ENKIDU

MIOVSKI GROUP

undefined

Glossar

Qualitativer Parameter Ein urteilsbasierter Wert für einen Schmerzpunkt — Kritikalität, Priorität, Dringlichkeit, Umfang, Stimmung, Evidenzdichte — den dieser Leitfaden deterministisch und belastbar macht.

Schmerzpunkt Ein Problem, eine Herausforderung, Lücke oder ein unerfülltes Bedürfnis, das ein Befragter ausdrücklich äußert und durch ein wörtliches Zitat belegt.

Rubrik-Anker Ein expliziter Deskriptor für eine bestimmte Stufe eines bestimmten Parameters, an dem der Agent die Transkriptbelege abgleichen muss.

Cohens Kappa (kappa) Eine Statistik, die die Übereinstimmung zwischen zwei Bewertern (hier zwei Agentenläufe oder Agent gegen menschlichen Kodierer) über den Zufall hinaus misst.

Fleiss' Kappa (kappa_F) Eine Verallgemeinerung von Cohens Kappa auf mehr als zwei Bewerter — zur Validierung der Mehrfachdurchlauf-Ensemble-Bewertung.

MCDA Multi-Criteria Decision Analysis — ein gewichtetes additives Modell, das mehrere unabhängig bewertete Parameter zu einem zusammengesetzten Prioritätswert kombiniert.

Bayessche Aktualisierung Progressive Verfeinerung einer Wahrscheinlichkeitsverteilung über mögliche Werte, während jedes neue Interview verarbeitet wird.

Nachweiskette Der strukturierte Datensatz, der jeden vergebenen Wert mit den konkreten Transkriptzitaten verbindet, die ihn rechtfertigen.

Modul (M1-M5) Eine Stufe der Pipeline — Aufnahme, Extraktion, Bewertung, MCDA-Berechnung und Audit-Paket-Erzeugung.

Prompt-Hash Ein SHA-256-Hash des exakt verwendeten System-Prompts, gespeichert zur präzisen Reproduktion eines Laufs.

Zusammenfassung

Die Analyse von Interviewtranskripten zur Extraktion und Priorisierung von Schmerzpunkten erfordert qualitative Werte — Kritikalität, Priorität, Dringlichkeit, Wirkung —, die von Natur aus subjektiv sind. Für einen gesteuerten, auditierbaren Prozess müssen drei Probleme zugleich gelöst werden: Subjektivität, Wiederholbarkeit und Nachvollziehbarkeit. Dieser Leitfaden beantwortet alle drei.

Er stellt vier Methoden zunehmender mathematischer Strenge vor — rubrikverankerte deterministische Bewertung, Mehrfachdurchlauf-Ensemble-Bewertung mit statistischer Konvergenz, gewichtete Multi-Kriterien-Entscheidungsanalyse und bayessche hierarchische Bewertung — jeweils mit Beispiel und Governance-Metrik. Anschließend folgt ein vollständiger Umsetzungsleitfaden für den Agenten: eine Fünf-Modul-Pipeline. Empfohlen wird ein Hybrid: rubrikverankerte Bewertung, die ein MCDA-Komposit speist, validiert durch Ensemble-Zuverlässigkeit.

> Die mathematischen Werkzeuge verwandeln eine typischerweise subjektive Übung in einen disziplinierten, auditierbaren Prozess – jeder Wert nachvollziehbar, reproduzierbar und belastbar.

1. Das statistische Dilemma

Subjektivität
Zwei Analysten, die dasselbe Transkript lesen, können unterschiedliche Kritikalitäten vergeben. Wie wird die Bewertung eines Agenten deterministisch und belastbar?

Wiederholbarkeit
Wenn dasselbe Transkript morgen erneut verarbeitet wird, liefert der Agent dieselben Ergebnisse? Welche mathematischen Garantien?

Nachvollziehbarkeit
Für Governance- und Ethik-Audits muss jeder Wert auf konkrete Belege zurückführbar sein. Wie baut man eine Audit-Kette?

2. Rahmen der qualitativen Parameter

Kritikalität — Skala 1-5
Grad, in dem der Schmerzpunkt die Mandatserfüllung oder die rechtliche Konformität bedroht. Nutzung: Risikoregister, Audit-Nachweis.

Priorität — P1-P4
Empfohlene Reihenfolge der Behebung nach Wirkung vs. Aufwand. Nutzung: Roadmap-Sequenzierung.

Häufigkeit — Anzahl (n) + Prozent
Wie oft der Punkt unabhängig genannt wird. Nutzung: statistisches Gewicht.

Umfang — Eng / Mäßig / Weit / Unternehmen
Zahl der betroffenen Direktionen oder Rollen. Nutzung: Investitionsbegründung.

