

# Offenheit als Risiko und Ressource: Eine Fallstudie zur Gestaltung einer gamifizierten KI-Anwendung gegen Diskriminierung

Nicola Marsden <sup>1</sup>

**Abstract:** Offenheit gilt in der Informatik als zentrales Prinzip – verbunden mit Transparenz, Teilhabe und Innovationsfähigkeit. Doch was bedeutet Offenheit, wenn digitale Technologien auf sensiblen Erfahrungen marginalisierter Gruppen basieren? Am Beispiel der KITE-II-App, einer gamifizierten Anwendung zur Unterstützung von Gründerinnen im Umgang mit geschlechtsspezifischer Diskriminierung, wird Offenheit als ambivalentes Gestaltungsprinzip untersucht. Die App kombiniert interaktive Gesprächssimulationen mit KI-gestütztem Feedback auf Basis eines kuratierten Prompt-Designs, einer strukturierten Bias-Wissensbasis und einer empathisch gerahmten Analyse. Anhand zentraler Designentscheidungen werden Spannungsfelder zwischen Sichtbarmachung und Schutz, Beteiligung und Deutungshoheit, Transparenz und narrativer Integrität herausgearbeitet. Offenheit wird in diesem Zusammenhang nicht als technisches Merkmal verstanden, sondern als ethisch gerahmte Gestaltungspraxis. Der Beitrag entwickelt Kriterien für eine verantwortungsvolle, relationale Offenheit, um zu einer kritisch-reflexiven Human-Computer Interaction beizutragen, die gesellschaftliche Machtverhältnisse nicht nur abbildet, sondern aktiv mitgestaltet.

**Keywords:** Openness, Diskriminierung, KI, Gamification, Partizipation, Gender, Ethik

## 1 Einleitung

Offenheit ist ein zentrales Ideal der digitalen Kultur. In der Human-Computer Interaction wird Offenheit oft mit Transparenz, Teilhabe und ethischer Rechenschaft gleichgesetzt. Doch wie lässt sich Offenheit in digitalen Anwendungen umsetzen, die mit sensiblen Erfahrungen arbeiten? Dieser Beitrag untersucht diese Frage anhand einer Fallstudie zur KITE-II-App, einer KI-gestützten, gamifizierten Anwendung für Gründerinnen (KITE steht für "KI-Thinktank female Entrepreneurship"). Die App nutzt das Potenzial spielerischer Szenarien für strategisches Empowerment marginalisierter Gruppen [Mc11]. Die App wurde entwickelt, um geschlechtsspezifische Diskriminierung im Gründungskontext sichtbar zu machen und kommunikative Handlungsspielräume strategisch zu erweitern [Re24]. Dabei treffen unterschiedliche Anforderungen aufeinander: Offenheit gegenüber Nutzer\*innen und Öffentlichkeit, Schutz vor Exposition und Fehlinterpretation, Partizipation ohne Repräsentationszwang.

Offenheit kann in solchen Kontexten nicht als universeller Wert, sondern als kontextspezifisch auszubalancierende Praxis verstanden werden. Der Beitrag analysiert die gestalterischen,

<sup>1</sup> Hochschule Heilbronn, Lab für Sozioinformatik, Max-Planck-Straße 39, 74081 Heilbronn, Germany, nicola.marsden@hs-heilbronn.de,  <https://orcid.org/0000-0003-1316-8147>

technischen und ethischen Implikationen dieser Haltung und zeigt, wie Offenheit verantwortungsvoll, begrenzt und machtsensibel realisiert werden kann. Damit leistet er einen Beitrag zur Diskussion über kritische, gesellschaftlich eingebettete Informatik.

Abschnitt 2 stellt die KITE-II-App vor, Abschnitt 3 analysiert Spannungsfelder von Offenheit, Abschnitt 4 reflektiert zentrale Gestaltungsentscheidungen und Abschnitt 5 schließt mit Kriterien für verantwortungsvolle Offenheit ab.

