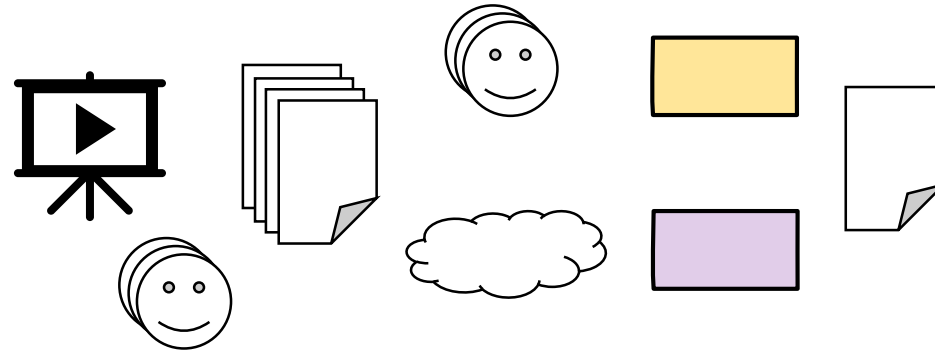
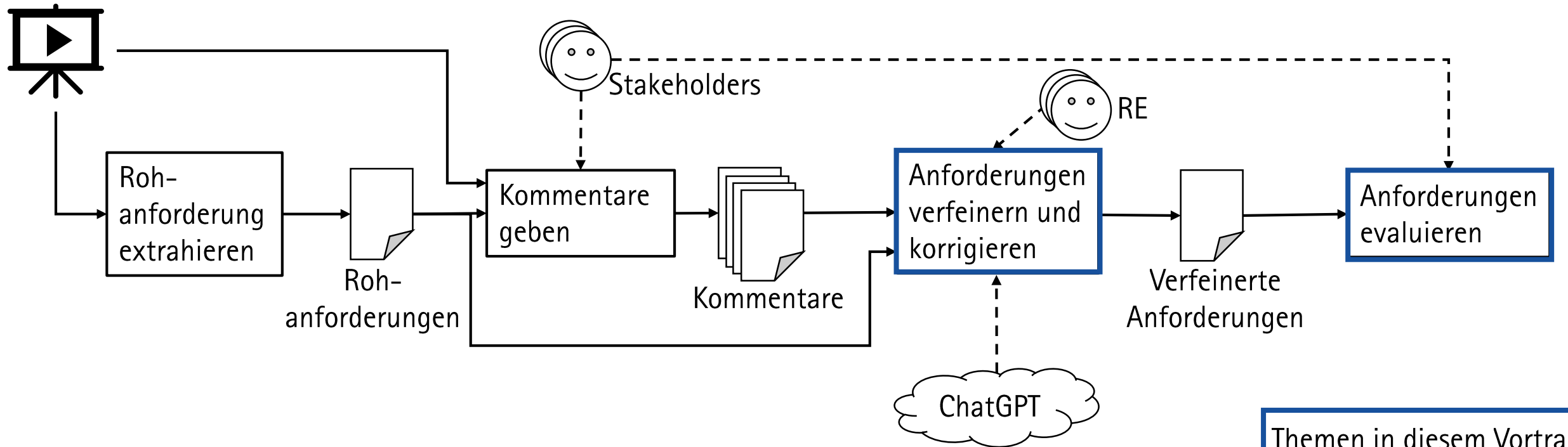


Verfeinerung der Anforderungen mit ChatGPT auf Basis von Stakeholder-Feedback



Jianwei Shi und Kurt Schneider
Jianwei.Shi@inf.uni-hannover.de

- Anforderungsingenieure (RE) ...
 - extrahieren Rohanforderungen von einem vorhandenen Visionsvideo
 - zeigen Stakeholders Rohanforderungen und das Visionsvideo
 - bitten um Kommentaren zu den Rohanforderungen
 - verfeinern und korrigieren die Rohanforderungen



Themen in diesem Vortrag

Einführung

- Ziel: Verfeinerung der Gherkin-Anforderung nach den Stakeholder-Kommentaren
- Warum Gherkin? Wegen leichter Übertragbarkeit in den Test-Code
- Formulieren in Gherkin-Anforderungen
 - **Angenommen** der Nutzer befindet sich auf der Anmeldeseite
 - **Wenn** der Nutzer auf den "Create" Link klickt
 - **Dann** sollte der Nutzer zum Registrierungsformular weitergeleitet werden
 - **Wenn** der Nutzer seine E-Mail und ein Passwort eingibt
 - **Und** das Passwort wiederholt
 - **Und** auf "Create" klickt
 - **Dann** sollte der Nutzer eine Bestätigungs-E-Mail erhalten
- Problem: Verfeinerung zeitaufwändig und anstrengend
- Lösung: Verfeinerung mit Hilfe von GPT