Dringlichkeit — Skala 1-5
Zeitliche Empfindlichkeit — Frist, rechtlicher Auslöser oder sich verschlechternder Zustand. Nutzung: Sprint-Planung.

Stimmungsintensität — -1,0 bis +1,0
Emotionales Gewicht der verwendeten Sprache. Nutzung: Risiko des Change-Managements.

Evidenzdichte — Spärlich / Mäßig / Reich
Umfang und Spezifität der Belege je Punkt. Nutzung: Konfidenzbewertung.

2.1 Warum diese Parameter für Governance wichtig sind

Jeder Parameter entspricht einer eigenen Governance-Anforderung. Kritikalität erfüllt risikobasierte Audit-Anforderungen. Häufigkeit liefert statistische Grundlage — ein von 8 von 10 genannter Punkt wiegt mehr als einer von 1 von 10. Umfang bestimmt, ob eine Lösung eine unternehmensweite Investition erfordert. Stimmungsintensität ist ein Frühindikator für Akzeptanzwiderstand.

3. Option A — Rubrikverankerte deterministische Bewertung

Konzept. Dieser Ansatz definiert vorab eine detaillierte Rubrik mit expliziten Ankerbeschreibungen für jede Stufe jedes Parameters. Der Agent muss aus diesen Ankern wählen und die konkreten Belege zitieren. Am einfachsten umzusetzen und zu auditieren.

3.1 Die Kritikalitäts-Rubrik

1 — Gering
Nur Unannehmlichkeit. Keine Auswirkung auf Mandat, Konformität oder Dienste. Nachhaltige Umgehungen.

2 — Mäßig
Messbare Ineffizienz. Zeit geht verloren, aber Fristen werden gehalten. Kein Konformitätsrisiko.

3 — Hoch
Wiederkehrende Störung. Fristen gefährdet. Moral betroffen. Datenqualitätsbedenken.

4 — Schwer
Direkte Bedrohung des Mandats. Konformitätsfristen verpasst oder unmittelbar. Datenintegrität beeinträchtigt.

5 — Kritisch
Systemisches Versagen. Öffentliche Sicherheit, rechtliche Pflichten oder Ministerzusagen unmittelbar gefährdet.

3.2 Mathematische Grundlage — Cohens Kappa für Wiederholbarkeit

Zur Validierung wiederholbarer Ergebnisse wird die Inter-Rater-Zuverlässigkeit mit Cohens Kappa gemessen; zwei Agentenläufe gelten als zwei Bewerter:

```
kappa = (P_o - P_e) / (1 - P_e) = (0.85 - 0.3625) / (1 - 0.3625) = 0.764
```

wobei P_o die beobachtete und P_e die zufällig erwartete Übereinstimmung ist. Ein Kappa von 0,764 bedeutet "beträchtliche Übereinstimmung"; für Governance wird kappa >= 0,80 angestrebt.

0,41 - 0,60 Mäßige Übereinstimmung

0,61 - 0,80 Beträchtliche Übereinstimmung

0,81 - 1,00 Nahezu perfekte Übereinstimmung

3.3 Nachvollziehbarkeit — die Nachweiskette

Für jeden Wert erzeugt der Agent einen strukturierten Datensatz: Punkt-ID, Parameter, Wert, passender Anker, Quellzitate {Interview, Sprecher, Absatz, Verbatim}, Konfidenz (0-1), Begründung, Zeitstempel, Modellversion und SHA-256-Hash des Prompts.

4. Option B — Ensemble-Bewertung mit statistischer Konvergenz

Konzept. Statt eines einzigen Durchlaufs wird dasselbe Transkript N-mal (typisch 5 bis 7) mit kontrollierter Temperaturvariation ausgeführt. Der Endwert ergibt sich aus der statistischen

Verteilung, nicht aus einem einzelnen Lauf.

```
Score_pk = median(S_pk)   |   CV_pk = sigma(S_pk) / mittel(S_pk)   |   Konfidenz = 1 - CV_pk
```

Beispiel: für "Veraltetes e-Permitting-Tool", bei 5 Läufen S = {4,4,3,4,4}: Median 4, CV = 0,118, Konfidenz 88,2 %. Konsensschwelle: übersteigt der Interquartilsabstand (IQR) 1,0, wird der Punkt zur menschlichen Prüfung markiert.

```
kappa_F = (Pbar - Pbar_e) / (1 - Pbar_e)   —   Ziel kappa_F >= 0,75
```

Fleiss' Kappa verallgemeinert Cohens Kappa auf N > 2 Läufe und validiert die Mehrbewerter-Übereinstimmung der Ensemble-Methode.

