ENKIDU

MIOVSKI GROUP

undefined

Glosario

Parámetro cualitativo Un valor basado en el juicio asignado a un punto de dolor — criticidad, prioridad, urgencia, alcance, sentimiento, densidad de evidencia — que esta guía vuelve determinista y defendible.

Punto de dolor Un problema, desafío, brecha o necesidad no satisfecha expresado explícitamente por un entrevistado y respaldado por cita textual.

Ancla de rúbrica Un descriptor explícito de un nivel concreto de un parámetro concreto, con el que el agente debe cotejar la evidencia.

Kappa de Cohen (kappa) Una estadística que mide el acuerdo entre dos evaluadores (aquí, dos ejecuciones del agente, o el agente frente a un codificador humano) más allá del azar.

Kappa de Fleiss (kappa_F) Una generalización del Kappa de Cohen a más de dos evaluadores — para validar la puntuación de conjunto multipase.

ADMC Análisis de Decisión Multicriterio — un modelo aditivo ponderado que combina varios parámetros puntuados de forma independiente en un único puntaje de prioridad.

Actualización bayesiana Refinamiento progresivo de una distribución de probabilidad sobre los puntajes posibles a medida que se procesa cada nueva entrevista.

Cadena de evidencia El registro estructurado que vincula cada puntaje asignado con las citas precisas de la transcripción que lo justifican.

Módulo (M1-M5) Una etapa del pipeline — ingestión, extracción, puntuación, cálculo ADMC y generación del paquete de auditoría.

Hash de prompt Un hash SHA-256 del prompt de sistema exacto usado, almacenado para reproducir con precisión una ejecución.

Resumen

Analizar transcripciones de entrevistas para extraer y priorizar puntos de dolor exige asignar valores cualitativos — criticidad, prioridad, urgencia, impacto — subjetivos por naturaleza. Para un proceso gobernado y auditable deben resolverse tres problemas a la vez: subjetividad, repetibilidad y trazabilidad. Esta guía responde a los tres.

Presenta cuatro métodos de rigor matemático creciente — puntuación determinista anclada en rúbrica, puntuación de conjunto multipase con convergencia estadística, análisis de decisión multicriterio ponderado y puntuación jerárquica bayesiana — cada uno con su ejemplo y su métrica de gobernanza. Luego ofrece una guía de implementación completa del agente: un pipeline de cinco módulos. El enfoque recomendado es híbrido: puntuación anclada en rúbrica que alimenta un compuesto ADMC, validado por la fiabilidad de conjunto.

> Las herramientas matemáticas transforman un ejercicio habitualmente subjetivo en un proceso analítico disciplinado y auditable — cada puntaje trazable, reproducible y defendible.

1. El dilema estadístico

Subjetividad
Dos analistas que leen la misma transcripción pueden asignar criticidades distintas. ¿Cómo hacer la puntuación del agente determinista y defendible?

Repetibilidad
Si la misma transcripción se procesa mañana, ¿el agente producirá los mismos resultados? ¿Qué garantías matemáticas?

Trazabilidad
Para las auditorías de gobernanza y ética, cada puntaje debe ser trazable a evidencia precisa. ¿Cómo se construye una cadena de auditoría?

2. Marco de parámetros cualitativos

Criticidad — escala 1-5
Grado en que el punto de dolor amenaza la entrega del mandato o el cumplimiento legal. Uso: registro de riesgos, evidencia de auditoría.

Prioridad — P1-P4
Orden de remediación recomendado según impacto vs esfuerzo. Uso: secuenciación de la hoja de ruta.

Frecuencia — recuento (n) + porcentaje
Cuántas veces se cita el punto de forma independiente. Uso: peso estadístico.

Alcance — Estrecho / Moderado / Amplio / Empresa
Número de direcciones o roles afectados. Uso: justificación de inversión.

