei *eher zu*, einer *neutral*). Die Freude am Ausprobieren neuer technischer Tools war ebenfalls stark ausgeprägt ($Mdn = 4,0$; drei *vollumfänglich*, vier *eher zu*, einer *neutral*). Die nachgelagerte Erhebung der KI-Affinität ergab, dass die Probanden im Allgemeinen stark mit KI-Anwendungen wie ChatGPT vertraut sind ($Mdn = 4,5$; vier *vollumfänglich*, vier *eher zu*). Bemerkenswert ist jedoch das hohe Maß an Skepsis gegenüber der Zuverlässigkeit dieser Tools. Der Aussage, KI-Tools würden Aufgaben generell fehlerfrei ausführen, stimmte kein einziger Proband zu ($Mdn = 1,5$; vier *überhaupt nicht*, einer *eher nicht*, drei *neutral*). Dennoch fiel die regelmäßige Nutzung von KI-Tools im Arbeits- oder Studienalltag hoch aus ($Mdn = 4,0$). Während zwei Probanden einer regelmäßigen Nutzung eher nicht zustimmten, gaben drei Personen an, eher zuzustimmen, und drei weitere stimmten *vollumfänglich* zu. Die Verteilung der soziodemografischen Merkmale sowie der Technik- und KI-Affinität erwies sich über die beiden Gruppen hinweg als weitgehend homogen, sodass eine Verzerrung der qualitativen Ergebnisse durch systematische Gruppenunterschiede weitestgehend ausgeschlossen werden konnte.

4.4 Erhebungsinstrumente

4.4.1 Quantitative Instrumente

Für die quantitative Flankierung zur Erfassung des Vertrauens wurde der Trust in Automation Fragebogen von Jian et al. (2000) genutzt, angepasst auf eine 5-Punkt-Likert-Skala. Die Anpassung der Likert-Skala erfolgte, um die Fragen einheitlich zu gestalten und um die Probanden durch ständig wechselnde Skalen nicht zu ermüden.

Im Vorfeld der Erhebung wurden demografische Informationen, wie Alter, Bildungsabschluss, Berufsfeld und Wohnsituation erhoben, um gegebenenfalls die quantitativen und qualitativen Ergebnisse der Studie in den Kontext setzen zu können. Außerdem wurden Fragen zur allgemeinen technischen Affinität, dem grundsätzlichen Interesse an einer Photovoltaikanlage auf 5-Punkt-Likert-Skalen und den bevorzugten Informationsquellen für teurere technische Anschaffungen gestellt, um z.B. negative Einschätzungen gegenüber moderner Technik oder Photovoltaik mit einzubeziehen.

Der Trust in Automation-Fragebogen wurde nach der Bearbeitung der Aufgabe in Anwesenheit der Versuchsleitung durchgeführt. Im Anschluss an den Fragebogen wurden drei Fragen zur allgemeinen KI-Affinität auf einer 5-Punkt-Likert-Skala gestellt, ebenfalls zur Einordnung der Vertrauensbewertung. Zum Abschluss wurde das Vertrauen der Probanden in Figma Make auf Basis des PV-Rechners, als Produkt von Figma Make und möglicher eigener Erfahrungen mit Figma Make auf einer 5-Punkt-Likert-Skala gemessen.

4.4.2 Qualitative Instrumente

Die Probanden wurden vor der Bearbeitung darum gebeten ihre Gedankengänge und Entscheidungen mit der Thinking-Aloud-Methode (Ericsson & Simon, 1984) zu erläutern und zu begründen, um nachvollziehen zu können, auf welche Weise der PV-Rechner genutzt und wahrgenommen wurde.

Während der Bearbeitung des Trust in Automation-Fragebogens wurden die Probanden darum gebeten, ihre Einschätzungen und Antworten verbal zu begründen.

Im Anschluss an den Fragebogen wurden den Probanden vier weitere Fragen gestellt, die in Kombination mit dem PPP-Modell (Lee & See, 2004), bzw. dem ABI-Modell (Mayer et al., 1995) entwickelt wurden, um die drei Vertrauensdimensionen zusätzlich zu den Begründungen des Fragebogens und der Thinking-Aloud-Methode abdecken zu können.

Im weiteren Verlauf der Studie wurden die Probanden über die KI-Genese des Rechners aufgeklärt, bzw. daran erinnert und darum gebeten, die Gedanken, die das in ihnen auslöste, sowie den Einfluss dieser Informationen auf ihr Vertrauen in die Anwendung im Kontext zu verbalisieren und zu begründen.

Zum Abschluss wurde das Vertrauen der Probanden in Figma Make auf Basis des PV-Rechners, als Produkt von Figma Make und möglicher eigener Erfahrungen mit Figma Make erfragt.

4.5 Datenauswertungsstrategie: Thematische Analyse nach Braun und Clarke

Die Auswertung der transkribierten Interviews erfolgte in Anlehnung an die Prinzipien der reflexiven Thematischen Analyse nach Braun und Clarke (2006). Ein zentrales Merkmal der reflexiven TA ist der bewusste Umgang mit der eigenen Subjektivität. Dies spiegelt sich in dieser Arbeit wider, durch das technische und konzeptionelle Vorwissen des Autors, der zugleich als Entwickler des KI-generierten Prototyps agierte. Für die Analyse wurde ein hybrider Ansatz aus induktivem und deduktivem Vorgehen gewählt. Aufgrund des starken Bezugs dieser Arbeit auf die Vertrauensdefinition nach Mayer et al. (1995) sowie deren Weiterentwicklung für die Automatisierung im PPP-Modell nach Lee und See (2004) erfolgte die Codierung maßgeblich theoriegeleitet. Die Dimensionen Purpose, Process und Performance sowie das auf diesen Dimensionen aufbauende Ziel der Appropriate Reliance, bei dem weder Overtrust noch Distrust dominieren, dienten als strukturierendes Gerüst. Gleichzeitig blieb der Analyseprozess offen für induktive Themen, die sich iterativ aus den individuellen Nutzererfahrungen hinsichtlich der KI-Genese ergaben. Die praktische Umsetzung orientierte sich an den sechs Phasen nach Braun und Clarke (2006), wurde jedoch stark an die theoriegeleitete Ausrichtung der Untersuchung adaptiert:

1. **Datenvertrautheit:** Im ersten Schritt erfolgte die intensive Durchsicht des Datenmaterials. Die Transkripte wurden mehrfach gelesen und die Aufnahmen gehört, um ein ganzheitliches Verständnis für die Aussagen und die Reaktionen der Probanden auf den PV-Rechner zu entwickeln.
2. **Codierung:** Anstatt das gesamte Material frei zu kodieren, wurde gezielt nach aussagekräftigen Zitaten und Datenankern gesucht, die die Interaktion mit dem System, das Vertrauen oder die Skepsis der Nutzer widerspiegeln.
3. **Theme-Bildung:** Die Theme-Bildung erfolgte deduktiv durch die Übernahme etablierter theoretischer Konstrukte. Das PPP-Modell nach Lee und See (2004) sowie das Konzept der Appropriate Reliance dienten als vordefinierte thematische Orientierung.
4. **Überprüfung der Themes:** In diesem Schritt wurden die zuvor identifizierten Zitate der theoretischen Begriffe zugeordnet. Es wurde geprüft, ob ein Zitat am besten die Dimension Process, Performance, Purpose oder die Appropriate Reliance abbildet, um eine trennscharfe Auswertung zu gewährleisten.

- 5. Definition der Themes:** Da die Namen der Themes durch die Literatur bereits feststanden, beschränkte sich diese Phase darauf, die theoretischen Definitionen von Lee und See (2004) sowie den spezifischen Kontext der vorliegenden Studie zu übertragen.
- 6. Ausarbeitung der Ergebnisse:** Im letzten Schritt wurden die zugeordneten Zitate theoriegeleitet ausgewertet, um zu analysieren, wie sich das Wissen um die KI-Genese auf die verschiedenen Vertrauensdimensionen der Nutzer auswirkte

5 Ergebnisse

5.1 Quantitative Vertrauensmessung

5.1.1 Trust in Automation Fragebogen

Zur quantitativen Erfassung des Systemvertrauens (TiA-Ergebnisse: siehe Tabelle 5.1) wurde der Trust in Automation-Fragebogen nach Jian et al. (2000) herangezogen, welcher für die vorliegende Studie auf eine 5-stufige Likert-Skala adaptiert wurde. Gemäß der Auswertungslogik des Instruments wurden die fünf negativ formulierten Misstrauens-Items zunächst umgepolt und anschließend mit den verbleibenden Vertrauens-Items zu einem individuellen Gesamtscore pro Gruppe ($N = 4$) bei einer Gesamtstichprobe von $N = 8$ aggregiert. Ein hoher Score bedeutet ein durchschnittlich höheres Vertrauen als ein niedriger Score. Durch die geringe Stichprobe ist der grundsätzliche Aussagegehalt dieser quantitativen Flankierung beschränkt.

Der Vergleich der Gruppenmittelwerte (siehe Tabelle 5.2) zeigt keine deutliche Diskrepanz zwischen den einzelnen Gruppen. Während die über die KI-Genese aufgeklärte Gruppe 1 einen Gesamtscore von $M = 3,15$ ($SD = 1,00$) erreicht hat, liegt dieser bei Gruppe 2, welche nicht über die KI-Herkunft aufgeklärt wurde, lediglich bei $M = 2,94$ ($SD = 1,23$). Gruppe 1 hat entgegen möglicher Erwartungen, dass das KI-Label das Vertrauen in die Anwendung negativ beeinflusst (Schilke & Reimann, 2025), ein höheres durchschnittliches Vertrauen in das Programm als die unaufgeklärte Gruppe, dies kann aber auch maßgeblich durch die geringe Stichprobe beeinflusst worden sein.

Auf individueller Ebene zeigt sich bei Proband G1T2 ein stark abweichendes, minimales Vertrauensniveau von $Score = 1,92$. Die anderen drei Nutzer lagen bei $Score = 3,33$, $Score = 3,00$ und $Score = 4,33$.

Gruppe 2 hat mit G2T2 ($Score = 1,75$) und G2T3 ($Score = 2$) zwei Nutzer, die dem System minimal vertrauen. G2T2 wurde im Laufe der Nutzung durch den Fragebogen motiviert, mehrere Eingaben zu ändern, um das Berechnungsverhalten

zu überprüfen. Da der Rechner diese geänderte Angabe nicht in die Ergebnisberechnungen einbezog, führte dies zu einem starken Vertrauensverlust, was den Probanden dazu veranlasste, die vorherige positive Einschätzung zu revidieren und den Fragebogen zu überarbeiten.

Die anderen beiden Probanden haben mit einem Score von *Score* = 4,08 (G2T1) und *Score* = 3,92 (G2T4) relativ hohes Vertrauen in den Rechner.

Teilnehmer	Trg	Hin	Mis	Skp	Sch	Ver	Sic	Int	Vrl	Zuv	VK	Vtr	MW
G1T1	3	4	5	2	5	4	3	2	3	3	1	5	3,33
G1T2	2	2	3	1	2	2	2	1	2	2	2	2	1,92
G1T3	5	5	4	3	3	2	2	2	2	2	2	4	3,00
G1T4	5	5	3	4	5	4	4	5	4	4	4	5	4,33
G2T1	4	4	5	5	4	4	4	4	4	4	4	3	4,08
G2T2	1	3	1	1	2	1	1	3	1	1	1	5	1,75
G2T3	2	1	2	2	2	2	3	3	1	3	2	1	2,00
G2T4	3	5	3	5	1	3	5	5	5	5	2	5	3,92
MW p. A.	3,13	3,63	3,25	2,88	3,00	2,75	3,00	3,13	2,75	3,00	2,25	3,75	
SA p. A.	1,46	1,51	1,39	1,64	1,51	1,16	1,31	1,46	1,49	1,31	1,31	1,31	
MW G1 p. A.	3,75	4,00	3,75	2,50	3,75	3,00	2,75	2,50	2,75	2,75	2,25	4,00	
SA G1 p. A.	1,50	1,41	0,96	1,29	1,50	1,15	0,96	1,73	0,96	0,96	1,26	1,41	
MW G2 p. A.	2,50	3,25	2,75	3,25	2,25	2,50	3,25	3,75	2,75	3,25	2,25	3,50	
SA G2 p. A.	1,29	1,71	1,71	2,06	1,26	1,29	1,71	0,96	2,06	1,71	1,26	1,91	

Tabelle 5.1: Übersicht der Itemwerte sowie Mittelwerte und Standardabweichungen

Gruppe	MW	SA
G1	3,15	1,00
G2	2,94	1,23

Tabelle 5.2: Gruppenmittelwerte und Standardabweichungen

Legende: Trg = Trügerisch, Hin = Hinterlistig, Mis = Misstrauisch, Skp = Skeptisch, Sch = Schädlich, Ver = Vertrauenswürdig, Sic = Sicher, Int = Integer, Vrl = Verlässlich, Zuv = Zuverlässig, VK = Vertrauen können, Vtr = Vertraut. MW = Mittelwert, SA = Standardabweichung, p. A. = pro Aussage.

5.1.2 Eigene Single Items

Die Aussage: *Ich habe das Gefühl der Rechner ist bei der Nutzung auf meine individuellen Bedürfnisse, mein möglicherweise fehlendes Vorwissen und meine Ziele als Nutzer eingegangen* hat mit $M = 4,5$ ($SD = 1$) in Gruppe 1 und $M = 4,25$ ($SD = 0,96$) in Gruppe 2 über beide Gruppen hinweg hohe Zustimmungswerte (Item-Ergebnisse siehe Tabelle 5.3).

