

**BACHELORARBEIT**

# **Künstliche Intelligenz und Immersion: Eine Nutzerstudie zu Large Language Models im Game Design**

---

vorgelegt am 14. März 2025  
Paul Sennack

Erstprüfer: Prof. Dr. Eike Langbehn  
Zweitprüfer: Simon Dewert

---

**HOCHSCHULE FÜR ANGEWANDTE  
WISSENSCHAFTEN HAMBURG**

Department Medientechnik  
Finkenau 35  
22081 Hamburg

## **Zusammenfassung**

Dynamische, durch Large Language Models (LLMs) generierte Inhalte finden zunehmend Anwendung im Game Design, insbesondere im Bereich der interaktiven Narration. Diese Arbeit untersucht den Einfluss narrativer LLM-Inhalte auf die Immersion in Videospielen, speziell in Bezug auf Dialogsysteme. Dazu wurde ein 2D-Spielprototyp entwickelt, in dem Spieler mit Nicht-Spieler-Charakteren (NPCs) interagieren konnten, deren Dialoge entweder statisch oder dynamisch wiedergegeben wurden. Erfahrungswerte der Spieler wurden mithilfe des Game Experience Questionnaire und des Presence Questionnaire gemessen und anschließend durch qualitative Rückmeldungen ergänzt. Die Ergebnisse zeigen, dass dynamische Dialogsysteme einen signifikanten Einfluss auf die Immersion der Spieler hatten. Spieler empfanden die Interaktion mit dynamischen NPCs als immersiver, wohingegen vor allem technische Limitierungen dieses immersive Spielerlebnis unterbrach. Diese Erkenntnisse unterstreichen das Potenzial, aber auch die Herausforderungen von LLM-generierten Inhalten im Game Design.

## **Abstract**

Dynamic content generated by Large Language Models (LLMs) is increasingly being used in game design, particularly in the field of interactive narration. This work examines the influence of narrative LLM content on immersion in video games, specifically in relation to dialogue systems. To this end, a 2D game prototype was developed in which players could interact with non-player characters (NPCs) whose dialogues were either static or dynamically generated. Player experiences were measured using the Game Experience Questionnaire and the Presence Questionnaire, and subsequently supplemented with qualitative feedback. The results show that dynamic dialogue systems had a significant impact on the players' immersion. Players found the interaction with dynamic NPCs to be more immersive, whereas technical limitations, in particular, interrupted this immersive gaming experience. These findings highlight the potential, but also the challenges, of LLM-generated content in game design.

## **Danksagungen**

Diesen Abschnitt möchte ich dafür nutzen, mich bei all denjenigen zu bedanken, die mich bei dieser Arbeit unterstützt haben.

Einen herzlichen Dank richte ich an meine Familie, meine Mutter, meinen Vater und meinen Bruder, für ihre bedingungslose Liebe und Unterstützung.

Einen Dank richte ich an meine Freunde, für ihre Zusprüche und Erfahrungen, die sie mit mir teilten.

Einen Dank richte ich an die Betreuer meiner Arbeit, für ihre Hilfestellung und Feedback.

Einen besonderen Dank richte ich an all die Freiwilligen, die sich Zeit genommen haben, an dieser Studie teilzunehmen. Namentlich danke ich Finn, Maik, Nicolas, John, Birgit, Finn, Sang und Kolja. Außerdem danke ich allen anonymen Teilnehmern.

Ohne euch wäre diese Arbeit nicht möglich gewesen.

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>III</b>
<b>Abkürzungsverzeichnis</b>	<b>IV</b>
<b>1 Einleitung</b>	<b>1</b>
<b>2 Grundlagen</b>	<b>2</b>
2.1 Immersion in Videospielen . . . . .	2
2.1.1 Definition und Bedeutung im Game Design . . . . .	2
2.1.2 Modelle und Theorien zur Immersion . . . . .	4
2.1.3 Einfluss von Narrativen auf die Immersion . . . . .	6
2.2 Grundlagen von Large Language Models . . . . .	8
2.2.1 Begriffserklärung und Einordnung . . . . .	8
2.2.2 Einsatz von KI in narrativen Strukturen . . . . .	10
2.2.3 Stärken und Schwächen dynamisch generierter Inhalte . . . . .	12
<b>3 Umsetzung</b>	<b>13</b>
3.1 Beschreibung des 2D-Spiels . . . . .	13
3.1.1 Wahl der Game Engine . . . . .	13
3.1.2 Aufbau der Spielumgebung . . . . .	14
3.1.3 Implementierung der Dialogsysteme . . . . .	15
3.2 Technische Umsetzung . . . . .	16
3.2.1 Auswahl des LLMs . . . . .	16
3.2.2 Begrenzungen und Kontrollmechanismen . . . . .	17
3.2.3 Herausforderungen und Lösungsansätze . . . . .	18
<b>4 Methodik</b>	<b>19</b>
4.1 Durchführung der Nutzerstudie . . . . .	19
4.1.1 Zielsetzung der Studie . . . . .	19
4.1.2 Demografische Merkmale der Stichprobe . . . . .	19
4.1.3 Aufbau der Nutzerstudie . . . . .	20
4.2 Vorstellung der Messinstrumente . . . . .	21
4.2.1 Quantitative Methoden: Fragebögen . . . . .	21
4.2.2 Qualitative Methoden: Offene Fragen . . . . .	22

<b>5</b>	<b>Analyse und Ergebnisse</b>	<b>23</b>
5.1	Analyse quantitativer Ergebnisse . . . . .	23
5.1.1	Core-Modul des GEQ - Allgemeine Spielerfahrung . . . . .	24
5.1.2	Social-Presence-Modul des GEQ . . . . .	25
5.1.3	Post-Game-Modul des GEQ . . . . .	26
5.1.4	Hauptkategorien des PQ - Wahrnehmung von Präsenz . . . . .	27
5.1.5	Subskalen des PQ . . . . .	28
5.1.6	Vergleich der statischen und dynamischen Dialoge . . . . .	29
5.1.7	Streuung und individuelle Unterschiede . . . . .	30
5.1.8	Signifikanz und Stärke des beobachteten Effekts . . . . .	32
5.2	Analyse qualitativer Ergebnisse . . . . .	32
<b>6</b>	<b>Diskussion</b>	<b>36</b>
6.1	Interpretation der Ergebnisse . . . . .	36
6.2	Kritische Reflexion und Herausforderungen . . . . .	37
<b>7</b>	<b>Fazit und Ausblick</b>	<b>40</b>
7.1	Wesentliche Erkenntnisse der Arbeit . . . . .	40
7.2	Implikationen fürs Game Design . . . . .	40
7.3	Ideen für weitere Forschung . . . . .	41
	<b>Literatur</b>	<b>42</b>

# Abbildungsverzeichnis

5.1	Mittelwerte und Standardabweichungen der Core-Kategorien im GEQ (eigene Darstellung, basierend auf den erhobenen Daten) . . . . .	24
5.2	Mittelwerte und Standardabweichungen der Social-Presence-Kategorien im GEQ (eigene Darstellung, basierend auf den erhobenen Daten) . . . . .	25
5.3	Mittelwerte und Standardabweichungen der Post-Game-Kategorien im GEQ (eigene Darstellung, basierend auf den erhobenen Daten) . . . . .	26
5.4	Mittelwerte und Standardabweichungen der Hauptkategorien im PQ (eigene Darstellung, basierend auf den erhobenen Daten) . . . . .	27
5.5	Mittelwerte und Standardabweichungen der Subskalen im PQ (eigene Darstellung, basierend auf den erhobenen Daten) . . . . .	28
5.6	Mittelwerte der Dialogsysteme im Vergleich (eigene Darstellung, basierend auf den erhobenen Daten) . . . . .	29
5.7	Streuung der Core- und Post-Game-Module des GEQ nach Kategorien (eigene Darstellung, basierend auf den erhobenen Daten) . . . . .	30
5.8	Streuung der Social-Presence-Module des GEQ nach Kategorien (eigene Darstellung, basierend auf den erhobenen Daten) . . . . .	31
5.9	Streuung des PQ nach Kategorien (eigene Darstellung, basierend auf den erhobenen Daten) . . . . .	32
5.10	Wortwolke mit häufig genannten Begriffen (eigene Darstellung, basierend auf den erhobenen Daten) . . . . .	35

# Abkürzungsverzeichnis

BERT	Bidirectional Encoder Representations from Transformers (dt. Bidirektionale Einbettungsrepräsentationen aus Transformern)
GEQ	Game Experience Questionnaire
GPT	Generative Pretrained Transformer (dt. generative vortrainierte Transformer)
KI	Künstliche Intelligenz
KNN	Künstliches neuronales Netz
LLM	Large Language Model (dt. großes Sprachmodell)
ML	Machine Learning (dt. maschinelles Lernen)
NLP	Natural Language Processing (dt. Verarbeitung natürlicher Sprache)
NPC	Non-Player-Character (dt. Nicht-Spieler-Charakter)
PaSSAGE	Player-Specific Stories via Automatically Generated Events (dt. spieterspezifische Geschichten durch automatisch generierte Ereignisse)
PQ	Presence Questionnaire
SD	Standard Deviation (dt. Standardabweichung)
SLM	Small Language Model (dt. kleines Sprachmodell)
VR	Virtual Reality (dt. virtuelle Realität)

# Gender-Hinweis

In dieser Arbeit wird aus Gründen der Lesbarkeit das generische Maskulin verwendet, welches sich auf alle vorkommenden Personenbezeichnungen und damit - sofern nicht anders angegeben - gleichermaßen auf alle Geschlechter bezieht.

# Kapitel 1

## Einleitung

Künstliche Intelligenz (KI) ist längst kein futuristisches Konzept mehr, sondern bereits fester Bestandteil des modernen Alltags. Der Fortschritt dieser Technologien hat in den letzten Jahren tiefgreifende Auswirkungen auf verschiedene kreative und technische Bereiche gezeigt. Besonders im Game Design (dt. Spielgestaltung) eröffnen KI-Modelle wie Large Language Models (LLMs; dt. große Sprachmodelle) neue Möglichkeiten, dynamisch angepasste und immersive Inhalte zu gestalten (Sweetser, 2024). So können sie neben der externen Inhaltsgenerierung direkt in Spiele integriert werden und interaktive Dialogsysteme sowie prozedural generierte Spielwelten unterstützen, die den Spielern das Eintauchen in virtuelle Welten ermöglichen.

Immersion beschreibt das Gefühl des vollständigen Eintauchens in eine andere Welt und ist ein zentraler Begriff im Game Design. Sie wird nicht nur durch visuelle und auditive Qualität beeinflusst, sondern hauptsächlich durch die narrative Kohärenz und die Interaktivität innerhalb der Spielwelt hervorgerufen (Jenkins, 2004). Hier bieten LLMs einen entscheidenden Vorteil: Sie können in Echtzeit kontextabhängige Inhalte erstellen, die den Handlungsverlauf eines Spiels dynamisch anpassen und somit individuellere Spielerfahrungen erzeugen. Der Einsatz von LLM-generierten Inhalten wirft jedoch auch kritische Fragen auf, die sowohl die technische Umsetzung als auch ethische Überlegungen betreffen. Zu den Herausforderungen gehören stereotypische Darstellungen, die Wahrung von kultureller Authentizität und die Automatisierung kreativer Prozesse. Darüber hinaus stellt sich die Frage, inwieweit LLMs die menschliche Kreativität überhaupt ergänzen oder gar ersetzen können, ohne dass die Qualität der Immersion darunter leidet.

Diese Arbeit beleuchtet diese Art der Inhaltsgenerierung in Bezug auf narratives Game Design und untersucht den Einfluss von LLM-generierten Inhalten auf die Immersion in Videospielen. Anhand einer Nutzerstudie werden Potenziale und Grenzen dargelegt, die den Wert von LLMs im Game Design unterstreichen.

# Kapitel 2

## Grundlagen

### 2.1 Immersion in Videospiele

Immersion wird von Spielern, Entwicklern und Forschern gleichermaßen als ein Kernpunkt in Videospiele angesehen und gilt häufig als ein entscheidender Faktor für den Spielspaß. Allgemein wird sie als das intensive Eintauchen in eine Spielwelt verstanden, wobei eine konkrete Definition häufig fehlt. Dieses Eintauchen kann durch Eigenschaften eines Spiels entweder gefördert oder gestört werden (Brown und Cairns, 2004). In der Forschung wird Immersion zudem oft als das Resultat einer gelungenen Spielerfahrung beschrieben – sie ist das Endprodukt, wenn alle Elemente eines Spiels harmonisch zusammenwirken und ein stimmiges Spielerlebnis bilden (Jennett et al., 2008).

#### 2.1.1 Definition und Bedeutung im Game Design

Spieler teilen ein gemeinsames Verständnis von Immersion, was jedoch keine einheitliche Erfahrung, sondern einen Grad der Bindung an ein Spiel beschreibt. Interessanterweise ist Immersion dabei keine notwendige Voraussetzung für den Spielspaß: Spieler entscheiden sich je nach ihrer aktuellen Stimmung bewusst für bestimmte Spiele, um spezifische Erfahrungen zu machen oder bestimmte emotionale Reaktionen hervorzurufen (Brown und Cairns, 2004). Ermi und Mäyrä (2005) ergänzen, dass die Auswahl eines Spiels oft auf der Erwartung basiert, dass es bestimmte emotionale oder kognitive Erlebnisse ermöglicht. So suchen Spieler gezielt nach Spielen, die ihren Bedürfnissen in einem bestimmten Moment entsprechen und optimale emotionale Reaktionen hervorrufen können (Ermi und Mäyrä, 2005), was verdeutlicht, dass Immersion zwar ein zentraler Bestandteil von Spielerfahrungen sein kann, aber hauptsächlich von den individuellen Motivationen und Situationen der Spieler abhängt.

Immersion findet Anwendung in zahlreichen Bereichen, wird jedoch am häufigsten im Zusammenhang mit Software, insbesondere Virtual Reality (VR; dt. virtuelle Realität) und Videospiele, diskutiert (Brown und Cairns, 2004). Auch in der Gestaltung von Benutzeroberflächen widmet man sich zunehmend der Frage, wie Gestaltungselemente noch immersiver in Anwendungen integriert werden können.

In der VR-Forschung wird die immersive Erfahrung auch durch das Konzept der Präsenz beschrieben. Präsenz definiert dabei den Grad, in dem die kognitiven und wahrnehmenden Systeme eines Spielers so beeinflusst werden, dass dieser den Eindruck gewinnt, sich an einem anderen Ort zu befinden (Patrick et al., 2000). In den Medienwissenschaften wird Präsenz ähnlich interpretiert: als das psychologische Gefühl, in einer computergenerierten Welt zu sein, anstatt lediglich einen Computer zu nutzen (Ermi und Mäyrä, 2005). Immersion und Präsenz werden dabei häufig synonym verwendet, obwohl sich die Begriffe in Nuancen unterscheiden: Während Präsenz stärker mit dem metaphorischen *Transport* in eine andere Welt assoziiert ist (McMahan, 2003), bezieht sich Immersion stärker auf die psychischen Prozesse, die das Spielerlebnis fördern (Ermi und Mäyrä, 2005). Immersion wird daher weitestgehend bevorzugt, da sie die kognitive und emotionale Einbindung der Spieler genauer beschreibt. Als Gemeinsamkeit zeigen Studien allerdings, dass eine erhöhte kognitive Anstrengung, wie beispielsweise durch gesteigerte Schwierigkeit im Spiel, sowohl die Immersion als auch das Gefühl der Präsenz verstärken kann (Ravaja et al., 2004).

Auch zeigt Immersion deutliche Parallelen zur Flow-Theorie von Csikszentmihalyi (2004), da beide Konzepte auf der Fokussierung der Aufmerksamkeit beruhen und sowohl die Wahrnehmung von Zeit als auch des Selbst beeinflussen können (Brown und Cairns, 2004; Csikszentmihalyi, 1990). Wie bei Flow (dt. Fluss) wird auch bei Immersion der gewünschte Zustand erreicht, wenn das Können und Wissen des Spielers mit der Herausforderung des Spiels übereinstimmt (Ermi und Mäyrä, 2005). Dieses Gleichgewicht ist besonders in Videospielen erlebbar, da sich die Schwierigkeit häufig dynamisch an den Fortschritt der Spieler anpasst und diese dadurch an die Grenzen ihrer Fähigkeiten bringen (Ermi und Mäyrä, 2005). Ein wesentlicher Unterschied zwischen Immersion und Flow liegt jedoch in der Art der Erfahrung: Während Immersion flüchtig sein kann und kurzfristig aufrechterhalten werden muss, wird Flow bei Ablenkung vollständig unterbrochen und lässt sich nur durch erneutes Eintauchen wiederherstellen (Brown und Cairns, 2004; Csikszentmihalyi, 1990).

In Spielrezensionen sowie Beschreibungen von Kindern findet man häufig realistische Spielwelten oder atmosphärische Sounds als zentrale Merkmale von Immersion (Ermi und Mäyrä, 2005). Auch wird sie oft mit einer gewissen Tiefe in Verbindung gebracht, die das Spielerlebnis intensiver und glaubwürdiger erscheinen lässt. Dennoch gibt es Spiele, die trotz hochwertigen Grafik- und Sounddesigns keine immersive Wirkung bieten (Swing, 2000). Daraus lässt sich schließen, dass audiovisuelle Elemente zwar eine wichtige Rolle spielen, aber nicht allein ausschlaggebend für Immersion sind. So können realistische Grafiken und Sounds eher als eine Art Voraussetzung statt als Bedingungen gesehen werden. Audiovisuelle Komponenten wirken vielmehr unterstützend, indem sie die Glaubwürdigkeit und Ästhetik eines Spiels verbessern, während Immersion durch ein Zusammenspiel mit anderen Faktoren wie Narration, Herausforderung oder Spielbarkeit entsteht (Ermi und Mäyrä, 2005).