Urgencia — escala 1-5
Sensibilidad temporal — plazo, disparador legal o condición que se degrada. Uso: planificación de sprints.

Intensidad de sentimiento — -1,0 a +1,0
Peso emocional del lenguaje empleado. Uso: riesgo de gestión del cambio.

Densidad de evidencia — Escasa / Moderada / Rica
Volumen y especificidad de la evidencia por punto. Uso: puntaje de confianza.

2.1 Por qué importan estos parámetros para la gobernanza

Cada parámetro corresponde a un requisito de gobernanza distinto. La criticidad satisface los requisitos de auditoría basados en riesgo. La frecuencia aporta base estadística — un punto citado por 8 de 10 pesa más que uno citado por 1 de 10. El alcance determina si una solución requiere inversión de empresa. La intensidad de sentimiento es un indicador anticipado de resistencia a la adopción.

3. Opción A — Puntuación determinista anclada en rúbrica

Concepto. Predefine una rúbrica detallada con descripciones de ancla explícitas para cada nivel de cada parámetro. El agente debe elegir entre esas anclas y citar la evidencia precisa. Es el enfoque más simple de implementar y auditar.

3.1 La rúbrica de criticidad

1 — Baja
Solo inconveniente. Sin impacto en mandato, cumplimiento o servicios. Soluciones alternativas sostenibles.

2 — Moderada
Ineficiencia medible. Se pierde tiempo, pero se cumplen los plazos. Sin riesgo de cumplimiento.

3 — Alta
Interrupción recurrente. Plazos en riesgo. Moral afectada. Preocupaciones de calidad de datos.

4 — Grave
Amenaza directa al mandato. Plazos de cumplimiento incumplidos o inminentes. Integridad de datos comprometida.

5 — Crítica
Falla sistémica. Seguridad pública, obligaciones legales o compromisos ministeriales en riesgo inmediato.

3.2 Base matemática — Kappa de Cohen para la repetibilidad

Para validar que la rúbrica produce resultados repetibles, la fiabilidad entre evaluadores se mide con el Kappa de Cohen; dos ejecuciones del agente se tratan como dos evaluadores:

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

donde P_o es la proporción de acuerdo observada y P_e el acuerdo esperado por azar. Un kappa de 0,764 indica «acuerdo sustancial»; para gobernanza, apuntar a kappa >= 0,80.

0,41 - 0,60 Acuerdo moderado

0,61 - 0,80 Acuerdo sustancial

0,81 - 1,00 Acuerdo casi perfecto

3.3 Trazabilidad — la cadena de evidencia

Para cada puntaje, el agente produce un registro estructurado: identificador del punto, parámetro, puntaje, ancla coincidente, citas fuente {entrevista, hablante, párrafo, verbatim}, confianza (0-1), razonamiento, marca de tiempo, versión del modelo y hash SHA-256 del prompt.

4. Opción B — Puntuación de conjunto con convergencia estadísti

Concepto. En lugar de una sola pasada, se ejecuta la misma transcripción N veces (típicamente 5 a 7) con variación controlada de temperatura. El puntaje final deriva de la distribución estadística, no

de una ejecución única.

```
Score_pk = mediana(S_pk)   |   CV_pk = sigma(S_pk) / media(S_pk)   |   Confianza = 1 - CV_pk
```

Ejemplo: para «Herramienta e-Permitting obsoleta», en 5 ejecuciones S = {4,4,3,4,4}: mediana 4, CV = 0,118, confianza 88,2 %. Umbral de consenso: si el rango intercuartílico (IQR) > 1,0, el punto se marca para revisión humana.

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

El Kappa de Fleiss generaliza el de Cohen a N > 2 ejecuciones y valida el acuerdo multievaluador del método de conjunto.