Die Aussage: *Der Einfluss von KI bei der Erstellung des PV-Rechners beeinflusst mein Vertrauen in das System positiv* wurde von beiden Gruppen sehr ähnlich bewertet. Gruppe 1 hat hier einen Mittelwert von $M = 2$ ($SD = 0,82$), Gruppe 2 liegt hier mit $M = 2,25$ ($SD = 1,26$) nicht weit entfernt.

Die abschließend bewertende Aussage: *Mein Vertrauen in die Ergebnisse von Figma Make ist hoch* wurde auch über beide Gruppen hinweg ähnlich beurteilt, Gruppe 1 kommt hier auf $M = 2,75$ ($SD = 0,96$) und Gruppe 2 auf $M = 3,25$ ($SD = 0,96$). Dies wurde von einigen Probanden direkt damit begründet, dass sie bisher noch keine Erfahrung mit Figma Make gemacht haben und lediglich den PV-Rechner als Ergebnis testen konnten.

Teilnehmer	Individuelle Bedürfnisse			KI-Einfluss auf Vertrauen			Vertrauen in Figma Make		
	Wert	MW	SA	Wert	MW	SA	Wert	MW	SA
G1T1	5			2			2		
G1T2	3			1			2		
G1T3	5			3			4		
G1T4	5	G1: 4,50	G1: 1,00	2	G1: 2,00	G1: 0,82	3	G1: 2,75	G1: 0,96
G2T1	5			4			4		
G2T2	4			2			2		
G2T3	3			1			3		
G2T4	5	G2: 4,25	G2: 0,96	2	G2: 2,25	G2: 1,26	4	G2: 3,25	G2: 0,96
MW	4,38			2,13			3,00		
SA	0,92			0,99			0,93		

Tabelle 5.3: Zusammengefasste Übersicht der drei Items mit Einzelwerten sowie Gruppenmittelwerten und Standardabweichungen

5.2 Qualitative Ergebnisse: Thematische Analyse

Zur Beantwortung der Forschungsfrage:

Wie beeinflusst die Offenlegung der KI-Genese einer Softwareanwendung die subjektive Wahrnehmung und das Vertrauen der Nutzer in das System?

wurden die Transkripte der Thinking-Aloud-Sitzungen sowie die anschließenden Interviews einer Thematischen Analyse nach Braun und Clarke (2006) unterzogen.

Ziel dieser Auswertung war es, zugrunde liegende Themes der Vertrauensbildung und -erosion zu identifizieren. Die Analyse brachte vier zentrale Themes hervor, die sich eng an den theoretischen Dimensionen des Vertrauens in automatisierte Systeme sowie an aktuellen Diskursen zur angemessenen KI-Skepsis orientieren. Diese vier Muster strukturieren das Erleben der Versuchsgruppen und werden in den folgenden Unterkapiteln anhand direkter Zitate detailliert dargelegt:

1. *Theme: Process - Die KI als undurchsichtige Blackbox und Stochastischer Papagei*
2. *Theme: Performance - Spannungsfeld zwischen Usability und Automation Bias*
3. *Theme: Purpose - Wahrgenommene Intention und Professionalität*
4. *Theme: Appropriate Reliance und Warranted Distrust - Die Ambivalenz der KI-Offenlegung*

5.2.1 Theme 1: Process - Die KI als undurchsichtige Blackbox und Stochastischer Papagei

Dieses Theme beschreibt, inwiefern die Probanden die internen Rechenwege der Anwendung nachvollziehen können und sich dadurch Prozessvertrauen bilden kann (Lee & See, 2004).

Vor Allem in Gruppe 1 zeigt sich ein starker Wunsch nach mehr Offenheit bei den Berechnungsprozessen und Algorithmen, was grundsätzlich bei mehreren Probanden zu einem gewissen Misstrauen geführt hat. Sie sind sich der Opazität des Systems bewusst und fordern Transparenz.

G1T1 (Score: 3.33) bringt das Problem auf den Punkt:

„[00:26:40] Skeptisch schon, also, ich würde auf jeden Fall gerne wissen wollen, so im Nachhinein, was für Formeln stecken dahinter, wie sind die dazugekommen? Weil gerade ist es eine riesen Blackbox und man weiß, dass das KI generiert ist. Denn, wenn ich jetzt wüsste, es hat sich ein Energieberater mal hingestellt und diese Formeln in sein Software-Tool gegossen, da würde man ihm wahrscheinlich mehr vertrauen, auch wenn der Energieberater auch Fehler machen kann“ (Interviewtranskript G1T1).

Dem Probanden sind die Blackbox-Charakteristiken von KI-Tools durchaus bewusst. Er spricht explizit an, dass er einem Menschen, z.B. einem Energieberater, der auch Fehler machen kann, vermutlich mehr Vertrauen schenken würde als einem KI-generierten Programm. Anhand dieser Aussage manifestiert sich der Kern des Blackbox-Problems: Die Intransparenz des PV-Rechners in Kombination mit dem Wissen um die KI-Genese verhindert, dass die Probanden den Prozess der Berechnung validieren können. Bemerkenswert ist hierbei die Asymmetrie in der Vertrauenszuschreibung. Während die Probanden bereit wären, die undurchsichtigen Rechenwege eines menschlichen Energieberaters aufgrund dessen angenommener professioneller Integrität zu akzeptieren, führt dieselbe Intransparenz bei der KI zu expliziter Skepsis. Das Fehlen einer erklärbaren algorithmischen Logik zwingt die aufgeklärte Gruppe 1 somit in eine Position der Unsicherheit, welche den Aufbau von Prozessvertrauen hemmt.

Diese Intransparenz des Berechnungsprozesses spiegelt sich auch bei Proband G1T2 wider, welcher anmerkt:

„[00:14:14] das Programm gibt dir einfach irgendwelche Werte, die du nicht überprüfen kannst, außer du rechnest es selber nach“ (Interviewtranskript G1T2).

Er klagt über fehlende Transparenz der Rechenwege, weswegen sein Vertrauen in den Rechner unter anderem in Gruppe 1 mit einem Gesamtscore von 1,92 am niedrigsten ist.

Besonders aufschlussreich ist auch die Meinung von Proband G1T4 (Score: 4.33). Initial schenkt er dem Rechner von allen Probanden über beide Gruppen hinweg am meisten Vertrauen, obwohl der Nutzer im Vorfeld über die KI-Genese aufgeklärt wurde. Auch im Laufe der Nutzung äußert der Nutzer kaum Zweifel an der Integrität des Rechners und gibt positive Rückmeldungen, insbesondere auch bezüglich einiger Usability Aspekte, wie den Tooltips, die dem Nutzer eine hohe Integrität suggerieren:

„[00:02:58] Das sind schon mal gute Infos, dass man überhaupt so für jemanden, der halt wenig Erfahrung mit Solaranlagen hat[...]ein bisschen ein Gefühl hat, wenn man so einen Rechner hat, worauf man achten soll. Also das finde ich schon gut und finde ich auch intuitiv mit den, Tooltips hier“ (Interviewtranskript G1T4).

Auch beim erneuten Hinweis auf die KI-Genese spricht er weiterhin positiv über den Einfluss von KI bei der Erstellung des Rechners:

„[00:24:49] Finde ich eigentlich völlig in Ordnung. Also, für mich gehts ja mehr darum, dass ich ein Tool mehr hab, was ich irgendwie schnell und zugänglich benutzen kann“ (Interviewtranskript G1T4).

Erst als G1T4 den Einfluss der KI auf die Erstellung des Rechners bewerten soll, beginnt er, Skepsis auszudrücken:

„[00:26:20] Ich würde eher sagen, ein bisschen weniger, weil man ja schon das Gefühl hat, man gibt das jetzt sag ich mal, so ein bisschen in die Hände von jemand Dummen... dann weiß man natürlich auch nicht, ob der PV-Rechner dann wirklich letztendlich ein durchdachtes Produkt ist und weil ich dann das Gefühl hätte, bei einem Menschen, der würde wirklich eher ein Produkt rausbringen, wenn er weiß, das ist jetzt da, wo ich es haben will und das ist ein abgeschlossenes Ganzes“ (Interviewtranskript G1T4).

Anhand dieser Diskrepanz wird deutlich, dass eine intuitiv gestaltete Benutzeroberfläche zwar temporär Vertrauen aufbauen kann, das bewusste Wissen um die KI-Genese jedoch einen nachträglichen, kritischen Reflexionsprozess auslösen kann. Der Proband spricht der KI die Fähigkeit ab, wie ein menschlicher Entwickler intendiert und zielgerichtet zu handeln, wodurch das Vertrauen in die Validität der Ergebnisse eingeschränkt wird.

Gruppe 2 betrachtet den Rechner im Gegensatz zu Gruppe 1 etwas ambivalenter. Es gab, ähnlich wie in Gruppe 1, Probanden, die von Beginn an den Wunsch äußerten, dass die Berechnungen offener kommuniziert werden, G2T2 sagt dazu folgendes:

„[00:26:06] was vielleicht noch interessant wäre, wäre halt vielleicht so ein bisschen die Rechnung nachvollziehen zu können, weil jetzt einfach mir die Werte angezeigt werden, die berechnet wurden. Aber irgendwie dahinter gehen zu können, okay, wir haben das so berechnet, dass man wirklich sieht, ah, okay, das ist der Prozess dahinter“ (Interviewtranskript G2T2).

Diese Aussage zeigt, dass eine fehlende Prozesstransparenz auch ohne das Wissen um die KI-Herkunft kritisch hinterfragt wird, sobald Nutzer die zugrunde liegende Berechnungslogik für ihre persönliche Entscheidungsfindung validieren möchten.

Dennoch wies Gruppe 2 auch mehrere Probanden auf, die dem System eher stark Vertrauen schenken. Das Statement *Das System ist trügerisch* beantwortet Proband G2T1 (Score: 4,08) folgendermaßen:

„[00:14:12] Finde ich jetzt gerade nicht so, weil es scheint mir relativ berechenbar über den Verlauf eines Jahres zu sein [...] [00:17:23] Ja, ich muss sagen, ich fand es erstmal generell sehr übersichtlich gestaltet. Ich finde jetzt auch zum Beispiel gerade die Seite, die wir jetzt hier aufgeschlagen haben, den Amortisierungsrechner übersichtlich gestaltet. Ich finde, die Daten sind erstmal übersichtlich aufbereitet“ (Interviewtranskript G2T1).

Hier wird deutlich, wie stark das übersichtliche und moderne Systemdesign dafür sorgen kann, dass Nutzer Affect-Based Trust entwickeln und dem System einen Vertrauensvorschuss geben. Die visuelle Aufbereitung baut kognitive Unsicherheiten ab und suggeriert eine hohe Berechenbarkeit des Systems.

Besonders prägnant manifestierte sich dieser Vertrauensvorschuss bei Proband G2T4 (Score 3,92). Dieser evaluierte die Systemarchitektur maßgeblich über Assoziationen mit bereits etablierten Finanz-Tools:

„[00:27:17] Ja, das, was ich schon gesagt habe. Es sieht ein bisschen aus wie Finanzfluss. Deswegen war es für mich direkt so, okay, ist irgendwie vertraut, kann man mit arbeiten“ (Interviewtranskript G2T4).

Diese Aussage belegt, dass das professionelle und moderne Design des Rechners, was stark an andere bekannte Tools erinnert, durch die Vertrautheit zu einem erhöhten Prozessvertrauen führen kann.

5.2.2 Theme 2: Performance - Spannungsfeld zwischen Usability und Automation Bias

Dieses Theme fokussiert sich auf die Vertrauensdimension der Performance (Lee & See, 2004) und beschreibt, wie die Probanden die Leistungsfähigkeit, Zuverlässigkeit, Validität und fachliche Kompetenz des Photovoltaik-Rechners bewerten. Da Zuverlässigkeit und Validität im Sinne der Richtigkeit des Rechenergebnisses selten durch den Nutzer überprüft werden können, dient in so einer Situation

oft professionelles Interface Design als Cue. Professionelles Design kann maßgeblichen Einfluss auf den Initial Trust des Nutzers in eine Anwendung haben (Zieglmeier & Lehene, 2021).

Wie stark die Bewertung der System-Performance jedoch vom situativen Risiko abhängt, zeigt sich an der ambivalenten Einstellung von Proband G1T1, welche sich bei der Antwort auf das Statement *Ich habe Vertrauen in das System* deutlich zeigt:

„[00:28:23] Also ich weiß halt nicht? Es kommt halt immer so ein bisschen auf die Situation drauf an, in der man gerade ist, wenn ich jetzt wirklich akut entscheiden müsste, ob ich mir eine Photovoltaikanlage kaufe oder nicht, dann könnte es sein, dass ich diese Frage auch anders beantworten werde. Weil ich dann vielleicht doch nochmal woanders nachschauen würde und es vielleicht vergleichen würde mit dem Ergebnis eines anderen Tools oder mit dem Ergebnis von einem Energieberater oder so [...] Aber wenn ich jetzt einfach nur rangehe mit der losen Überlegung, okay, könnte sich das für mich rechnen in ein paar Jahren? Ich habe jetzt keinen akuten Bedarf. [...] Dann habe ich schon Vertrauen, dass die Ergebnisse grundsätzlich erstmal soweit passen. So, und da sage ich dann kommt hin und das sieht ja ganz gut aus und das stimmt mich positiv. [...] Aber so ist es natürlich auch das positive Ergebnis, von dem man sich so ein bisschen beschwichtigen lässt“ (Interviewtranskript G1T1).