5. Option C — Gewichtete Multi-Kriterien-Entscheidungsanalyse (M

Konzept. Dieser Ansatz berechnet einen zusammengesetzten Prioritätswert aus mehreren unabhängig bewerteten Parametern mittels eines gewichteten additiven Modells. Die Gewichte legen die Beteiligten vor der Analyse fest, sodass die Priorisierung organisatorische Werte widerspiegelt.

```
Prioritaet(p) = sum_{k=1..K} w_k x normwert_k(p)   |   normwert = (roh - min) / (max - min)
```

Empfohlene Gewichtskonfiguration
Kritikalität 0,30 ; Häufigkeit 0,20 ; Umfang 0,20 ; Dringlichkeit 0,15 ; Stimmung 0,10 ; Evidenzdichte 0,05 (Summe = 1,0).

```
Beispiel: Prioritaet = 0.225 + 0.140 + 0.200 + 0.075 + 0.020 + 0.050 = 0.710  -> P2 (Hoch)
```

Klassifizierungsschwellen
0,75-1,00 P1 (Sofort) ; 0,50-0,74 P2 (Hoch) ; 0,25-0,49 P3 (Mittel) ; 0,00-0,24 P4 (Niedrig).

Sensitivitätsanalyse. Eine Monte-Carlo-Analyse variiert die Gewichte um +/-10 % über 1000 Simulationen; ändert sich die Prioritätsklasse, wird der Punkt als gewichtsempfindlich markiert und mit den Beteiligten erörtert.

6. Option D — Bayessche hierarchische Bewertung

Konzept. Die mathematisch strengste Option. Statt eines Punktwerts pflegt der Agent eine Wahrscheinlichkeitsverteilung über mögliche Werte, die mit dem Satz von Bayes aktualisiert wird, während Interviews verarbeitet werden.

```
P(Score=c | Evidenz) = P(Evidenz | Score=c) x P(Score=c) / P(Evidenz)
```

```
Nach Interview n: Pn(c|En) proportional zu P(En|c) x P(n-1)(c|E(n-1))
```

Beispiel — "Veraltetes e-Permitting-Tool": uniformer Prior; nach Interview 1 ist die MAP-Schätzung Wert 4 (Schwer) bei 55 %; Interview 2 konzentriert den Posterior auf 68 %. Jedes Interview verfeinert die Verteilung.

Governance-Vorteile. Jeder Zwischen-Posterior wird gespeichert und bildet eine vollständige Audit-Spur. Überspannt das 90-%-Kredibilitätsintervall mehr als 2 Stufen, wird der Punkt zur menschlichen Klärung markiert.

7. Vergleich der Optionen

Komplexität
A Rubrik: gering. B Ensemble: mittel. C MCDA: mittel-hoch. D Bayessch: hoch.

Strenge / Aufwand
A: mäßig, 1-2 Tage. B: hoch, 3-5 Tage. C: hoch, 1 Woche. D: höchste, 2 Wochen.

Governance-Metrik
A: Cohens kappa >= 0,80. B: Fleiss' kappa >= 0,75. C: Sensitivitätsstabilität >= 90 %. D: 90-%-KI-Breite <= 2 Stufen.

7.1 Empfohlener Ansatz

Für die unmittelbaren Bedürfnisse von ECCC wird ein Hybrid aus A und C empfohlen: die rubrikverankerte Bewertung (A) bewertet jeden Parameter, dann erzeugt MCDA (C) die zusammengesetzten Werte. Das Ensemble (B) kommt als Validierungsschicht hinzu; der bayessche Ansatz (D) eignet sich, wenn der Interviewkorpus wächst.

8. Den Agenten bauen — Umsetzungsleitfaden

Der Agent besteht aus fünf Modulen, die eine Pipeline bilden. Jedes Modul ist ein eigener Claude-API-Aufruf mit spezialisiertem System-Prompt, strukturiertem Ausgabeschema und Audit-Protokollierung.

M1 — Transkript-Aufnahme
Zerlegt Rohtranskripte in strukturierte Segmente (Sprecher, Abschnitt, Absätze).

M2 — Schmerzpunkt-Extraktion
Identifiziert einzelne Schmerzpunkte mit Zitaten.

M3 — Bewertungsmaschine
Wendet die Rubrik an, um jeden Parameter mit Nachweisketten und Konfidenz zu bewerten.

M4 — MCDA-Berechnung
Berechnet die zusammengesetzte Priorität (deterministisches Modul, ohne LLM-Aufruf).

M5 — Audit-Paket-Erzeugung
Stellt die vollständige Audit-Spur für die Governance-Prüfung zusammen (JSON + XLSX).

Hinweis: Die vollständigen System-Prompts (M1-M3) und die Python-Listings (M4, Orchestrator, Ensemble, Kappa) befinden sich in der vollständigen englischen Ausgabe dieses Leitfadens.