5. Opción C — Análisis de decisión multicriterio (ADMC)

Concepto. Calcula un puntaje de prioridad compuesto a partir de varios parámetros puntuados de forma independiente, mediante un modelo aditivo ponderado. Los pesos los fijan las partes interesadas antes del análisis, asegurando que la priorización refleje los valores de la organización.

```
Prioridad(p) = sum_{k=1..K} w_k x puntaje_normalizado_k(p)   |   normalizado = (bruto - min) / (max - min)
```

Configuración de pesos recomendada
Criticidad 0,30 ; Frecuencia 0,20 ; Alcance 0,20 ; Urgencia 0,15 ; Sentimiento 0,10 ; Densidad de evidencia 0,05 (suma = 1,0).

```
Ejemplo: Prioridad = 0.225 + 0.140 + 0.200 + 0.075 + 0.020 + 0.050 = 0.710  ->  P2 (Alta)
```

Umbrales de clasificación
0,75-1,00 P1 (Inmediata) ; 0,50-0,74 P2 (Alta) ; 0,25-0,49 P3 (Media) ; 0,00-0,24 P4 (Baja).

Análisis de sensibilidad. Un análisis de Monte Carlo varía los pesos +/-10 % en 1000 simulaciones; si la clase de prioridad cambia, el punto se marca como sensible a los pesos y se discute con las partes interesadas.

6. Opción D — Puntuación jerárquica bayesiana

Concepto. La opción más rigurosa. En vez de un puntaje puntual, el agente mantiene una distribución de probabilidad sobre los puntajes posibles, actualizada con el teorema de Bayes a medida que se procesan las entrevistas.

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

```
Tras la entrevista n: Pn(c|En) proporcional a P(En|c) x P(n-1)(c|E(n-1))
```

Ejemplo — «Herramienta e-Permitting obsoleta»: prior uniforme; tras la entrevista 1 la estimación MAP es el puntaje 4 (Grave) al 55 %; la entrevista 2 concentra el posterior al 68 %. Cada entrevista refina la distribución.

Ventajas de gobernanza. Cada posterior intermedio se almacena, creando una pista de auditoría completa. Si el intervalo creíble al 90 % abarca más de 2 niveles, el punto se marca para arbitraje humano.

7. Comparación de opciones

Complejidad
A Rúbrica: baja. B Conjunto: media. C ADMC: media-alta. D Bayesiana: alta.

Rigor / esfuerzo
A: moderado, 1-2 días. B: alto, 3-5 días. C: alto, 1 semana. D: máximo, 2 semanas.

Métrica de gobernanza
A: kappa de Cohen >= 0,80. B: kappa de Fleiss >= 0,75. C: estabilidad de sensibilidad >= 90 %. D: ancho de IC al 90 % <= 2 niveles.

7.1 Enfoque recomendado

Para las necesidades inmediatas de ECCC se recomienda un híbrido de las opciones A y C: la puntuación anclada en rúbrica (A) puntúa cada parámetro, y luego el ADMC (C) produce los puntajes compuestos. El conjunto (B) se añade como capa de validación; el bayesiano (D) conviene si el corpus de entrevistas crece.

8. Construir el agente — guía de implementación

El agente se compone de cinco módulos que forman un pipeline. Cada módulo es una llamada distinta a la API de Claude con un prompt de sistema especializado, un esquema de salida estructurado y registro de auditoría.

M1 — Ingestión de la transcripción
Analiza las transcripciones en bruto en segmentos estructurados (hablante, sección, párrafos).

M2 — Extracción de puntos de dolor
Identifica los puntos distintos con citas.

M3 — Motor de puntuación
Aplica la rúbrica para puntuar cada parámetro con cadenas de evidencia y confianza.

M4 — Cálculo ADMC
Calcula la prioridad compuesta (módulo determinista, sin llamada LLM).

M5 — Generación del paquete de auditoría
Compila la pista de auditoría completa para la revisión de gobernanza (JSON + XLSX).

Nota: los prompts de sistema completos (M1-M3) y los listados de código Python (M4, orquestador, conjunto, Kappa) figuran en la edición inglesa completa de esta guía.