Diese Aussage illustriert prägnant, wie Nutzer die mangelnde Überprüfbarkeit der System-Performance kompensieren. Da G1T1 die tatsächliche Validität der Berechnung nicht prüfen kann, macht er sein Vertrauen vom Grad seiner eigenen Verwundbarkeit abhängig. In einem Szenario mit geringem Risiko, ohne akuten Entscheidungsdruck, ist er bereit, dem System Initial Trust entgegenzubringen. Die wahrgenommene Systemkompetenz wird künstlich dadurch gestützt, dass das angezeigte Ergebnis für den Nutzer finanziell positiv und damit wünschenswert ist. Sobald sich jedoch das situative Risiko erhöht, durch eine akute Entscheidungssituation, reicht dieser Cue nicht mehr aus. In diesem Fall würde der Proband das Ergebnis durch die Performance eines menschlichen Experten, wie z.B. einem Energieberater oder weiteren ähnlichen Amortisationsrechnern überprüfen wollen, was auf eine grundlegende Skepsis gegenüber der alleinigen Leistungsfähigkeit des Rechners hindeutet.

Neben dem situativen Risiko kann die Bewertung der System-Performance maßgeblich durch das fachliche Vorwissen der Nutzer geprägt sein. Proband G1T2, der durch seinen beruflichen Hintergrund im Bereich von Stromtarif-Vergleichsportalen

bereits rudimentäres Vorwissen zu Photovoltaik besitzt, betrachtete die Ergebnisse mit Skepsis. Da die Ergebnisse seiner Erfahrung widersprachen, sprach er dem System die Validität ab:

„[00:05:43] Also die Stromproduktion ist sehr hoch. [...] : [00:14:14] Und er gibt dir einfach Werte, also das Programm gibt dir einfach irgendwelche Werte, die du nicht überprüfen kannst, außer du rechnest es selber nach. Wenn du selber nachrechnen würdest und die gleichen Werte rauskommen, dann wäre es sicher. Dann würde ich da sagen, 5 [Bewertung des Statements: *Das System bietet Sicherheit*]. Aber so, stellt er mir einfach nur dar, was ich erwarten könnte, wobei das, wie gesagt, für mich unrealistische Zahlen sind. Das scheint mir sehr unrealistisch zu sein.[...] Deswegen bin ich da ein bisschen skeptisch und deswegen glaube ich auch, dass es wenig Sicherheit bietet“ (Interviewtranskript G1T2).

Anhand dieses Zitates wird deutlich, wie bereichsspezifisches Vorwissen als Schutzschild gegen unreflektierten Overtrust wirken kann. Der Proband gleicht das Ergebnis mit seinen eigenen Erfahrungswerten ab. Da die ausgegebenen Werte der eigenen Erfahrung auf Basis seines Vorwissens widersprechen und das System gleichzeitig keine Überprüfung der Rechenwege zulässt, wird die Kompetenz zur Erzeugung valider Ergebnisse in Frage gestellt. Die vermeintlich zu hohe Stromerzeugungsschätzung blockiert somit den Aufbau von Cognitive Trust.

Ein gänzlich anderes Bild zeichnete sich bei Proband G2T1 ab:

„[00:17:23] Ja, ich muss sagen, ich fand es erstmal generell sehr übersichtlich gestaltet. Ich finde jetzt auch zum Beispiel gerade die Seite, die wir jetzt hier aufgeschlagen haben, den Amortisierungsrechner übersichtlich gestaltet. [...] [00:17:57] Und da fühle ich mich erstmal wohl in der Benutzung, also ich fühle mich sicher“ (Interviewtranskript G2T1).

Der Proband nutzt unterbewusst die professionelle Optik des Systems und die wahrgenommene Übersichtlichkeit als Cue für Sicherheit und Validität. Daraus wird hoher Affect-Based Trust generiert, welcher beim Nutzer ein relativ hohes Gesamtvertrauen von 4,08 auslöst. Der Proband entwickelt unbegründeten Overtrust, da die empfundene Sicherheit bei der UI-Bedienung fälschlicherweise mit der Richtigkeit der algorithmischen Berechnungslogik gleichgesetzt wird.

Ein ähnliches Muster zeigt sich bei Proband G2T4, der das Statement: *Das System bietet Sicherheit* folgendermaßen beantwortet und begründet:

„[00:21:27] Das würde ich schon sagen, weil du da halt wirklich an die Hand genommen wirst und durchgeführt wirst. Deswegen, wenn ich damit jetzt was ausrechnen würde, würde ich mich schon sicher fühlen, wenn ich so naiv bin und das nicht mit anderen Sachen vergleiche“ (Interviewtranskript G2T4).

Anhand dieses Zitats wird deutlich, wie stark eine hohe Usability, in diesem Fall die strukturierte Nutzerführung, Vertrauen erzeugen kann. Der Nutzer entwickelt durch die reibungslose Bedienung zunächst Affect-Based Trust in die Systemkompetenz. Gleichzeitig offenbart der Nachsatz jedoch eine verbleibende Skepsis: Dem Probanden ist bewusst, dass er die tatsächliche Validität der Berechnungen mangels Fachwissen nicht überprüfen kann und unreflektierten Overtrust in die bloße Systemausgabe, ohne eine weitere Prüfung mit anderen Systemen, ein Risiko darstellt.

Dennoch zeigte sich auch bei den unaufgeklärten Probanden der Gruppe 2, dass ein initiales, durch professionelles Design aufgebautes Vertrauen äußerst fragil sein kann, sobald die Ergebnisse aktiv validiert werden. Die Konfrontation mit dem Fragebogen-Statement zur Systemzuverlässigkeit motivierte Proband G2T2 dazu, die Zuverlässigkeit des Rechners durch die Eingabe von Extremwerten (Erhöhung der Nutzungsdauer) gezielt auf die Probe zu stellen. Als die erwartete Anpassung der Ergebnisse ausblieb, entdeckte er mangelnde Validität in den Berechnungen:

„[00:30:48] Über 25 Jahre ist das offensichtlich gleich, aber über die Nutzungsdauer, da hat sich auch nichts geändert. Das ist auch noch dasselbe, und ich habe jetzt 50 Jahre eingegeben, als Nutzungsdauer. Auch da sehe ich ja jetzt keinen Unterschied. Dann würde ich auf jeden Fall sagen, nein, auf das System ist kein Verlass“ (Interviewtranskript G2T2).

Diese ausbleibende Anpassung hat den Probanden zu einer vollständigen Neubewertung des Fragebogens bewogen, was deutlich macht, wie abrupt ein durch professionelles Design hergeleiteter Affect-Based Trust in Skepsis umschlagen kann. Die Usability und Optik verlieren in dem Moment ihre vertrauensstiftende Wirkung, in dem die Erwartungshaltung des Nutzers an die Systemreaktivität gebrochen wird. Die sich nicht auf das Ergebnis auswirkende Parameterveränderung offenbart die Fehler des Systems und belegt, dass eine unzureichende Systemzuverlässigkeit das initiale Affect-Based Trust sofort zerstören kann.

5.2.3 Theme 3: Purpose - Wahrgenommene Intention und Professionalität

Dieses Theme widmet sich der Vertrauensdimension Purpose nach Lee und See (2004). Hierbei steht die Frage im Zentrum, welche Absichten und Motive die Probanden hinter der Bereitstellung des Photovoltaik-Rechners vermuten. Dies äußert sich in der Mensch-Computer-Interaktion häufig durch wahrgenommene Hilfsbereitschaft, Wohlwollen und wie gut sich der Nutzer bei der Erreichung seines Ziels durch das System unterstützt fühlt.

Die Purpose Dimension erwies sich in der Auswertung als die von fast allen Probanden am positivsten wahrgenommene Vertrauensfacette. Sowohl in der aufgeklärten Gruppe 1, als auch in der unaufgeklärten Gruppe 2 äußerten die Nutzer das Gefühl, vom System gut beraten und angeleitet zu werden.

Exemplarisch dafür steht Nutzer G1T1, der sich durch die strukturierte Aufbereitung der Informationen insbesondere als Laie ohne Vorwissen exzellent unterstützt fühlte:

„ [00:34:00] Aber für mich auf jeden Fall, als Digital Native, das war super entspannt. Also, super übersichtlich, clicky, bunty, wunderbar. Ich glaube auch meine Mutter wird damit klarkommen. Also, ich bin mit dem System vertraut. Ich finde auch viele Sachen, die für mich als Laie Sinn machen sind drin und viele Parameter.[...] [00:37:45] Ich finde auf jeden Fall, dass viele Parameter, über die man sich vielleicht im Vorfeld auch gar nicht so sehr Gedanken macht, mit dabei sind und die Erklärung finde ich auf jeden Fall sehr wichtig und gut. Also, gerade dass der Link da vom HTW-Berlin mit drin ist, da wurde sich auf jeden Fall Gedanken gemacht. Wenn du so Energieverbrauch hast, wie berechne ich den jetzt? Dann darf ich wieder auf die Google-Suche gehen, in eine Recherche gehen, um einen Wert in einem Rechner auszufüllen. Und wenn das dann schon da ist, diese Recherche halt schon vorgeleistet wurde, ist das gut, dann ist da auf jeden Fall gut eingegangen worden auf meine Bedürfnisse“ (Interviewtranskript G1T1).

Dieses Zitat belegt eindrucksvoll, welchen großen Einfluss Prinzipien wie Progressive Disclosure und integrierte Tooltips auf die Vertrauensbildung haben können. Da das System die Nutzer bei fachfremden Eingaben nicht alleinlässt, sondern ihnen durch Erklärungen und externe Referenzen die eigene Recherchearbeit proaktiv abnimmt, wird dem Rechner eine ehrliche und uneigennützige Unterstützungsabsicht unterstellt. Die Probanden interpretieren diese hohe Nutzerzentrierung als

direktes Signal für Wohlwollen der Anwendung. Sie fühlen sich sicher durch den Prozess geleitet, um ihr Ziel, die Wirtschaftlichkeitsprüfung eines Angebots für eine PV-Anlage, zu erreichen, wodurch ein massives Purpose-Vertrauen aufgebaut wird.

Dieses Muster bestätigt sich bemerkenswerterweise selbst bei dem Probanden mit dem geringsten Gesamtvertrauen (Score = 1,92) der Gruppe 1. Obwohl Proband G1T2 die Performance des Rechners aufgrund seines Vorwissens äußerst kritisch bewertete, sprach er der Purpose-Dimension ein hohes Maß an Wohlwollen zu. Er lobte explizit die nutzerzentrierte Führung, die ihm ausführlichere Recherchearbeit abnahm:

„[00:20:21] Ansonsten sind viele der Fragen vom Rechner eigentlich darauf ausgelegt, dass ich mich als Laie auch damit beschäftigen kann, dass ich nicht zu viel recherchieren muss, was ich will oder was ich brauche. Sondern dass ich einfach sagen kann, ja, ich hab so ein Dach, ich will die so lange benutzen. Das sind ungefähr die Kosten“ (Interviewtranskript G1T2).

Anhand dieses Zitates wird ersichtlich, dass selbst initial skeptische Nutzer in der anwenderfreundlichen Gestaltung einen klaren Mehrwert sehen. Die intuitive, unterstützende Führung erzeugte ein Gefühl von Wohlwollen. Gleichzeitig verdeutlicht dieser Kontrast jedoch auch die Grenzen des Purpose-Vertrauens. Eine positive, hilfsbereite Ansprache und ein professionelles Design allein reichen nicht aus, um einen durch mangelnde Performance begründeten Vertrauensverlust vollständig zu kompensieren.

Auch der Warnhinweis am Ende der Ergebnisübersicht der Anwendung wird positiv hervorgehoben:

„[00:17:02] Was gut ist, ist der wichtige Hinweis ganz unten, dass es eine vereinfachte Schätzung ist und tatsächliche Erträge variieren können. Außerdem, dass das Programm einen unabhängigen Betrieb oder einen Berater empfiehlt. Das ist gut“ (Interviewtranskript G1T2).

Dieser Hinweis fungiert als proaktives Erwartungsmanagement. Indem das System seine eigenen Grenzen transparent offenlegt und explizit an menschliche Fachexperten verweist, beweist es in den Augen des Nutzers Aufrichtigkeit und Wohlwollen. Es wird der Eindruck erweckt, dass die Anwendung keine manipulativen Verkaufsabsichten verfolgt oder eine fehlerfreie Berechnung vortäuscht. Diese ehrliche Kommunikation der Systemgrenzen schützt die Anwendung vor dem Verdacht der Täuschung und festigt das Purpose-Vertrauen maßgeblich, da

dem Nutzer signalisiert wird, dass sein Schutz vor finanziellen Fehlentscheidungen über dem reinen Selbstzweck des Tools steht.

Ein scharfer Kontrast zeigte sich jedoch bei Proband G2T3, der mit einem Score von 1,75 das geringste Vertrauen der gesamten Studie aufwies. Obwohl er der unaufgeklärten Gruppe angehörte, stand er der Purpose-Dimension hochgradig skeptisch gegenüber. Er bezeichnete die statische farbliche Gestaltung der Ergebnisseite als potenziell manipulativ, da der Amortisationszeitraum stets in grüner Schrift präsentiert wurde, unabhängig von der tatsächlichen Rentabilität des Angebots:

„[00:44:58] Ich stimme voll umfänglich zu, dass das System sich hinterhältig verhält. Weil es negative Angebote nicht als solche deklariert. Eine grüne Zahl, eine grüne Farbe, der Farbgebung ist bei sowas sehr wichtig, eine grüne Farbe ist das grüne Zeichen Go, du kannst gehen. Wenn die Amortisationszeit jetzt auf einmal 30 Jahre wäre, die Nutzung aber nur 25, ja. und die Zahl dann immer noch grün ist, kann ein Nutzer, der sehr auf sowas reagiert, das tatsächlich bewusst übersehen“ (Interviewtranskript G2T3).

