ENKIDU

MIOVSKI GROUP

undefined

Glossaire

Paramètre qualitatif Une valeur fondée sur le jugement attribuée à un point de friction — criticité, priorité, urgence, portée, sentiment, densité de preuves — que ce guide rend déterministe et défendable.

Point de friction Un problème, défi, lacune ou besoin non satisfait exprimé explicitement par une personne interrogée et appuyé par une citation verbatim.

Ancre de grille Un descripteur explicite d'un niveau précis d'un paramètre précis, auquel l'agent doit associer les preuves de la transcription.

Kappa de Cohen (kappa) Une statistique mesurant l'accord entre deux évaluateurs (ici, deux exécutions de l'agent, ou l'agent contre un codeur humain) au-delà du hasard.

Kappa de Fleiss (kappa_F) Une généralisation du Kappa de Cohen à plus de deux évaluateurs — pour valider la notation d'ensemble multi-passes.

ADMC Analyse décisionnelle multicritère — un modèle additif pondéré combinant plusieurs paramètres notés indépendamment en un score de priorité composite.

Mise à jour bayésienne Affinage progressif d'une distribution de probabilité sur les scores possibles à mesure que chaque nouvelle entrevue est traitée.

Chaîne de preuves L'enregistrement structuré reliant chaque score attribué aux citations précises de la transcription qui le justifient.

Module (M1-M5) Une étape du pipeline — ingestion, extraction, notation, calcul ADMC et génération du dossier d'audit.

Empreinte de prompt Un hachage SHA-256 du prompt système exact utilisé, stocké pour reproduire précisément une exécution.

Résumé

Analyser des transcriptions d'entrevues pour extraire et prioriser les points de friction exige d'attribuer des valeurs qualitatives — criticité, priorité, urgence, impact — par nature subjectives. Pour un processus gouverné et auditable, trois problèmes doivent être résolus à la fois : la subjectivité, la reproductibilité et la traçabilité. Ce guide répond aux trois.

Il présente quatre méthodes de rigueur mathématique croissante — notation déterministe ancrée sur grille, notation d'ensemble multi-passes à convergence statistique, analyse décisionnelle multicritère pondérée et notation hiérarchique bayésienne — chacune avec son exemple et sa métrique de gouvernance. Il fournit ensuite un guide de mise en ?uvre complet de l'agent : un pipeline en cinq modules, de l'ingestion à la génération du dossier d'audit. L'approche recommandée est hybride : notation ancrée sur grille alimentant un composite ADMC, validée par la fiabilité d'ensemble.

> `Les outils mathématiques transforment un exercice habituellement subjectif en un processus analytique discipline et auditable — chaque score traçable, reproductible et défendable.`

1. Le dilemme statistique

Subjectivité
Deux analystes lisant la même transcription peuvent attribuer des scores de criticité différents. Comment rendre la notation d'un agent déterministe et défendable ?

Reproductibilité
Si la même transcription est retraitée demain, l'agent produira-t-il les mêmes résultats ? Quelles garanties mathématiques ?

Traçabilité
Pour les audits de gouvernance et d'éthique, chaque score doit être traçable à des preuves précises. Comment bâtir une chaîne d'audit ?

2. Cadre des paramètres qualitatifs

Criticité — échelle 1-5
Degré auquel le point de friction menace la réalisation du mandat ou la conformité légale. Usage : registre des risques, preuve d'audit.

Priorité — P1-P4
Ordre de remédiation recommandé selon impact vs effort. Usage : séquençage de la feuille de route.

Fréquence — nombre (n) + pourcentage
Combien de fois le point est cité indépendamment. Usage : poids statistique.

Portée — Étroite / Modérée / Large / Entreprise
Nombre de directions ou de rôles touchés. Usage : justification d'investissement.

Urgence — échelle 1-5
Sensibilité temporelle — échéance, déclencheur légal ou condition qui se dégrade. Usage : planification des sprints.

Intensité de sentiment — -1,0 à +1,0
Poids émotionnel du langage employé. Usage : risque de gestion du changement.

Densité de preuves — Faible / Modérée / Riche
Volume et spécificité des preuves par point. Usage : score de confiance.

2.1 Pourquoi ces paramètres comptent pour la gouvernance

Chaque paramètre correspond à une exigence de gouvernance distincte. La criticité satisfait les exigences d'audit fondées sur le risque. La fréquence fournit un ancrage statistique — un point cité par 8 personnes sur 10 pèse plus qu'un cité par 1 sur 10. La portée détermine si une solution exige un investissement d'entreprise. L'intensité de sentiment est un indicateur avancé de résistance à l'adoption.

3. Option A — Notation déterministe ancrée sur grille

Concept. Cette approche prédéfinit une grille détaillée avec des descriptions d'ancrage explicites pour chaque niveau de chaque paramètre. L'agent est contraint de choisir parmi ces ancres et doit citer la preuve précise. C'est l'approche la plus simple à mettre en ?uvre et à auditer.

3.1 La grille de criticité

1 — Faible
Simple inconvénient. Aucun impact sur le mandat, la conformité ou les services. Contournements durables.