9. Implementación de la API

Requisitos: una clave de API de Anthropic, Python 3.10+ y el SDK anthropic. El orquestador ejecuta M1-M5 a temperatura 0,0 para el determinismo, capturando en cada llamada el hash del prompt, la marca de tiempo y la versión del modelo. Una validación de conjunto (opción B) y un cálculo del Kappa de Cohen validan el umbral kappa >= 0,80. Los listados completos están en la edición inglesa.

10. Cumplimiento de la auditoría de gobernanza y ética

Cada decisión debe ser trazable por una cadena ininterrumpida: transcripción fuente -> párrafo analizado (M1) -> punto + cita (M2) -> puntaje + ancla + cadena de evidencia (M3) -> prioridad + descomposición (M4) -> paquete de auditoría con hashes y marcas de tiempo (M5).

Reproducibilidad
temperatura 0,0, hash SHA-256 de los prompts, versión del modelo fijada.

Trazabilidad y transparencia
Cada puntaje remite a citas textuales; los prompts de sistema versionados se incluyen en el paquete.

Detección de sesgos y supervisión humana
El conjunto detecta inconsistencias; la regla conservadora (nivel inferior en empates) evita la inflación; los elementos marcados exigen arbitraje humano registrado.

Ética del análisis de entrevistas
Anonimización de hablantes (identificadores por rol); no confundir autoridad jerárquica con calidad de evidencia; calibrar el sentimiento para lenguaje atenuado.

11. Recorrido práctico — datos de entrevista del SCF

Herramienta e-Permitting obsoleta (ColdFusion). Criticidad 4, Urgencia 4, Alcance 3, Sentimiento -0,5, Evidencia 2, Frecuencia 2/10.

```
Prioridad = 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
```

Confusión sobre la seguridad de la información. Criticidad 3, Urgencia 3, Alcance 4 (Empresa), Sentimiento -0,4, Evidencia 3, Frecuencia 3/10.

```
Prioridad = 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
```

Burocracia de proceso excesiva (40 pasos para publicar una página). Criticidad 3, Urgencia 3, Alcance 3, Sentimiento -0,6, Evidencia 3, Frecuencia 2/10.

```
Prioridad = 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. Hoja de ruta de implementación

Fase 1 — Calibración de la rúbrica (1 sem.)
Rúbrica finalizada con anclas aprobadas; configuración de pesos firmada por los DG.

Fase 2 — Desarrollo (2 sem.)
Módulos M1-M5 implementados y probados; prompts versionados.

Fase 3 — Piloto (1 sem.)
Procesar 5 transcripciones; calcular los kappa; validar contra codificadores humanos.

Fase 4 — Calibración (1 sem.)
Afinar las anclas donde kappa < 0,80; ajustar los prompts; repetir el piloto.

Fase 5 — Producción (2 sem.)
Procesar todas las transcripciones; generar los paquetes de auditoría completos.

Fase 6-7 — Revisión e informe (2 sem.)
Presentar los paquetes al comité de ética; producir el informe de prioridades para DG/ADM.

13. Conclusión

El enfoque híbrido recomendado — puntuación anclada en rúbrica alimentada en el ADMC — ofrece el equilibrio óptimo entre repetibilidad, trazabilidad y tiempo de implementación. El pipeline completo garantiza que cada puntaje y cada prioridad puedan trazarse a evidencia precisa, reproducirse con entradas idénticas y defenderse en auditoría.

Las herramientas matemáticas — Kappa de Cohen, modelo aditivo ponderado, actualización bayesiana y análisis de sensibilidad de Monte Carlo — transforman un ejercicio subjetivo en un proceso disciplinado y auditable. Es la convicción de los marcos analíticos de MIOVSKI GROUP: un puntaje vale solo por la cadena de evidencia que lo respalda.

FIN DEL DOCUMENTO