Anhand dieses Datenfragmentes wird deutlich, wie der unpassende Einsatz von UI-Methodiken das Purpose-Vertrauen massiv beschädigen kann. Da die erlernte Signalfarbe nicht dynamisch an die tatsächliche Rentabilität der Offerte gekoppelt ist, empfindet der Proband das Design als gezielte Irreführung. Er unterstellt der Anwendung eine bewusste Täuschungsabsicht, die darauf abzielt, dem Nutzer selbst unrentable Angebote als positiv zu verkaufen. Das fehlen von Wohlwollen führt hier zu einem eklatanten Vertrauensverlust, da der Nutzer davon ausgeht, dass das System mögliche Eigeninteressen bewusst oder fahrlässig über den Schutz des Anwenders vor finanziellen Fehlentscheidungen stellt.

Dennoch erweist sich das Purpose-Vertrauen auch in Gruppe 2 als überwiegend positiv. Ein bemerkenswertes Beispiel hierfür liefert Proband G2T2, der während der Durchführung des Fragebogens Inkonsistenzen in der Berechnung festgestellt hat und aus diesem Grund einen starken Vertrauensverlust erlitt, ist. Dennoch ist er vom Wohlwollen und der Hilfsbereitschaft des Systems überzeugt:

„[00:42:29] Auf die individuellen Bedürfnisse ja, also wenn du jetzt z.B. sagst, mit verschiedenen Dachseiten oder so, das ist sehr gut, dass man das mit einfließen lassen kann und halt auch so die Finanzierungsparameter, dass man das da selbstständig eintragen kann, das fand ich alles an sich sehr gut. Fehlendes Vorwissen kann man sich durch diese Infoboxen auf jeden Fall holen“ (Interviewtranskript G2T2).

Die vom System angebotene Granularität der Eingabemöglichkeiten, wie etwa die Konfiguration verschiedener Dachseiten, sowie die proaktive Bereitstellung von Fachwissen durch Infoboxen, werden vom Nutzer als Indikatoren für Wohlwollen interpretiert. Das transparente Angebot zur Wissensvermittlung schützt das System in diesem Fall davor, dass der Nutzer die fehlerhaften Berechnungen als bewusste, hinterhältige Täuschungsabsicht interpretiert.

5.2.4 Theme 4: Appropriate Reliance und Warranted Distrust - Die Ambivalenz der KI-Offenlegung

Das abschließende Theme bildet den analytischen Höhepunkt der Untersuchung und fokussiert sich auf das Konzept der Appropriate Reliance nach Lee und See (2004) und des Warranted Distrust nach Visser et al. (2023). Während die vorherigen Themes die isolierte Wahrnehmung der Systemoberfläche und Funktionalität behandelten, wird hier der direkte Störfaktor der KI-Genese beleuchtet.

Proband G1T1 äußert sich bei der Antwort auf das Statement: *Ich stehe dem System skeptisch gegenüber* reflektiert und mit einem Warranted Distrust gegenüber dem Rechner, explizit mit Bezug auf die KI-Genese der Anwendung:

„[00:26:40] Ich glaube, mit dem Vorwissen, dass das KI generiert ist. [...] Weil gerade ist es eine riesen Blackbox und man weiß, dass das KI generiert ist. Denn, wenn ich jetzt wüsste, es hat sich ein Energieberater mal hingesetzt und diese Formeln in sein Software-Tool gegossen, da würde man ihm wahrscheinlich mehr vertrauen, auch wenn der Energieberater auch Fehler machen kann.[...] Aber trotzdem weiß man auch, was so KI-Tools mittlerweile können und gerade solche Rechnungen, das ist ja keine komplexe Mathematik “ (Interviewtranskript G1T1).

Der Proband kommuniziert hier sehr offen, dass explizit das Wissen über die KI-Genese und die damit verbundene Intransparenz die Skepsis gegenüber dem System erzeugt. Er ist zwar dazu bereit, der KI eine grundlegende Kompetenz für einfache mathematische Operationen zuzuschreiben, weigert sich jedoch, sich auf die Ergebnisse blind zu verlassen. Stattdessen würde er einem menschlichen Fachexperten wie einem Energieberater, deutlich eher Vertrauen schenken, selbst wenn ihm dessen prinzipielle Fehleranfälligkeit bewusst ist. Dies illustriert gesunden Warranted Distrust zum Schutz vor unreflektiertem Overtrust.

Bei Proband G1T4 wird besonders deutlich, dass ein KI-Hinweis einen Einfluss auf den Umgang mit einer Anwendung haben kann. Er verdeutlicht, dass allein

die Offenlegung der KI-Genese als präventiver Störfaktor fungiert, der ihn vor einem unreflektierten Overtrust bewahrt. Auf die Frage nach dem Einfluss des KI-Wissens auf sein Vertrauen in das System erklärte er:

„[00:27:31] Deswegen muss ich da schon sagen, ich nutze das Tool gerne, lass mich auch, irgendwie in die Richtung lenken von den Preisen, aber ich würde dann dennoch irgendwie selber auf kleine Rechnungen oder so zurückgreifen, dass ich halt immer noch mal so ein bisschen gegenchecke. Würde ich aber nicht machen, wenn ich nicht wüsste, dass es KI ist?“ (Interviewtranskript G1T4)

Anhand dieses Datenfragments lässt sich die Schutzfunktion des KI-Labels gegen den Automation Bias belegen. Ohne das Wissen um den generativen Hintergrund der Anwendung würde der Nutzer möglicherweise von der professionellen Benutzeroberfläche auf eine fehlerfreie Berechnungslogik im Hintergrund schließen und dem System Overtrust entgegenbringen. Die transparente Kommunikation der KI-Herkunft bricht diesen Automatismus jedoch auf, was den beschriebenen Effekt des Transparenz-Paradoxons illustriert. Das Wissen um die KI-Genese transformiert das potenzielle blinde Vertrauen in Warranted Distrust, das durch Wachsamkeit und Skepsis geprägt ist. Der Proband würde das System zwar weiterhin als Entscheidungshilfe nutzen, etabliert jedoch bewusst manuelle Kontrollmechanismen, um die Ergebnisse aktiv zu validieren.

In der unaufgeklärten Gruppe 2 zeigt sich anhand von G2T4, wie stark sich initiales Vertrauen und daraus folgend Overtrust entwickeln kann, wenn es keinen Hinweis gibt, dass die Anwendung KI generierter Herkunft ist. Vor der Offenlegung kommt der Proband auf einen relativ hohen Vertrauensscore von 3,92. Wie unter Process beschrieben hat sich dieser Proband von der professionellen Optik, die ihn z.B. an bekannte Tools erinnert, positiv beeinflussen lassen.

Bei der Offenlegung der KI-Genese zeigt sich dann, dass der Proband extrem überrascht und überwältigt wurde:

„[00:31:48] Ich hoffe du hast das alles nochmal kontrolliert, dass da nicht irgendwelche Scheiße drin ist, ansonsten würde ich jetzt, wenn ich nochmal zurückgehen könnte, die Fragen vielleicht anders beantworten, was Skepsis angeht“ (Interviewtranskript G2T4).

Hieran wird deutlich, dass das Wissen über die KI-Herkunft eines Systems maßgeblichen Einfluss auf die Vulnerabilität haben kann, die der Anwender bei der Nutzung zulässt. Er erleidet nach der Offenlegung einen sofortigen Vertrauensverlust. Dies spricht dafür, dass der Proband mit einer Offenlegung vor der Studie eine deutlich kritischere Grundhaltung gezeigt hätte. Dies stützt die Effekte des

Transparenz-Paradoxons: Die Offenlegung der KI schützt den Nutzer davor, dem Automation Bias zu erliegen, und transformiert das aus dem Design abgeleitete Vertrauen in schützenden Warranted Distrust.

Auch Proband G2T1 hat hohen Initial Trust (4,08) in die Anwendung, basierend auf der Übersichtlichkeit und der gut aufbereiteten Ergebnisse:

„[00:16:04] Sieht für mich erstmal vertrauenserweckend aus. [...] [00:16:08] Ich finde, das System bietet auch Sicherheit [...] [00:17:23] ich fand es erstmal generell sehr übersichtlich gestaltet. Ich finde jetzt auch zum Beispiel gerade die Seite, die wir jetzt hier aufgeschlagen haben, den Amortisierungsrechner übersichtlich gestaltet. Ich finde, die Daten sind erstmal übersichtlich aufbereitet“ (Interviewtranskript G2T1).

Im starken Gegensatz zu dem massiven Vertrauensverlust bei G2T4 reagierte G2T1 auf die nachträgliche Offenlegung der generativen KI-Herkunft jedoch deutlich ausgewogener und weniger alarmistisch. Auf das Statement: *der Einfluss von KI bei der Erstellung des PV-Rechners beeinflusst mein Vertrauen in das System positiv* reagiert er:

„[00:20:05] Finde ich für mich sehr ausgewogen. Ich stehe dem generell wieder nicht besonders skeptisch und auch nicht besonders euphorisch gegenüber. Und dann finde ich das erstmal eine sinnvolle Ergänzung. Bin aber immer froh, wenn noch ein realer Mensch, in Anführungsstrichen, darüber guckt und das Ganze begleitet und prüft und, ja, vor allem überprüft. Aber die Kombination aus sowohl einer KI als auch einem realen Mensch, der mit seiner eigenen Expertise das Ganze so kombiniert, finde ich jetzt erstmal eigentlich eher einen, positiven Eindruck, den ich dadurch gewinne“ (Interviewtranskript G2T1).

Anhand dieses Datenankers zeigt sich, dass die Offenlegung nicht zwangsläufig zu einem vollständigen Vertrauensverlust führen muss, sondern eine Vertrauenskalibrierung anstoßen kann. Der Proband bewahrt sich eine grundsätzliche Offenheit gegenüber der Technologie, auch wenn er darauf hinweist, dass er die Überprüfung durch einen Menschen für richtig hält.

Zum Abschluss dieses Segments gibt es auch noch das Beispiel von Proband G2T3. Dieser wies bereits vor der Offenlegung das geringste Gesamtvertrauen auf und legte starken Warranted Distrust an den Tag. Er testete das System aktiv auf seine Grenzen und identifizierte das Verhalten des Rechners, insbesondere durch die statische und irreführende Farbkodierung der Ergebnisseite, als potenziell trügerisch. Im Gegensatz zu dem durch die Offenlegung schockierten Probanden

G2T4 reagierte G2T3 auf die Information der KI-Genese vollkommen überrascht und sah seine Skepsis bestätigt:

„ [00:52:40] Hat man gemerkt. Woran ich das gemerkt habe? Dass es wenig auf individuelle menschliche Bedürfnisse eingegangen ist, sondern sehr, sehr allgemein gehalten war. [...] [00:52:40] Ich glaube, dass wenn komplett, das komplett von Menschenhand gemacht worden wäre, diese Farbgebung mit, mit Farbvariablen betrachtet worden wäre. [00:52:40] Und sowas tut KI zumindest nicht zwingend, außer wenn du es der KI explizit sagst“ (Interviewtranskript G2T3).

Anhand dieses Zitates wird deutlich, dass aufmerksame Nutzer die Defizite autonom generierter Benutzeroberflächen durchaus erkennen und auf eine fehlende menschliche Instanz im Designprozess zurückführen. Der Proband dekonstruiert hier intuitiv das Problem des Vibecodings: Er erkennt, dass generative KI-Modelle die nutzerzentrierte Bedeutung von UI-Heuristiken, wie z.B. dynamische Ampelfarben zur Warnung vor schlechten Angeboten, nicht von sich aus verstehen. Die KI imitiert lediglich Oberflächenstrukturen, ohne deren funktionalen Zweck im Sinne der Mensch-Computer-Interaktion zu adaptieren, es sei denn, diese Parameter werden der KI durch präzises Prompting durch einen menschlichen Entwickler vorgegeben. Das Wissen um die KI-Genese löst bei diesem Probanden somit keinen plötzlichen Vertrauensverlust aus, sondern validiert lediglich seinen Vorab etablierten Distrust.

6 Diskussion und Limitationen

6.1 Beantwortung der Forschungsfrage

Der Zweck dieser Studie ist zu klären, wie sich die Offenlegung der KI-Genese des PV-Amortisationsrechners auf die einzelnen Vertrauensdimensionen als auch auf Warranted Trust, bzw. Warranted Distrust der Probanden ausgewirkt hat.

Ziel der vorliegenden Untersuchung war es zu analysieren, inwiefern die Offenlegung der KI-Genese eines Photovoltaik-Amortisationsrechners die spezifischen Vertrauensdimensionen der Nutzer (Purpose, Process, Performance) beeinflusst (Lee & See, 2004). Die qualitativen Ergebnisse verdeutlichen, dass die Transparenz über die algorithmische Herkunft einen massiven Einfluss auf die kognitive und emotionale Wahrnehmung der Benutzeroberfläche haben kann, aber nicht zwangsläufig haben muss. Im Zentrum der nachfolgenden Diskussion steht daher die theoriegeleitete Einordnung der Befunde. Es soll kritisch reflektiert werden, wie die Offenlegung der KI-Genese künstlich erzeugten Overtrust abbaut und stattdessen eine Verschiebung hin zu Calibrated Trust und einem schützenden Warranted Distrust initiiert, welcher in Appropriate Reliance resultiert (Lee & See, 2004; Visser et al., 2023).