## 2 KITE II als Fallstudie

Gründungssituationen sind häufig von strukturellen Ungleichheiten geprägt, sei es in Kreditverhandlungen, Gesprächen mit Investor\*innen oder bei der Preisgestaltung gegenüber Kund\*innen. Frauen erleben in diesen Kontexten spezifische Herausforderungen, die nicht auf ein Defizit an Kompetenz oder Ambition zurückzuführen sind, sondern auf tief verwurzelte geschlechtsspezifische Biases [TS21]. Digitale Unterstützungsangebote adressieren solche Machtasymmetrien bislang kaum. Vorherrschende Formate betonen individuelle Anpassungsleistungen, etwa in Form von „Resilienz“ oder „Durchsetzungskraft“, und blenden strukturelle Dimensionen systematisch aus. Das öffentlich geförderte Projekt KITE II setzt hier an [Ma25; SB22]: Aufbauend auf dem Vorgängerprojekt KITE I, in dem typische Diskriminierungsmuster im Gründungskontext identifiziert und in szenariobasierte Formate überführt wurden [SB22], entstand im Projekt KITE II eine gamifizierte, KI-gestützte Anwendung.

Ziel ist es, diskriminierende Gesprächsdynamiken zu erkennen und ihnen strategisch zu begegnen. Dabei setzt KITE II nicht auf normative Reaktionsmuster, sondern auf strukturreflexives Empowerment. Spieler\*innen durchlaufen interaktive Dialogszenarien aus dem Gründungsalltag und erhalten daraufhin eine qualitative, KI-gestützte Rückmeldung, die auf einer kuratierten Wissensbasis zu Biases beruht. Entwickelt wurde die App in einem interdisziplinären, iterativ-partizipativen Co-Design-Prozess unter Einbezug der bundesweiten Gründerinnenagentur sowie eines Fachbeirats mit Expertinnen aus Gründungspraxis, Gleichstellung und Digitalisierung. Neben Gründerinnen selbst richtet sich die App auch an Personen, die mit Gründerinnen arbeiten, etwa Coaches, Investierende oder Gründungsberater\*innen, sowie an die Fachöffentlichkeit und Bildungseinrichtungen, die sich mit gendergerechter Innovationspraxis befassen.

Die Offenheit der Anwendung ist bewusst gewählt: KITE II ist ohne Registrierung zugänglich, speichert keine personenbezogenen Daten und erlaubt eine Perspektivewahl beim Einstieg. Dabei bleibt die inhaltliche Rahmung strikt auf ein machtkritisches, reflexives Lernziel fokussiert. Die App versteht sich so als Beitrag zu einer digitalen Bildung, die nicht nur Wissen vermittelt, sondern zur strukturell informierten Reflexion befähigt, und zwar in einem geschützten, partizipativ entwickelten Erfahrungsraum.

Vergleichbare KI-gestützte Feedbacksysteme werden auch in anderen Kontexten wie DEI-Schulungen, Performance-Analysen und inklusiver Führungskommunikation eingesetzt,

etwa zur Analyse diskriminierungssensibler Sprache oder zur Förderung reflektierter Kommunikation [Ko25]. Studien zeigen jedoch, dass der Einsatz algorithmischer Rückmelde-systeme auch mit wahrgenommenem Kontrollverlust einhergehen kann, insbesondere wenn die Funktion und der Nutzen nicht transparent kommuniziert werden [SZ24]. Eine klare Zielkommunikation sowie hybride Modelle, die KI und menschliche Begleitung kombinieren, können die Akzeptanz erhöhen und Widerständen vorbeugen.

## 2.1 Aufbau der App

Die Anwendung beginnt mit einem Einstiegsdialog, der die Nutzer\*innen inhaltlich abholt und instruiert. Zentral ist dabei die Auswahl der eigenen Perspektive: Gründerinnen erhalten eine Einführung in Zielsetzung und Nutzung der App, andere Rollen (z. B. Coach, Investor\*in, Berater\*in) werden zur Perspektivübernahme angeleitet. Diese betrifft ausschließlich den Einstiegsdialog; die Szenarien selbst bleiben identisch.