- Ein Visionsvideo wiederverwenden
 - über eine Webseite, wo man Lebenslauf editieren kann
- Sechs Stück Rohanforderungen
- Nachrichten an ChatGPT
 1. Struktur der ersten Nachricht:
 - a) Software-Beschreibung
 - b) Gherkin-Rohanforderungen
 2. Struktur der zweiten Nachricht:
 - a) Rollenzuteilung
 - b) Kommentar zu den Rohanforderungen
 - c) Aufgabestellung
- Zero-Shot Learning (Kein Beispiel einer verfeinerten Anforderung gegeben)

1a

Auf der Webseite CVParser kann man sein LinkedIn-Profil in einen Lebenslauf umwandeln. Diese Webseite wird nun beschrieben.

Beschreibung:

Zum Anfang ist man nicht angemeldet und befindet sich auf der Seite zum Anmelden.

Hier kann man auf einen "create link" klicken, wodurch man zu einem Registrierungsformular weitergeleitet wird. ...

2a

Du bist ein Requirements Engineer und

2c

sollst die Testspezifikation in Gherkin verbessern.

2b

Zu Szenario 1 wurde folgender Kommentar geschrieben:
"Für mich wäre wünschenswert, dass noch die Möglichkeit bestehe, sich mit einem Google Konto oder Facebook Konto anzumelden. Des Weiteren schlage ich, den Code per SMS zu erhalten."

2c

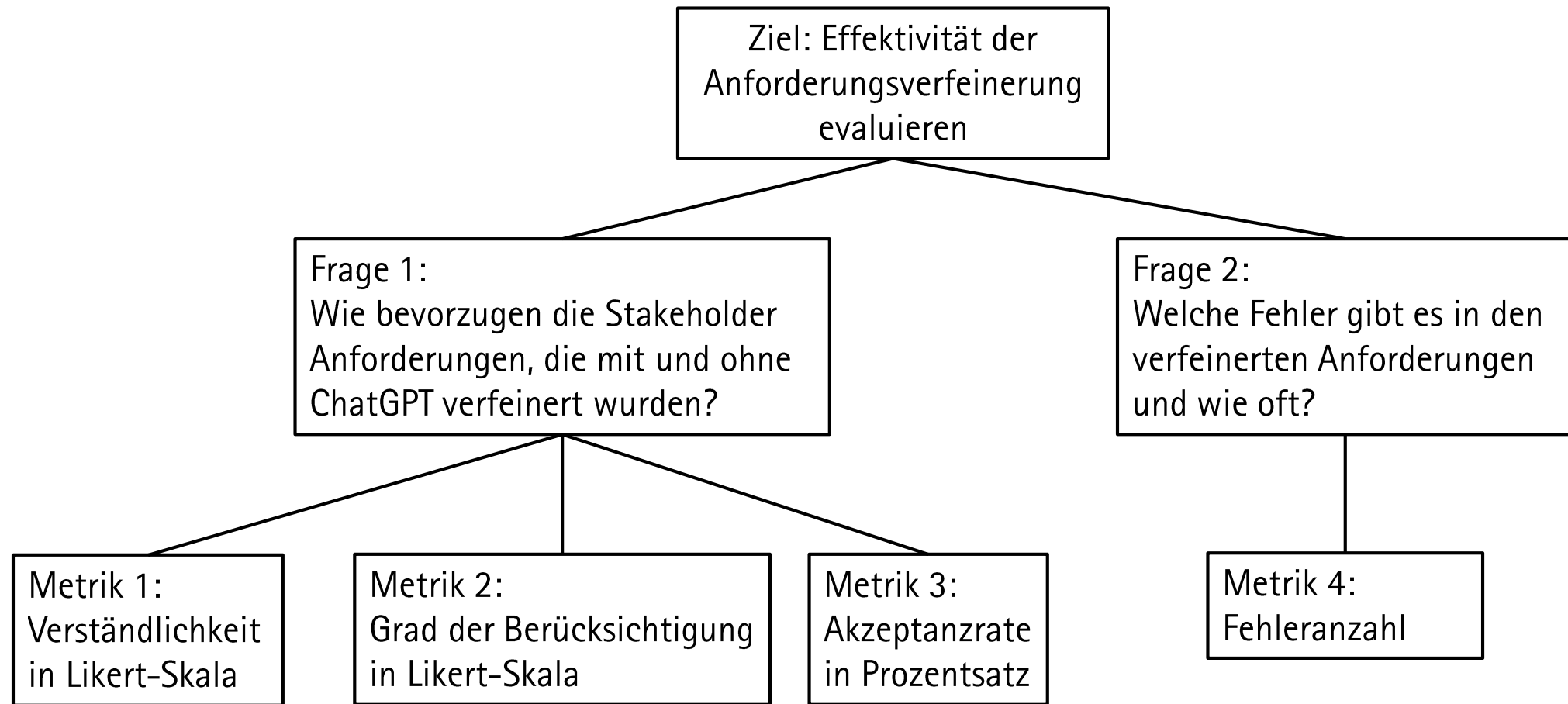
Verbessere die Testspezifikation entsprechend.