Eine ebenso wichtige Rolle spielt die Bedienbarkeit eines Spiels, insbesondere im Hinblick auf Steuerung und Eingabemethoden. Laut Brown und Cairns (2004) ist die grundlegende Form von Immersion nicht erreichbar, wenn es Probleme bei der Bedienbar-

keit gibt, wobei sie betonen, dass die Eingabemethoden so gestaltet sein müssen, dass sie für den Spieler *unsichtbar* werden (Brown und Cairns, 2004). Diese Unsichtbarkeit bedeutet, die Steuerung intuitiv und mühelos funktionieren zu lassen, damit Spieler vollständig in das Spiel eintauchen können. Bedienbarkeit bildet somit eine Basis, auf der Spieler die Spielwelt intuitiv erkunden und sich aktiv in sie einbringen können. Ermi und Mäyrä (2005) unterstreichen die Bedeutung der aktiven Teilnahme für immersive Erlebnisse. Spieler müssen selbst tätig werden und sich aktiv ins Spielgeschehen einbringen, um eine starke emotionale und kognitive Bindung zu erfahren. Diese aktive Beteiligung bestimmt im Wesentlichen, wie das immersive Zusammenspiel zwischen Spieler und Spielumgebung abläuft, selbst in Spielen mit weniger beeindruckender Grafik (Ermi und Mäyrä, 2005).

Immersive Erfahrungen sind komplexe Zusammenspiele aus Gedanken, Gefühlen und Handlungen und sind daher stark kontextabhängig. Sie entstehen nicht als vorprogrammierte Ereignisse, sondern als das Ergebnis einer einzigartigen Interaktion zwischen dem Spieler und dem Spiel. Jeder Spieler bringt eigene Wünsche, Erwartungen und Erfahrungen mit, die maßgeblich beeinflussen, wie die Spielwelt erlebt und interpretiert wird (Ermi und Mäyrä, 2005). Die gleiche Aktivität kann in einem bestimmten Kontext als positiv wahrgenommen werden, während sie in einem anderen unattraktiv erscheinen könnte (DeJean, 2002). Diese Kontextabhängigkeit unterstreicht, dass Spielerlebnisse nicht standardisiert sind, sondern von externen Umständen, persönlichen Präferenzen und individuellem Zugang beeinflusst werden.

### 2.1.2 Modelle und Theorien zur Immersion

Die genaue Bedeutung von Immersion ist schwer zu greifen und bleibt oft vage. Obwohl der Begriff regelmäßig verwendet wird, ist unklar, ob dabei stets das gleiche Konzept gemeint ist. Wie Brown und Cairns (2004) betonen, scheint es zwar ein gemeinsames Verständnis von Immersion in der Gaming-Community (dt. Spielgemeinschaft) zu geben, doch bleibt offen, welche Elemente dieses Phänomen ausmachen und wodurch es hervorgerufen wird (Brown und Cairns, 2004). Ebenso heben Ermi und Mäyrä (2005) hervor, dass der Begriff zwar weit verbreitet ist, jedoch häufig unspezifisch und ohne Bezug zu konkreten Erfahrungen genutzt wird (Ermi und Mäyrä, 2005). Jennett et al. (2008) schließen daraus, dass es notwendig ist, Immersion eindeutiger zu definieren und systematisch zu messen, um die Erfahrungen besser vergleichen und einordnen zu können (Jennett et al., 2008).

Im Rahmen einer qualitativen Studie entwickelten Brown und Cairns (2004) eine *grounded theory* (Corbin und Strauss, 2014), die die Eigenschaften von Immersion im Kontext von Videospiele beschreibt. Ziel dieser Studie war es, auf Basis von Interviews mit Spielern eine robuste Grundlage zu schaffen, die als Grundlage oder Referenz für zukünftige Konzepte dienen kann. Dabei wird Immersion als graduelles Erlebnis verstanden, das sich durch verschiedene Stufen der Selbstbeteiligung charakterisieren lässt und je nach Kontext und Spielerfahrung gemindert oder verstärkt wird (Brown und Cairns, 2004). Brown und Cairns (2004) identifizieren drei Stufen der Immersion, die durch unterschiedliche Hürden (engl. „barriers“) voneinander abgegrenzt werden: Bindung (engl.

„engagement“), Vertiefung (engl. „engrossment“) und vollständige Immersion (engl. „total immersion“). Diese Stufen sind nicht statisch, sondern repräsentieren den Grad des Eintauchens in das Spiel. Die Hürden, die Spieler und Spiele überwinden müssen, sind dabei auf menschliche und spielspezifische Faktoren, wie Konzentration oder Atmosphäre, zurückzuführen. Das Überwinden einer Hürde ermöglicht den Übergang zur nächsten Stufe, garantiert ihn jedoch nicht (Brown und Cairns, 2004).

Bindung ist die erste Stufe und beschreibt den niedrigsten Grad der Immersion. Um sie zu erreichen, muss der Spieler bereit sein, Zeit, Mühe und Aufmerksamkeit zu investieren. Die Bedienbarkeit muss dafür intuitiv sein, damit der Spieler ein Interesse am Spiel entwickeln kann. Ist die Hürde einmal überwunden, verspürt der Spieler den Wunsch, weiterzuspielen. Die zweite Stufe, Vertiefung, bindet den Spieler emotional an das Spiel. Sie kann erreicht werden, wenn Spielelemente, wie Grafik und Handlung, ansprechend kombiniert sind. Aufgrund der bis dahin investierten Zeit und Aufmerksamkeit ist der emotionale Einsatz hoch und Zweifel an der Glaubwürdigkeit der Spielwelt werden zunehmend abgelegt. Die Bedienbarkeit muss dafür reibungslos funktionieren, um dem Spieler zu ermöglichen, sich völlig auf die Erfahrung einzulassen. Die höchste Stufe bildet die vollständige Immersion, eine seltene und flüchtige Erfahrung, bei der der Spieler absolut von der realen Welt abgetrennt ist. Auf dieser Stufe richten sich die gesamte Aufmerksamkeit, alle Gedanken und Gefühle auf das Spiel. Die Überwindung der Hürde ermöglichen hier die Empathie und die Atmosphäre. Empathie entsteht durch die emotionale Bindung an die Charaktere oder die Spielwelt, während Atmosphäre durch eine glaubwürdige und kohärente Spielwelt geschaffen wird. Vollständige Immersion ist instabil und kann durch Ablenkung leicht unterbrochen werden (Brown und Cairns, 2004).

Ein zentraler Aspekt des Modells ist die Rolle der Aufmerksamkeit, die sich in visuelle, auditive und mentale Komponenten aufteilen lässt. Die Intensität und Anzahl der Aufmerksamkeitsquellen beeinflussen, wie tief der Spieler in die Immersion eintauchen kann. Spieler in höheren Stufen der Immersion verlieren dabei zunehmend das Gefühl für Zeit und ihre reale Umgebung, da ihre gesamte Aufmerksamkeit auf das Spiel gerichtet ist (Brown und Cairns, 2004).

Das Modell von Brown und Cairns hat die Forschung zur Immersion maßgeblich beeinflusst, da es erstmals eine klare Struktur für die Analyse von Spielerfahrungen bietet. Es zeigt, dass Immersion nicht binär zu verstehen ist, sondern dynamisch und von individuellen und kontextuellen Faktoren beeinflusst wird. Gleichzeitig legt das Modell den Grundstein für weitere theoretische Ansätze für Forscher und Spieleentwickler.

Ermi und Mäyrä (2005) verfolgten eine andere Herangehensweise an die Immersion, als sie ein Modell entwickelten, das die zugrundeliegenden Komponenten von Spielerfahrungen systematisch darstellt. Ihr Ziel war es, die Vielschichtigkeit von Spielerlebnissen zu strukturieren und Immersion als eine der Schlüsseldimensionen dieser Erfahrungen genauer zu analysieren. Ihr Ansatz hebt besonders die Interaktivität hervor, da ihre Probanden die Möglichkeit, Entscheidungen zu treffen und direkten Einfluss auf das Spielgeschehen zu nehmen, als einer der immersivsten Aspekte genannt wurde (Ermi und Mäyrä, 2005). Das sogenannte SCI-Modell teilt die Spielerfahrung in drei

Dimensionen auf: sensorische Immersion (engl. „sensory immersion“), herausforderungsbezogene Immersion (engl. „challenge-based immersion“) und imaginative Immersion (engl. „imaginative immersion“). Diese Dimensionen haben Überschneidungen und tragen in unterschiedlicher Intensität zu einem immersiven Spielerlebnis bei (Ermi und Mäyrä, 2005).

Sensorische Immersion bezieht sich auf die audiovisuellen Aspekte eines Spiels und ermöglicht das Eintauchen über die Sinne. Elemente wie realistische Grafiken, atmosphärische Sounds und gut funktionierende Kameraeinstellungen werden als entscheidend für das Gefühl des Eintauchens in die Sinneswahrnehmung genannt. Herausforderungsbezogene Immersion entsteht durch das Zusammenwirken der Fähigkeiten des Spielers und den Herausforderungen des Spiels. Dieses Gefühl kommt auf, wenn motorische und mentale Fähigkeiten gleichermaßen gefordert werden, etwa durch strategisches Denken, schnelle Reaktionen oder das Lösen logischer Probleme. Die Freude am Spiel ist hierbei eng mit dem Erfolgserlebnis und dem Fortschritt im Spiel verbunden: Unsicherheiten über den Ausgang einer Herausforderung verstärken die Spannung und fördern die Immersion. Videospiele fördern diesen Zustand besonders, da sie die Herausforderung dynamisch an das Können der Spieler anpassen können. Die imaginative Immersion beschreibt das Eintauchen in die narrative oder kreative Dimension eines Spiels. Sie entsteht, wenn Spieler sich mit Charakteren identifizieren oder sich in die Geschichten vertiefen. Diese Dimension bietet den Spielern die Möglichkeit, ihre Vorstellungskraft zu nutzen, Empathie für die Figuren zu entwickeln oder sich in Szenarien auszuleben, die im realen Leben nicht möglich oder akzeptabel wären (Ermi und Mäyrä, 2005).

Das SCI-Modell liefert eine strukturierte Analyse der immersiven Spielerfahrung und zeigt, wie unterschiedliche Dimensionen zusammenwirken, um ein intensives Spielerlebnis zu erschaffen. Dabei verdeutlicht es, dass Immersion nicht allein von einem einzelnen Faktor abhängt, sondern von einer Kombination aus Sinneswahrnehmungen, Herausforderung und narrativer Elemente. Wichtig ist dabei, dass die Dimensionen nicht isoliert für sich betrachtet werden müssen: Gut integrierte Spiele verbinden sensorische, herausforderungsbezogene und imaginative Elemente, um ein immersives Gesamterlebnis zu erzeugen. Andererseits können so auch Spiele, die einen Aspekt weniger abdecken, als immersiv wahrgenommen werden.

Im Gegensatz zum Modell von Brown und Cairns (2004), das Immersion als einen graduellen Prozess beschreibt, legt das SCI-Modell den Fokus auf das Zusammenwirken unterschiedlicher Dimensionen. Während Brown und Cairns (2004) die Hürden und Übergänge zwischen den Stufen betonen, zeigt das SCI-Modell, dass Immersion auf verschiedenen Ebenen gleichzeitig erlebt werden kann.

### **2.1.3 Einfluss von Narrativen auf die Immersion**

Immersive Erlebnisse beruhen nicht zwangsläufig auf einer narrativen Handlung, können durch diese allerdings verstärkt werden. Ein zentraler Unterschied zwischen Videospiele und traditionellen Medien besteht in der Interaktivität, die es Spielern ermöglicht, direkten Einfluss auf die Handlung zu nehmen. Narrationen in Spielen stellen Spieler

nicht linear als passive Konsumenten dar, sondern eröffnen emotional und intellektuell fesselnde Welten, mit denen sie interagieren können (Jenkins, 2004). Diese Interaktivität hebt Spiele als ein Medium hervor, das über die bloße Wiedergabe von Geschichten hinausgeht. Frühere Kritiken sahen Narrative und Gameplay als vermeintliche Gegensätze, da Narrative darauf abzielen, eine Geschichte zu erzählen, während Gameplay auf die Autonomie der Spieler fokussiert ist (Adams, 1999; Juul, 2005). Jenkins (2004) widerspricht dieser Sichtweise und argumentiert, dass Spiele die Grenzen traditioneller Erzählformen erweitern. Durch ihre abstrakte und experimentelle Natur erschaffen sie Möglichkeiten, Geschichten zu erzählen und Spieler einzubinden. Er beschreibt dieses Konzept als *narrative Architektur*. Anstatt Geschichten linear zu präsentieren, entwerfen Spieleentwickler Räume, in denen Spieler mit narrativen Elementen interagieren und ihre eigenen Geschichten ausspielen können (Jenkins, 2004).

Unter dem Begriff des *Environmental Storytelling* (dt. Umgebungsnarration) fasst Jenkins (2004) zusammen, wie sich narrative Elemente in der Spielwelt nutzen lassen, um Spieler auf unterschiedliche Weise immersiv in die Handlung einzubinden. Es umfasst vier zentrale narrative Bereiche: assoziative Räume (engl. „evocative spaces“), interaktive Erzählungen (engl. „enacting stories“), narrative Einbettung (engl. „embedding narratives“) und spontan entstehende Erzählungen (engl. „emergent narratives“) (Jenkins, 2004).

Assoziative Räume greifen auf die vorherigen Erfahrungen der Spieler zurück und erweitern diese. Bekannte Charaktere, Orte oder Geschichten - wie beispielsweise in Lizenzspielen von Star Wars oder Harry Potter - bauen auf der emotionalen Bindung des Spielers an ein Franchise auf. Dabei wird keine neue Geschichte geliefert, sondern das bestehende Wissen des Spielers bereichert. Interaktive Erzählungen erlauben es Spielern, aktiv an der Handlung teilzunehmen, indem sie narrative Ereignisse ausspielen. Sie erledigen Aufgaben oder lösen Rätsel, wobei die spezifische Ausführung oder Reihenfolge dabei weniger entscheidend ist. Bei diesem Ansatz liegt lediglich ein grob definiertes Ziel vor, welches Spieler auf ihre Weise erreichen können. Narrative Einbettung vermittelt die Geschichte direkt durch die Spielwelt: Spieler sammeln Informationen, um die Handlung zu rekonstruieren oder vor auszusehen. Handlungsrelevante Gegenstände, wiederkehrende Orte oder Rückblicke lenken die Narration und ermuntern Spieler zur Erkundung der Spielwelt. Spontan entstehende Erzählungen liefern die größte Freiheit und Flexibilität: Hier gibt es kaum oder gar keine vorgefertigte Geschichte, stattdessen entsteht die Handlung durch die Interaktion der Spieler mit der Spielwelt. Sie setzen eigene Prioritäten oder kombinieren Spielelemente auf unvorhersehbare Weisen, wodurch individuelle Geschichten entstehen (Jenkins, 2004). Beispiele dafür sind Sandbox-Spiele wie Minecraft, in denen Spieler ihre eigene Fantasien ausleben können.

Jenkins' Konzept der narrativen Architektur verdeutlicht, dass Narrative in Spielen weit über das Erzählen von Geschichten hinausgeht. Mit den vier Bereichen des Environmental Storytelling bietet er eine Grundlage, die ein Grundgerüst für immersive Spielerlebnisse darstellt.

## 2.2 Grundlagen von Large Language Models

Künstliche Intelligenz (KI) ist ein umfassender Begriff und bezeichnete ursprünglich den Versuch, menschliche Intelligenz mit Maschinen nachzubilden. Ab den 1980er-Jahren entwickelte sich das Feld in verschiedene Richtungen: Während der ursprüngliche Ansatz weiterhin verfolgt wird, entsteht mit dem Machine Learning (ML; dt. maschinelles Lernen) ein neuer Forschungsbereich, der sich auf das datengetriebene Lösen komplexer Probleme konzentrierte (Langley et al., 2011). Trotz dieser unterschiedlichen Ansätze und der unscharfen Definition von KI werden die Begriffe KI und ML umgangssprachlich oft synonym verwendet. ML bezeichnet den Ansatz, Probleme mithilfe von Algorithmen zu lösen, die zu komplex sind, um durch einfache Regeln beschrieben zu werden. Dabei werden basierend auf vorhandenen Daten Muster erkannt, die sich zur Vorhersage unbekannter Daten und auf neue Probleme verallgemeinern lassen. Obwohl ML besonders in der Computerlinguistik für Sprach- und Texterkennung eingesetzt wird, kommt es auch in anderen technischen Bereichen zum Einsatz.

### 2.2.1 Begriffserklärung und Einordnung

Machine Learning (ML) basiert auf einem Trainingsprozess, bei dem ein mathematisches Modell durch einen Lernalgorithmus angepasst wird, sodass es anschließend Vorhersagen treffen, Empfehlungen geben oder auf Eingaben reagieren kann (Döbel et al., 2018). Dieses Verfahren erfordert häufig große Mengen an Trainingsdaten, aus denen der Algorithmus Muster ableitet. Die Trainingsdaten können unterschiedliche Formate haben und ermöglichen es dem Algorithmus, die Modellparameter so anzupassen, dass das Modell nach Abschluss des Lernprozesses Lösungen für unbekannte Datensätze generieren kann. Das Lernen erfolgt in drei Hauptformen, die je nach Methode unterschiedliche Vorteile bieten und auf den jeweiligen Anwendungszweck des Modells abgestimmt sind. Die häufigste Methode ist das überwachte Lernen, bei dem die Trainingsdaten mit als korrekt geltenden Lösungen, sogenannten *Labels* (dt. Beschriftungen), versehen sind (Frochte, 2020). Soll ein Modell beispielsweise lernen, das Bild einer Blume mit dem Wort *Blume* zu verknüpfen, wird ihm die korrekte Zuordnung mitgeteilt. Beim unüberwachten Lernen sind den Trainingsdaten keine Labels zugeordnet; im Gegensatz zum überwachten Lernen muss der Algorithmus selbstständig Strukturen und Muster erkennen. Dieser Vorgang nennt sich *Clustern* und wird beispielsweise vorbereitend auf überwachtes Lernen eingesetzt (Ghahramani, 2003). Beim bestärkenden Lernen beobachtet das Modell seine Umgebung und reagiert autonom darauf. Diese sogenannten *Agenten* erhalten je nach Aktion verschiedene Belohnungen in Form von Werten und entwickeln über zahlreiche Iterationen eine Strategie zur Problemlösung (Kaelbling et al., 1996). Da der Agent üblicherweise nach dem Prinzip von Versuch und Irrtum handelt, hängt der Erfolg stark vom Kontext ab und erfordert oft erhebliche Rechenleistung sowie Zeit. Ein typisches Beispiel ist ein Saugroboter, der seinen Reinigungsweg optimiert, um die aufgenommene Staubmenge zu maximieren.