9. API-Umsetzung

Voraussetzungen: ein Anthropic-API-Schlüssel, Python 3.10+ und das anthropic-SDK. Der Orchestrator führt M1-M5 bei Temperatur 0,0 für Determinismus aus und erfasst bei jedem Aufruf den Prompt-Hash, den Zeitstempel und die Modellversion. Eine Ensemble-Validierung (Option B) und eine Cohens-Kappa-Berechnung validieren die Schwelle kappa >= 0,80. Die vollständigen Listings stehen in der englischen Ausgabe.

10. Konformität mit dem Governance- und Ethik-Audit

Jede Entscheidung muss über eine lückenlose Kette nachvollziehbar sein: Quelltranskript -> analysierter Absatz (M1) -> Punkt + Zitat (M2) -> Wert + Anker + Nachweiskette (M3) -> Priorität + Zerlegung (M4) -> Audit-Paket mit Hashes und Zeitstempeln (M5).

Reproduzierbarkeit
Temperatur 0,0, SHA-256-Hashing der Prompts, fixierte Modellversion.

Nachvollziehbarkeit und Transparenz
Jeder Wert verweist auf wörtliche Zitate; die versionierten System-Prompts liegen dem Paket bei.

Verzerrungserkennung und menschliche Aufsicht
Das Ensemble erkennt Inkonsistenzen; die konservative Regel (niedrigere Stufe bei Gleichstand) verhindert Inflation; markierte Elemente erfordern protokollierte menschliche Klärung.

Ethik der Interviewanalyse
Anonymisierung der Sprecher (rollenbasierte Kennungen); hierarchische Autorität nicht mit Beleg-Qualität verwechseln; Stimmung für zurückhaltende Sprache kalibrieren.

11. Praktischer Durchgang — CWS-Interviewdaten

Veraltetes e-Permitting-Tool (ColdFusion). Kritikalität 4, Dringlichkeit 4, Umfang 3, Stimmung -0,5, Evidenz 2, Häufigkeit 2/10.

```
Prioritaet = 0.30(0.75)+0.20(0.20)+0.20(0.67)+0.15(0.75)+0.10(0.25)+0.05(0.50) =
0.562 -> P2
```

Verwirrung um die Informationssicherheit. Kritikalität 3, Dringlichkeit 3, Umfang 4 (Unternehmen), Stimmung -0,4, Evidenz 3, Häufigkeit 3/10.

```
Prioritaet = 0.30(0.50)+0.20(0.30)+0.20(1.00)+0.15(0.50)+0.10(0.30)+0.05(1.00) =
0.565 -> P2
```

Übermäßige Prozessbürokratie (40 Schritte, um eine Seite zu veröffentlichen). Kritikalität 3, Dringlichkeit 3, Umfang 3, Stimmung -0,6, Evidenz 3, Häufigkeit 2/10.

```
Prioritaet = 0.30(0.50)+0.20(0.20)+0.20(0.67)+0.15(0.50)+0.10(0.20)+0.05(1.00) =
0.469 -> P3
```

12. Umsetzungs-Roadmap

Phase 1 — Rubrik-Kalibrierung (1 Wo.)
Finalisierte Rubrik mit genehmigten Ankern; Gewichtskonfiguration von den DGs abgezeichnet.

Phase 2 — Entwicklung (2 Wo.)
Module M1-M5 implementiert und getestet; Prompts versioniert.

Phase 3 — Pilot (1 Wo.)
5 Transkripte verarbeiten; Kappa berechnen; gegen menschliche Kodierer validieren.

Phase 4 — Kalibrierung (1 Wo.)
Anker verfeinern, wo kappa < 0,80; Prompts justieren; Pilot wiederholen.

Phase 5 — Produktion (2 Wo.)
Alle Transkripte verarbeiten; vollständige Audit-Pakete erzeugen.

Phase 6-7 — Prüfung und Bericht (2 Wo.)
Pakete dem Ethikgremium vorlegen; Prioritätenbericht für DG/ADM erstellen.

13. Fazit

Der empfohlene Hybrid-Ansatz — rubrikverankerte Bewertung, in MCDA eingespeist — bietet das optimale Gleichgewicht aus Wiederholbarkeit, Nachvollziehbarkeit und Umsetzungszeit. Die vollständige Pipeline stellt sicher, dass jeder Wert und jede Priorität auf konkrete Belege zurückgeführt, mit identischen Eingaben reproduziert und im Audit verteidigt werden können.

Die mathematischen Werkzeuge — Cohens Kappa, gewichtetes additives Modell, bayessche Aktualisierung und Monte-Carlo-Sensitivitätsanalyse — verwandeln eine subjektive Übung in einen disziplinierten, auditierbaren Prozess. Es ist die Überzeugung der analytischen Rahmen von MIOVSKI GROUP: Ein Wert ist nur so viel wert wie die Nachweiskette hinter ihm.

ENDE DES DOKUMENTS