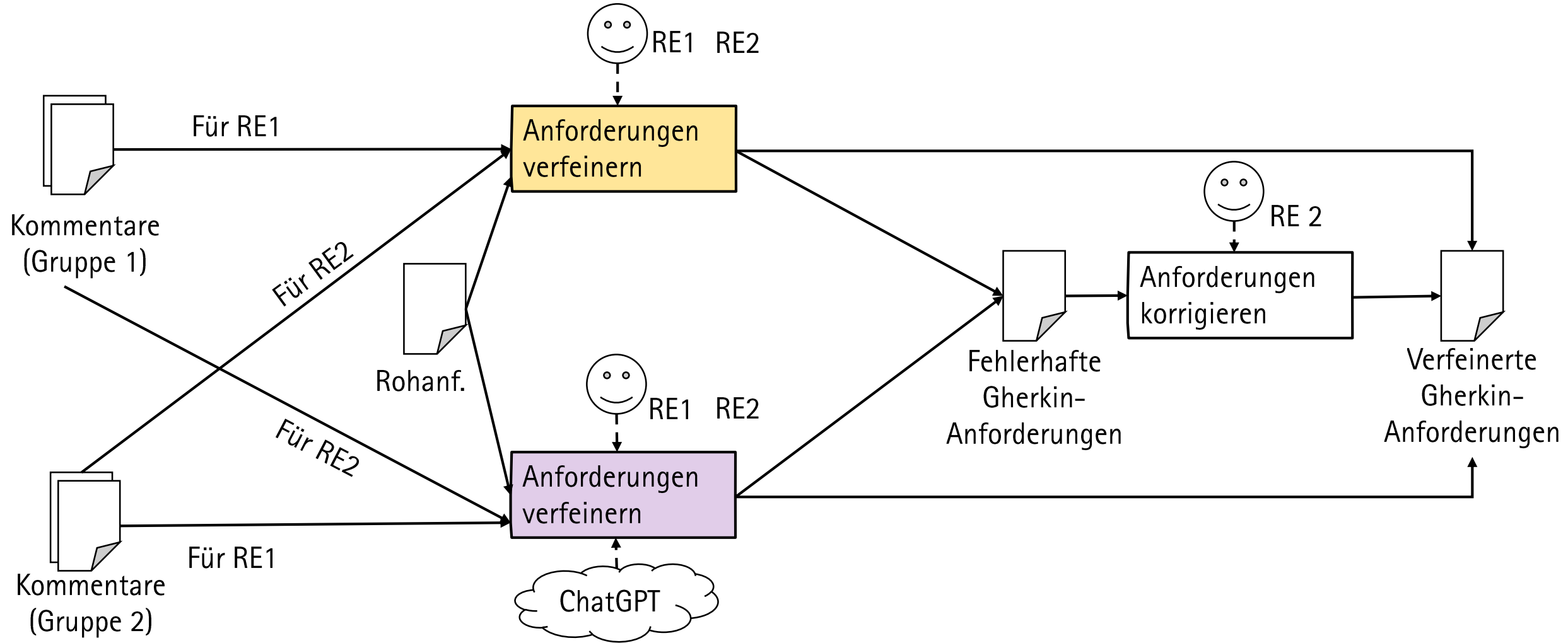
Anforderungen mit Hilfe von ChatGPT verfeinern