2 — Modérée
Inefficacité mesurable. Du temps est perdu, mais les échéances sont tenues. Aucun risque de conformité.

3 — Élevée
Perturbation récurrente. Échéances à risque. Moral affecté. Préoccupations de qualité des données.

4 — Grave
Menace directe au mandat. Échéances de conformité manquées ou imminentes. Intégrité des données compromise.

5 — Critique
Défaillance systémique. Sécurité publique, obligations légales ou engagements ministériels à risque immédiat.

3.2 Base mathématique — Kappa de Cohen pour la reproductibilité

Pour valider que la grille produit des résultats reproductibles, la fiabilité inter-évaluateurs est mesurée par le Kappa de Cohen ; deux exécutions de l'agent sont traitées comme deux évaluateurs :

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

où P_o est la proportion d'accord observée et P_e l'accord attendu par hasard. Un kappa de 0,764 indique un « accord substantiel » ; pour la gouvernance, viser kappa >= 0,80.

0,41 - 0,60	Accord modéré
0,61 - 0,80	Accord substantiel
0,81 - 1,00	Accord presque parfait

3.3 Traçabilité — la chaîne de preuves

Pour chaque score, l'agent produit un enregistrement structuré : identifiant du point, paramètre, score, ancre associée, citations sources {entrevue, locuteur, paragraphe, verbatim}, confiance (0-1), raisonnement, horodatage, version du modèle et empreinte SHA-256 du prompt.

4. Option B — Notation d'ensemble à convergence statistique

Concept. Au lieu d'une seule passe, on exécute la même transcription N fois (typiquement 5 à 7) avec variation contrôlée de température. Le score final découle de la distribution statistique, non d'une exécution unique.

```
Score_pk = mediane(S_pk)   |   CV_pk = sigma(S_pk) / moyenne(S_pk)   |   Confiance = 1 - CV_pk
```

Exemple : pour « Outil e-Permitting obsolète », sur 5 exécutions S = {4,4,3,4,4} : médiane 4, CV = 0,118, confiance 88,2 %. Seuil de consensus : si l'écart interquartile (IQR) > 1,0, le point est signalé pour révision humaine.

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

Le Kappa de Fleiss généralise celui de Cohen à N > 2 exécutions et valide l'accord multi-évaluateurs de la méthode d'ensemble.

5. Option C — Analyse décisionnelle multicritère (ADMC)

Concept. Cette approche calcule un score de priorité composite à partir de plusieurs paramètres notés indépendamment, via un modèle additif pondéré. Les poids sont fixés par les parties prenantes avant l'analyse, assurant que la priorisation reflète les valeurs organisationnelles.

```
Priorite(p) = sum_{k=1..K} w_k x score_normalise_k(p)   |   score_normalise = (brut - min) / (max - min)
```

Configuration de poids recommandée

Criticité 0,30 ; Fréquence 0,20 ; Portée 0,20 ; Urgence 0,15 ; Sentiment 0,10 ; Densité de preuves 0,05 (somme = 1,0).

```
Exemple : Priorite = 0.225 + 0.140 + 0.200 + 0.075 + 0.020 + 0.050 = 0.710  ->  P2 (Élevée)
```

Seuils de classification

0,75-1,00 P1 (Immédiate) ; 0,50-0,74 P2 (Élevée) ; 0,25-0,49 P3 (Moyenne) ; 0,00-0,24 P4 (Faible).

Analyse de sensibilité. Une analyse Monte-Carlo fait varier les poids de +/-10 % sur 1000 simulations ; si la classe de priorité change, le point est signalé comme sensible aux poids et discuté avec les parties prenantes.

6. Option D — Notation hiérarchique bayésienne

Concept. L'option la plus rigoureuse. Au lieu d'un score ponctuel, l'agent maintient une distribution de probabilité sur les scores possibles, mise à jour par le théorème de Bayes à mesure que les entrevues sont traitées.

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

```
Après l'entrevue n : Pn(c|En) proportionnel à P(En|c) x P(n-1)(c|E(n-1))
```

Exemple — « Outil e-Permitting obsolète » : prior uniforme, après l'entrevue 1 l'estimation MAP est le score 4 (Grave) à 55 % ; l'entrevue 2 concentre le posterior à 68 %. Chaque entrevue affine la distribution.

Avantages de gouvernance. Chaque posterior intermédiaire est stocké, créant une piste d'audit complète. Si l'intervalle crédible à 90 % couvre plus de 2 niveaux, le point est signalé pour arbitrage humain.

7. Comparaison des options

Complexité
A Grille : faible. B Ensemble : moyenne. C ADMC : moyenne-élevée. D Bayésienne : élevée.

Rigueur / effort
A : modérée, 1-2 jours. B : élevée, 3-5 jours. C : élevée, 1 semaine. D : maximale, 2 semaines.

Métrique de gouvernance
A : kappa de Cohen >= 0,80. B : kappa de Fleiss >= 0,75. C : stabilité de sensibilité >= 90 %. D : largeur d'IC à 90 % <= 2 niveaux.