6.1.1 Process

Die qualitativen Ergebnisse bezüglich der Vertrauensdimension Process verdeutlichen, dass die frühzeitige Offenlegung der KI-Genese (Gruppe 1) eine tiefgreifende kritische Auseinandersetzung mit den systemischen Abläufen erzeugte. Entsprechend der Definition von Lee und See (2004) bezieht sich Prozessvertrauen maßgeblich darauf, inwiefern die zugrunde liegenden Algorithmen eines Systems für den Nutzer verständlich und nachvollziehbar sind. Genau diese Nachvollziehbarkeit wurde von den Probanden G1T1, G1T2 und G1T3 in Frage gestellt. Das Bewusstsein um den KI-Ursprung führte zu fundamentalen Bedenken hinsichtlich der Validität der Berechnungslogik, weshalb die Nutzer explizit nach Transparenz

verlangten, um die Ergebnisse manuell verifizieren zu können. Proband G1T4 reflektierte diesen Umstand proaktiv und gab an, dass er die Berechnungen ohne das Wissen um die KI-Genese deutlich weniger hinterfragt hätte. Dies belegt, dass die mangelnde Transparenz generativer KI, oft als Blackbox-Problem beschrieben (Zednik, 2019), erst durch das KI-Label in den mentalen Fokus der Nutzer rückte und den Wunsch nach manueller Überprüfbarkeit auslöste.

Die qualitativen Ergebnisse der Process-Dimension der unaufgeklärten Gruppe 2 zeigen allerdings auch nicht, dass Probanden der Berechnungslogik ohne Hinweis auf die KI-Herkunft mit einem starken Overtrust begegnen müssen. Ausschließlich bei Proband G2T1 (4,08) wird keinerlei Skepsis den Berechnungen gegenüber geäußert. Bemerkenswert ist der Fall von G2T2: Angeregt durch den Fragebogen erkundete der Proband die Eingabemöglichkeiten des Systems, deckte funktionale Schwächen auf und passte daraufhin seine initial höhere Perceived Trustworthiness an die nun erkannte, niedrige Actual Trustworthiness des fehlerhaften Systems an. Der infolgedessen niedrige Gesamtscore (2) illustriert eine dynamische Vertrauenskalibrierung durch direkte Interaktion.

Dieses explorative Verhalten legt jedoch auch nahe, dass die kognitive Wachsamkeit der Probanden durch die methodische Rahmung der Studie (Thinking-Aloud-Methode und gezielte Usability-Fragen) künstlich erhöht worden sein könnte. In realistischen Nutzungskontexten ist eine derart tiefe kritische Reflexion nicht zwingend zu erwarten. Dies gilt insbesondere deshalb, weil auch etablierte, von Menschen programmierte Informationssysteme, wie etwa Rendite- oder Finanzrechner, wie der von Proband G2T4 erwähnte Rechner von Finanzfluss.de, in der Regel lediglich ihre Eingabeparameter erklären, ihre interne Berechnungslogik jedoch ebenfalls als Blackbox belassen.

Darüber hinaus verdeutlichen die Aussagen der Probanden G1T2 und G2T3 die zentrale Rolle von bestehendem Vorwissen. Wie der Literaturüberblick von Visser et al. (2023) zusammenfasst, fungiert Domänenwissen als starker Faktor für das Vertrauen in automatisierte und KI-basierte Systeme. Probanden mit Vorwissen im Bereich der Photovoltaik glichen die Systemergebnisse mit ihren eigenen Erwartungshaltungen ab. Sobald eine Diskrepanz auftrat, waren sie in der Lage, das unvollständige oder fehlerhafte System als solches zu identifizieren. Dies bestätigt Beobachtungen, wonach Experten durch ihr Fachwissen weniger anfällig für unreflektierten Overtrust sind, da mangelndes Wissen Laien oftmals daran hindert, Systemfehler überhaupt erst zu erkennen.

Diese Aussagen deuten darauf hin, dass die Offenlegung die Probanden der Gruppe 1 in einen besonderen Zustand kognitiver Wachsamkeit versetzte, der ohne diesen Hinweis unwahrscheinlicher wäre (Schilke & Reimann, 2025). Theoretisch

lässt sich dies als Paradebeispiel für das von Visser et al. (2023) beschriebene Warranted Distrust einordnen. Sie betonen, dass Distrust nicht bloß die Abwesenheit von Vertrauen ist, sondern eine eigenständige, schützende Dimension darstellt, die durch Skepsis und kritische Reflexion gekennzeichnet ist.

Durch dieses Konzept könnte sich auch die scheinbare Diskrepanz zu den quantitativen Fragebogenergebnissen erklären: Obwohl Gruppe 1 quantitativ ein tendenziell höheres Systemvertrauen als Gruppe 2 aufwies, belegen die qualitativen Aussagen ein hohes Maß an Skepsis. Dieser vermeintliche Widerspruch löst sich auf, wenn man die KI-Offenlegung als kalibrierendes Element versteht. Durch die transparente Kommunikation der Systemherkunft sank das Gefühl der Vulnerabilität der Nutzer, da sie wussten, worauf sie sich einlassen. Anstatt einem unreflektierten Overtrust zu erliegen (Lee & See, 2004), passten die Probanden ihre Perceived Trustworthiness an die tatsächliche mögliche Realität des Systems an (Schlicker & Langer, 2021). Die von den Probanden geäußerte Bereitschaft, die KI-Ergebnisse lediglich als Richtwert zu nutzen und sich final eine Zweitmeinung bei einem Energieberater oder durch eigene Berechnungen einzuholen, zeugt folglich nicht von einem vollständigen Distrust. Vielmehr demonstriert sie ein angemessenes Verhältnis aus Warranted Trust und Warranted Distrust, bei dem die Berechnungen mit einer angemessenen, schützenden Skepsis bewertet werden (Visser et al., 2023). Dies schafft eine optimale Voraussetzung für das Ziel der Appropriate Reliance, also dem optimalen Verhältnis aus Vertrauen und Misstrauen dem System gegenüber und der daraus resultierenden Handlung, um weder blindem Vertrauen in das System zu erliegen und eine Fehlentscheidung auf Basis davon zu tätigen, aber gleichzeitig auch fähige Systeme nicht vollständig abzulehnen, nur weil KI bei der Erstellung eine Rolle gespielt hat (Lee & See, 2004).

6.1.2 Performance

Die qualitativen Ergebnisse und Erkenntnisse bezüglich der Vertrauensdimension Performance ähneln sehr stark den Ergebnissen der Process Dimension.

Die Ergebnisse der aufgeklärten Gruppe 1 zeigen, dass das Wissen um die KI-Genese nicht zwingend zu einer vollständigen Systemablehnung führen muss, sondern vielmehr eine funktionale Skepsis im Sinne eines Warranted Distrust erzeugt (Visser et al., 2023). Sowohl G1T1 als auch G1T3 und G1T4 äußerten zwar Skepsis gegenüber der absoluten Validität der Ergebnisse, sprachen den generativen KI-Modellen jedoch grundlegende Kompetenzen in der Generierung einfacher Algorithmen und mathematischer Operationen zu. Diese Einschätzung

fürhte zu der Aussage der Probanden, den Ergebnissen aufgrund der KI-Herkunft nicht blind zu vertrauen, sondern diese durch Zweitmeinungen überprüfen zu wollen. Dies demonstriert eine erfolgreiche Vermeidung von Overtrust und die Etablierung einer funktionalen Appropriate Reliance (Lee & See, 2004).

Interessant ist hierbei die Selbstreflexion von Proband G1T1: Das für ihn positiv ausfallende Ergebnis hinsichtlich der Amortisationszeit erzeugte einen vertrauensvollen Eindruck, dem er gerne glauben schenken wollte. Durch sein Vorwissen im Bereich von Immobilieninvestitionen konnte er die berechneten 11,6 Jahre als vermeintlich realistischen Zeitraum einordnen und somit für sich validieren. Im starken Kontrast dazu steht Proband G1T2, welcher den Ergebnissen des Systems keinerlei Vertrauen schenkte. Aufgrund seines domänenspezifischen Vorwissens im Bereich der Photovoltaik stufte er die prognostizierte Stromerzeugung als deutlich zu hoch ein. In diesem Fall konnte auch die professionelle Benutzeroberfläche keinen positiven Einfluss auf das Vertrauen üben. Dies bestätigt, dass bereichsspezifisches Vorwissen als kognitives Schutzschild fungiert (Visser et al., 2023), da Inkonsistenzen in der Systemleistung zuverlässiger aufgedeckt werden können.

In der unaufgeklärten Gruppe 2 zeichnet sich ein ambivalenteres Bild ab, welches die Diskrepanz zwischen Actual und Perceived Trustworthiness verdeutlicht. Proband G2T1 fühlte sich in der Bedienung sehr sicher. Die übersichtliche Aufbereitung der Daten erzeugte bei ihm ein hohes Maß an Vertrauen. Hier zeigt sich eindeutig, wie UI-Cues eine künstlich erhöhte Perceived Trustworthiness generieren können, die stark von der Actual Trustworthiness der Berechnungslogik abweichen kann (Schlicker & Langer, 2021).

Ein ähnliches Verhaltensmuster offenbarte G2T4. Da der Rechner seiner Auffassung nach etablierten Tools, wie dem WISO-Steuertool oder Finanzfluss.de, optisch ähnelte, entwickelte er durch diese wahrgenommene Vertrautheit ein hohes Vertrauen in die Ergebnisse. Zwar merkte der Proband selbstreflektiert an, dass ein ungeprüftes Vertrauen naiv sei, doch belegt dieser Fall, dass ein professionelles Design als Cue für eine hohe Systemkompetenz herangezogen wird und auf Affect-Based Trust einzahlen kann (Schlicker & Langer, 2021). Auch bei Proband G2T2 war initial ein hohes Vertrauen vorhanden. Da man annehmen kann, dass in einer realistischen Nutzungssituation, ohne die kognitive Wachsamkeit, die durch die Thinking-Aloud-Methode und die Fragebögen forciert wurde, die Eingaben nicht derart kritisch auf die Probe gestellt worden wären, ist davon auszugehen, dass auch dieser Nutzer der professionellen Optik gefolgt wäre und einen hohen Affect-Based Trust beibehalten hätte (Zieglmeier & Lehene, 2021).

Lediglich Proband G2T3 äußerte sich von Beginn an extrem skeptisch. Ähnlich wie G1T2 verfügte dieser Nutzer über domänenspezifisches PV-Vorwissen, welches

in seinem familiären Umfeld aufgebaut wurde. Dieses Wissen führte dazu, dass er mehr Personalisierung und Variabilität in der Konfiguration erwartete. Die fehlende Berücksichtigung von Finanzierungskrediten im vorgegebenen Test-Szenario interpretierte er als potenziellen Versuch des Systems, eine Schwäche in der Kreditberechnung zu vertuschen. Zudem empfand er die visuelle Darstellung der Ergebnisse als hochgradig irreführend, da jedes Endergebnis durch eine grüne Farbgebung gerahmt wurde. Es ist folglich davon auszugehen, dass dieser Proband aufgrund seines Vorwissens auch ein rein von Menschen programmiertes System mit derselben ausgeprägten Wachsamkeit betrachtet hätte.

Zusammenfassend verdeutlicht die Gegenüberstellung beider Gruppen in der Performance-Dimension die weitreichenden Gefahren von KI-generierten Benutzeroberflächen (Vibecoding) im Kontext des Automation Bias. Wie die Ergebnisse der unaufgeklärten Gruppe zeigen, kann ein professionelles User Interface für Laien als Ersatz für eigentlich nicht überprüfbare Systemkompetenz genutzt werden (Schlicker & Langer, 2021). Die UI-Cues treiben die subjektive Perceived Trustworthiness künstlich in die Höhe und verschleiern mögliche Defizite der tatsächlichen Berechnungslogik. Die empirischen Daten zeigen, dass diese trügerische Design-Fassade durch zwei Faktoren durchbrochen werden kann: Zum einen durch ausgeprägtes domänenspezifisches Vorwissen, welches, wie bei den Probanden G1T2 und G2T3 beobachtet, als unabhängiges kognitives Schutzschild fungiert, da die Nutzer unplausible Ergebnisse oder fehlende Systemparameter selbstständig entlarven können. Zum Anderen durch die proaktive Offenlegung der KI-Genese. Das Wissen um den KI-Ursprung kann für Laien als notwendiger Reibungspunkt dienen, der den durch die UI ausgelösten Automation Bias unterbricht und eine kritische Reflexion der Systemleistung erzwingt (Schilke & Reimann, 2025). Folglich wirkt die KI-Offenlegung im Bereich der Performance nicht als reiner Vertrauenszerstörer, sondern als essenzielles kalibrierendes Instrument. Sie ist hilfreich, um Overtrust in durch generative KI erzeugte Systeme zu verhindern und im Sinne von Lee und See (2004) Appropriate Reliance überhaupt erst zu ermöglichen.

6.1.3 Purpose

Die qualitativen und quantitativen Ergebnisse bezüglich der Vertrauensdimension Purpose (Lee & See, 2004) zeigen über die fast komplette Stichprobe hinweg ein positives Bild. Dies spiegelt sich in den quantitativen Daten wider: Das Statement *Das System verhält sich hinterhältig* erreichte mit einem Durchschnittswert von $M = 3,63$ (umgepolter Wert) den zweithöchsten Zustimmungsgrad für Vertrauen. Noch eindrücklicher belegt das Single-Item zur Nutzerzentrierung: *Ich habe das*

Gefühl der Rechner ist bei der Nutzung auf meine individuellen Bedürfnisse, mein möglicherweise fehlendes Vorwissen und meine Ziele als Nutzer eingegangen mit dem höchsten Gesamtscore von $M = 4,38$ diese Tendenz. Insbesondere Nutzer ohne tiefgreifendes Fachwissen fühlten sich über weite Strecken gut unterstützt und entwickelten nicht das Gefühl, das System würde sich böse verhalten.