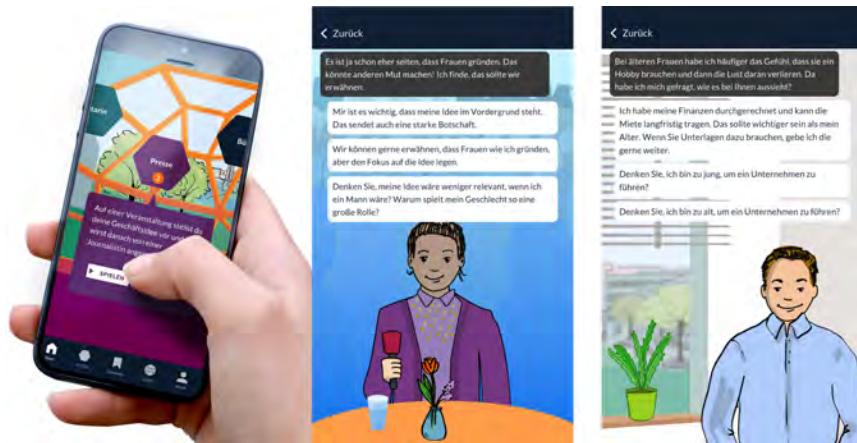


Abb. 1: Die Founder's Bubble (links), Dialogausschnitte aus dem Gespräch mit der Presse (Mitte) und dem Vermieter (rechts)

Nach dem Einstieg betreten Nutzer\*innen ein Foyer mit allen verfügbaren Dialogszenarien (Visual Novels). Diese sind jederzeit pausier- und abbrechbar. Nach der Auswahl eines Szenarios verlaufen die interaktiven Gespräche über verzweigte Antwortpfade in Ich-Perspektive. Die spielende Person wird nicht visuell dargestellt, wodurch Merkmale wie Alter oder Hautfarbe bewusst ausgespart bleiben. Der Fokus liegt auf geschlechtsspezifischer Diskriminierung. Die Gesprächspartner\*innen sind animiert, intersektional divers gestaltet und reagieren dynamisch auf das Spielverhalten.

## 2.2 Spielkonzept und Szenarien

Das Spielkonzept der KITE-II-App basiert auf dialogisch strukturierten Visual Novels, die reale Gründungssituationen simulieren und dabei typische Herausforderungen sichtbar machen. Die Integration spieltypischer Elemente wie verzweigter Entscheidungspfade außerhalb eines Spiels qualifiziert KITE II als gamifizierte Anwendung [De11]: Spieler\*innen navigieren durch interaktive Gesprächsverläufe, in denen sie jeweils zwischen verschiedenen Antwortoptionen wählen, und so den Verlauf des Dialogs aktiv mitgestalten. Insgesamt stehen sieben zentrale Szenarien zur Verfügung, die basierend auf der Forschung in KITE I [SB22] unterschiedliche Stationen und Konstellationen im Gründungsprozess abbilden: Kreditanfrage bei einer Bank, Terminvereinbarung zur Unternehmensgründung, Interview mit einer Journalistin, Besichtigung potenzieller Büroräume, Eltern über die Gründungsabsicht informieren, Verhandlung mit einer potenziellen Kundin und das Gespräch mit einem Angel Investor.

Die Szenarien sind so konzipiert, dass sie unterschiedliche Diskriminierungsmuster und Gesprächsdynamiken sichtbar machen – ein didaktisches Potenzial, das auch in der Forschung zu Visual Novels als reflektive Lernform hervorgehoben wird [CCM21]. Alle Dialoge folgen einer modularen Struktur mit mehreren Antwortpfaden und situativen Eskalationspunkten. Auf diese Weise wird eine differenzierte Reflexion über Handlungsoptionen, Gesprächsverläufe und mögliche Biases ermöglicht.

## 2.3 Feedback-Mechanik: KI-gestützte Gesprächsanalyse

Die anschließende KI-Analyse basiert auf einem strukturierten Prompt-Design [Ma25]. Ziel des Feedbackdesigns war die Förderung strategischer Handlungsfähigkeit, ohne mit normativen Erwartungen zu überfrachten. Spieler\*innen sollen Gespräche strategisch gestalten können, ohne Risiken durch Bias-Adressierung einzugehen. Daher werden explizite Hinweise auf stereotype Äußerungen im Feedback vorsichtig und kontextsensibel formuliert. Zur Qualitätssicherung der KI-Auswertung wurde ein mehrstufiges Prompting entwickelt, das unter anderem auf Debiasing-Ansätzen für LLMs basiert [Ga23].