1. Bis zu fünf Verfeinerungen anfordern (in Anlehnung zu Chen et al. [1])
2. Nicht mehr anfordern, wenn RE eine gute Antwort schon bekommt
3. Die beste Antwort auswählen
4. Diese Antwort korrigieren

[1] B. Chen et al., 'On the Use of GPT-4 for Creating Goal Models: An Exploratory Study', in *2023 IEEE 31st International Requirements Engineering Conference Workshops (REW)*, Hannover, Germany: IEEE, Sep. 2023, pp. 262–271. doi: [10.1109/REW57809.2023.00052](https://doi.org/10.1109/REW57809.2023.00052).

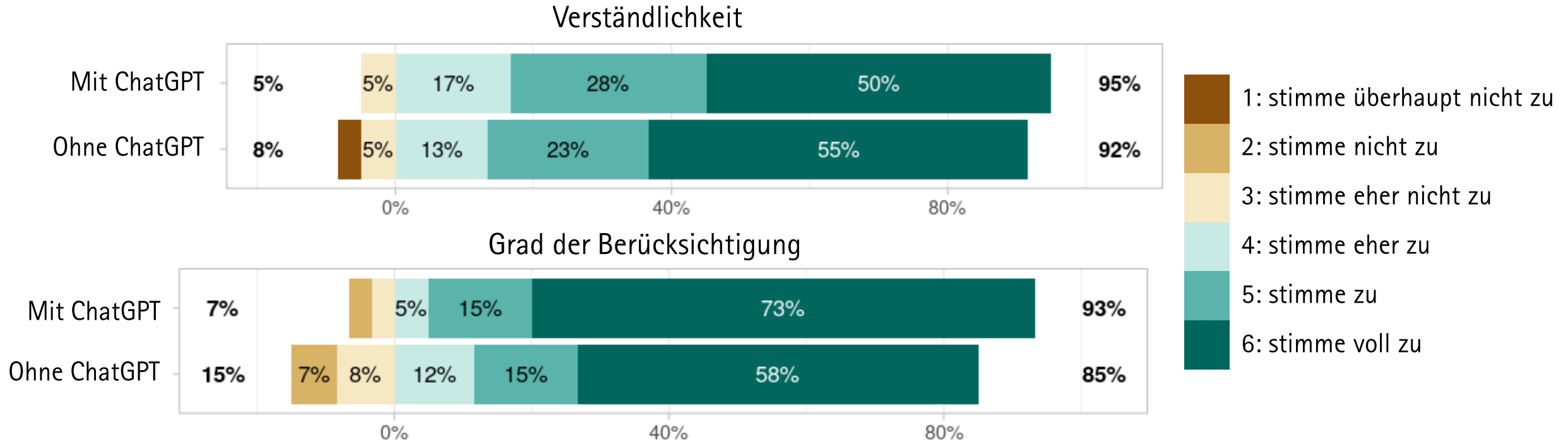
Studienentwurf: GQM Paradigma





- 60 Kommentare gesammelt von 12 Probanden
- 60 verfeinerte Anforderungen ohne ChatGPT
- 60 verfeinerte Anforderungen mit ChatGPT
- Für jede verfeinerte Anforderung:
Jeder Proband durfte seine Präferenz
auf einer Likert-Skala von 1 bis 6 (siehe rechts)
angeben:
 - Die Verbesserung ist gut formuliert und leicht zu verstehen.
 - Die Verbesserung hat meinen Kommentar gut berücksichtigt.
- Dann Akzeptanz für die verfeinerten Anforderungen angeben
(akzeptieren/ablehnen)

1: stimme überhaupt nicht zu
2: stimme nicht zu
3: stimme eher nicht zu
4: stimme eher zu
5: stimme zu
6: stimme voll zu