Der Einsatz eines trainierten Modells hängt von seinem Verwendungszweck ab. Generative Modelle können nicht nur Vorhersagen treffen, sondern auch neue Daten generie-

ren, die den Trainingsdaten ähneln. Eine dieser Modellklassen sind künstliche neuronale Netze (KNN). Sie gehören zu den nicht-symbolischen Ansätzen des ML und zeichnen sich durch ihre hochkomplexen Internen aus, die eine Nachvollziehbarkeit ihrer Entscheidungen mittels einfacher Formeln erschweren (Langley et al., 2011). KNN basieren auf neuronalen Netzwerken, bei denen jeder Knotenpunkt eine eigenständige Entscheidung trifft, sodass das Netzwerk als Ganzes zu einer Lösung gelangt. Jedes dieser künstlichen Neuronen nutzt eine Aktivierungsfunktion, die bestimmt, wie auf eingehende Signale reagiert wird. Die Gewichte, die die Relevanz eines Eingangssignals bestimmen, werden während des Trainings iterativ angepasst, sodass das Modell eine Lösung erstellen kann. Eine der ersten bedeutenden Entwicklungen in diesem Bereich war die McCulloch-Pitts-Zelle, die 1943 von Warren McCulloch und Walter Pitts als erstes künstliches binäres Neuron entwickelt wurde. Das Modell konnte einfache binäre Entscheidungen treffen und logische UND-, ODER- und NICHT-Gatter der booleschen Algebra darstellen. Die gewichteten Eingangssignale wurden dabei aufsummiert und in ein Ausgangssignal verwandelt, falls ihre Summe einen vorgegebenen Schwellenwert überschreitet (Rojas, 2013). Neuronale Netze sind entsprechend der Position und Funktion ihrer Neuronen in drei Schichten aufgebaut: Die Eingangsschicht empfängt die Eingabedaten und die Ausgangsschicht gibt die berechnete Lösung aus. Die dazwischenliegenden, versteckten, Schichten bestimmen die Tiefe des Netzwerks und ermöglichen mit mehreren Verarbeitungsebenen die Lösung komplexer Probleme. Das *Deep Learning* (dt. tiefes Lernen) beschreibt Methoden zum Trainieren tiefer neuronaler Netze (Hardesty, 2017). Der Begriff wurde erstmals 1986 von Rina Richter im Zusammenhang mit ML verwendet. In ihrem Artikel beschrieb sie, wie ein unzureichender Lösungsweg dokumentiert werden kann, um in zukünftigen Versuchen schneller zur korrekten Lösung zu gelangen (Dechter, 1986).

Die Verarbeitung von Informationen in einem künstlichen neuronalen Netz (KNN) wird durch dessen Architektur bestimmt. Einige Architekturen sind speziell auf die Verarbeitung von Bilddaten ausgelegt, während andere die Grundlage für Sprachmodelle wie Generative Pretrained Transformers (GPTs; dt. generative vortrainierte Transformer) bilden. Diese ermöglichen es, Texte zu generieren, zusammenzufassen oder zu übersetzen. Eine der bedeutendsten Entwicklungen in diesem Bereich ist die 2017 von Google vorgestellte Transformer-Architektur. In ihrem einflussreichen Paper *Attention Is All You Need* beschrieben Vaswani et al. (2017) den für diese Architektur einzigartigen Aufmerksamkeitsmechanismus, der es ermöglicht, bestimmten Sequenzen innerhalb eines Datensatzes mehr Gewicht zu verleihen als anderen (Vaswani et al., 2017). Dieses Verfahren reduziert den Rechenaufwand erheblich und bildet die Grundlage für moderne Sprachmodelle (Géron, 2023; Levy, 2024). Transformer verarbeiten Texte, indem sie diese in kleine Segmente unterteilen und in numerischer Form darstellen. Dieser Prozess wird als Tokenisierung bezeichnet, wobei sogenannte Token entstehen – kleinere Textabschnitte, die mithilfe von Worteinbettungen in mathematische Vektoren umgewandelt werden. Dieses Vorgehen basiert auf Word2Vec, einem Verfahren zur Darstellung von Wörtern als mehrdimensionale Vektoren (Mikolov et al., 2013). Durch die Einbettung in einen mehrdimensionalen Raum kann das Modell dann eine sinnvolle Vorhersage über das nächste Wort in einer Sequenz treffen (Karani, 2018). Der mehrdimensionale Raum repräsentiert das Vokabular des Modells, die Summe aller möglichen Wortrepräsentationen.

tionen. Die Einführung des Aufmerksamkeitsmechanismus ermöglichte es Modellen, semantische Zusammenhänge in Texten zu erfassen. Dadurch kann ein Modell nicht nur Texte syntaktisch analysieren, sondern ihnen auch eine inhaltliche Bedeutung zuweisen – ein Ansatz, der die kognitive Aufmerksamkeit des Menschen nachahmt (Vaswani et al., 2017). Dieses tiefgehende Textverständnis ist eine zentrale Voraussetzung für Anwendungen im Natural Language Processing (NLP; dt. Verarbeitung natürlicher Sprache), einem Forschungsbereich, das automatisierte Analysieren, Interpretieren und Generieren menschlicher Sprache beschreibt (Radford et al., 2018). Transformer nutzen zusätzlich die Multi-Head-Attention-Technik, eine Methode, bei der verschiedene Token innerhalb einer Sequenz unterschiedliche Gewichtungen erhalten (Lopardo, 2020). Dies verstärkt das Modellverständnis für den Kontext einer Eingabe und ermöglicht kohärentere Antworten.

Auch Large Language Models (LLMs; dt. große Sprachmodelle) basieren auf der Transformer-Architektur und unterscheiden sich von herkömmlichen Sprachmodellen sowohl im Umfang als auch im Einsatzzweck. Während klassische Sprachmodelle für spezifische Anwendungsbereiche konzipiert sind, verarbeiten LLMs umfangreiche Datensätze und ermöglichen vielseitigere Antworten. Eine Unterkategorie der LLMs bilden die Small Language Models (SLMs; dt. kleine Sprachmodelle), die trotz geringerer Modellgröße eine ähnliche Qualität in der Textgenerierung anstreben (Rojo-Echeburúa, 2024). Zu den ersten LLMs gehörten die Modelle Bidirectional Encoder Representations from Transformers (BERT; dt. Bidirektionale Einbettungsrepräsentationen aus Transformern) von Google sowie die GPTs von OpenAI, die beide im Jahr 2018 veröffentlicht wurden (Devlin et al., 2019; Radford et al., 2018). BERT ist ein Encoder-Modell und primär für die Verarbeitung von Texteingaben konzipiert, wobei es kontextualisierte Repräsentationen von Wörtern erstellt, indem es Abhängigkeiten innerhalb einer Sequenz erfasst (Vaswani et al., 2017). Im Gegensatz dazu wurde GPT als Modell entwickelt, das speziell für die Generierung von kohärenten Textsequenzen optimiert ist. Es legte den Grundstein für moderne LLMs wie OpenAIs ChatGPT, das 2022 veröffentlicht wurde und die öffentliche Wahrnehmung von KI maßgeblich prägte (Toews, n. d.). LLMs haben sich insbesondere in der automatischen Textgenerierung bewährt, da sie durch das Training ein Verständnis für Wort- und Satzstrukturen erworben haben und daher besonders gut für Anwendungen im Bereich NLP geeignet sind (Goldberg, 2017).

### **2.2.2 Einsatz von KI in narrativen Strukturen**

Die Integration von KI in den Entwicklungsprozess von Videospielen hat besonders im Bereich narrativer Inhalte und Dialogsysteme bedeutende Fortschritte gemacht. Während herkömmliche Spiele in der Regel auf statische, vordefinierte Dialogbäume setzen, ermöglichen moderne KI-Technologien dynamische Anpassung von Narrativen durch Spielerentscheidungen.

Ein frühes Beispiel für den Einsatz von KI in narrativen Systemen ist das interaktive Drama *Façade* von Mateas und Stern (2005). Anhand dieses Spiels stellten sie das Konzept der prozeduralen Autorschaft vor, indem sie statische Textbausteine mit dynamisch generierten Dialogstrukturen kombinierten. Ziel war es, ein narratives System zu entwickeln,

das nicht nur auf vordefinierten Dialogen basiert, sondern sich in Echtzeit an das Verhalten der Spieler anpasst. Das Konzept der prozeduralen Autorschaft beschreibt dabei die Praxis, keine festen, linearen Handlungsstränge zu entwerfen, sondern stattdessen Systeme und Algorithmen zu programmieren, die es Spielern ermöglichen, eine eigene, dynamische Erzählweise zu erleben (Mateas und Stern, 2005).

Ein wesentlicher Vorteil dieser Methodik liegt in ihrer Flexibilität: Statt einer mit festgelegten Handlung reagiert das Spiel dynamisch auf die Aktionen der Spieler. Dadurch wird nicht nur die Immersion verstärkt, sondern auch eine höhere Wiederspielbarkeit ermöglicht. Während in klassischen Erzählformen die Reihenfolge der Ereignisse vor der Veröffentlichung eines Spiels festgelegt wird, eröffnet die prozedurale Generierung die Möglichkeit, Geschichtsstrukturen in Echtzeit entstehen zu lassen (Thue et al., 2007). Allerdings treten auch beim Einsatz solcher dynamischen Narrativen Herausforderungen auf. Ein häufig genannter Kritikpunkt an *Faade* war, dass Spieler oft nicht erkennen konnten, wie sich ihre Handlungen konkret auf die Geschichte auswirken. Die Orientierung innerhalb der narrativen Struktur fiel wegen fehlender Fortschrittsbalken und Wegpunkte schwer. Stattdessen verlagert sich die Orientierung auf die Beziehungen zwischen den Charakteren, wodurch sich die Spielwelt weniger greifbar anfühlen kann (Mateas und Stern, 2005). Eine weitere Herausforderung besteht in der Fehlinterpretation von Eingaben der Spieler. *Faade* nutzt naturliche Sprache als Eingabemethode, was zwar realistische Konversationen ermoglichen soll, aber auch dazu fuhrt, dass etwa 30% der Eingaben nicht korrekt interpretiert werden und so Frustration hervorgerufen werden kann (Mateas und Stern, 2005).

Mit der Weiterentwicklung von LLMs wie OpenAIs GPT-Reihe oder Microsofts Phi-Modellen haben sich die Moglichkeiten fur dynamische Dialogsysteme weiter verbessert. LLMs sind in der Lage, flussige und kontextbezogene Gesprache zu generieren, indem sie Zusammenhange in naturlicher Sprache verstehen und darauf reagieren konnen (Kumaran et al., 2023). Besonders im Bereich Interactive Storytelling (dt. interaktives Geschichtenerzahlen) haben sich KI-Modelle als nutzlich erwiesen, um die Reihenfolge und Inhalte einer Geschichte dynamisch an das Spielverhalten anzupassen (Thue et al., 2007).

Ein Beispiel fur eine fortschrittliche Anwendung dieser Technologien ist *Player-Specific Stories via Automatically Generated Events* (PaSSAGE; dt. spieler-spezifische Geschichten durch automatisch generierte Ereignisse). In diesem System wird Player Modeling (dt. Spielermodellierung) genutzt, um die Geschichte anhand Praferenzen der Spieler zu steuern. Das bedeutet, dass die Narrative auf Basis des Spielstils angepasst werden, sodass Spieler Inhalte erleben, die auf ihre Handlungen abgestimmt sind (Thue et al., 2007). Wahrend Entwickler klassischer Spiele darauf angewiesen sind, eine groe Anzahl an Nebenquests zu entwerfen, um eine breite Spielerschaft anzusprechen, konnte ein adaptives System wie PaSSAGE Quests und narrative Elemente gezielt an die Praferenzen der Spieler anpassen, wodurch das Spielerlebnis dynamischer und immersiver werden konnte (Thue et al., 2007).

Die Forschung zeigt, dass insbesondere story-driven (dt. handlungsbasierte) Spiele eine hohe Immersion ermoglichen, da sie den Spielern das Gefuhl geben, die Kontrolle uber

die Erzählung zu haben. Dabei spielen Dialoge eine zentrale Rolle, da sie die Handlung vorantreiben (Kumaran et al., 2023). Moderne LLMs bieten in diesem Bereich noch nie dagewesene Möglichkeiten, da sie nicht nur zuvor geschriebene Inhalte variieren, sondern auch neue Inhalte basierend auf Spielerinteraktionen generieren können.

### 2.2.3 Stärken und Schwächen dynamisch generierter Inhalte

Der Einsatz von LLMs zur Erzeugung dynamischer Inhalte in Videospielen bringt technologische Vorteile, jedoch auch einige Herausforderungen mit sich. Im Vergleich zu klassischen, vorab geschriebenen Dialogen ermöglichen generierte Dialoge eine höhere Flexibilität, erfordern aber gleichzeitig eine sorgfältige Kontrolle, um Kohärenz und Qualität sicherzustellen. Ein zentraler Vorteil von LLMs ist ihre Fähigkeit, innerhalb von Sekunden neue Handlungsstränge zu generieren. Während traditionelle Spieleentwicklungen darauf angewiesen sind, umfangreiche Narrationen manuell zu entwerfen, können KI-Systeme komplexe interaktive Szenen in kürzester Zeit erstellen. SCENECRAFT, ein KI-gestütztes Werkzeug zur automatisierten Handlungserstellung, demonstrierte dieses Potenzial basierend auf einfachen Anweisungen der Autoren. So erstellte es innerhalb weniger Minuten interaktive Szenen (Kumaran et al., 2023). Diese Effizienz reduziert den Entwicklungsaufwand erheblich, da weniger vorprogrammierte Inhalte benötigt werden. Allerdings gibt es technische Herausforderungen: LLMs sind zwar leistungsfähig, aber ihre Rechenanforderungen können erheblich sein, besonders bei großen Modellen. Kleinere Modelle sind zwar ressourcenschonender, tendieren jedoch zur Erzeugung weniger kohärenter und weniger authentischer Dialoge (Klinkert et al., 2024). Dies stellt Entwickler vor die Herausforderung, ein Gleichgewicht zwischen Leistung und Qualität zu finden. Neben den technischen Aspekten bringen dynamisch generierte Inhalte auch inhaltliche Vorteile, insbesondere im Hinblick auf Immersion und Interaktion. LLMs sind in der Lage, überzeugend menschliches Verhalten zu imitieren, wodurch sie Non-Player-Characters (NPCs; dt. Nicht-Spieler-Charaktere) organischer und weniger mechanisch erscheinen lassen (Klinkert et al., 2024). Während vorgefertigte Dialoge oft statisch wirken und sich über Sitzungen hinweg wiederholen, reagieren LLMs dynamisch auf Entscheidungen der Spieler, was ein individuelleres und abwechslungsreicheres Erlebnis liefert. Ein wichtiger Aspekt dabei ist das Kontextbewusstsein der Modelle. NPCs können sich je nach Spielkontext unterschiedlich verhalten, wodurch ihre Antworten realistischer, natürlicher und situationsabhängiger wirken (Klinkert et al., 2024). Dies verstärkt das Gefühl, dass die Spielwelt lebendig ist. Allerdings haben aktuelle Modelle Schwierigkeiten, diese Kohärenz über längere Zeiträume aufrechtzuerhalten. Über längere Interaktionen hinweg können Widersprüche und Inkonsistenzen auftreten, da das Modell nicht immer in der Lage ist, den gesamten vorherigen Dialogverlauf zu berücksichtigen. Ein weiteres Problem ist die Unvorhersehbarkeit dynamischer Dialoge. Während statische Dialoge exakt definiert und getestet werden können, können LLMs gelegentlich unerwartete oder unpassende Antworten generieren (Klinkert et al., 2024), was zu Immersionsbrüchen führen kann. Besonders problematisch sind dabei ethische und sicherheitsrelevante Aspekte, da LLMs potenziell unangemessene oder fehlerhafte Inhalte erzeugen können, wenn sie nicht ausreichend reguliert werden.

# Kapitel 3

## Umsetzung

### 3.1 Beschreibung des 2D-Spiels

Der entwickelte Prototyp ist ein 2D-Spiel, in dem der Spieler eine Spielfigur durch ein kleines Dorf steuern und mit verschiedenen NPCs interagieren kann. Der Schwerpunkt liegt auf dem Dialogsystem, das entweder vorab definierte, statische Gespräche oder dynamisch generierte Dialoge mit den Dorfbewohnern ermöglicht. Daran sollen die Auswirkungen von LLM-generierten Dialogen auf die Immersion untersucht werden.