Während der Entwicklung wurde deutlich, wie sensibel das Modell auf unpräzise Kontexte reagiert. So wurde in frühen Tests beispielsweise das Anbieten eines Kaffees durch einen Bankberater fälschlich als sexistische Mikroaggression klassifiziert – ein Effekt, der sich durch gezieltes Tagging und Präzisierung der Wissensbasis vermeiden ließ. Die systematische Verknüpfung von inhaltlicher Kontextualisierung, Beispielstellen im Dialog und explizitem Bias-Tagging erwies sich als entscheidend, um Halluzinationen des LLM zu minimieren und die inhaltliche Qualität des Feedbacks zu sichern. Die Rückmeldung ist in Du-Form gehalten, empathisch formuliert und auf Bestärkung ausgerichtet. Sie soll eine differenzierte Reflexion des Gesprächsverhaltens ermöglichen und zugleich praktische Hinweise zur Weiterentwicklung geben.

## 2.4 Wissensbasis: Systematisierung von Biases

Die Analyse der KITE-II-App basiert auf einer strukturierten Wissensbasis, die rund 20 Biases systematisiert und in vier übergeordnete Kategorien gliedert. Die erste Kategorie umfasst **strukturelle wirtschaftliche Barrieren**, etwa den Gender Pay Gap oder den Risk Aversion Bias, die Gründerinnen in finanziierungsbezogenen Situationen benachteiligen können. Die zweite Kategorie adressiert **gesellschaftliche Erwartungen und soziale Normen**, darunter Phänomene wie Tokenism oder benevolent sexism, die in zwischenmenschlichen Interaktionen subtil, aber wirksam zur Reproduktion von Ungleichheit beitragen. Eine dritte Gruppe bildet die Kategorie **Wahrnehmung und Führungsrollen**, in der unter anderem Erwartungshaltungen bezüglich Familienplanung oder stereotype Vorstellungen über Führungsstile verortet sind. Schließlich identifiziert die vierte Kategorie **psychologische und kommunikative Barrieren**, etwa Mikroaggressionen oder den sogenannten Tightrope Bias, der sich in widersprüchlichen Erwartungshaltungen an das Verhalten von Frauen in Verhandlungssituationen äußert.

Jeder Bias ist in der Wissensbasis prägnant definiert, mit Beispielen aus den Visual Novels illustriert und durch weiterführende wissenschaftliche Quellen kontextualisiert. Die Wissensbasis dient sowohl als Referenz für die KI-gestützte Analyse als auch als Lernangebot für Nutzer\*innen, die ihr Verständnis vertiefen möchten. Durch die gezielte Auswahl, didaktische Aufbereitung und intersektionale Fundierung wird eine nachvollziehbare und anschlussfähige Deutungsgrundlage geschaffen, die differenzierte Reflexion fördert und zugleich die analytische Präzision des KI-Feedbacks absichert.

## 3 Spannungsfelder von Offenheit im Kontext sensibler Erfahrungen

Offenheit wird in der Informatik häufig als inhärent positiv konnotiert, meist verbunden mit Idealen wie Transparenz, Reproduzierbarkeit und Partizipation. In der Gestaltung von Anwendungen, die auf sensiblen Erfahrungen marginalisierter Gruppen beruhen, ist Offenheit jedoch nicht nur Chance, sondern auch Risiko. Die KITE-II-App zeigt exemplarisch, dass Offenheit entlang mehrerer Dimensionen kritisch reflektiert und kontextabhängig begrenzt werden muss. Im Folgenden systematisieren wir zentrale Spannungsfelder, die sich im Entwicklungsprozess ergeben haben.

**Sichtbarkeit vs. Schutzbedarf:** Die Sichtbarmachung von Diskriminierung gilt als Voraussetzung für gesellschaftliche Veränderung. Gleichzeitig kann sie für Betroffene mit Re-Traumatisierung oder ungewollter Wiedererkennbarkeit einhergehen, insbesondere, wenn Narrative aus ihrem situativen Kontext herausgelöst und in öffentlich zugängliche Formate überführt werden. In KITE II wurde daher bewusst auf die Veröffentlichung konkreter Gesprächsverläufe verzichtet. Stattdessen erfolgt eine Verallgemeinerung typischer Gesprächsmuster, verbunden mit kontrollierter Narration durch vordefinierte Pfade.

**Partizipation vs. Deutungshoheit:** Partizipative Designprozesse gelten als Mittel zur Demokratisierung technologischer Entwicklung. Gleichzeitig werfen sie Fragen nach Repräsentation und Deutungshoheit auf [Dr20]: Wer wird beteiligt, wer nicht? Wer entscheidet, welche Perspektiven als authentisch gelten? Die in KITE entstandenen Inhalte wurden deshalb nicht als repräsentativ verallgemeinert, sondern als Ausgangspunkt für multiperspektivische Reflexion gestaltet.