Die implementierten UI-Elemente, wie erklärende Tooltips, Progressive Disclosure, das moderne Design und der abschließende Warnhinweis zur Genauigkeit der Berechnung lösten bei den Probanden ein starkes Gefühl von Wohlwollen und Glaubwürdigkeit aus (Mayer et al., 1995). Entsprechend den Erkenntnissen von Springer und Whittaker (2018) reduziert das Progressive Disclosure den Cognitive Load und baut somit effektiv Vertrauen auf. Darüber hinaus fungiert der Warnhinweis als proaktives Erwartungsmanagement, da das System Aufrichtigkeit beweist und sich vor dem Verdacht manipulativer Verkaufsabsichten schützt. Lediglich Proband G2T3 begegnete der Anwendung, wie bereits beschrieben, von Beginn an mit einer extremen kognitiven Wachsamkeit, weshalb die mehrheitlich positiv wahrgenommene Oberfläche keinen Vertrauensvorschluss generieren konnte. Dass auch Proband G1T2 in diesem Bereich nicht vollständig überzeugt war, lässt den Schluss zu, dass domänenspezifisches Vorwissen den Einfluss von vertrauensbildenden UI-Cues selbst in der Purpose-Dimension stark abschwächen kann (Visser et al., 2023).

Wie differenziert Nutzer diese Vertrauensdimensionen bewerten, zeigt sich prägnant am Fall von Proband G2T2. Obwohl dieser während des Prozesses Fehler in der Berechnung aufdeckte und daraufhin seinen Fragebogen hinsichtlich der Leistungsfähigkeit und Integrität massiv nach unten korrigierte, bewertete er das Statement zur Hinterhältigkeit weiterhin mit einer neutralen 3. Auch das Single-Item in Bezug auf die individuellen Bedürfnisse wurde mit einem hohen Wert von 4 beurteilt. Dies bestätigt die theoretische Grundannahme von Lee und See (2004) sowie Mayer et al. (1995), wonach Vertrauen ein strikt multidimensionales Konstrukt darstellt: Ein System kann als fachlich fehleranfällig, aber gleichzeitig als wohlwollend und nutzerzentriert wahrgenommen werden.

Bei der Interpretation dieser hohen Purpose-Werte muss jedoch das Forschungsdesign des Vibecodings kritisch eingeordnet werden. Das von den Nutzern gelobte Progressive Disclosure sowie die Forderung nach Usability-Standards und moderner Optik entsprangen keiner inhärenten Nutzerzentrierung der generativen KI, sondern wurden durch die Prompts des Entwicklers explizit induziert. Die vertrauensbildenden Cues wurden dem System folglich künstlich von außen vorgegeben. Das zugrunde liegende Sprachmodell agierte hierbei lediglich als stochastischer Papagei (Bender et al., 2021), der etablierte Design-Muster ohne eigenes empa-

thisches Verständnis reproduzierte. Der KI kann somit keine umfassende eigene intentionale Nutzerzentrierung attestiert werden. Dass die KI den semantischen und ökonomischen Kontext ihrer Ausgaben oft nicht durchdringt, zeigt sich prägnant an der von Proband G2T3 kritisierten irreführenden grünen Farbgebung für potenziell unrentable Ergebnisse. Dies belegt, dass unreflektiertes Vibecoding Design-Standards blind anwendet, was bei aufmerksamen Nutzern die Illusion des Wohlwollens durchbrechen kann und Warranted Distrust erzeugt (Visser et al., 2023).

Zusammenfassend offenbart die Diskussion der Purpose-Dimension das manipulative Potenzial KI-generierter Benutzeroberflächen. Durch gezieltes Vibecoding lassen sich etablierte HCI-Standards präzise implementieren, so dass sie beim Nutzer eine künstlich überhöhte Perceived Trustworthiness im Sinne eines vermeintlichen Wohlwollens erzeugen (Schlicker & Langer, 2021). Diese Illusion birgt die akute Gefahr eines Overtrust da die Nutzer der fehleranfälligen Blackbox, deren Existenz sie sich ohne KI-Hinweis nicht einmal bewusst sind, fälschlicherweise eine ehrliche Unterstützungsabsicht zuschreiben. Die vorliegenden qualitativen Ergebnisse verdeutlichen jedoch, dass das Ziel der Appropriate Reliance nach Lee und See (2004) nur dann erreicht werden kann, wenn diese emotionale Illusion durchbrochen wird (Schilke & Reimann, 2025). Dies gelingt in der Praxis entweder durch Vorwissen, eine hohe kognitive Wachsamkeit bei der Nutzung oder als direkter Effekt der bewussten Offenlegung der KI-Genese. Das Wissen um den generativen Ursprung der Anwendung wirkt hierbei als zwingend notwendiges Korrektiv: Es kann das durch die UI künstlich induzierten Affect-Based Trust in schützenden Warranted Distrust nach Visser et al. (2023) transformieren. Erst durch diese Etablierung einer funktionalen Skepsis gegenüber den tatsächlichen Absichten des Systems wird verhindert, dass Nutzer der wahrgenommenen Hilfsbereitschaft der Anwendung blind folgen und infolgedessen unreflektierte finanzielle Fehlentscheidungen treffen.

6.2 Probleme und Limitationen der Studie

Im Rahmen der vorliegenden Untersuchung ergaben sich einige Limitationen, Herausforderungen und Probleme, die es kritisch zu reflektieren gilt.

Ein zentrales Problem ist sowohl die Stichprobengröße als auch die Zusammensetzung der Studie. Die Stichprobengröße ist aufgrund des qualitativen Charakters der Studie und mangelnder Ressourcen auf $N = 8$ begrenzt. Die Rekrutierung erfolgte primär als Gelegenheitsstichprobe. Dies führte zu einer stark homogenen

Kohorte, die ausschließlich aus männlichen, technisch und digital affinen Personen in einer Altersspanne von Mitte 20 bis Mitte 30 bestand. Die Generalisierbarkeit der Ergebnisse auf die Gesamtbevölkerung oder auf weniger technikaffine Nutzergruppen ist folglich stark eingeschränkt. Zukünftige Forschungsarbeiten sollten die hier identifizierten Tendenzen durch Studiendesigns mit repräsentativen, diverseren Stichproben statistisch validieren.

Ein weiterer methodischer Einflussfaktor ergibt sich aus der gewählten Thinking-Aloud-Methode (Ericsson & Simon, 1984). Die ständige physische oder virtuelle Präsenz des Studienleiters, der zugleich als Entwickler der Anwendung fungierte, kann das Verhalten der Probanden beeinflusst haben. Es ist nicht auszuschließen, dass Effekte der sozialen Erwünschtheit auftraten. Zudem kann die Aufforderung, eigene Gedanken kontinuierlich zu verbalisieren, die Nutzer zu einer unnatürlich hohen kognitiven Reflexionsleistung animiert haben, was das Risiko eines unreflektierten Übervertrauens, z.B. durch Automation Bias, künstlich verringert haben könnte. Künftige Studien könnten dies umgehen, indem sie die Interaktion in unmoderierten Remote-Settings untersuchen, um ein natürlicheres Nutzungsverhalten zu erfassen.

Bei der Datenerhebung kam es zudem zu einem technischen Ausfall, durch den die Audioaufzeichnung des Interviews von Proband G1T3 verloren ging. Um den Datenverlust zu kompensieren, wurde umgehend ein detailliertes Gedächtnisprotokoll angefertigt. Obwohl die Kernaussagen gesichert werden konnten, schränkt dies die Zitiertiefe und die Analysemöglichkeiten für dieses spezifische Interview ein.

Auch bei der quantitativen Flankierung der Daten musste ein methodischer Kompromiss eingegangen werden. Der genutzte Trust in Automation-Fragebogen (Jian et al., 2000) wurde zugunsten einer einheitlichen Umfragestruktur und zur kognitiven Entlastung der Probanden von einer 7-stufigen auf eine 5-stufige Likert-Skala transformiert. Diese Anpassung schränkt die Konstruktvalidität und die direkte Vergleichbarkeit der quantitativen Werte mit anderen Studien ein. Zukünftige Mixed-Methods-Ansätze sollten auf validierte, standardisierte Skalen im Originalformat zurückgreifen, um die Aussagekraft des Fragebogens und die Vergleichbarkeit mit anderen Studien, die diesen Fragebogen nutzen, zu gewährleisten.

Abschließend muss die technische Eingrenzung der Studie betrachtet werden. Da explizit die Wahrnehmung und das Nutzervertrauen in eine durch Vibecoding generierte Software untersucht wurden, fand keine Evaluation der Actual Trustworthiness des PV-Amortisationsrechners statt. Die Studie misst ausschließlich die Perceived Trustworthiness der Nutzer.

7 Fazit und Ausblick

Ziel der vorliegenden Bachelorarbeit war es, zu untersuchen, inwiefern die Offenlegung der generativen KI-Herkunft einer Softwareanwendung das Nutzervertrauen und die Bewertung der Systemergebnisse beeinflusst. Anhand der Evaluation eines durch Vibecoding erstellten Photovoltaik-Amortisationsrechners konnte gezeigt werden, dass das Design der Benutzeroberfläche eine entscheidende und potenziell irreführende Rolle in der Mensch-Computer-Interaktion einnehmen kann. Die qualitativen Ergebnisse beantworten die zentrale Forschungsfrage, indem sie zeigen, dass die proaktive KI-Offenlegung das durch die Benutzeroberfläche künstlich generierte Vertrauen zwar massiv senkt, dieser Vertrauensverlust jedoch eine notwendige Schutzfunktion für den Anwender erfüllt, welche dem Nutzer seine eigene Vulnerabilität offenbart und diesem so eine Appropriate Reliance ermöglichen, welche weder zu Overtrust, noch zu Distrust neigt.

Die Synthese der Vertrauensdimensionen Purpose, Performance und Process (Lee & See, 2004) offenbart die weitreichenden Gefahren von KI-generierten Benutzeroberflächen. Die qualitativen Ergebnisse der unaufgeklärten Kontrollgruppe verdeutlichen, dass ein professionelles Interface-Design, gepaart mit etablierten Usability-Standards wie dem Progressive Disclosure nach Springer und Whittaker (2018), von Laien als Ersatzindikator für Systemkompetenz herangezogen werden kann. Obwohl das zugrundeliegende Sprachmodell lediglich als stochastischer Papagei agiert (Bender et al., 2021) und kein semantisches Verständnis für die eigentliche ökonomische Rentabilität besitzt, induzierte das Design eine hohe Perceived Trustworthiness und die Illusion von Wohlwollen. Da die tatsächliche Validität der Berechnungen für den Nutzer eine Blackbox darstellt (Schlicker & Langer, 2021), führte diese UI-Fassade bei einigen unaufgeklärten Nutzern zu einem initialen Overtrust.

Dieser Overtrust wurde lediglich durch Vorwissen, intensives Testen der Anwendung oder die Offenlegung der KI-Herkunft angemessen angepasst. Die aufgeklärten Probanden reagierten auf das KI-Label mit starker kognitiver Wachsamkeit und dem expliziten Wunsch, die Berechnungslogik des Systems manuell oder durch menschliche Fachexperten zu verifizieren. Dieser Effekt bestätigt das in der Literatur beschriebene Transparenz-Paradoxon (Buder et al., 2024; Schilke &

Reimann, 2025), wonach Transparenz über algorithmische Prozesse nicht zwingend zu mehr Vertrauen, sondern häufig zu einem sofortigen Vertrauensabbau führt. Der beobachtete Vertrauensverlust ist im Kontext generativer KI jedoch kein negativer Effekt, sondern die zwingende Voraussetzung für Appropriate Reliance. Das KI-Label transformiert das unreflektierte Affect-Based Trust in einen schützenden Warranted Distrust nach Visser et al. (2023). Erst diese funktionale Skepsis ermöglicht eine Appropriate Reliance (Lee & See, 2004), bei der Anwender die KI-generierten Ergebnisse nicht blind übernehmen, sondern als Schätzung kritisch hinterfragen.

Aus diesen Erkenntnissen leiten sich Implikationen für die Praxis und die zukünftige Forschung ab. Durch den aufkommenden Trend des Vibecodings sinken die technischen Hürden für die Softwareentwicklung drastisch. Zudem steigt die Verfügbarkeit von KI-Tools und Agenten, wodurch auch Nutzer mit weniger IT und Domänen-Kompetenz in der Lage sind, Software zu erzeugen. Es ist davon auszugehen, dass das Internet zeitnah mit optisch hochprofessionellen, aber inhaltlich geringfügig validierten oder vollständig ungeprüften Web-Anwendungen geflutet wird. Um Anwender in Hochrisiko-Bereichen, wie bei finanziellen Investitionsentscheidungen, vor unreflektierten Fehlentscheidungen zu schützen, deuten die Ergebnisse dieser Arbeit darauf hin, dass eine Kennzeichnungspflicht für KI-generierte Anwendungen sehr hilfreich wäre. Zukünftige Forschungsarbeiten sollten an diese qualitativen Befunde anknüpfen und den schützenden Effekt der KI-Offenlegung auf den Automation Bias durch quantitative, experimentelle Designs mit größeren, diverseren Stichproben statistisch validieren. Letztlich zeigt diese Untersuchung: In einer Zeit, in der KI-Modelle fehlerfrei simulieren können, vertrauenswürdig auszusehen, wird Warranted Distrust zu einer wichtigen Dimension in der Begegnung mit Software.