#### 3.1.1 Wahl der Game Engine

Für die Umsetzung des Prototyps wurde die Godot Engine gewählt. Diese Entscheidung basierte auf verschiedenen technischen und praktischen Faktoren. Ein zentraler Punkt war die einfache Verbreitung und Implementierung des Spiels, da der Prototyp innerhalb kurzer Zeit eine möglichst immersive Erfahrung bieten sollte. Godot ist eine Open-Source-Engine (dt. quelloffen), die sich durch ihre hohe Flexibilität und eine aktive Community (dt. Gemeinschaft) auszeichnet. Insbesondere das modulare Node-System ermöglichte eine effiziente Strukturierung der Spielinhalte und erleichterte die Implementierung beider Dialogsysteme des Prototyps. Ein weiterer entscheidender Aspekt war die Performance (dt. Leistung) der Engine. Godot zeichnet sich durch einen geringen Ressourcenverbrauch aus, bietet eine hohe Kompatibilität und erlaubt das Exportieren in verschiedene Plattformen wie Windows und Android. Dies war besonders relevant, da LLM-gestützte Systeme teilweise eine erhöhte Rechenleistung erfordern. Neben den technischen Vorteilen spielte auch die Zukunftssicherheit der Engine eine Rolle. Godot gewinnt zunehmend an Beliebtheit, insbesondere unter Studierenden und Indie-Entwicklern, da es eine niedrige Einstiegshürde bietet und von einer engagierten Community unterstützt wird. Die Entwicklung des Prototyps wurde durch eine umfangreiche Dokumentation und ein robustes Plugin-System unterstützt.

Die Implementierung des Spiels erfolgte vollständig in dem visuellen Editor und GDScript, einer speziell für die Godot Engine entwickelten Skriptsprache. Die Wahl dieser Sprache basierte auf mehreren Faktoren: Zum einen ist GDScript stark an Python angelehnt, was sie intuitiv und leicht erlernbar macht. Dies ermöglichte eine effiziente Entwicklung und schnelle Anpassung des Codes. Zum anderen weist die objektorientierte

Struktur Parallelen zu anderen modernen Programmiersprachen auf, was das Entwickeln übersichtlicher und performanter gestaltet. Ein weiterer entscheidender Vorteil ist die nahtlose Integration in die Godot Engine. Während alternative Sprachen wie C# oder VisualScript ebenfalls unterstützt werden, ermöglicht GDScript eine direkte Anbindung an die Engine. Dies wirkt sich positiv auf die Performance und Flexibilität aus und erleichtert insbesondere die Integration von Plugins.

Bei der Wahl der Game Engine wurden alternative Tools in Betracht gezogen, insbesondere Unity, das zwar verbreitet und leistungsstark, jedoch nicht optimal für dieses Projekt war. Die Implementierung eines vergleichbaren 2D-Dialogsystems hätte in Unity einen höheren Entwicklungsaufwand erfordert, da die Engine primär auf 3D-Anwendungen ausgelegt ist. Zudem hätte der höhere Speicher- und Leistungsbedarf von Unity potenziell längere Ladezeiten verursacht, was der performanten Umsetzung hätte entgegenwirken können.

### 3.1.2 Aufbau der Spielumgebung

Die Spielwelt repräsentiert ein kleines mittelalterliches Fantasy-Dorf, das sich in einem Wald befindet und von fünf NPCs bewohnt wird. Sie ist frei erkundbar und als halblinesares Gebiet konzipiert, was bedeutet, dass Spieler sich innerhalb eines festgelegten Rahmens bewegen, aber selbst entscheiden können, in welcher Reihenfolge sie mit den NPCs interagieren. Dadurch wird verhindert, dass sich ein strikter Ablauf entwickelt, der den Spieler in eine bestimmte Richtung zwingt. Gleichzeitig ermöglicht diese Struktur die Analyse der Spielerpräferenzen in Bezug auf die Dialogsysteme.

Die Welt ist in drei zentrale Bereiche unterteilt: Beim Start des Spiels gelangt der Spieler zunächst in einen Einführungsbereich, in dem die grundlegende Steuerung anhand einer Anzeige vorgestellt wird. Das Dorf bildet den Hauptbereich des Spiels und beherbergt vier NPCs mit statischen Dialogen. Zusätzlich gibt es ein angrenzendes Waldgebiet, in dem sich der NPC mit dynamisch generierten, LLM-basierten Dialogen befindet. Zwei versperrte Wege vermitteln darüber hinaus den Eindruck einer größeren Spielwelt und deuten eine potenzielle Fortsetzung der Geschichte an, wodurch Spieler den Prototyp in eine umfassendere Handlung einordnen können und er so natürlicher wirkt.

Das grafische Design folgt einem stilisierten, bewusst nicht realistischen Ansatz. Diese stilistische Entscheidung schafft eine klare Abgrenzung zur realen Welt, was laut Ermi und Mäyrä (2005) die Immersion verstärken kann (Ermi und Mäyrä, 2005). Dennoch wurden hochwertige Assets verwendet und die Umgebung detailreich gestaltet, um eine ansprechende Spielwelt zu erschaffen. Zusätzlich unterstützt eine Soundkulisse mit typischen Waldgeräuschen die Atmosphäre, wodurch das Spielgeschehen authentischer wirkt und der Spieler seine Sinne stärker auf die Spielerfahrung fokussieren kann. Als gestalterische Inspiration dienten die Pokémon-Spielreihe sowie World of Warcraft. Diese Spiele bieten Spielern gerade durch ihren stilisierten Grafikstil ein hohes Maß an Immersion, da ihre zeitlose, unverwechselbare Gestaltung eine direkte Vergleichbarkeit mit der realen Welt verhindert.

Die Dorfgemeinschaft besteht aus fünf NPCs, die jeweils eine eigene Persönlichkeit und Hintergrundgeschichte besitzen:

- Bjorn – Ein Einsiedler und Alchemist, der abgeschieden im Wald lebt und seine magischen Fähigkeiten vor den Dorfbewohnern verbirgt.
- Don – Ein ehemaliger Freund von Bjorn, der sich von der Magie abgewandt hat und nun als Dorfältester fungiert.
- Lisa – Eine Gärtnerin mit einem besonderen Interesse an Heilkräutern, die Bjorn gelegentlich hilft.
- Gary – Ein misstrauischer Jäger, der überzeugt ist, dass Bjorn dunkle Magie praktiziert und ihn deshalb beobachtet.
- Jim – Ein pragmatischer Bauer, der sich nicht in die Konflikte des Dorfes einmischt und sich auf seine eigene Arbeit konzentriert.

Diese Charaktere wurden gezielt entworfen, um eine dynamische soziale Struktur innerhalb des Dorfes zu simulieren. Ihre unterschiedlichen Beziehungen zueinander sowie ihre emotionalen Antworten – insbesondere in Bezug auf Bjorn – sollen die Immersion verstärken und der Spielwelt eine immersive Tiefe verleihen.

Die Steuerung der Spielfigur wurde gezielt auf eine intuitive Benutzererfahrung ausgelegt. Die Interaktion im Spiel erfolgt über eine Kombination aus Tastatur- und Mauseingaben, wobei die Bewegung des Charakters über die Tastatur gesteuert wird, während Dialoge mit NPCs durch einen Tastendruck initiiert werden. Statische Dialoge werden über Mausklicks auf Auswahlmöglichkeiten navigiert, während sich bei dynamischen Dialogen ein Texteingabefeld öffnet, in dem der Spieler seine eigenen Antworten formulieren kann. Damit wurden einerseits intuitive Steuerungsmöglichkeiten geschaffen, andererseits jedoch auch etablierte Branchenschemata berücksichtigt: Die Bewegungssteuerung entspricht gängigen Konventionen vieler Spiele, während die Eingabe der dynamischen Dialoge zusätzlich durch Mausklicks erfolgen kann. Zudem können die Dialogfenster über ein standardisiertes X-Symbol oder die Escape-Taste geschlossen werden, was eine vertraute Benutzererfahrung bietet.

### **3.1.3 Implementierung der Dialogsysteme**

Die statischen Dialoge wurden mithilfe eines regelbasierten Dialogsystems umgesetzt, das auf einer JSON-Datei basiert. Jeder NPC besitzt einen eigenen Dialogbaum, dessen Zweige den Gesprächsverlauf steuern. Die Dialoge sind in mehrere Sequenzen unterteilt, in denen der Spieler zwischen verschiedenen Antwortmöglichkeiten wählen kann. Je nach Entscheidung nimmt der Dialog einen unterschiedlichen Verlauf. Jeder Dialogbaum beginnt mit einem Einstiegspunkt, von dem aus der Spieler zwischen mehreren Optionen wählen kann. Jede Option ist mit einem Status verknüpft, der den nächsten Dialog bestimmt. Sobald eine bestimmte Auswahl getroffen wurde und eine vordefinierte Dialogtiefe erreicht ist, wird ein neuer Dialogzweig aktiviert. Ab diesem Punkt kann

der Spieler nicht mehr zum vorherigen Zweig zurückkehren. Dadurch setzen NPCs in späteren Interaktionen bestimmte Informationen voraus und reagieren auf vorherige Konversationsstränge. Jeder Zweig enthält mindestens eine Option, mit der das Gespräch beendet wird, bevor – falls vorhanden – der nächste Dialogabschnitt freigeschaltet wird. Diese Dialogstruktur gewährleistet, dass jeder NPC emotional, handlungsfördernd, charakterspezifisch und sozial antworten kann, wodurch die Gespräche kohärent bleiben.

Im Gegensatz zu den statischen Dialogen nutzt das dynamische Dialogsystem ein LLM, das in Echtzeit generierte Antworten liefert. Dies ermöglicht es Spielern, freie Texteingaben zu machen, auf die die KI reagiert, wodurch der Gesprächsverlauf offener und weniger vorhersehbar ist. Das dynamische Dialogsystem kommt ausschließlich bei dem NPC Bjorn zum Einsatz. Damit er weitestgehend kohärente und inhaltlich sinnvolle Antworten geben kann, wird ihm zu Beginn jeder Interaktion ein sogenannter Prompt übermittelt. Dieser enthält Hintergrundinformationen, die seine Antworten in eine bestimmte Richtung lenken. Dazu gehören ein Charakterprofil, das Bjorns Persönlichkeit und seine Verhaltensweise beschreibt, sowie Informationen über das Dorf und die anderen NPCs, um kontextbezogene Antworten zu ermöglichen. Darüber hinaus gibt der Prompt formale Anweisungen zur Struktur und Länge der Antworten vor und enthält inhaltliche Einschränkungen, um das Gespräch in einem sinnvollen Rahmen zu halten. Während die Umgebung und zuvor geführte Unterhaltungen keine direkte Rolle in der Generierung der Antworten spielen, wird innerhalb einer laufenden Interaktion der gesamte Gesprächsverlauf gespeichert und dem Modell übergeben. Dadurch kann Bjorn auf frühere Aussagen des Spielers Bezug nehmen, sich an gestellte Fragen erinnern und darauf basierend antworten.

Die beiden Dialogsysteme unterscheiden sich grundlegend: Während statische Dialoge eine hohe Vorhersagbarkeit und Kontrolle ermöglichen, bieten dynamische Dialoge eine größere Flexibilität und Interaktivität. Statische Dialoge tragen durch ihre klare Erzählstruktur zur Konsistenz der Spielwelt bei, während dynamische Dialoge eine individuellere und angepasste Spielerfahrung bieten. Eine zentrale Frage besteht darin, ob die zusätzliche Freiheit der dynamischen Dialoge auch zu einer erhöhten Immersion führt oder ob unpassende Antworten diese beeinträchtigen können.

## 3.2 Technische Umsetzung

Die Implementierung eines LLMs für die dynamischen Dialoge im Prototyp wurde so konzipiert, dass es direkt über Godot ausgeführt wird und sämtliche Berechnungen lokal auf dem Computer des Spielers erfolgen. Dadurch entfällt die Notwendigkeit einer externen API, was nicht nur Latenzzeiten reduziert, sondern auch sicherstellt, dass das Modell unabhängig und ohne Internetverbindung funktioniert.

### 3.2.1 Auswahl des LLMs

Für die Umsetzung der dynamischen Dialoge wurde das Modell Phi-3 Mini von Microsoft gewählt. Mit einer Größe von etwa 1,4 GB ist es kompakt genug, um effizient auf einer

CPU ausgeführt zu werden, während es gleichzeitig qualitative Antworten generiert. Die geringe Quantisierung beim Überführen in das GGUF-Format reduziert den Speicherbedarf, geht jedoch mit leichten Genauigkeitsverlusten einher. Da es sich um ein Instruct Model handelt, wurde es speziell für klassische Eingabe-Ausgabe-Szenarien optimiert, wodurch es sich besonders gut für interaktive Rollenspiele eignet und die natürliche Kommunikation zwischen Spieler und NPC erleichtert. Ein weiterer entscheidender Faktor für die Wahl dieses Modells war dessen aktuelle Trainingsbasis. Phi-3 Mini wurde mit einer Kombination aus öffentlich zugänglichen Datensätzen, synthetisch generierten Trainingsdaten für Mathematik und logisches Denken sowie hochqualitativen Chat-Daten trainiert. Dadurch zeichnet es sich durch seine Leistung in den Bereichen Sprachverständnis, logisches Schlussfolgern und Konsistenz aus. Trotz seiner vergleichsweise geringen Größe von 3,8 Milliarden Parametern schneidet es in Benchmarks gut gegen größere Modelle ab und bleibt dabei ressourcenschonend. Zusätzlich verfügt es über eingebaute Sicherheitsmechanismen, die unangemessene oder unerwünschte Inhalte herausfiltern. Die Integration des Modells in das Spiel erfolgte über das Godot-LLM-Plugin, das eine direkte Einbindung eines LLMs in die Engine ermöglicht. Das Plugin stellt spezielle Nodes zur Verfügung, über die Parameter wie Top-K, Seed-Werte oder die maximale Kontextgröße konfiguriert werden können. Dadurch konnte das Verhalten des Modells gezielt angepasst werden, um Kohärenz in den generierten Antworten sicherzustellen, die gleichzeitig jedoch kreativ bleiben. Die Kommunikation zwischen Spiel und Modell erfolgt über einfache Zeichenketten (engl. „strings“): Spieler geben Eingaben ein, die an das LLM weitergeleitet werden, woraufhin die generierten Antworten in dem Dialogfenster des Spiels angezeigt werden. Neben der aktuellen Eingabe verarbeitet das Modell den gesamten bisherigen Gesprächsverlauf der Sitzung, um eine zusammenhängende Unterhaltung zu führen. Die erste Nachricht, die an das Modell übermittelt wird, enthält einen Initialisierungsprompt, der dem NPC zuvor genannten Verhaltensweisen vorgibt. In diesem Prompt wird auch festgelegt, dass Bjorn als zurückgezogener, rätselhafter Alchemist mit einer knappen Satzlänge antwortet.

### 3.2.2 Begrenzungen und Kontrollmechanismen

Da das Modell in Echtzeit Antworten generiert, mussten verschiedene Maßnahmen ergriffen werden, um unangemessene Ausgaben zu unterbinden. Eine zentrale Herausforderung dynamischer Dialogsysteme besteht darin, dass generierte Antworten unvorhersehbar sein oder vom eigentlichen Thema abweichen können. Um dies zu verhindern, wurde das Modell durch eine Kombination aus Formatierungsvorgaben, Filtermechanismen und Längenbeschränkungen kontrolliert und in seinen Ausgaben eingeschränkt. Der erste Prompt, der an das Modell übergeben wird, enthält daher auch Formatierungshinweise und eine Beispielstruktur, nach der Antworten generiert werden sollen. Zusätzlich wurde ein Blacklist-basiertes Filtersystem implementiert, das bestimmte Formatierungsfehler automatisch erkennt und korrigiert. Dadurch wird verhindert, dass unnatürlich wirkende Antworten die immersive Spielerfahrung stören. Sollte eine generierte Antwort eine bestimmte Länge überschreiten, wird das Modell automatisch neu initialisiert. Auch auf technischer Ebene stellten sich Herausforderungen, die bei der Implementierung

berücksichtigt werden mussten. Phi-3 Mini ist ein vergleichsweise kompaktes Modell, wodurch es nicht über die Kapazität verfügt, umfangreiches Weltwissen zu speichern. Stattdessen ist es größtenteils auf den vorliegenden Kontext angewiesen. Ein Vorteil der lokalen Implementierung besteht darin, dass keine externen Rate Limits (dt. Ratenbeschränkungen) existieren, wodurch das Modell ohne Einschränkungen auf Eingaben reagieren kann. Gleichzeitig bedeutet dies jedoch, dass sämtliche Rechenleistung vom Computer des Spielers bereitgestellt werden muss. Diese Limitierungen wurden bewusst akzeptiert, um in der Nutzerstudie (siehe Kapitel 4) potenziell leistungsschwache Rechner zu berücksichtigen.

### **3.2.3 Herausforderungen und Lösungsansätze**

Die Implementierung des dynamischen Dialogsystems stellte eine Reihe von Herausforderungen dar, die durch gezielte Anpassungen schrittweise gelöst wurden. Eine der größten Schwierigkeiten bestand in der Einbindung eines geeigneten Plugins: Das Godot-LLM-Plugin erwies sich gegenüber vergleichbaren Plugins als stabilste Lösung. Allerdings traten Fehlverhalten wie Satzwiederholungen erst nach mehreren Dialogsequenzen auf, weswegen diese Probleme erst spät in der Entwicklungsphase erkannt und behoben wurden. Ein weiteres Problem stellten inkohärente Antworten des Modells dar, die dazu führten, dass manche Dialoge nicht immersiv oder mechanisch wirkten. Dies geschah insbesondere dann, wenn das Modell Formatierungsangaben ignorierte oder Nutzereingaben falsch interpretierte. Um dem entgegenzuwirken, wurde der Initialisierungsprompt überarbeitet, sodass das Modell konsistentere Antworten generieren konnte.