**Transparenz vs. narrative Integrität:** In KI-Systemen wird Transparenz häufig durch Offenlegung von Trainingsdaten, Modellparametern oder Entscheidungslogiken operationalisiert. In der KITE-II-App wäre eine solche Offenheit problematisch: Vollständige Transparenz würde nicht nur potenzielle Wiedererkennbarkeit erzeugen, sondern auch die Schutzzräume untergraben. Stattdessen wurde auf strukturelle Transparenz gesetzt, etwa durch nachvollziehbares Prompting und eine offen zugängliche Wissensbasis.

**Offenheit in der Nutzung vs. Begrenzung in der Gestaltung:** Die Anwendung ist bewusst niedrigschwellig gestaltet: Sie ist ohne Anmeldung nutzbar, speichert keine personenbezogenen Daten und steht zum Projektabschluss einer breiten Öffentlichkeit offen. Gleichzeitig ist die Offenheit der Nutzung durch eine begrenzte Dialogstruktur gerahmt: Alle Gesprächsverläufe sind kuratiert, Antwortoptionen vordefiniert und die Interaktion mit generativer KI bewusst begrenzt. Genau diese begrenzte Offenheit ermöglicht einen geschützten Erfahrungsraum.

**Didaktische Offenheit vs. ethische Rahmung:** In KITE II sollen Nutzer\*innen explizit eigene Wege der Auseinandersetzung mit Diskriminierung entwickeln. Diese Rahmung knüpft an Konzepte reflektiven Game Designs an, die gezielte Irritationen zulassen, ohne Vorgaben zu machen [Kh18]. Gleichzeitig gibt die App mit ihrer Bias-Wissensbasis, ihrer Rollenvergabe im KI-Prompt („Du bist eine Geschlechterforscherin“) und ihrer empathischen Tonalität eine normative Rahmung vor. Diese Rahmung versteht sich jedoch nicht als moralisches Korsett, sondern als machtbewusste Strukturierung eines Reflexionsraums.

Die Reflexion dieser Spannungsfelder zeigt: Offenheit in sensiblen digitalen Anwendungen muss relational gedacht werden. Sie ist kein Selbstzweck, sondern eine Gestaltungshaltung, die immer wieder neu austariert werden muss – zwischen Sichtbarkeit und Schutz, Beteiligung und Begrenzung, Freiheit und Verantwortung.

## 4 Bias-aware by Design: Reflexion der Gestaltungsentscheidungen

Die Gestaltung folgt dem Prinzip „Bias-aware by Design“: Diskriminierung wird nicht retrospektiv adressiert, sondern ist von Anfang an zentrales Designkriterium. Technische, soziale und ethische Dimensionen greifen dabei ineinander. Im Folgenden skizzieren wir zentrale Elemente dieses Gestaltungsansatzes.

**Partizipative Validierung:** Die Dialogszenarien entstanden im Austausch mit Gründerinnen, Gleichstellungsexpert\*innen und Fachpersonen. Ziel war nicht Repräsentation, sondern die

Entwicklung multiperspektivischer Muster und eines dynamischen Problemverständnisses [Dr20]. Die geteilten Erfahrungen wurden als Ausgangspunkt für generalisierbare Muster genutzt, nicht als repräsentative Einzelfälle.

**Kuratiertes Interaktionsdesign:** Statt offener KI-Systeme wie Chatbots nutzt KITE II für die Dialoge strukturierte Entscheidungsgraphen. Die Interaktion bleibt kontrolliert, um einen Raum zu sichern, der Schutz vor retraumatisierenden Inhalten und unvorhersehbaren Reaktionen bietet. Die generative KI wird erst in der Feedbackphase eingesetzt – und dort unter enger semantischer Steuerung.

**Feedback als empowernde Intervention:** Die KI-Rückmeldung adressiert Spieler\*innen direkt, anerkennt Leistungen und benennt blinde Flecken. Sie zielt nicht auf Bewertung, sondern auf Ermächtigung, was auch in Serious Games als ethisch reflektierende, transformative Praxis beschrieben wird [Mc11]. Konfrontative Reaktionen werden nicht eingefordert und Spielräume für strategisches Verhalten bleiben gewahrt.