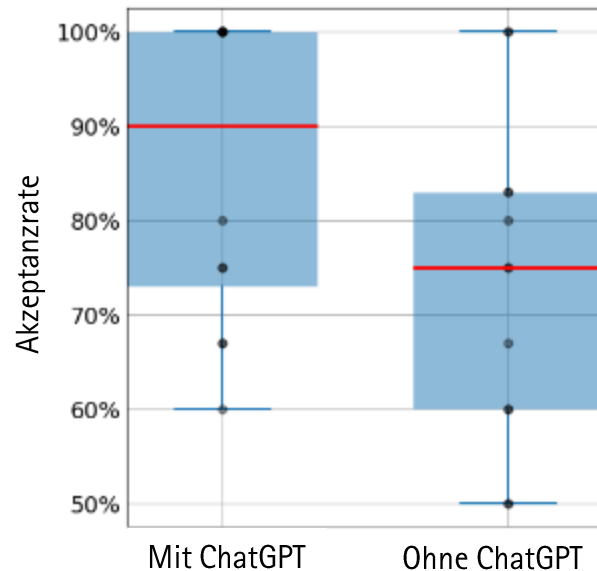


Null-hypothesen: Es gibt keinen Unterschied in ...	Verständlichkeit	Grad der Berücksichtigung
Effektgröße (Rank-biserial coefficient)	0,030 (weniger als small effect)	0,172 (small effect)
p (Mann Whitney U Test, $\alpha=0,05$)	0,754 (nicht signifikant)	0,054 (nicht signifikant)
Power ($1-\beta$ err prob)	0,054	0,182

Achtung: Untere Tabelle zeigt nur einen Datenausschnitt.

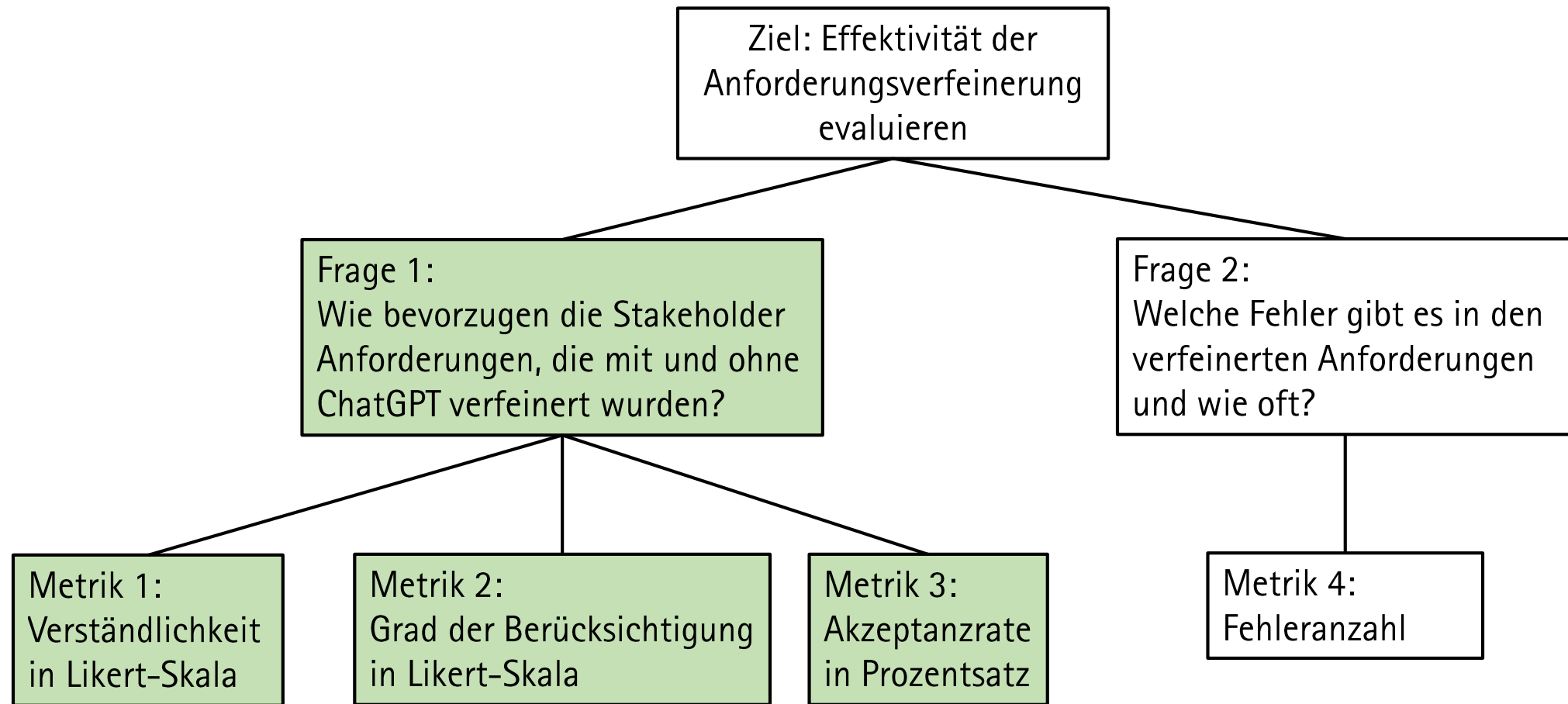
	Ohne ChatGPT verfeinert							Mit ChatGPT verfeinert						
Akzeptanz zu Szenario ...	1	2	3	4	5	6	Akzeptanz-rate	1	2	3	4	5	6	Akzeptanz-rate
Proband 104	Ja	Ja		Ja	Ja	Ja	100%	Ja	Ja		Ja	Nein	Nein	60%
Proband 102	Nein	Ja	Ja		Ja		75%	Ja	Nein	Ja		Ja		75%