Literatur

- Andrej Karpathy [@karpathy]. (2025, 2. Februar). *There's a new kind of coding i call „vibe coding“, where you fully give in to the vibes, embrace exponentials, and forget that the code even exists. it's possible because the LLMs (e.g. cursor composer w sonnet) are getting too good. also i just talk to composer with SuperWhisper* [Twitter]. Verfügbar 17. April 2026 unter <https://x.com/karpathy/status/1886192184808149383>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Arx, S. v., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2021, 16. August). On the opportunities and risks of foundation models. <https://doi.org/10.48550/arXiv.2108.07258>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Buder, F., Hesel, N., & Dietrich, H. (2024). Beyond the buzz: Creating marketing value with generative AI. *NIM Marketing Intelligence Review*, 16(1), 50–55. <https://doi.org/10.2478/nimmir-2024-0008>
- Dijkstra, E. W. (1982). On the Role of Scientific Thought [Book Title: Selected Writings on Computing: A personal Perspective], 60–66. https://doi.org/10.1007/978-1-4612-5695-3_12
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data* [Pages: 426]. The MIT Press.
- Figma. (2025, 6. Mai). *Introducing Figma Make: A new way to test, edit, and prompt designs*. Verfügbar 17. April 2026 unter <https://www.figma.com/blog/introducing-figma-make/>
- Figma, I. (2025a). *Figma Make: Gestalten mit KI-gestützten Design-Tools* [Figma]. Verfügbar 8. Januar 2026 unter <https://www.figma.com/de-de/make/>
- Figma, I. (2025b, 24. April). *Figma-KI-Bericht 2025: Perspektiven aus der Sicht von Designer*innen und Entwickler*innen* | *Figma Blog* [Figma]. Verfügbar 17.

- April 2026 unter <https://www.figma.com/de-de/blog/figma-2025-ai-report-perspectives/>
- Fowler, M. (2018, 20. November). *Refactoring: Improving the design of existing code* [Google-Books-ID: 2H1_DwAAQBAJ]. Addison-Wesley Professional.
- Fraunhofer-IPA. (n. d.). *KI Agent – Einführung und Umsetzung* [Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA]. Verfügbar 9. Februar 2026 unter <https://www.ipa.fraunhofer.de/de/loesungen/digitalisierung-und-ki/strategie-und-smart-services/ki-agent.html>
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Pedreschi, D., & Giannotti, F. (2018, 21. Juni). A survey of methods for explaining black box models. <https://doi.org/10.48550/arXiv.1802.01933>
- Gulati, S., McDonagh, J., Sousa, S., & Lamas, D. (2024). Trust models and theories in human–computer interaction: A systematic literature review. *Computers in Human Behavior Reports*, 16, 100495. <https://doi.org/10.1016/j.chbr.2024.100495>
- Hunt, A., & Thomas, D. (2011). *The pragmatic programmer: From journeyman to master* (26. print). Addison-Wesley.
- Jian, J.-Y., Bisantz, A., & Drury, C. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics*, 4, 53–71. https://doi.org/10.1207/S15327566IJCE0401_04
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
- Martin, R. C. (2003). *Agile software development: Principles, patterns, and practices* [Google-Books-ID: 0HYhAQAAIAAJ]. Pearson Education.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, 20(3), 709–734. Verfügbar 17. April 2026 unter <http://www.jstor.org/stable/258792>
- McKinsey & Company. (2025, November). *The State of AI: Global Survey 2025* | McKinsey. Verfügbar 17. April 2026 unter <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>
- PVGIS24. (n. d.). *pvgis.com - PVGIS24 Solarrechner für Photovoltaik*. Verfügbar 17. April 2026 unter <https://pvgis.com/de>
- rechnerphotovoltaik.de. (n. d.). *Photovoltaik - Rechner, Informationen, Kosten, Planung*. Verfügbar 17. April 2026 unter <https://www.rechnerphotovoltaik.de/>
- Sayali, C., Heling, E., & Cools, R. (2023). Learning progress mediates the link between cognitive effort and task engagement.
- Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*, 188, 104405. <https://doi.org/10.1016/j.obhdp.2025.104405>

- Schlicker, N., & Langer, M. (2021). Towards warranted trust: A model on the relation between actual and perceived system trustworthiness [Conference Name: MuC '21: Mensch und Computer 2021 ISBN: 9781450386456]. *Mensch und Computer 2021*, 325–329. <https://doi.org/10.1145/3473856.3474018>
- Shore, J. (2004). Fail fast [software debugging]. *IEEE Software*, 21(5), 21–25. <https://doi.org/10.1109/MS.2004.1331296>
- Springer, A., & Whittaker, S. (2018). Progressive disclosure: Designing for effective transparency.
- Theobald, S., & Brandt, D. S. (2025, 27. November). *Vibe Coding verstehen: Definition, Potenziale und Risiken des KI-Trends - Blog des Fraunhofer IESE* [Fraunhofer IESE]. Verfügbar 8. Januar 2026 unter <https://www.iese.fraunhofer.de/blog/vibe-coding-definition-potenziale-risiken/>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017, 12. Juni). Attention is all you need. <https://doi.org/10.48550/arXiv.1706.03762>
- Visser, R., Peters, T. M., Scharlau, I., & Hammer, B. (2023, 4. Dezember). Trust, distrust, and appropriate reliance in (x)AI: A survey of empirical evaluation of user trust. <https://doi.org/10.48550/arXiv.2312.02034>
- Zednik, C. (2019). Solving the black box problem: A normative framework for explainable artificial intelligence. *Philosophy & Technology*, 34(2), 265–288. <https://doi.org/10.1007/s13347-019-00382-7>
- Zieglmeier, V., & Lehene, A. M. (2021). Designing trustworthy user interfaces. *33rd Australian Conference on Human-Computer Interaction*, 182–189. <https://doi.org/10.1145/3520495.3520525>

Anhang

Studienergebnisse

KI-Akzeptanz:						Antworten Aufgaben:							
Teilnehmer	Gruppe	KI-Vertrautheit	KI-Zuverlässig	# KI-Nutzung	# Figma Vertrauen	KI-Einfluss (Positiv)	Amortisation	# Gewinn (25J)	Monat max.	Unsicherheiten	Indiv. Bedürfniss		
G1T1	Gruppe 1	5	2	5	2	2	11,6	36993	Juni	Nein	5		
G1T2	Gruppe 1	4	1	2	2	1	11,6	18993	Juni, Juli	Nein	3		
G1T3	Gruppe 1	5	1	5	4	3	11,6	18993	Juli	Nein	5		
G1T4	Gruppe 1	5	1	4	3	2	11,6	18993	Juli	Nein	5		
G2T1	Gruppe 2	5	3	4	4	4	0	36975	Juni	Nein	5		
G2T2	Gruppe 2	4	3	5	2	2	11,6	18993	Juni/Juli	Nein	4		
G2T3	Gruppe 2	4	3	4	3	1	11,6	36993	Juni & Juli	Ja	3		
G2T4	Gruppe 2	4	1	2	4	2	11,6	18993	Juni / Juli	Nein	5		
Nicht umgepolte Ergebnisse:													
Teilnehmer	Trügerisch	# Hinterhältig	Misstrauisch	# Skeptisch	# Schädlich	# Vertrauen	# Sicherheit	# Integer	# Verlass	Zuverlässig	Vertrautheit		
G1T1	3	2	1	4	1	4	3	2	3	3	5		
G1T2	4	4	3	5	4	2	2	1	2	2	2		
G1T3	1	1	2	3	3	2	2	2	2	2	4		
G1T4	1	1	3	2	1	4	4	5	4	4	5		
G2T1	2	2	1	1	2	4	4	4	4	4	3		
G2T2	5	3	5	5	4	1	1	3	1	1	5		
G2T3	4	5	4	4	4	2	3	3	1	3	1		
G2T4	3	1	3	1	5	3	5	5	5	5	5		
Demographie etc:													
Teilnehmer	Alter	Bildung	Beruf	Wohnsituation	Infoquelle	# Interesse PV	Einst. Technik	# Einarbeitung	# Neue Tools ausprob.				
G1T1	18-29	Master+	IT	Mieter	Online/Video/Hersteller	3	5	5	5				
G1T2	18-29	Bachelor	Marketing	Eltern	Fachzeitschriften/Online	5	5	5	4				
G1T3	18-29	Bachelor	Informatik	Eltern	Online/Video/Foren	5	5	5	5				
G1T4	18-29	Ausbildung	IT	Mieter	Online/Social/Freunde	4	5	3	4				
G2T1	30-39	Abitur	Touristik	Mieter	Online/Video/KI-Chat	3	5	4	4				
G2T2	18-29	Bachelor	Consulting	Mieter	Video/Foren/KI-Chat	5	5	4	3				
G2T3	30-39	Ausbildung	E-Commerce	Eltern	Herstellerwebseiten	3	5	4	5				
G2T4	18-29	Techniker	Bau	Mieter	Herstellerwebseiten	4	5	5	4				

Instruktionsbogen

Arbeitsblatt

PV/Solar-Anlage Wirtschaftlichkeitsrechner – Studieninformationen

Vielen Dank für die Teilnahme an der Studie.

Das von dir Gesagte wird zu Zwecken der Analyse aufgezeichnet und im Nachhinein transkribiert. Antworte bitte möglichst ehrlich und offen. Der Zweck der Studie ist die Evaluation des Photovoltaik/Solar-Amortisations/Wirtschaftlichkeitsrechners, nicht wie schnell du die Aufgabe löst. Lass dir so viel Zeit, wie du benötigst. Halte dich nicht zurück mit Kritik und deinen Gedankengängen.

Wichtige Links:

[Umfrage](#) | [Rechner](#)

Aufgabenstellung

Stell dir vor, du planst eine PV-Anlage für dein Einfamilienhaus. Nutze den Rechner, um zu prüfen, ob sich die Anlage wirtschaftlich lohnt.

Bitte sprich während der Bearbeitung laut aus, was du denkst. Beschreibe deine Überlegungen, Entscheidungsgründe und eventuelle Unsicherheiten.

Alle nötigen Informationen befinden sich im folgenden Cheat Sheet. Bitte trage die Informationen in die dafür vorgesehenen Felder und folge den Anweisungen des Rechners. Für Fragen steht der Studienleiter während der gesamten Zeit zur Verfügung.

Die oben angezeigten Links führen dich zum PV-Rechner und zur auszufüllenden Umfrage. Dort gibt es ein paar demografische Fragen. Im Anschluss an die Bearbeitung gibt es einen Fragebogen, fülle diesen bitte ehrlich aus. In der Umfrage gibt es außerdem drei Felder zur Beantwortung dreier Fragen. Du kannst für das Zusammenrechnen der Module gerne deinen Handytaschenrechner oder den Windowstaschenrechner verwenden.

Im Anschluss an die Aufgabe würden wir mit dir gerne noch ein 10-20 Minütiges Interview zur Nutzung des Rechners durchführen.

Eingabedaten für den Rechner:

Schritt 1: Standort

Parameter	Wert
Standort der PV-Anlage	21217 Seevetal

Schritt 2: Dachkonfiguration

Dachseite 1

Parameter	Wert
Dachausrichtung	Südwest
Montageart	Aufdach
Modultyp	Monokristallin
Dachneigung	45 Grad
Installierte Leistung	19 × 0,4 kWp Module (zusammenrechnen)
Systemverluste	14 Prozent

Dachseite 2

Parameter	Wert
Dachausrichtung	Südost
Montageart	Aufdach
Modultyp	Monokristallin
Dachneigung	45 Grad
installierte Leistung	10 × 0,4 kWp Module (zusammenrechnen)
Systemverluste	14 Prozent

Schritt 3: Nutzungsdauer und Degradation

Parameter	Wert
Geplante Nutzungsdauer	25 Jahre
Leistungsabnahme pro Jahr	0,4 Prozent

Schritt 4: Kosten und Inflation

Parameter	Wert
Gesamtkosten	18.000 Euro
Zusätzliche Einmalkosten	0 Euro
Jährliche Betriebskosten	180 Euro
Erwartete Inflationsrate	2,5 Prozent

Schritt 5: Stromverbrauch und Vergütung

Parameter	Wert
Garantierte Vergütungsdauer	20 Jahre
Einspeisevergütung	7,9 Cent/kWh
Jährlicher Stromverbrauch	3.500 kWh
Aktueller Strompreis	35 Cent/kWh
Eigenverbrauchsquote	35 Prozent

Schritt 6: Finanzierung

Parameter	Wert
Kreditbetrag	0 Euro
Zinssatz	0 Prozent
Kreditlaufzeit	0 Jahre

Fragebogen

Finale Umfrage zur Usability und Akzeptanz eines PV-Rechners (Bache...

<https://docs.google.com/forms/d/1RSKy3a1EOrrvg4vMFZhNLs9Pcrjr...>

Finale Umfrage zur Usability und Akzeptanz eines PV-Rechners (Bachelorarbeit)

Diese Umfrage dient der Datenerhebung für eine Bachelorarbeit, die sich mit der Usability und der Akzeptanz eines Photovoltaik-Rechners auseinandersetzt. Alle Angaben werden anonym und vertraulich behandelt.

* Gibt eine erforderliche Frage an

1. Einverständniserklärung: Ich bin damit einverstanden, dass meine Antworten und die während der Aufgabenbearbeitung (Thinking Aloud) gemachten Äußerungen zu Analyse Zwecken aufgezeichnet, transkribiert und anonymisiert in der Bachelorarbeit verwendet werden dürfen. *

Markieren Sie nur ein Oval.

- ☐ Ja, ich stimme zu
- ☐ Nein, ich stimme nicht zu

Abschnitt 2:

Bitte Beantworte diesen Abschnitt vor der Bearbeitung der Aufgabe

2. 1. Altersgruppe *

Markieren Sie nur ein Oval.