**Wissensbasiertes Prompting:** Die Analyse erfolgt strukturiert, mit expliziter Rollenzuweisung und Bias-Wissensbasis. So werden Halluzinationen vermieden und die Qualität der Analyse stabilisiert, insbesondere bei subtilen Gesprächsdynamiken.

**Tagging als epistemisches Interface:** Bias-Zuweisungen erfolgen manuell auf Basis wissenschaftlicher Forschung und nicht durch das Modell. So bleibt die epistemische Kontrolle über sensible Bewertungen beim menschlichen Designteam.

**Fehler als Ressource:** Fehlerhafte Interpretationen des LLM (z. B. im Kaffee-Beispiel) wurden nicht nur korrigiert, sondern als Reflexionsanlass in den Designprozess integriert und dienen so als Anlass zur kontinuierlichen Auseinandersetzung mit den Grenzen automatisierter Analyse.

Die gestalterische Leitlinie „Bias-aware by Design“ zeigt sich in KITE II nicht in einem einzelnen Feature, sondern im Zusammenspiel architektonischer, inhaltlicher und partizipativer Elemente. Sie zielt auf eine reflexive, machtbewusste Gestaltung von KI-Systemen, und zwar nicht durch Neutralität, sondern durch explizite Positionierung.

## 5 Diskussion: Kriterien verantwortungsvoller Offenheit

Offenheit gilt in der HCI-Community als Voraussetzung für Transparenz, Teilhabe und ethische Rechenschaftspflicht [BWD21]. Doch in Anwendungen, die auf sensiblen Erfahrungen marginalisierter Gruppen basieren, kann sie ambivalent oder sogar kontraintuitiv wirken. Die Fallstudie KITE II macht deutlich, dass Offenheit stets kontextabhängig und relational zu denken ist.

**1. Offenheit braucht Schutzräume.** Freie Zugänglichkeit steht oft im Widerspruch zum Schutzbedürfnis vulnerabler Gruppen. Dabei ist eine gewisse narrative Distanz zentral, um

sichere Reflexionsräume zu ermöglichen [Ia22]. In KITE II wurde auf die Veröffentlichung konkreter Gesprächsverläufe verzichtet. Stattdessen werden typische Muster generalisiert und durch vordefinierte Pfade kontrolliert dargestellt.

**2. Offenheit ist keine Repräsentation.** Partizipation ersetzt das Repräsentiert-Sein nicht [Ma25]. Auch ein partizipativ entwickeltes System kann nur Ausschnitte zeigen. Offenheit sollte als Einladung zur Kontextualisierung verstanden werden – nicht als Behauptung universeller Gültigkeit.

**3. Offenheit muss gerahmt werden.** Offenheit erfordert bewusste Steuerung: Wer spricht mit wem, aus welcher Position, in welchem Modus? KITE II nutzt explizite Rollenverteilungen, strukturierte Prompts und eine dokumentierte Wissensbasis, und zwar nicht als Einschränkung, sondern als Voraussetzung für Schutz und Nachvollziehbarkeit. Die Kombination aus Prompt-Engineering und expliziter Rollenzuweisung zielt auf eine epistemisch kontrollierte Offenheit [Ga23].

**4. Offenheit darf nicht machtblind sein.** Maximale Transparenz kann bestehende Machtasymmetrien verstärken. Offenheit ohne machtkritische Reflexion kann Diskriminierung reproduzieren [Bl20] – etwa durch Exponierung verletzlicher Erfahrungen. KITE II positioniert sich explizit parteilich: mit Fokus auf Empowerment, intersektionaler Analyse und begrenzter Offenheit.

**5. Offenheit ist ethische Praxis – keine Voreinstellung.** Verantwortungsvolle Offenheit ergibt sich nicht aus Lizenzmodellen oder Open-Source-Strukturen. Sie erfordert kontinuierliches Aushandeln, etwa zwischen Sichtbarkeit und Schutz oder zwischen Gestaltungsfreiheit und Verantwortung. In KITE II zeigt sich dies etwa im iterativen Prompt-Design und in der Reflexion über technologische Grenzziehung.