# Kapitel 4

## Methodik

### 4.1 Durchführung der Nutzerstudie

#### 4.1.1 Zielsetzung der Studie

Diese Nutzerstudie untersucht die Frage, wie narrative Inhalte, die durch LLMs generiert wurden, die Immersion in Videospiele beeinflussen. Der Schwerpunkt liegt auf dem Vergleich zwischen dynamisch generierten und vorgefertigten Dialogen. Die zugrunde liegende Hypothese besagt, dass LLM-generierte Dialoge die Immersion positiv beeinflussen, da sie dynamischer, flexibler und anpassungsfähiger auf Spielerinteraktionen reagieren können. Diese Annahme stützt sich auf frühere Forschungsergebnisse, die darauf hindeuten, dass adaptive und kontextabhängige Inhalte immersive Spielerfahrungen fördern und ist insbesondere im Kontext des technologischen Wandels in der Spieleindustrie von Bedeutung. Der Einsatz von KI spielt eine zunehmend größere Rolle im Game Design, zur Automatisierung der Inhaltserstellung, zur Unterstützung bei der Programmierung oder zur Entwicklung narrativer Erlebnisse. Die Studie trägt zur aktuellen Diskussion bei, inwieweit LLMs eine sinnvolle Ergänzung oder gar eine Alternative zu traditionellen, vorab geschriebenen Dialogsystemen darstellen können.

#### 4.1.2 Demografische Merkmale der Stichprobe

Die Studie wurde mit zehn freiwilligen Probanden durchgeführt, wobei auf eine möglichst heterogene Stichprobe geachtet wurde. Um potenzielle Einflussfaktoren auf die Spielerfahrung zu erfassen, wurden im Fragebogen personenbezogene Daten erhoben. Diese dienten sowohl der Beschreibung der Stichprobe als auch der Interpretation der Ergebnisse im Hinblick auf erfahrungsbedingte Faktoren, die Einfluss auf die Studie hätten haben können. Zu ihnen gehörten:

- Vorname (zur anonymisierten Zuordnung der Daten)
- Alter und Geschlecht
- Allgemeine Spielerfahrung
- Selbsteinschätzung der Kompetenz in Videospiele

- Erfahrungen mit Immersion
- Individuelle Merkmale der Immersionserfahrung
- Vertrautheit mit KI und ML

### 4.1.3 Aufbau der Nutzerstudie

Der Hauptaspekt der Studie war der Vergleich zwischen statischen und dynamischen Dialogen und deren Einfluss auf die Immersion der Spieler. Aufgrund der geringen Anzahl an Probanden wurde darauf verzichtet, die beiden Dialogsysteme in separaten Levels zu testen. Stattdessen wurden sie innerhalb einer einheitlichen Spielwelt kombiniert, sodass die Spieler in einem kleinen Dorf frei umhergehen und mit NPCs interagieren konnten. Um an der Studie teilnehmen zu können, mussten die Probanden folgende Voraussetzungen erfüllen:

- Einen Windows-PC zum Ausführen des Prototyps
- Englischkenntnisse, da die Dialoge auf Englisch formuliert waren
- Kopfhörer oder Lautsprecher, um möglichst viele Sinne abzudecken

Für die Durchführung war eine Gesamtzeit von 30 Minuten vorgesehen, aufgeteilt in 10 Minuten Spielzeit und 20 Minuten für die Beantwortung des Fragebogens. Die Teilnehmer durften je nach Bedarf allerdings über die geplante Spielzeit hinaus weiterspielen. Dies ermöglichte eine zusätzliche Analyse, da längere Spielzeiten als Interaktionsbedürfnis und persönliche Präferenz interpretiert werden konnten. Zudem wurden die Interaktionszeiten mit allen NPCs protokolliert, um nachträglich zu untersuchen, ob sich die Spieler in bestimmten Dialogen länger aufhielten als in anderen. Die gesamte Studie wurde remote (dt. von zu Hause) durchgeführt. Alle Teilnehmer erhielten einen Downloadlink, der ein Paket mit dem Spielprototyp, einer schriftlichen Anleitung in Form eines PDFs und dem Fragebogen enthielt. Die mitgelieferte Anleitung enthielt eine kurze Einführung in die Studie sowie eine Schritt-für-Schritt-Beschreibung der Durchführung. Sie stellte sicher, dass alle Teilnehmer unter vergleichbaren Bedingungen spielten und an den Ablauf herangeführt wurden. Zusätzlich enthielt sie einen Troubleshooting-Bereich, in dem mögliche technische Probleme sowie deren Lösungen aufgeführt waren. Dies war hilfreich, um eventuellen Stabilitätsproblemen des Prototyps direkt entgegenwirken zu können. Der Fragebogen war in mehrere Abschnitte unterteilt und online abrufbar. Die Reihenfolge wurde bewusst wie folgt gewählt:

- Game Experience Questionnaire (GEQ) - Core-Modul
- GEQ - Social-Presence-Modul für statische NPC
- GEQ - Social-Presence-Modul für dynamische NPC
- GEQ - Post-Game-Modul

- Presence Questionnaire (PQ) - ohne haptische Fragen
- Offene Fragen
- Personenbezogene Fragen

Zu Beginn enthielt der Fragebogen eine kurze Einleitung, die die Teilnehmer darauf hinwies, die Fragen in der vorgegebenen Reihenfolge zu bearbeiten und die Antwortoption auszuwählen, die ihren persönlichen Einschätzungen am nächsten kamen. Zum Schluss wurden die Probanden gebeten, die Protokolldatei hochzuladen, die während des Spielens automatisch erstellt wurde.

## 4.2 Vorstellung der Messinstrumente

Eine Nutzerstudie ist eine wissenschaftliche Methode zur systematischen Erhebung von Daten über das Verhalten, die Wahrnehmung und die Präferenzen von Nutzern. Sie ermöglicht eine Analyse der Interaktion von Personen mit einem System und liefert Erkenntnisse darüber, wie Nutzererfahrungen optimiert werden können. Zur Auswertung kommen sowohl quantitative als auch qualitative Methoden zum Einsatz. Quantitative Verfahren erlauben eine objektive, statistische Analyse, während qualitative Methoden eine Untersuchung individueller Erlebnisse ermöglichen. In Kombination liefern sie mathematische Befunde mit kontextbezogenen Einblicken und Hintergründe spezifischer Bewertungen.

### 4.2.1 Quantitative Methoden: Fragebögen

Der Game Experience Questionnaire (GEQ; W. A. IJsselsteijn et al., 2013) wurde entwickelt, um Spielerfahrungen in Videospiele strukturiert und reproduzierbar zu erfassen. Er wurde sowohl im Labor als auch in Feldstudien validiert und hat sich als zuverlässiges Instrument zur Messung von Spielerlebnissen erwiesen (W. IJsselsteijn et al., 2008). Aufgrund seiner wissenschaftlichen Fundierung eignet er sich für aussagekräftige Ergebnisse in empirischen Studien. Der Fragebogen ist modular aufgebaut, sodass einzelne Abschnitte je nach Studienziel verwendet werden können. Die enthaltenen Fragen sind dabei verschiedenen Kategorien zugeordnet, darunter Immersion, Herausforderung und soziale Interaktion, wodurch eine Analyse der Spielerfahrung möglich wird. In dieser Studie wurde der GEQ insbesondere eingesetzt, um neben der allgemeinen Spielerfahrung gezielt die Unterschiede zwischen den statischen und dynamischen Dialogsystemen zu erfassen.

Der PQ wurde entwickelt, um das Empfinden von Präsenz in virtuellen Umgebungen systematisch zu erfassen (Witmer und Singer, 1998). Ursprünglich wurde er speziell für VR-Anwendungen konzipiert, doch seine Anwendbarkeit erstreckt sich auch auf Videospiele im Allgemeinen. Da in dieser Studie keine VR-Elemente enthalten sind, wurden unter anderem haptische Fragen aus dem Fragebogen entfernt, und der Fragebogen auf studienrelevante Aspekte angepasst.

Beide Fragebögen verwendeten eine Likert-Skala, die die Bewertung in feineren Abstufungen als binäre Ja/Nein-Antworten zu erfassen (Gritsch, 2012). Die Skalen beider Fragebögen unterscheiden sich dabei leicht: Der PQ nutzt eine Skala von eins bis fünf, die eine Abstufung von sehr wenig bis sehr viel Zustimmung erlaubt. Der GEQ hingegen verwendet eine Skala von null bis vier, wodurch zusätzlich die Auswahlmöglichkeit *überhaupt nicht* gegeben ist. In der Auswertung (siehe Kapitel 5) wurden die Werte zur besseren Vergleichbarkeit normiert.

#### 4.2.2 Qualitative Methoden: Offene Fragen

Neben den standardisierten Fragebögen wurden offene Fragen eingesetzt, um die Gedankengänge der Probanden zu berücksichtigen. Diese Fragen ermöglichten eine detaillierte Erfassung der subjektiven Spielerfahrungen und lieferten qualitative Einblicke. Der Abschnitt mit den offenen Fragen begann mit einer allgemeinen Einschätzung des Spielerlebnisses, bevor er spezifisch auf die wahrgenommene Immersion und anschließend auf die Bewertung der Dialogsysteme einging. Die Teilnehmer hatten dabei die Möglichkeit, ihre Antworten ausführlich zu begründen, da keine Mindestantwortenlänge vorgegeben wurde.

# Kapitel 5

## Analyse und Ergebnisse

Die folgende Analyse untersucht die im Rahmen der Nutzerstudie erhobenen Daten, die den Einfluss von LLM-generierten Dialogen auf die Immersion der Spieler darstellen. Dabei werden zuerst die quantitativen Ergebnisse präsentiert, gefolgt von einer inhaltlichen Analyse der qualitativen Rückmeldungen der Teilnehmer. Dieses Kapitel dient der deskriptiven Auswertung der Daten; eine inhaltliche Interpretation der Ergebnisse und deren Einordnung in bestehende Forschung erfolgt im anschließenden Kapitel (siehe Kapitel 6). Der Vergleich der Immersionswerte zwischen den beiden Dialogsystemen wird anhand der Ergebnisse des GEQ und des PQ durchgeführt und mithilfe visueller Balkendiagramme und Boxplots verdeutlicht. Zur Bestimmung von Tendenzen und zur Analyse der Streuung werden Mittelwerte und Standardabweichungen berechnet. Zur statistischen Überprüfung signifikanter Unterschiede zwischen den beiden Bedingungen wird ein t-Test durchgeführt. Im Anschluss werden die qualitativen Daten aus den offenen Fragen ausgewertet. Hierzu erfolgt eine Kategorisierung der häufigsten Themen und Begriffe, die mithilfe einer Wortwolke visualisiert werden. Besonders prägnante oder repräsentative Zitate aus den Teilnehmerantworten werden dabei hervorgehoben, um bestimmte Punkte hervorzuheben.

### 5.1 Analyse quantitativer Ergebnisse

Die Berechnungen der Mittelwerte und Standardabweichungen für die einzelnen Kategorien zeigen, dass die Teilnehmer insgesamt ein hohes Maß an Immersion erlebt haben. Besonders ausgeprägt waren neben der Immersion die Kategorien Flow, Competence (dt. Kompetenz) und Positive Affect (dt. positive Wirkung), während Tension (dt. Anspannung) vergleichsweise niedrig bewertet wurde. Dies legt nahe, dass die Gestaltung der Spielwelt und die Interaktionsmöglichkeiten die Bedingungen für eine immersive Erfahrung erfüllten.

### 5.1.1 Core-Modul des GEQ - Allgemeine Spielerfahrung

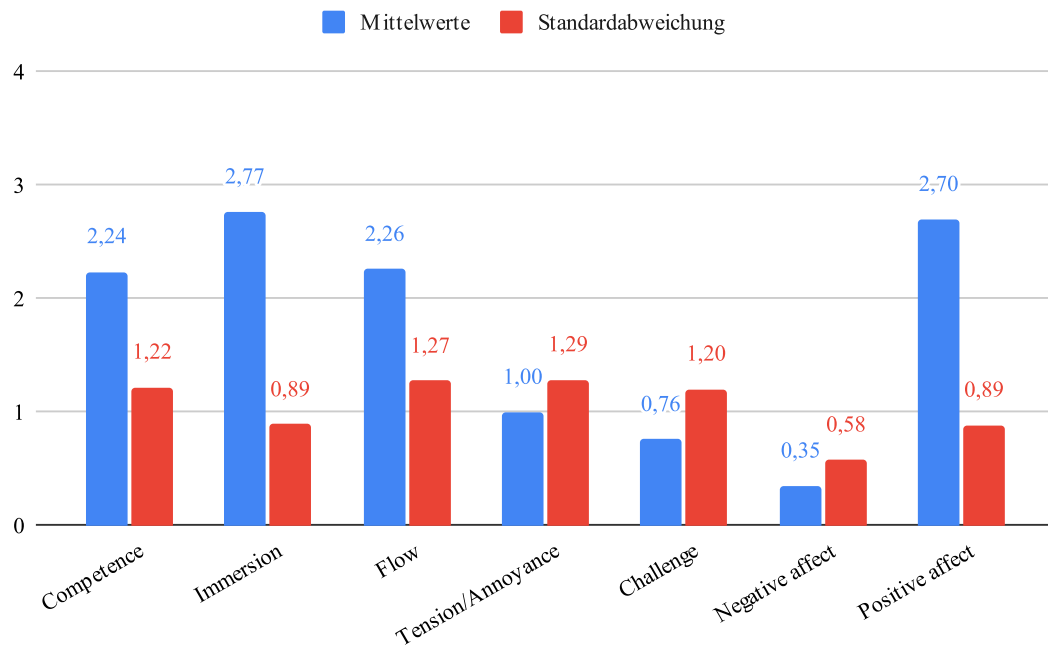


Abbildung 5.1: Mittelwerte und Standardabweichungen der Core-Kategorien im GEQ (eigene Darstellung, basierend auf den erhobenen Daten)

Im Core-Modul des GEQ (siehe Abbildung 5.1) erreichte die Kategorie Flow einen Mittelwert von 2.26 ( $SD = 1.27$ ), was auf eine starke Einbindung der Spieler in das Spielgeschehen hindeutet. Competence lag mit 2.24 ( $SD = 1.22$ ) auf einem ähnlichen Niveau, woraus sich schließen lässt, dass sich die Spieler gut in der Spielwelt zurechtfinden. Die Werte für Tension/Annoyance (dt. Verärgerung) fielen mit 1.00 ( $SD = 1.29$ ) vergleichsweise niedrig aus, was darauf hinweist, dass das Spiel nicht als frustrierend oder stressig wahrgenommen wurde. Die Bewertung der Wirkung zeigt eine deutliche Tendenz zur positiven Spielerfahrung: Positive Affect wurde mit 2.70 ( $SD = 0.89$ ) deutlich höher bewertet als Negative Affect (dt. negative Wirkung) mit 0.35 ( $SD = 0.58$ ). Dies legt ein als angenehm empfundenes Spielerlebnis nahe. Die höchsten Werte erreichte die zusammengefasste Kategorie Immersion, bestehend aus Sensory und Imaginative Immersion, mit einem Mittelwert von 2.77 ( $SD = 0.89$ ). Dies unterstreicht die hohe Intensität der immersiven Erfahrung über die gestellten Fragen hinweg.

### 5.1.2 Social-Presence-Modul des GEQ

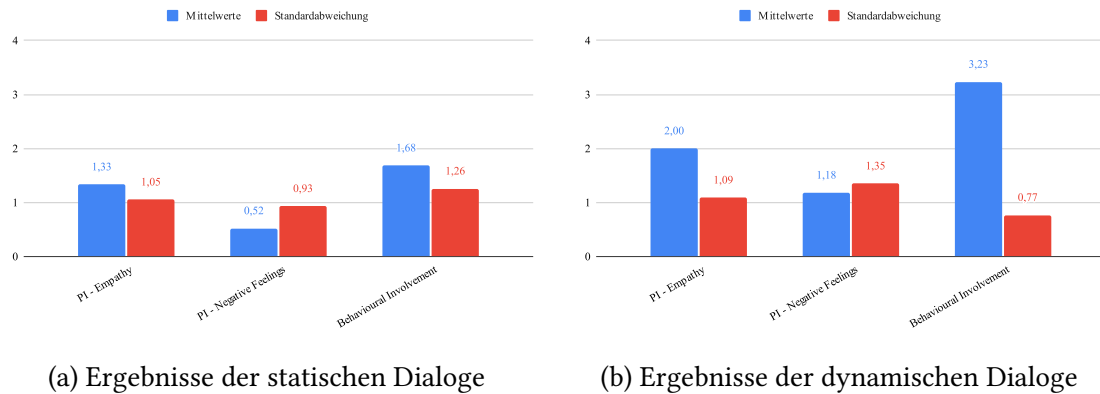


Abbildung 5.2: Mittelwerte und Standardabweichungen der Social-Presence-Kategorien im GEQ (eigene Darstellung, basierend auf den erhobenen Daten)

Das Social-Presence-Modul (siehe Abbildung 5.2), das für beide Dialogsysteme separat erhoben wurde, zeigt über beide Datensätze hinweg eine ähnliche Rangordnung der Kategorien. Behavioural Involvement (dt. verhaltensbezogene Einbindung) erzielte in beiden Bedingungen die höchsten Werte, gefolgt von Empathy (dt. Empathie), während Negative Feelings (dt. negative Gefühle) als niedrigste Kategorie des Psychological Involvement (dt. psychologische Einbindung) bewertet wurde. Auffällig ist, dass die Werte in allen Kategorien bei den dynamischen Dialogen (siehe Abbildung 5.2b) im Vergleich zu den statischen Dialogen (siehe Abbildung 5.2a) durchgängig höher ausfallen. Insbesondere Behavioural Involvement zeigt einen deutlichen Anstieg von 1.68 (SD = 1.26) auf 3.23 (SD = 0.77) und hat sich damit nahezu verdoppelt. Dies deutet darauf hin, dass dynamische Dialoge nicht nur zu einer verstärkten sozialen Einbindung führten, sondern auch intensivere emotionale Reaktionen hervorriefen – sowohl im positiven als auch im negativen Sinne.