7.1 Approche recommandée

Pour les besoins immédiats d'ECCC, un hybride des options A et C est recommandé : la notation ancrée sur grille (A) note chaque paramètre, puis l'ADMC (C) produit les scores composites. L'ensemble (B) s'ajoute comme couche de validation ; le bayésien (D) convient si le corpus d'entrevues croît.

8. Construire l'agent — guide de mise en ?uvre

L'agent se compose de cinq modules formant un pipeline. Chaque module est un appel distinct à l'API Claude avec un prompt système spécialisé, un schéma de sortie structuré et une journalisation d'audit.

M1 — Ingestion de la transcription
Analyse les transcriptions brutes en segments structurés (locuteur, section, paragraphes).

M2 — Extraction des points de friction
Identifie les points distincts avec citations.

M3 — Moteur de notation
Applique la grille pour noter chaque paramètre avec chaînes de preuves et confiance.

M4 — Calcul ADMC
Calcule la priorité composite (module déterministe, sans appel LLM).

M5 — Génération du dossier d'audit
Compile la piste d'audit complète pour la revue de gouvernance (JSON + XLSX).

Nota : les prompts système complets (M1-M3) et les listages de code Python (M4, orchestrateur,

ensemble, Kappa) figurent dans l'édition anglaise complète de ce guide.

9. Mise en ?uvre de l'API

Prérequis : une clé d'API Anthropic, Python 3.10+ et le SDK anthropic. L'orchestrateur exécute M1-M5 à température 0,0 pour le déterminisme, en capturant à chaque appel l'empreinte du prompt, l'horodatage et la version du modèle. Une validation d'ensemble (option B) et un calcul du Kappa de Cohen valident le seuil kappa >= 0,80. Les listages complets sont dans l'édition anglaise.

10. Conformité à l'audit de gouvernance et d'éthique

Chaque décision doit être traçable par une chaîne ininterrompue : transcription source -> paragraphe analysé (M1) -> point + citation (M2) -> score + ancre + chaîne de preuves (M3) -> priorité + décomposition (M4) -> dossier d'audit avec hachages et horodatages (M5).

Reproductibilité
température 0,0, hachage SHA-256 des prompts, version du modèle épinglée.

Traçabilité et transparence
Chaque score renvoie à des citations verbatim ; les prompts système versionnés sont inclus au dossier.

Détection de biais et supervision humaine
L'ensemble détecte les incohérences ; la règle conservatrice (niveau inférieur en cas d'égalité) évite l'inflation ; les éléments signalés exigent un arbitrage humain journalisé.

Éthique de l'analyse d'entrevues
Anonymisation des locuteurs (identifiants par rôle) ; ne pas confondre autorité hiérarchique et qualité de preuve ; calibrer le sentiment pour un langage attenué.

11. Parcours pratique — données d'entrevue du SCF

Outil e-Permitting obsolète (ColdFusion). Criticité 4, Urgence 4, Portée 3, Sentiment -0,5, Preuves 2, Fréquence 2/10.

```
Priorite = 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
```

Confusion sur la sécurité de l'information. Criticité 3, Urgence 3, Portée 4 (Entreprise), Sentiment -0,4, Preuves 3, Fréquence 3/10.

```
Priorite = 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
```

Bureaucratie de processus excessive (40 étapes pour publier une page). Criticité 3, Urgence 3, Portée 3, Sentiment -0,6, Preuves 3, Fréquence 2/10.

```
Priorite = 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. Feuille de route de mise en ?uvre

Phase 1 — Calibration de la grille (1 sem.)
Grille finalisée avec ancres approuvées ; configuration de poids signée par les DG.

Phase 2 — Développement (2 sem.)
Modules M1-M5 implémentés et testés ; prompts versionnés.

Phase 3 — Projet pilote (1 sem.)
Traiter 5 transcriptions ; calculer les kappa ; valider contre des codeurs humains.

Phase 4 — Calibration (1 sem.)
Affiner les ancres là où kappa < 0,80 ; ajuster les prompts ; refaire le pilote.

Phase 5 — Production (2 sem.)
Traiter toutes les transcriptions ; générer les dossiers d'audit complets.

Phase 6-7 — Revue et rapport (2 sem.)
Présenter les dossiers au comité d'éthique ; produire le rapport de priorités pour DG/SMA.

13. Conclusion

L'approche hybride recommandée — notation ancrée sur grille alimentée dans l'ADMC — offre l'équilibre optimal entre reproductibilité, traçabilité et temps de mise en ?uvre. Le pipeline complet garantit que chaque score et chaque priorité peuvent être tracés à des preuves précises, reproduits avec des entrées identiques et défendus en audit.

Les outils mathématiques — Kappa de Cohen, modèle additif pondéré, mise à jour bayésienne et analyse de sensibilité Monte-Carlo — transforment un exercice subjectif en un processus discipline et auditable. C'est la conviction des cadres analytiques de MIOVSKI GROUP : un score ne vaut que la chaîne de preuves derrière lui.

FIN DU DOCUMENT