Offenheit in sensiblen Technologien sollte nicht als technisches Ziel, sondern als ethisch gerahmte Haltung begriffen werden. Für die Gestaltung von KI-Anwendungen heißt das, Offenheit nicht nur technisch oder politisch zu denken, sondern als eine soziale Entscheidungspraxis, die situativ, relational und ethisch gerahmt ist. Zukünftige Forschung sollte ausloten, wie solche Prinzipien auf andere Domänen übertragbar sind und welche Formen kontextsensibler Offenheit technologisch, didaktisch und sozial tragfähig sind. Die KITE-II-App versteht sich in diesem Sinne nicht als abgeschlossenes Produkt, sondern als explorativer Beitrag zu einer kritischen, sozial eingebetteten Informatik.

## 6 Förderung und Danksagung

Diese Arbeit wurde durch das Bundesministerium für Bildung, Familie, Senioren, Frauen und Jugend (BMBFSFJ) im Rahmen des Projekts „KI-Thinktank Female Entrepreneurship II – KITE II“ unter dem Förderkennzeichen 3923406K04 innerhalb der Förderlinie „KI für Gemeinwohl“ gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei der Autorin.

KITE II entsteht in einem interdisziplinären Team. Die Autorin dankt Rebecca Biebl, Florian Diller, Claudia Herling, Julian Leidel, Tim Reichert und Yvonne Tang, mit denen sie an der Entwicklung von KITE II arbeitet.

## Literaturverzeichnis

- [Bl20] Blodgett, S. L. et al.: Language (technology) is power: A critical survey of "bias in NLP. arXiv preprint, 2020, eprint: 2005.14050, <https://doi.org/10.48550/arXiv.2005.14050>.
- [BWD21] Ballou, N.; Warriar, V. R.; Deterding, S.: Are you open? A content analysis of transparency and openness guidelines in HCI journals. Proceedings of the 2021 CHI Conference on human factors in computing systems, S. 1–10, 2021.
- [CCM21] Camingue, J.; Carstensdottir, E.; Melcer, E. F.: What is a visual novel? In: Proceedings of the ACM on Human-Computer Interaction. Bd. 5 CHI PLAY. CHI PLAY, S. 1–18, 2021.
- [De11] Deterding, S. et al.: From game design elements to gamefulness: defining "gamification". Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments, S. 9–15, 2011.
- [Dr20] Draude, C. et al.: Situated algorithms: a sociotechnical systemic approach to bias. Online Information Review 44 (2), S. 325–342, 2020.
- [Ga23] Gallegos, I. et al.: Zero-shot self-debiasing in large language models. arXiv preprint, 2023, eprint: 2302.04023, <https://doi.org/10.48550/arXiv.2302.04023>.
- [Ia22] Iacovides, I. et al.: Close but not too close: Distance and relevance in designing games for reflection. Proceedings of the ACM on Human-Computer Interaction 6, 2022.
- [Kh18] Khaled, R.: Questions over answers: Reflective game design. In: Playful Disruption of Digital Media. Springer, S. 3–27, 2018.
- [Ko25] Kondra, S. et al.: AI and Diversity, Equity, and Inclusion (DEI): Examining the Potential for AI to Mitigate Bias and Promote Inclusive Communication. Journal of Artificial intelligence and Machine Learning 3 (1), S. 1–8, 2025.
- [Ma25] Marsden, N. et al.: Navigating Bias: Using LLMs to Analyze Discrimination in Entrepreneurial Game Dialogues. HCI International, 2025.
- [Mc11] McGonigal, J.: Reality is Broken: Why Games Make Us Better and How They Can Change the World. Penguin Press, New York, 2011.
- [Re24] Reichert, T. et al.: Empowering Female Founders with AI and Play: Integration of a Large Language Model into a Serious Game with Player-Generated Content. HCI International LNCS 14731, Part II, S. 69–83, 2024.
- [SB22] Schirmacher, A.; Bey, K. v. d.: KITE-Thinktank Female Entrepreneurship – Diskriminierungs-Bekämpfung für Gründerinnen durch gezielten Kompetenzaufbau beim Erkennen und Bewältigen von Diskriminierungsmustern, <https://www.kite-bga.de/experten/>, Interne Projektreport, 2022.
- [SZ24] Schlund, R.; Zitek, E. M.: Algorithmic versus human surveillance leads to lower perceptions of autonomy and increased resistance. Communications Psychology 2 (53), S. 1–9, 2024.
- [TS21] Tonoyan, V.; Strohmeyer, R.: Gender role (in-) congruity and resource-provider gender biases: a conceptual model. International Journal of Gender and Entrepreneurship 13 (3), S. 225–242, 2021.