### 5.1.3 Post-Game-Modul des GEQ

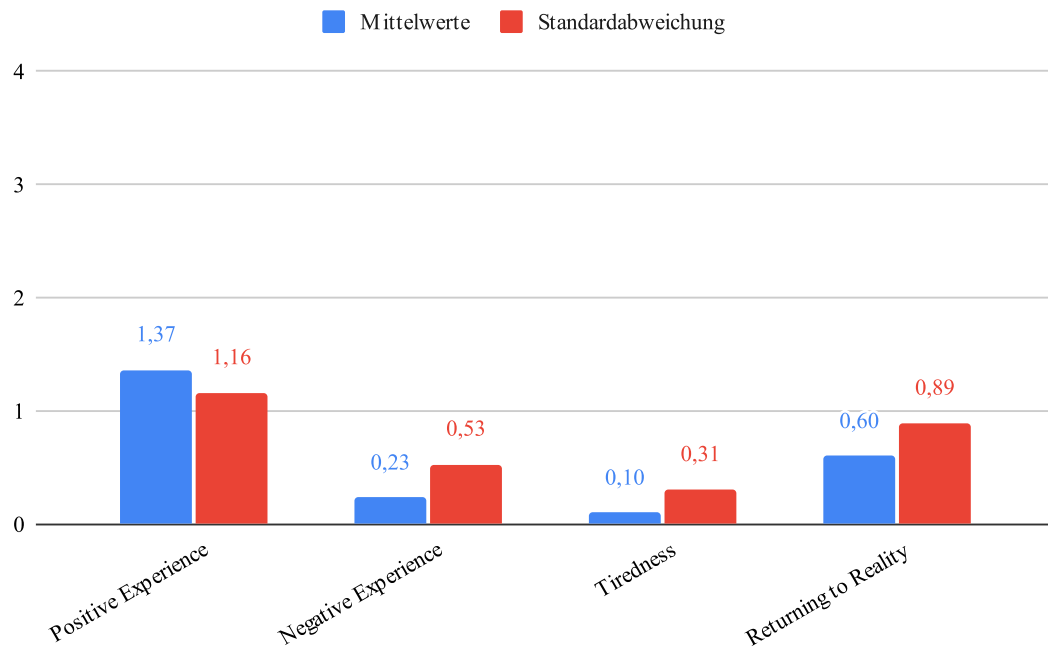


Abbildung 5.3: Mittelwerte und Standardabweichungen der Post-Game-Kategorien im GEQ (eigene Darstellung, basierend auf den erhobenen Daten)

Im Post-Game-Modul (siehe Abbildung 5.3) fielen die Werte in allen Kategorien vergleichsweise niedrig aus. Die höchste Bewertung erhielt Positive Experience (dt. positive Erfahrung) mit einem Mittelwert von 1.37 (SD = 1.16). Demgegenüber stand Negative Experience (dt. negative Erfahrung) mit einem deutlich geringeren Mittelwert von 0.23 (SD = 0.53), was darauf hindeutet, dass die Spieler ihre Erfahrung rückblickend überwiegend positiv bewerteten. Der Wert für Tiredness (dt. Ermüdung) lag mit 0.10 (SD = 0.31) ebenfalls auf einem sehr niedrigen Niveau, was allerdings auch auf die geringe Anzahl entsprechender Fragen im Fragebogen zurückzuführen ist.

### 5.1.4 Hauptkategorien des PQ - Wahrnehmung von Präsenz

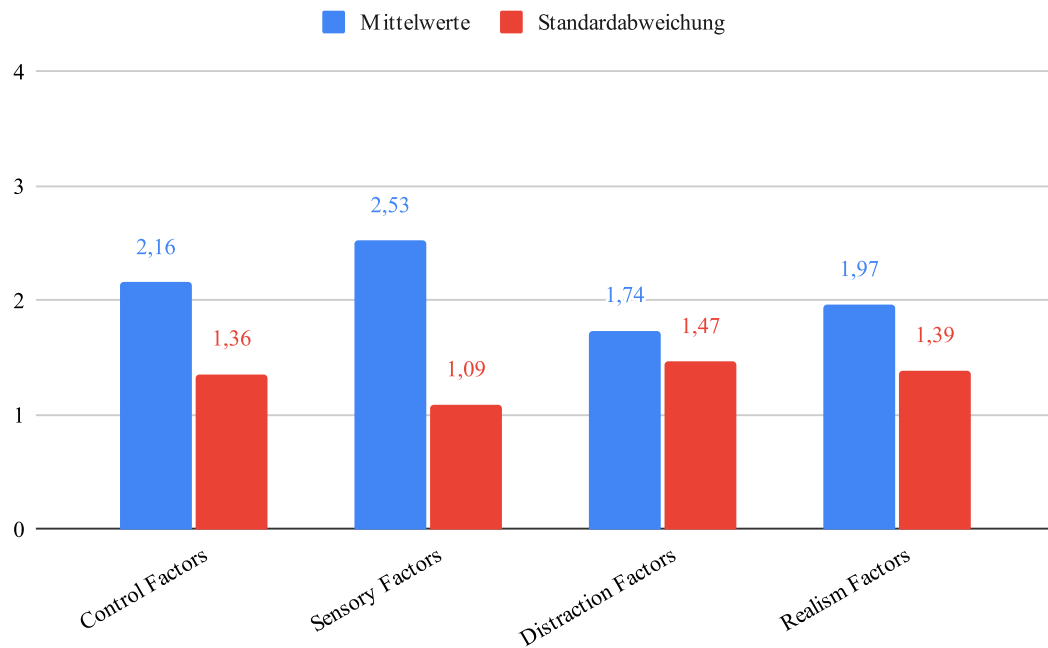


Abbildung 5.4: Mittelwerte und Standardabweichungen der Hauptkategorien im PQ (eigene Darstellung, basierend auf den erhobenen Daten)

Die Ergebnisse der Hauptkategorien des PQ (siehe Abbildung 5.4) deuten darauf hin, dass die Spieler insgesamt ein ausgeprägtes Gefühl der Einbindung in die Spielwelt erlebten. Besonders hervorzuheben ist die hohe Bewertung der sensorischen Faktoren (engl. „sensory factors“) mit einem Mittelwert von 2.53 (SD = 1.09). Sie legen nahe, dass die Spieler ihre Aufmerksamkeit auf die Wahrnehmung der virtuellen Umgebung richten konnten. Darüber hinaus erreichten die Kontrollfaktoren (engl. „control factors“) einen Mittelwert von 2.16 (SD = 1.36), was darauf hindeutet, dass die Spieler einen gewissen Grad der Kontrolle in der Spielwelt wahrgenommen haben. Die Ablenkungsfaktoren (engl. „distraction factors“) wiesen mit einem Mittelwert von 1.74 (SD = 1.47) einen vergleichsweise niedrigen Wert auf, was nahelegt, dass externe oder interne Störungen die Immersion der Spieler nur begrenzt beeinträchtigten. Die Realismusfaktoren (engl. „realism factors“) wurden mit einem Mittelwert von 1.97 (SD = 1.39) bewertet und lagen damit im mittleren Bereich. Dies könnte darauf hinweisen, dass die Spielwelt als glaubhaft wahrgenommen, aber nicht unbedingt als realistisch empfunden wurde.

### 5.1.5 Subskalen des PQ

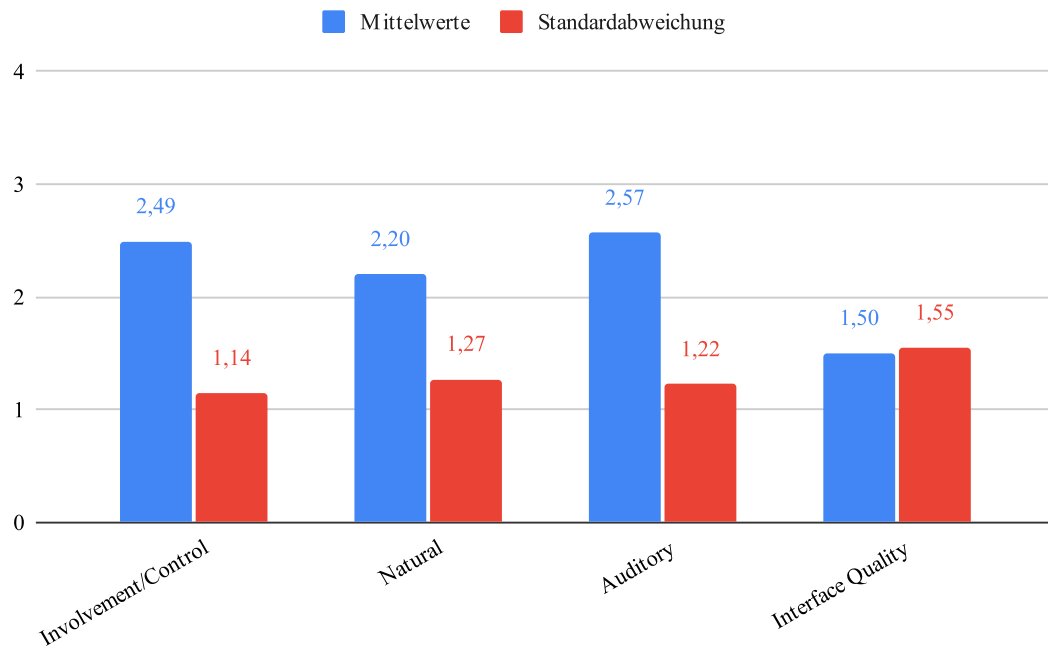


Abbildung 5.5: Mittelwerte und Standardabweichungen der Subskalen im PQ (eigene Darstellung, basierend auf den erhobenen Daten)

Die Ergebnisse der Subskalen (siehe Abbildung 5.5) unterstreichen die Erkenntnisse aus den Hauptkategorien des PQ. Besonders deutlich wird dies bei der auditiven (engl. „auditory“) Kategorie, die mit einem Mittelwert von 2.57 (SD = 1.22) vergleichsweise hoch bewertet wurde. Dies legt nahe, dass die akustische Gestaltung des Spiels von den Spielern positiv wahrgenommen wurde. Demgegenüber stand die Bewertung der Interfacequalität (engl. „interface quality“) mit einem Mittelwert von 1.50 (SD = 1.55). Das deutet darauf hin, dass die Bedienbarkeit oder das visuelle Feedback der Benutzeroberfläche undeutlich war. Die Bewertung der Natürlichkeit (engl. „natural“) mit einem Mittelwert von 2.20 (SD = 1.27) sagt eine überwiegend natürliche Wahrnehmung der Spielumgebung aus. Die Kategorie Involvement/Control (dt. Beteiligung/Kontrolle) (M = 2.49, SD = 1.14) deutet darauf hin, dass die Spieler zwar eine gewisse Kontrolle über das Geschehen empfanden, jedoch noch Verbesserungspotenzial hinsichtlich der Interaktionsmöglichkeiten bestand.

### 5.1.6 Vergleich der statischen und dynamischen Dialoge

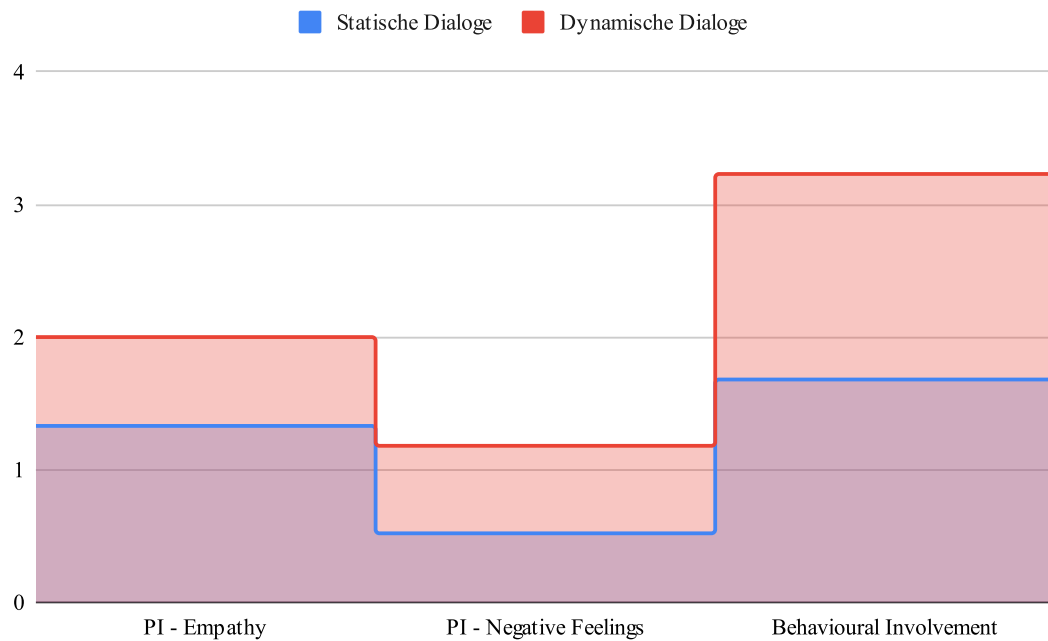


Abbildung 5.6: Mittelwerte der Dialogsysteme im Vergleich (eigene Darstellung, basierend auf den erhobenen Daten)

Die Vergleichswerte der beiden Dialogsysteme aus dem Social-Presence-Modul (siehe Abbildung 5.6) zeigen deutliche Unterschiede: Die Mittelwerte des dynamischen Dialogsystems liegen durchgängig höher als die des statischen Systems, während die allgemeine Verteilung der Werte erhalten bleibt. Besonders markant ist der Unterschied im Behavioural Involvement ( $M = 1.68$  zu  $M = 3.23$ ), was darauf hindeutet, dass die dynamischen Dialoge einen signifikant stärkeren Einfluss auf das Verhalten der Spieler hatten. Ebenfalls auffällig ist die deutliche Zunahme der Negative Feelings ( $M = 0.52$  zu  $M = 1.18$ ), während der Anstieg der Empathie ( $M = 1.33$  zu  $M = 2.00$ ) vergleichsweise moderat ausfiel. Dies zeigt, dass dynamische Dialoge zwar wesentlich mehr Einfluss auf das Verhalten der Spieler ausübte, die Veränderungen jedoch nicht ausschließlich als positiv wahrgenommen wurden.

### 5.1.7 Streuung und individuelle Unterschiede

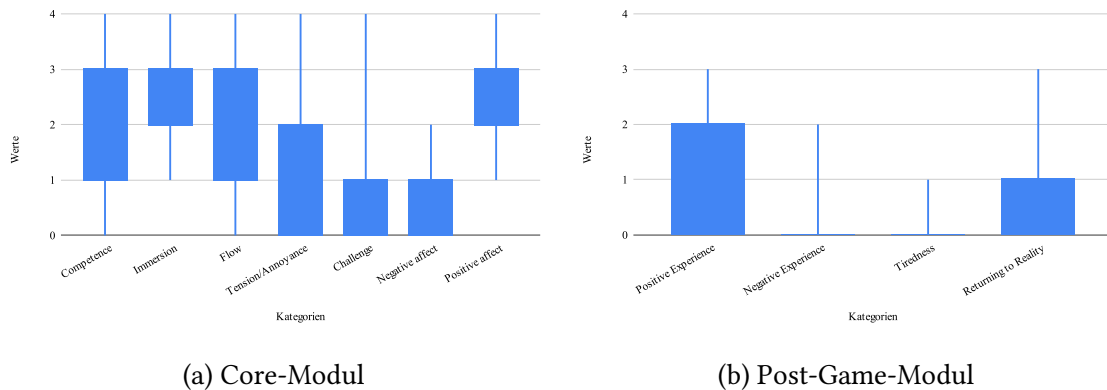


Abbildung 5.7: Streuung der Core- und Post-Game-Module des GEQ nach Kategorien (eigene Darstellung, basierend auf den erhobenen Daten)

Die Boxplots (siehe Abbildung 5.7) veranschaulichen eine vergleichsweise große Streuung der Werte über alle Kategorien der Core- und Post-Game-Module des GEQ (siehe Abbildungen 5.7a; 5.3), die insbesondere auf wenige extreme Antworten zurückzuführen ist. Besonders ausgeprägt zeigt sich dies in den Kategorien Tension/Annoyance und Challenge (dt. Herausforderung), die vereinzelte Ausreißer aufweisen: Während sich drei Teilnehmer als leicht gereizt beschrieben, gaben sieben an, keine Reizung verspürt zu haben. Ebenso berichteten zwei Teilnehmer von Frustration, während die übrigen kaum oder gar keinen Frust empfanden. Ein Teilnehmer empfand das Spiel als besonders anspruchsvoll und musste sich entsprechend anstrengen, während die Mehrheit nur eine geringe bis moderate kognitive Belastung angab. Demgegenüber weisen Kategorien, die negative Aspekte der Spielerfahrung abbilden, wie Negative Affect, Negative Experience und Tiredness, eine geringe Streuung auf. Auch die positiven Kategorien, darunter Immersion, Positive Affect und Positive Experience, liegen in einem relativ geringen Abstand zueinander. Dies deutet darauf hin, dass die Mehrheit der Teilnehmer das Spielerlebnis ähnlich bewertet hat und es insgesamt als positiv empfunden wurde. Etwas größere Unterschiede zeigen sich in den Kategorien Flow, Competence und Returning to Reality (dt. Rückkehr zur Realität), was möglicherweise auf individuelle Unterschiede wie unterschiedliche Erfahrungswerte im Umgang mit KI zurückzuführen ist.