- ☐ 18-29
- ☐ 30-39
- ☐ 40-49
- ☐ 50-59
- ☐ 60 oder älter
- ☐ Keine Angabe

3. 2. Höchster Bildungsabschluss *

Markieren Sie nur ein Oval.

- ☐ Weniger als Abitur/Fachabitur
- ☐ Abitur/Fachabitur
- ☐ Berufsausbildung
- ☐ Handwerksmeister
- ☐ Bachelor-Abschluss
- ☐ Master-Abschluss oder höher (z.B. Promotion)
- ☐ Betriebswirt o.Ä.
- ☐ Fachwirt o.Ä.
- ☐ Keine Angabe
- ☐ Sonstiges:

4. 3. Berufsfeld: *

5. 4. Wohnsituation *

Markieren Sie nur ein Oval.

- ☐ Eigentümer (Wohnung)
- ☐ Eigentümer (Haus)
- ☐ Mieter (Wohnung)
- ☐ Mieter (Haus)
- ☐ Wohnen bei den Eltern/Verwandten
- ☐ Studentenwohnheim/WG
- ☐ Andere
- ☐ Keine Angabe

6. 5. Bevorzugte Informationsquelle für teurere technische Anschaffungen *

Wählen Sie alle zutreffenden Antworten aus.

- ☐ Fachzeitschriften oder technische Magazine
- ☐ Online-Vergleichsportale
- ☐ YouTube-Videos oder andere Video-Plattformen
- ☐ Herstellerwebseiten
- ☐ Soziale Medien
- ☐ Empfehlungen von Freunden oder Bekannten
- ☐ Foren oder Online-Communities
- ☐ Artikel in Zeitungen oder Online-Medien
- ☐ Fachhändler o.Ä.
- ☐ KI-Chatbots
- ☐ Sonstiges: _____

7. 6. Ich bin grundsätzlich daran interessiert in Zukunft eine PV-Anlage zu erwerben/mieten (unabhängig ihrer aktuellen Wohnsituation) *

Markieren Sie nur ein Oval.

1 2 3 4 5

Stir ☐ ☐ ☐ ☐ ☐ Stimme vollumfänglich zu

8. 7. Ich habe generell eine positive Einstellung gegenüber neuen technologischen Entwicklungen. *

Markieren Sie nur ein Oval.

1 2 3 4 5

Stir ☐ ☐ ☐ ☐ ☐ Stimme vollumfänglich zu

9. 8. Es fällt mir im Alltag leicht, mich zügig in neue Software oder Webanwendungen einzuarbeiten. *

Markieren Sie nur ein Oval.

1 2 3 4 5

Stir ☐ ☐ ☐ ☐ ☐ Stimme vollumfänglich zu

10. 9. Ich probiere gerne neue technische Tools oder Funktionen aus, wenn sich mir die Gelegenheit bietet. *

Markieren Sie nur ein Oval.

1 2 3 4 5

Stir ☐ ☐ ☐ ☐ ☐ Stimme vollumfänglich zu

Abschnitt 3 Aufgaben:

Beantworte während/nach der Berechnung die folgenden Fragen:

11. Ergebnis 1: Wie lange dauert es, bis sich die PV-Anlage amortisiert hat (in Jahren)? *

12. Ergebnis 2: Wieviel Gewinn hat die Anlage über 25 Jahre erzielt (in Euro)? *

13. Ergebnis 3: Was ist der Monat mit der höchsten Stromproduktion? *

Abschnitt 4 Fragebogen:

Fülle bitte nach der Aufgabe diesen Fragebogen aus. Begründe bitte deine Entscheidungen

14. Inwieweit treffen die folgenden Aussagen auf das System zu? Begründe deine Entscheidungen *

Markieren Sie nur ein Oval pro Zeile.

	1 (Stimme überhaupt nicht zu)	2	3	4	5 (Stimme vollumfänglich zu)
Das System ist trügerisch.	☐	☐	☐	☐	☐
Das System verhält sich hinterhältig.	☐	☐	☐	☐	☐
Ich bin misstrauisch gegenüber den Absichten, Handlungen oder Ergebnissen des Systems.	☐	☐	☐	☐	☐
Ich stehe dem System skeptisch gegenüber.	☐	☐	☐	☐	☐
Die Handlungen des Systems werden schädliche oder nachteilige Folgen haben.	☐	☐	☐	☐	☐
Ich habe Vertrauen in das System.	☐	☐	☐	☐	☐

Das System bietet Sicherheit.	☐	☐	☐	☐	☐
Das System ist integer.	☐	☐	☐	☐	☐
Auf das System ist Verlass.	☐	☐	☐	☐	☐
Das System ist zuverlässig.	☐	☐	☐	☐	☐
Ich kann dem System vertrauen.	☐	☐	☐	☐	☐
Ich bin mit dem System vertraut.	☐	☐	☐	☐	☐

15. Gab es Momente, in denen du unsicher warst, wie du weitermachen sollst? *
Was hast du dann getan?

Markieren Sie nur ein Oval.

- ☐ Ja
☐ Nein

16. Sind dir Elemente der Oberfläche (Farben, Layout, Beschriftungen, Erklärungen) im Zusammenhang mit Zuverlässigkeit, Sicherheit oder Seriosität aufgefallen? Begründe bitte

17. Was ist deiner Ansicht nach die Hauptaufgabe, die dieser Rechner erfüllen soll?

18. Ich habe das Gefühl der Rechner ist bei der Nutzung auf meine individuellen Bedürfnisse, mein möglicherweise fehlendes Vorwissen und meine Ziele als Nutzer eingegangen. *

Markieren Sie nur ein Oval.

1 2 3 4 5

Stimm ☐ ☐ ☐ ☐ ☐ Stimme vollumfänglich zu

19. Sind Aufgabe und Interview abgeschlossen? *
- (Nach dem Interview gibt es noch ein paar Fragen)

Markieren Sie nur ein Oval.

- ☐ Ja: Gruppe 1 *Fahren Sie mit Frage 20 fort*
- ☐ Ja: Gruppe 2 *Fahren Sie mit Frage 23 fort*

Gruppe 1: Abschnitt 5 nach dem Interview

20. Gruppe 1: Du weißt, dass Oberfläche und Berechnungen mit Hilfe eines KI-Tools (Figma Make) erstellt wurden. Welche Gedanken löst dieses Wissen bei dir aus?

21. Der Einfluss von KI bei der Erstellung des PV-Rechners beeinflusst mein Vertrauen in das System positiv *

Markieren Sie nur ein Oval.

- 1 2 3 4 5
-
- Stirr ☐ ☐ ☐ ☐ ☐ Stimme vollumfänglich zu

22. Kennst du Figma Make und hast du schonmal mit Figma Make gearbeitet? *

Markieren Sie nur ein Oval.

- ☐ Ja, ich habe schonmal damit gearbeitet
- ☐ Ja, ich kenne es, aber habe noch nie damit gearbeitet
- ☐ Nein

Fahren Sie mit Frage 26 fort

Gruppe 2: Abschnitt 5 nach dem Interview

23. Gruppe 2: Sowohl die Oberfläche, welche du gerade bedient hast, als auch die Rechenformeln im Hintergrund wurden vollständig von dem KI-Tool Figma Make generiert.

Was geht dir dabei durch den Kopf?

24. Der Einfluss von KI bei der Erstellung des PV-Rechners beeinflusst mein Vertrauen in das System positiv *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Stir	☐	☐	☐	☐	☐	Stimme vollumfänglich zu

25. Kennst du Figma Make und hast du schonmal mit Figma Make gearbeitet? *

Markieren Sie nur ein Oval.

- ☐ Ja, ich habe schonmal damit gearbeitet
- ☐ Ja, ich kenne es, aber habe noch nie damit gearbeitet
- ☐ Nein

Fahren Sie mit Frage 26 fort

Abschnitt 6 Technik

26. Ich bin mit Künstlicher Intelligenz und deren Anwendungen (wie z. B. ChatGPT) im Allgemeinen vertraut *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Stirr	☐	☐	☐	☐	☐	Stimme vollumfänglich zu

27. Ich halte KI-Tools generell für sehr zuverlässig, wenn es darum geht, Aufgaben fehlerfrei auszuführen *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Stirr	☐	☐	☐	☐	☐	Stimme vollumfänglich zu

28. Ich nutze KI-Tools oder -Technologien regelmäßig für mein Studium oder meine Arbeit *

Markieren Sie nur ein Oval.

1 2 3 4 5

Stir ☐ ☐ ☐ ☐ ☐ Stimme vollumfänglich zu

29. Mein Vertrauen in die Ergebnisse von Figma Make ist hoch *

Markieren Sie nur ein Oval.

1 2 3 4 5

Stir ☐ ☐ ☐ ☐ ☐ Stimme vollumfänglich zu

Dieser Inhalt wurde nicht von Google erstellt und wird von Google auch nicht unterstützt.

Google

Formulare

Prompts

Prompts für den PV-Rechner:

Prompt 1: Kannst du mir eine benutzerfreundliche Website bauen, die Eingabefelder enthält für unterschiedliche KPIs, die am Ende etwas berechnen sollen. Gerne mit progressive disclosure als Grundphilosophie mit den folgenden Überschriften und Unterpunkte. die Überschriften sollen einen abgegrenzten progressive disclosure verschachtelten Bereich darstellen und die Unterpunkte sind Eingabe Felder:

Ort:

- Standort (API einbinden?) Berechnung aufschlüsseln:
https://joint-research-centre.ec.europa.eu/photovoltaic-geographical-information-system-pvgis/getting-started-pvgis/api-non-interactive-service_en

Photovoltaiksystem:

- Montageart
- Dachneigung
- Dachausrichtung
- Modultyp (Erklärungen zu Modultypen)
- Installierte Leistung
- Systemverluste

Geplante Nutzungsdauer:

- Nutzungsdauer in Jahren
- Degradation pro Jahr

Kosten:

- Gesamtkosten PV-Anlage
- Sonstige kosten
- Betriebskosten pro Jahr
- Inflationsrate pro Jahr

Einspeiservergütung:

- Dauer in Jahren
- Einspeiservergütung in ct pro kwh

Eigener Stromverbrauch:

- Eigener Jahresstromverbrauch
- Aktueller eigener Strompreis inklusive Grundpreis
- Eigenverbrauch in Prozent (Solarisator als Quelle nutzen/einbinden)

Finanzierung:

- Fremdkapital
- Zinssatz
- Laufzeit

Prompt 2:

kannst du eine Version machen, die einen nach und nach mit kleinen kurzen Einführungstexten durchleitet und bei der jede KPI einen kleinen Introtext bekommt

Prompt 3:

Das ist schon richtig gut, jetzt richte dich beim Theme noch ein bisschen mehr an die Discord-Farbpalette und gib mir am Ende bei den Ergebnissen so Monatsdurchschnittsdiagramme bei der Produktion an

Prompt 4:

kannst du das so verändern, dass man Schritt 2 mehrmals anketten kann, falls man mehrere Dachseiten mit Modulen belegen möchte

Prompt 5:

jetzt füge noch bei den Infolämpchen eine Erweiterung hinzu, so dass man auf die Lampe klicken kann und dort mehr Text hat, der eine ausführlichere Erklärung erlaubt

Prompt 6:

Bei Eigenverbrauchsquote gerne einmal einen Link einfügen: <https://solar.htw-berlin.de/rechner/unabhaengigkeitsrechner/> zum selber errechnen der Eigenverbrauchsquote. Es können gerne wie in der Version vorher kleine Infotexte stehen bleiben, aber wenn man auf die Info Lampe klickt, öffnen sich große Textfelder, in denen ausführlichere Infos stecken

Prompt 7:

kannst du mir noch ein hübsches Logo einfügen, was zu einem PV Rechner passt, was oben in der Mitte zentriert angezeigt wird, damit es professionell aussieht

Vor dem ersten Pre-Test:

Prompt 8:

Bitte lass die Sonne da, aber entferne die Animation. Ersetze die Überschrift durch das Logo, behalte die Unterschrift "Lassen Sie sich Schritt für Schritt durch die Berechnung führen" bitte bei. Entferne die Schrift Photovoltaik KPI Calculator und erweitere die Schrift PV-Rechner zu PV-Amortisationsrechner

Nach dem Pre-Test:

Prompt 9:

<https://www.gp-joule.com/de/>

Kannst du dich beim Farbschema etwas an gp joule orientieren. heller als jetzt. Außerdem nimm bitte die Logos wieder aus der Überschrift

Prompt 10:

Lass die Überschrift noch kleines bisschen offizieller und professioneller aussehen

Prompt 11: ich bin noch nicht zufrieden. Orientier dich bei der Überschrift bitte ein wenig an dieser Website: <https://pvgis.com/de> es soll aber stilistisch rein passen. Bitte füg außerdem irgendwo, wo es passt einen neu starten Button hinzu. Auch am Ende nach der Berechnung

Prompt 12:

Bitte füge auf der letzten Seite wieder eine Zurückfunktion ein, neu berechnen, soll aber bleiben und nenne den Button zum fertig stellen einer Dachseite: Dachseite abschließen

Prompt 13:

auf der letzten Seite fehlt der zurück Button. Bitte einfügen

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit mit dem Titel

Wahrnehmung KI-generierter Software: Auswirkungen der Offenlegung von KI-Genese auf das Nutzervertrauen

selbstständig und nur mit den angegebenen Hilfsmitteln verfasst habe. Alle Passagen, die ich wörtlich aus der Literatur oder aus anderen Quellen wie z. B. Internetseiten übernommen habe, habe ich deutlich als Zitat mit Angabe der Quelle kenntlich gemacht.

Hamburg, 23. April 2026