Leere Zellen: Probanden haben keine Kommentare zu entsprechenden Rohanforderungen geschrieben.



Null-Hypothese: es gibt keinen Unterschied in Akzeptanzrate.	Ohne ChatGPT	Mit ChatGPT
Durchschnitt	74%	85%
Standardabweichung	17%	16%
Effektgröße d	0,705 (medium effect)	
p (Mann Whitney U Test, $\alpha=0,05$)	0,134 (nicht signifikant)	
Power (1- β err prob)	0,363	

Studienentwurf: GQM Paradigma



Quelle	Verfeinerte Anforderungen vor Korrektur (Kommentar in eckigen Klammern)	Fehlertyp
3.5	Dann sollte der Nutzer die Möglichkeit haben, neue Kategorien anzulegen oder bestehende zu bearbeiten Und nach dem Anpassen auf "Confirm" klicken können Dann sollten die Änderungen, einschließlich neuer Kategorien, erfolgreich gespeichert werden	Formulierungsfehler Syntaxfehler, Formulierungsfehler
3.5	Dann sollte eine Fehlermeldung angezeigt werden und die Eingabeversuche um eins erhöht werden.	Fehlplatzierung
RE	Szenario 6: Anzeigen von Tutorials [Angenommen-Satz fehlt] Wenn [wer?] auf die Seite "About" klickt	Syntaxfehler Grammatikfehler
4o	Szenario 6: Anzeigen von Tutorials Angenommen der Nutzer ist auf der Seite "About" Wenn die Seite geladen wird Dann sollte eine Headline "Über CVParse" angezeigt werden	Formulierungsfehler

Im Nov. 2023 durchgeführt	Syntaxfehler	Fehlplatzierung des Schlüsselworts "und"	Formulierungsfehler	Fehler in der deutschen Grammatik
ChatGPT-3.5	16	12	13	0
RE	6	0	1	4

Im Aug. 2024 durchgeführt, und nur bei den Fällen, wo ChatGPT-3.5 Fehler gemacht hatte	Syntaxfehler	Fehlplatzierung des Schlüsselworts "und"	Formulierungsfehler	Fehler in der deutschen Grammatik
ChatGPT-4o	6	1	3	0

- GPT-3.5 machte mehr Fehler als RE
- GPT-4o beseitigte Fehler von 3.5, aber machte noch Fehler
- Diese Fehler kann ein RE leicht beheben

- Verfeinerungen von GPT enthalten Fehler
- RE müssen auf jeden Fall kontrollieren und ggf. korrigieren
- GPT schrieb auch gute Antworten
- RE lernen von den Antworten (z. B. detaillierte vollständige Anforderungen)
- Effizienz könnte ein Vorteil von GPT sein
- In der Studie: RE wählte die beste Verfeinerung von GPT aus
- In der realen Welt könnte es sein: teil der Verfeinerung kommt von RE, teil von GPT
- Anwendbar für Projekte, wo Projektmitwirkende
 - Gherkin-Anforderungen oder strukturierte Anforderung verwalten
 - Stakeholder-Feedback in natürlicher Sprache sammeln können

Fazit

- Anwendung von GPT für Anforderungsverfeinerung
- Zwei Versionen verglichen
 - Durch GPT verfeinert → Auswahl durch RE → Korrektur durch RE
 - Durch RE verfeinert → Korrektur durch RE
- Präferenzunterschiede nicht signifikant
- Empfehlung:
 - GPT nutzen und davon lernen
 - GPT-Antworten kritisch betrachten