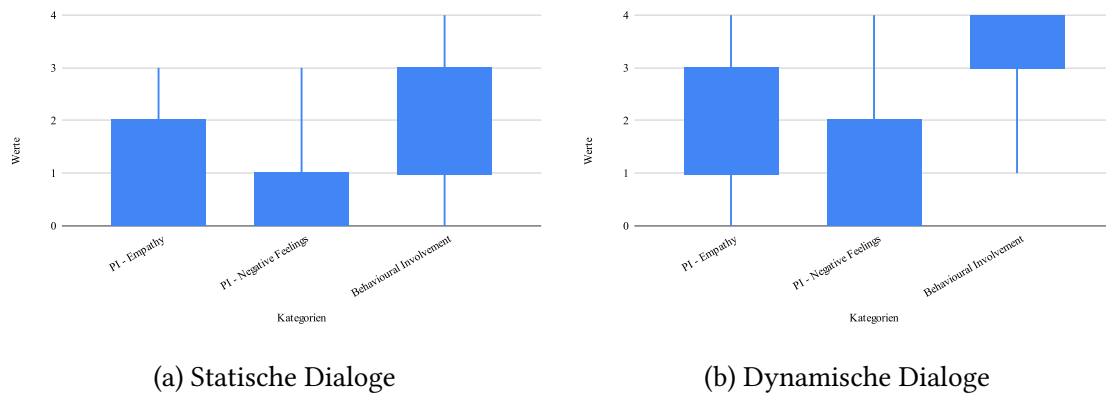


Abbildung 5.8: Streuung der Social-Presence-Module des GEQ nach Kategorien (eigene Darstellung, basierend auf den erhobenen Daten)

Die Boxplots der Dialogsysteme im Social-Presence-Modul des GEQ (siehe Abbildung 5.8) zeigen eine generelle Verschiebung nach oben bei den dynamischen Dialogen. Insbesondere die Kategorien Empathy und Negative Feelings weisen nun eine größere Streuung auf und decken nahezu den gesamten Wertebereich ab. Gleichzeitig zeigt sich beim Behavioural Involvement ein erhöhter Minimalwert, was darauf hindeutet, dass die Teilnehmer durchweg eine stärkere Verhaltenseinbindung bei den dynamischen Dialogen empfanden. Die Betrachtung der Standardabweichungen verdeutlicht zudem, dass die Werte bei den Negative Feelings hohe Unterschiede ausweisen. Die Abweichung bei Behavioural Involvement im Vergleich zu den statischen Dialogen fällt geringer aus. Dies legt nahe, dass die Wahrnehmung negativer Emotionen unter den Teilnehmern stärker schwankte, während die Einschätzung der aktiven Beteiligung konsistenter ausfiel. Besonders auffällig ist, dass Fragen, die sich auf die Interaktion mit den NPCs bezogen, bei den dynamischen Dialogen signifikant höhere Werte aufwiesen. Die Spieler passten ihr Verhalten individueller an die Antworten der NPCs an und richteten ihre Aufmerksamkeit stärker darauf. Emotionale Reaktionen wurden ebenfalls intensiver erlebt: Die Spieler gaben an, dass ihre Stimmung stärker durch die des dynamischen NPCs beeinflusst wurde als bei den statischen Dialogen. Sie berichteten zudem, dass der dynamische NPC positive Emotionen widerspiegelte – beispielsweise erwiderte er gute Laune. Neben positiven Emotionen wie Freude wurden ihm in Einzelfällen auch negativ konnotierte Emotionen wie Rachsucht oder Schadenfreude zugeschrieben.

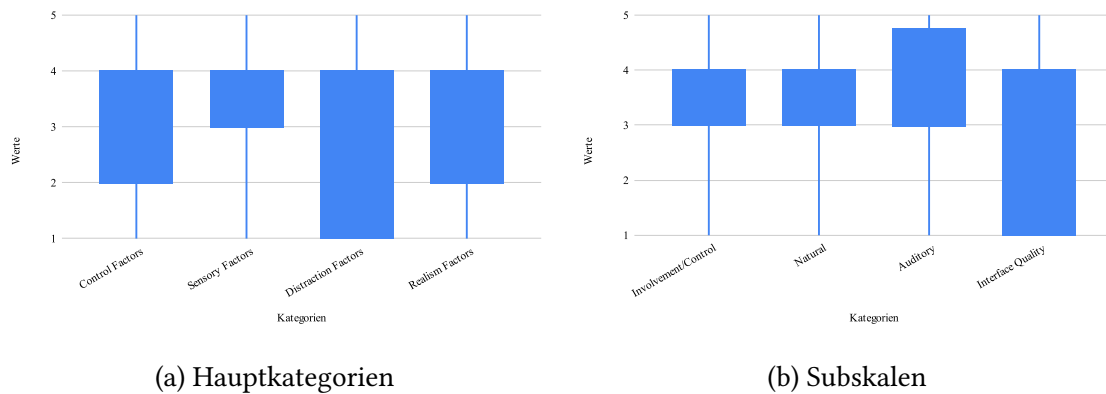


Abbildung 5.9: Streuung des PQ nach Kategorien (eigene Darstellung, basierend auf den erhobenen Daten)

Die Ergebnisse des PQ (siehe Abbildung 5.9) zeigen eine heterogene Verteilung, insbesondere bei subjektiven Wahrnehmungsfaktoren. Während sensorische Aspekte, etwa die auditive und natürliche Wahrnehmung, durchgehend positiv bewertet wurden, weisen Kategorien, die auf individueller Erfahrung und persönlicher Interpretation beruhen, eine größere Streuung auf.

### 5.1.8 Signifikanz und Stärke des beobachteten Effekts

Zum Vergleich der beiden Dialogsysteme wurden t-Tests durchgeführt, um zu prüfen, ob die Unterschiede in den Immersionswerten statistisch signifikant sind. Die Analyse ergab p-Werte von unter 0.005 ( $p < 0.005$ ) in allen Vergleichskategorien, was darauf hinweist, dass die beobachteten Unterschiede mit hoher Wahrscheinlichkeit nicht zufällig entstanden sind. Dies spricht dafür, dass die Art des Dialogsystems in dieser Studie einen signifikanten Einfluss auf die Immersion hatte. Zur Bestimmung der Effektgröße wurde Cohen's d berechnet. Die Werte lagen in allen drei Kategorien bei  $d = 0.6$ , was gemäß gängiger Interpretationen auf einen mittelstarken Effekt hinweist. Dies kann so interpretiert werden, dass die Wahl zwischen statischen und dynamischen Dialogsystemen einen nachweisbaren moderaten Einfluss auf die Immersion hatte.

## 5.2 Analyse qualitativer Ergebnisse

Während die quantitativen Ergebnisse darauf hindeuten, dass dynamische Dialoge tendenziell immersiver wahrgenommen wurden, liefern die offenen Antworten wichtige Erkenntnisse über die zugrunde liegenden Ursachen. In diesem Abschnitt folgt daher eine inhaltliche Analyse der offenen Fragen, um die subjektiven Eindrücke der Teilnehmer vorzustellen. Die Antworten wurden zunächst inhaltlich kategorisiert, wobei wiederkehrende Themen identifiziert wurden. Besonders häufig wurden Aspekte wie Immersion, Dialogqualität, Natürlichkeit sowie die Präferenzen für statische oder dynamische Dialoge genannt. Die wahrgenommene Natürlichkeit der Dialoge zeigte eine klare Tendenz: Dynamische Dialoge wurden als deutlich natürlicher als statische emp-

funden. Die Teilnehmer sahen den Grund dafür vor allem in der größere Freiheit der Interaktion und der Ungewissheit über die Antworten. Insbesondere emotionale Eigenschaften wie Zuneigung oder Misstrauen gegenüber dem Spieler trugen zur Immersion bei. Einschränkungen in der Natürlichkeit wurden hingegen vor allem durch technische Limitierungen verursacht: Generische oder sich wiederholende Antworten, Interaktionsabbrüche oder Missverständnisse des NPC führten zu unnatürlichen Interaktionen. Statische Dialoge wurden überwiegend als langweilig und einseitig beschrieben, jedoch nicht als grundsätzlich unnatürlich. Besonders zu Beginn der Spielerfahrung wurden sie von den Teilnehmern als informativer wahrgenommen. Erst im direkten Vergleich mit den dynamischen Dialogen erschienen sie eingeschränkter und uninteressanter.

*„Die statischen Dialoge waren etwas mühselig, was mein Eintauchen in die Welt gemindert hat, allerdings habe ich so auch sehr effizient viel über die Personenkonstellation erfahren. Mehr als wenn ich zuerst mit Bjorn, dem dynamischen NPC, gesprochen hätte.“ (P5)*

Die Wahrnehmung der Dialogqualität wurde in erster Linie anhand der Kohärenz und Konsistenz der Gespräche bewertet. Dynamischen Dialogen wurden positive Eigenschaften wie Abwechslung, Unvorhersehbarkeit und Spannung zugeschrieben. Allerdings wiesen die Teilnehmer auch auf Schwächen hin: ein begrenzter Wortschatz, eine gewisse Oberflächlichkeit sowie gelegentliche Fehlinterpretationen trugen dazu bei. Im Gegensatz dazu wurden statische Dialoge als repetitiv, aber informativ beschrieben. Da sie vom Spieler nicht aktiv beeinflusst werden konnten, wurde ihre interaktiven Qualitäten als limitierend wahrgenommen. Dennoch spielten sie eine entscheidende Rolle für die Vermittlung von Informationen. In Bezug auf den Einfluss der Dialoge auf die Immersion wurden sowohl positive als auch negative Effekte identifiziert. Besonders die Interaktion mit dem dynamischen NPC wurde von den Teilnehmern als immersiv und bereichernd beschrieben. Ihm wurden charakterliche Eigenschaften, individuelle Verhaltensweisen sowie eine aussagekräftige Persönlichkeit zugeschrieben, die ihr Spielerlebnis prägten. Viele Teilnehmer empfanden ihn als realistischer und nahbarer als die statischen NPCs.

*„Dass ich Bjorn wirklich alles fragen konnte, was ich wollte, und er immer geantwortet hat und dabei sogar humorvoll reagieren konnte. Außerdem hatte ich das Gefühl, dass ich von Bjorn bessere Antworten bekommen habe, welche mir mehr für einen Fortschritt des Spiels geholfen haben.“ (P2)*

Technische Einschränkungen stellten die wesentliche Quelle für negative Rückmeldungen dar. Teilnehmer bemängelten insbesondere gelegentliche Verzögerungen in den Antwortzeiten sowie Interaktionsabbrüche, die das Spielerlebnis beeinträchtigen konnten. Allerdings wurden diese negativen Aspekte durch die positiven Eigenschaften des dynamischen Systems weitgehend ausgeglichen. Insgesamt wurde die Interaktion mit dem dynamischen NPC als hochgradig immersiv wahrgenommen. Im direkten Vergleich zu den dynamischen Dialogen wurden die statischen Dialoge von den Teilnehmern als weniger immersiv empfunden. Interessanterweise wurde diese Ansicht häufig erst nach der Interaktion mit dem dynamischen NPC klar: Während statische Dialoge zunächst

als funktional und informativ wahrgenommen wurden, erschienen sie nach dem Erleben der dynamischen Variante als eingeschränkt und weniger ansprechend. Dennoch wurden auch positive Aspekte des statischen Systems genannt. Besonders das World Building wurde hervorgehoben, da die vorgegebenen Dialoge gezielt dazu dienten, Hintergrundinformationen über die Spielwelt und die Beziehungen zwischen den NPCs zu vermitteln. Diese Informationen hätten die Teilnehmer nach eigenen Aussagen durch den dynamischen NPC allein nicht in Erfahrung gebracht. Die Präferenz für das dynamische Dialogsystem war unter den Teilnehmern eindeutig: Alle gaben an, dass sie das dynamische System bevorzugten. In den qualitativen Rückmeldungen wurden überwiegend positive Argumente für dynamische Dialoge angeführt, während kaum Kritik am statischen System geäußert wurde. Diese Präferenz wurde durch die Interaktionszeiten, aus den im Fragebogen hochgeladenen Log-Dateien bestätigt: Etwa die Hälfte der gesamten Spielzeit wurde mit der Interaktion mit dem dynamischen NPC verbracht, während die Gespräche mit den statischen NPCs im Durchschnitt nur etwa halb so lange dauerten. Dieses Verhalten deutet darauf hin, dass die Spieler von der Offenheit und Individualität der dynamischen Dialoge besonders angezogen wurden.

*„In den statischen [Dialogen] habe ich nur eine Auswahl, während ich in den dynamischen [Dialogen] selbst den Kontext erzeuge, auf dem Bjorns Antwort beruht. Mein Kontext wandelt selbst generische Antworten in neue Blickwinkel auf die Situation.“ (P8)*

Um die häufigsten Begriffe aus den Antworten visuell darzustellen, wurde eine Wortwolke (siehe Abbildung 5.10) erstellt. Diese zeigt, welche Begriffe besonders oft im Zusammenhang mit den Dialogen und der Immersion verwendet wurden. Begriffe mit hoher Relevanz sind größer dargestellt und können darauf hindeuten, welche Aspekte für die Spielerfahrung besonders prägend waren.



Abbildung 5.10: Wortwolke mit häufig genannten Begriffen (eigene Darstellung, basierend auf den erhobenen Daten)

Auffällig ist, dass Begriffe wie *dynamischen*, *antworten* und *statischen* häufig vorkamen, was darauf hindeutet, dass sich die Teilnehmer gezielt mit den beiden Dialogsystemen auseinandergesetzt haben. Die häufige Nennung von *bjorn* verdeutlicht zudem, dass der dynamische NPC eine zentrale Rolle in dem Fragebogen spielte. Ebenso treten Begriffe wie *wiederholt* und *repetitiv* auf, was womöglich auf inkohärente Verhaltensweisen der generierten Inhalte hindeutet. Zusätzlich lassen sich Hinweise auf die Immersion innerhalb der Dialoge erkennen: Begriffe wie *interessant*, *glaubwürdig* und *gestärkt* deuten darauf hin, dass sich die Teilnehmer intensiv mit dem Vergleich der beiden Dialogsysteme beschäftigt haben. Die Anzahl der positiv konnotierten Begriffe legt nahe, dass sich die Teilnehmer mehr auf die Formulierung der positiven Erfahrungen konzentrierten, anstatt negative Aspekte auszuführen.

# Kapitel 6

## Diskussion

### 6.1 Interpretation der Ergebnisse

Die Ergebnisse der quantitativen Analyse belegen ein insgesamt hohes Maß an Immersion bei den Teilnehmern. Besonders hervorzuheben ist die Kategorie Immersion im Core-Modul des GEQ, die mit  $M = 2.77$  ( $SD = 0.89$ ) den höchsten Wert erreichte. Dies deutet darauf hin, dass die grundlegenden Bedingungen für eine immersive Spielerfahrung gegeben waren. Auch die Kategorien Flow und Competence erzielten hohe Mittelwerte, was auf eine tiefgehende immersive Spielerfahrung schließen lässt. Während statische Dialoge oft als kohärenter wahrgenommen wurden, führten dynamische Dialoge dazu, dass Spieler stärker in die Interaktion eingebunden waren. Die Kategorie Behavioural Involvement war mit  $M = 3.23$  ( $SD = 0.77$ ) bei den dynamischen Dialogen signifikant höher als bei den statischen Dialogen mit  $M = 1.68$  ( $SD = 1.26$ ), was darauf hindeutet, dass Spieler im dynamischen Dialogsystem stärker mit dem NPC interagierten und den Eindruck hatten, dass ihre Eingaben einen direkten Einfluss auf die Unterhaltung hatten. Dies deckt sich mit den Erkenntnissen von Ermi und Mäyrä (2005), die betonen, dass eine starke imaginative Immersion entsteht, wenn Spieler sich emotional mit einem Spiel verbinden können (Ermi und Mäyrä, 2005). Die ursprüngliche Hypothese, dass dynamische Dialoge eine stärkere Immersion fördern, konnte durch die Ergebnisse der Studie weitgehend bestätigt werden. Dynamische Dialoge zeigten besonders im Bereich der sozialen Interaktion eine deutliche positive Veränderung. Gleichzeitig zeigen die qualitativen Rückmeldungen jedoch, dass dieser Effekt nicht ausschließlich positiv war. Während viele Teilnehmer die Freiheit und Offenheit der dynamischen Dialoge lobten, kritisierten einige eine gewisse Ziellosigkeit oder unpassende Antworten.

*„Bjorn selbst hat Spaß gemacht. Allerdings lieferte er schnell verallgemeinerte Antworten oder Basisantworten, wenn ihm dazu nichts einfiel.“ (P7)*

Die Ergebnisse zeigen, dass ein immersives Spielerlebnis nicht nur durch Interaktivität und Freiheit entsteht, sondern neben Kohärenz und narrative Konsistenz auch von Spielelementen wie Grafik-, Sound- und Interfacedesign abhängt. Dies deckt sich ebenfalls mit den Ergebnissen von Ermi und Mäyrä (2005), da Immersion sowohl durch audiovisuelle als auch durch narrative und interaktive Elemente gefördert wird (Ermi und

Mäyrä, 2005). Die qualitativen Rückmeldungen verdeutlichen, dass dynamische Dialoge von den Teilnehmern überwiegend als natürlicher und lebendiger wahrgenommen wurden als statische Dialoge. Besonders wurde dabei die freie Texteingabe hervorgehoben, die maßgeblich zu einer immersiveren Interaktion mit der Spielwelt beitragen würde. Diese Wahrnehmung spiegelt sich auch in den Ergebnissen des Social-Presence-Moduls wider, wo Behavioural Involvement bei den dynamischen Dialogen deutlich höher lag.

*„Die statischen Dialoge wirkten repetitiv und als ob man auf einer festen Spur fährt, die dynamischen Dialoge wirkten so, als könnte ich eine freie Konversation führen.“ (P4)*

Gleichzeitig wurden auch Einschränkungen dynamischer Dialoge erkennbar. Während die Grundstruktur der Antworten überwiegend als natürlich empfunden wurde, berichteten einige Teilnehmer von unpassenden Aussagen des NPCs, die zu Inkonsistenzen führten und in manchen Fällen die Immersion beeinträchtigten.

*„[Bei den dynamischen Dialogen] hat Bjorn sich ständig wiederholt oder sozusagen an mir vorbeigeredet und nicht das beantwortet, was ich eigentlich wissen wollte.“ (P2)*

Diese Beobachtungen stehen im Einklang mit den Erkenntnissen von Mateas und Stern (2005) zu *Façade*, wonach Spieler hohe Erwartungen an dynamische Dialogsysteme haben und Frustration empfinden, wenn ihre Eingaben nicht wie erwartet interpretiert werden (Mateas und Stern, 2005). Ein weiteres häufig angesprochenes Thema war das Ausgleich zwischen Interaktivität und Struktur. Während einige Teilnehmer die Freiheit genossen, berichteten andere, dass die fehlende Führung zu Unsicherheit oder Orientierungslosigkeit führte. Dies bestätigt die Erkenntnisse von Kumaran et al. (2023), die darauf hinweisen, dass dynamische Inhalte zwar Flexibilität bieten, jedoch klare narrative Wegweiser benötigen, um ein kohärentes Spielerlebnis zu gewährleisten (Kumaran et al., 2023).

*„[...] Alternativ würde ich auch nur dynamische Dialoge interessant finden, allerdings müsste man dann auch auf irgendeine Art und Weise Hinweise bekommen, was man bei der Person so ansprechen kann, sonst kommt man nicht auf die Idee, was man so fragen könnte.“ (P2)*

## 6.2 Kritische Reflexion und Herausforderungen

Eine wesentliche methodische Einschränkung der Studie lag in der geringen Stichprobengröße. Mit nur zehn Teilnehmern waren die Ergebnisse zwar richtungsweisend, jedoch nicht repräsentativ für eine breitere Spielerschaft. Zudem nahmen alle Probanden freiwillig an der Studie teil, was zu einer potenziellen Verzerrung der Ergebnisse durch höhere Affinität zum Forschungsgegenstand geführt haben könnte. Zwar waren die unterschiedlichen Vorkenntnisse im Umgang mit KI deutlich, jedoch hatten die meisten Teilnehmer bereits umfangreiche Erfahrungen mit Videospielen gemacht. Ein weiteres potenzielles

Problem stellte eine mögliche Verzerrung der Ergebnisse durch Erwartungseffekte seitens der Teilnehmer dar. Da die Studie explizit den Vergleich zwischen statischen und dynamischen Dialogen untersuchte, könnten einige Teilnehmer stärker auf diese Unterschiede geachtet haben, als es in einer natürlichen Spielumgebung der Fall gewesen wäre. Zusätzlich könnte die Präsentation beider Dialogsysteme innerhalb derselben Spielwelt einen Reihenfolgeeffekt begünstigt haben: Trotz der halblinearen Struktur des Levels interagierten die meisten Teilnehmer zuerst mit den statischen NPCs, bevor sie auf den dynamischen NPC trafen. Diese direkte Gegenüberstellung könnte die Wahrnehmung der Teilnehmer dahingehend beeinflusst haben, ihre Aussagen polarisiert darzustellen. Allerdings wird dabei eine deutlichere Herausarbeitung der Unterschiede der Beiden Dialogsysteme begünstigt. Aufgrund der geringen Stichprobengröße wurde sich letztendlich für diese Version entschieden.

*„Die statischen Dialoge wirkten repetitiv und als ob man auf einer festen Spur fährt, die dynamischen Dialoge wirkten so, als könnte ich eine freie Konversation führen.“ (P4)*

Auch der Einfluss externer Variablen und Umstände stellte eine Herausforderung dar. Um möglichst viele Teilnehmer zu locken, wurde die Studie so konzipiert, dass sie eigenständig durchgeführt werden konnte. Dies reduzierte jedoch die Kontrolle über Ablenkungen oder Unterbrechungen während der Durchführung erheblich. Zudem waren die Teilnehmer auf ihre eigene Hardware angewiesen, was bedeutet, dass leistungsbedingte Ladezeiten oder unerwartete Systemabbrüche nicht vollständig ausgeschlossen werden konnten. Eine kontrollierte Laborumgebung hätte dieser Herausforderung entgegenwirken können. Eine der größten Herausforderungen dieser Studie lag in der Messbarkeit von Immersion. Bereits Brown und Cairns (2004) weisen darauf hin, dass Immersion ein schwer fassbares Konzept ist (Brown und Cairns, 2004). Zudem könnte die Interpretation der Messinstrumente selbst fehleranfällig gewesen sein: Likert-Skalen beruhen auf Selbstberichten, die durch soziale Erwünschtheit oder individuelle Selbsteinschätzungseffekte beeinflusst sein könnten. Darüber hinaus wurde der Fragebogen zur besseren Verständlichkeit ins Deutsche übersetzt, was – wenn auch in geringem Maße – zu möglichen Interpretationsfehlern geführt haben könnte (W. A. IJsselsteijn et al., 2013). Auch die offenen Fragen sind subjektiv: Zwar boten sie wertvolle Einblicke in die Wahrnehmung der Teilnehmer, jedoch variierten sie in der Ausführlichkeit der Antworten erheblich. Eine technische Herausforderung ergab sich aus den Limitierungen des LLMs. Während die meisten Antworten innerhalb einer akzeptablen Zeitspanne generiert wurden, kam es gelegentlich zu Verzögerungen, die den Spielfluss unterbrachen. Zudem konnte das gewählte Modell, obwohl es sich grundsätzlich als geeignet erwies, nicht in allen Bereichen optimiert werden. Aufgrund der kurzen Entwicklungszeit von nur zwei Wochen war es nicht möglich, alle Aspekte des Dialogsystems in der Tiefe auszuarbeiten, wie es für eine marktreife Anwendung erforderlich wäre. Beispielsweise hätten weiterführende Tests zur Feinabstimmung der generierten Antworten oder eine präzisere Fehlerkorrektur die Konsistenz und Kohärenz der Dialoge weiter verbessern können. Klinkert et al. (2024) weisen darauf hin, dass kleinere Sprachmodelle Schwierigkeiten haben, langfristig konsistente Konversationen aufrechtzuerhalten, was durch den Einsatz größerer Modelle

möglicherweise zu einer immersiveren Spielerfahrung geführt hätte (Klinkert et al., 2024). Insgesamt hätte eine größere Stichprobe unter kontrollierteren Bedingungen zu statistisch aussagekräftigeren Ergebnissen führen können. Eine weiterführende Studie mit einer größeren Teilnehmerzahl könnte die Erkenntnisse dieser Arbeit weiter validieren und den Einfluss von dynamischen Dialogsystemen auf die Immersion vertiefen.

# Kapitel 7

## Fazit und Ausblick

### 7.1 Wesentliche Erkenntnisse der Arbeit

Die vorliegende Arbeit untersuchte, wie LLM-generierte Inhalte die Immersion in einem Videospiel beeinflussen, insbesondere im Vergleich zu statischen Dialogsystemen. Die dazu durchgeführte Nutzerstudie zeigte, dass dynamisch generierte Dialoge eine signifikante Wirkung auf die Nutzererfahrung hatten. Die Ergebnisse des GEQ und des PQ belegten, dass Spieler bei der Interaktion mit dem dynamischen NPC ein erhöhtes Maß an Behavioural Involvement empfanden. Dies deutet darauf hin, dass sie sich stärker mit dem NPC auseinandersetzten und sich tiefer auf die Konversation einließen. Die qualitative Analyse bestätigte das, da Teilnehmer angaben, dynamische Dialoge würden dank ihrer Natürlichkeit und Unvorhersehbarkeit zu einer immersiveren Spielerfahrung beitragen. Gleichzeitig wurden jedoch auch Schwächen deutlich: Inkonsistenzen in den generierten Antworten und Fehlinterpretationen der Nutzereingaben führten zu Frustration oder Immersionsbrüchen. Dennoch bevorzugten alle Teilnehmer das dynamische Dialogsystem gegenüber den statischen Alternativen. Die Forschungsfrage der Arbeit - wie narrative LLM-Inhalte die Immersion in einem Videospiel beeinflussen - kann somit damit beantwortet werden, dass dynamische Dialogsysteme die Immersion signifikant beeinflussen - sowohl positiv als auch negativ.

### 7.2 Implikationen fürs Game Design

Die Ergebnisse dieser Arbeit liefern aussagekräftige Erkenntnisse für die Entwicklung immersiver Spiele. Besonders im Bereich von interaktiver Narration könnten LLMs genutzt werden, um flexiblere und dynamischere Dialogsysteme zu schaffen. Traditionelle, statische Dialoge bieten den Vorteil einer geleiteten Handlung mit hoher narrativer Qualität, während dynamische Dialoge eine stärkere Individualisierung und Variation ermöglichen. Die Herausforderung besteht darin, ein gleichermaßen narrative Kohärenz als auch spielerische Freiheit zu gewährleisten. Potenziale solcher Dialogsysteme liegen in der Personalisierung der Spielererfahrung, ihrer Flexibilität und ihrer Skalierbarkeit. LLMs ermöglichen es, dass NPCs individuell auf den Spielstil und die Entscheidungen der Spieler reagieren, was die Widerspielbarkeit und Immersion erhöht. Zudem

können solche Dialoge prozedural an neue Kontexte angepasst werden, statt manuell geschrieben und vorprogrammiert zu werden. Die Herausforderungen umfassen dagegen technische Limitierungen, die spielerische Erwartungshaltung und erhöhten Test- und Kontrollaufwand. Kleinere Modelle haben Schwierigkeiten, komplexe Kontexte langfristig aufrechtzuerhalten, Eingaben können fehlinterpretiert werden und die Implementierung dynamischer Systeme erfordert umfangreiche Tests, um sicherzustellen, dass generierte Dialoge sinnvoll und qualitativ hochwertig bleiben. Trotz dieser Herausforderungen zeigen die Ergebnisse, dass LLMs einen bedeutenden Fortschritt im Bereich des Game Design darstellen und das Potenzial haben, zukünftiges Game Design maßgeblich zu beeinflussen.

### 7.3 Ideen für weitere Forschung

Diese Studie wurde an einem kleinen Prototypen mit fünf NPC durchgeföhrt, ließe sich jedoch weiter skalieren. In zukünftigen Arbeiten könnte untersucht werden, wie dynamische Dialoge größere, komplexere Spielwelten beeinflussen. Beispielsweise könnten LLM-generierte Inhalte in Rollenspielen wesentlich realistischere NPCs geschaffen werden, die ihre Umgebung kontextsensitiv wahrnehmen und darauf reagieren. Ein weiterer Forschungsansatz wäre der Einfluss dynamischer Dialoge auf kulturelle Wahrnehmungen. Unterschiedliche kulturelle Prägungen könnten beeinflussen, wie natürliche Sprache und Verhalten wahrgenommen werden: Möglicherweise variieren die Prioritäten zwischen Interaktivität, Entscheidungsfreiheit und narrativer Struktur je nach kulturellem Hintergrund. Ein wichtiges Forschungsfeld betrifft ethische und moralische Fragestellungen sowie den sicheren Umgang mit KI-gestützten Systemen. Glaubwürdige, aber ethisch fragwürdige Ausgaben stellen bereits heute eine Herausforderung dar - die Kontrolle über generative KI bleibt ein zentrales Thema zukünftiger Forschung. Zudem könnte sich die Forschung verstärkt mit narrativer Optimierung befassen. Kohärenzverbesserungen, umfangreichere Speichermechanismen und neue technologische Entwicklungen tragen kontinuierlich dazu bei, KI authentischer zu gestalten. Forschung in diesem Bereich könnte die in dieser Studie identifizierten Herausforderungen adressieren und dynamische Narration weiter immersiver gestalten. Ebenso könnte weiterführende Forschung zur Immersion dazu beitragen, KI präziser an die Anforderungen und Bedürfnisse der Spieler anzupassen, um technische Limitierungen zu kompensieren. Während audiovisuelle Aspekte eines Spiels maßgeblich zur Wahrnehmung von Immersion beitragen, können Fortschritte im Umgang mit NLP genutzt werden, um die menschlichen Sinneswahrnehmung noch stärker zu beeinflussen. Schließlich könnten die in dieser Studie untersuchten Konzepte in Verbindung mit anderen KI-Technologien dazu beitragen, immersivere und fortschrittlichere Spielmechaniken zu entwickeln, indem sich diese Systeme gegenseitig regulieren oder erweitern.

# Literatur

- Adams, E. (1999). Three problems for interactive storytellers. *Designer's Notebook Column, Gamasutra*, 144.
- Brown, E., & Cairns, P. (2004). A grounded investigation of game immersion. *CHI'04 extended abstracts on Human factors in computing systems*, 1297–1300.
- Corbin, J., & Strauss, A. (2014). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience* (Bd. 1990). Harper & Row New York.
- Dechter, R. (1986). Learning while searching in constraint-satisfaction problems. *University of California, Computer Science Department, Cognitive Systems ...*
- DeJean, P.-H. (2002). *Difficulties and pleasure*. Taylor & Francis, ISBN-13.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 4171–4186.
- Döbel, I., Leis, M., Vogelsang, M., Neustroev, D., Petzka, H., Riemer, A., Rüping, S., Voss, A., Wegele, M., & Welz, J. (2018). Maschinelles Lernen: Kompetenzen, Forschung, Anwendung. *Fraunhofer-Gesellschaft in Kooperation mit Forschungszentrum Maschinelles Lernen im Fraunhofer-Cluster of Excellence Cognitive Internet Technologies Fraunhofer-Allianz Big Data & Künstliche Intelligenz*.
- Ermi, L., & Mäyrä, F. (2005). Fundamental components of the gameplay experience: Analysing immersion. *Proceedings of DiGRA 2005 Conference: Changing Views: Worlds in Play*.
- Frochte, J. (2020). *Maschinelles Lernen: Grundlagen und Algorithmen in Python*. Carl Hanser Verlag GmbH Co KG.
- Géron, A. (2023). *Praxiseinstieg Machine Learning mit Scikit-Learn, Keras und TensorFlow: Konzepte, Tools und Techniken für intelligente Systeme*. o'Reilly.
- Ghahramani, Z. (2003). Unsupervised learning. In *Summer school on machine learning* (S. 72–112). Springer.
- Goldberg, Y. (2017). *Neural network methods in natural language processing*. Morgan & Claypool Publishers.
- Gritsch, S. (2012). Die Likert-Skala–Meinungen abbilden. *ergopraxis*, 5(01), 16–17.
- Hardesty, L. (2017). Explained: neural networks. *MIT News*, 14.

- IJsselsteijn, W., Van Den Hoogen, W., Klimmt, C., De Kort, Y., Lindley, C., Mathiak, K., Poels, K., Ravaja, N., Turpeinen, M., & Vorderer, P. (2008). Measuring the experience of digital game enjoyment. *Proceedings of measuring behavior, 2008*(2008), 88–89.
- IJsselsteijn, W. A., De Kort, Y. A., & Poels, K. (2013). The game experience questionnaire. *Technische Universiteit Eindhoven*.
- Jenkins, H. (2004). Game design as narrative architecture. *Computer, 44*(3), 118–130.
- Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International journal of human-computer studies, 66*(9), 641–661.
- Juul, J. (2005). Games telling stories. *Handbook of computer game studies*, 219–226.
- Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of artificial intelligence research, 4*, 237–285.
- Karani, D. (2018). Introduction to word embedding and word2vec. *Towards Data Science, 1*.
- Klinkert, L. J., Buongiorno, S., & Clark, C. (2024). Evaluating the efficacy of LLMs to emulate realistic human personalities. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, 20*(1), 65–75.
- Kumaran, V., Rowe, J., Mott, B., & Lester, J. (2023). Scenecraft: automating interactive narrative scene generation in digital games with large language models. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, 19*(1), 86–96.
- Langley, P., et al. (2011). The changing science of machine learning. *Machine learning, 82*(3), 275.
- Levy, S. (2024). 8 google employees invented modern AI. here's the inside story [Section: tags]. *Wired*. Verfügbar 26. Januar 2025 unter <https://www.wired.com/story/eight-google-employees-invented-modern-ai-transformers-paper/>
- Lopardo, A. (2020, 8. Januar). *Word2vec to transformers* [Medium]. Verfügbar 26. Januar 2025 unter <https://towardsdatascience.com/word2vec-to-transformers-caf5a3daa08a>
- Mateas, M., & Stern, A. (2005). Procedural authorship: A case-study of the interactive drama *Façade*. *Digital Arts and Culture (DAC), 61*.
- McMahan, A. (2003). Immersion, engagement, and presence: A method for analyzing 3-D video games. In *The video game theory reader* (S. 67–86). Routledge.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems, 26*.
- Patrick, E., Cosgrove, D., Slavkovic, A., Rode, J. A., Verratti, T., & Chiselko, G. (2000). Using a large projection screen as an alternative to head-mounted displays for virtual environments. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems, 478–485*.
- Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al. (2018). Improving language understanding by generative pre-training.(2018).

- Ravaja, N., Salminen, M., Holopainen, J., Saari, T., Laarni, J., & Järvinen, A. (2004). Emotional response patterns and sense of presence during video games: Potential criterion variables for game design. *NordiCHI*, 4, 339–347.
- Rojas, R. (2013). *Neural networks: a systematic introduction*. Springer Science & Business Media.
- Rojo-Echeburúa, A. (2024, 14. November). *Kleine Sprachmodelle: Ein Leitfaden mit Beispielen*. Verfügbar 11. März 2025 unter <https://www.datacamp.com/blog/small-language-models>
- Sweetser, P. (2024). Large language models and video games: A preliminary scoping review. *Proceedings of the 6th ACM Conference on Conversational User Interfaces*, 1–8.
- Swing, E. (2000). Adding immersion to collaborative tools. *Proceedings of the fifth symposium on Virtual reality modeling language (Web3D-VRML)*, 63–68.
- Thue, D., Bulitko, V., Spetch, M., & Wasylishen, E. (2007). Interactive storytelling: A player modelling approach. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 3(1), 43–48.
- Toews, R. (n. d.). *The next generation of large language models* [Forbes] [Section: AI]. Verfügbar 26. Januar 2025 unter <https://www.forbes.com/sites/robtoews/2023/02/07/the-next-generation-of-large-language-models/>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3), 225–240.

# Eigenständigkeitserklärung

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

**Künstliche Intelligenz und Immersion:  
Eine Nutzerstudie zu Large Language Models im Game Design**

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 Angaben der Quelle kenntlich gemacht.

Hamburg, 14. März 2025