



## Evaluation Certifiante



# Rapport Stratégique : Exploration des Usages et des Implications Stratégiques

## Bloc optionnel 2 - CC2.3



*Réalisé par :*

**SGHIOURI** Mohammed

**SOBGUI** Ivan Joel

**BOTI** Armel Cyrille

**BEN LOL** Oumar

**DIVENGI KIBANGUDI BUNKEMBO** Nagui

**ELOUMOU MBOUDOU** Pascal Aurele

**Date : 19 Novembre 2025**

## Tables de matières

<b>1. Cartographie des Usages Métier.....</b>	<b>1</b>
1.1. Pour les Agents SAV.....	1
1.2. Pour les Managers.....	2
1.3. Pour la Direction.....	2
1.4. Pour les Autres Services.....	2
<b>2. Impacts Économiques, Organisationnels et Sociaux.....</b>	<b>2</b>
2.1. Impacts Économiques.....	2
2.2. Impacts Organisationnels.....	3
2.3. Impacts Sociaux.....	4
<b>3. Risques Éthiques et Réglementaires.....</b>	<b>4</b>
3.1. Traitement Automatique de Données Sensibles.....	4
3.2. Erreurs de Classification.....	4
3.3. Déshumanisation du Service.....	4
3.4. Stratégie RGPD et Souveraineté.....	5
<b>5. Scénarios d'Évolution.....</b>	<b>5</b>
4.1. Court Terme (1 an).....	5
4.2. Moyen Terme (3 ans).....	5
4.3. Long Terme (5 ans).....	6
4.4. Plan de Continuité d'Activité (PCA) et Fallback.....	6
4.5. Optimisation par Tokens (Smart Batching).....	6
<b>Conclusion.....</b>	<b>7</b>

# 1. Cartographie des Usages Métier

## 1.1. Pour les Agents SAV

**Filtrage Intelligent par Priorité** : Traitement des tweets par ordre de **Priorité (Score 0 à 2)** calculée automatiquement par les modèles LLM. Les agents se concentrent sur les tweets High Priority, les tweets Low Priority sont traités en différé ou routés automatiquement.

**Routage Thématique Automatique** : La colonne main\_topic aiguille automatiquement vers les services compétents (Réseau → technique, Facturation → comptable, etc.). Bénéfices : désengorgement des services, réduction des transferts, spécialisation des agents.

## 1.2. Pour les Managers

**Tableau de Bord Dynamique** : Vision temps réel de l'activité SAV (volume, sentiments, priorités, temps de réponse, taux de résolution). Alertes automatiques en cas de pic d'activité.

**Monitoring de Crise** : Détection précoce des signaux faibles avant l'escalade. Alerte automatique si >30% de priorité haute en 1h. Une hausse brutale du volume associée à un sentiment négatif signale une panne nationale.

## 1.3. Pour la Direction

**Pilotage de la Satisfaction Client** : Net Sentiment Score (NSS) pour l'évolution de la satisfaction globale, tendances à long terme, benchmark concurrentiel.

**Benchmark Social Media** : Comparaison des performances SAV sur Twitter avec d'autres canaux (volume, satisfaction, coût par résolution).

## 1.4. Pour les Autres Services

**Service Communication** : Détection de bad buzz, analyse de sentiment global, identification d'influenceurs.

**Service Produit** : Identification des dysfonctionnements récurrents, extraction des besoins clients, analyse des retours produits.

**Service Juridique** : Détection de réclamations légales, traçabilité pour conformité réglementaire, analyse de risques.

## 2. Impacts Économiques, Organisationnels et Sociaux

### 2.1. Impacts Économiques

**Réduction des Coûts Opérationnels** : Automatisation du tri manuel des tweets. Réduction de 30-40% du temps consacré aux tâches administratives. Réduction des transferts entre services. Traitement prioritaire des cas urgents réduit les risques d'escalade.

**Stratégie IA Multi-Modèles & Comparatif FinOps** (Coût par Token)

Nous adoptons une stratégie de **"Modèle le moins cher suffisant"**. Sachant qu'il est inutile d'utiliser un modèle coûteux pour une tâche simple.

**Tableau Comparatif des Coûts (Ordre de priorité) :**

Priorité	Fournisseur	Modèle	Coût	Cas d'usage
1 (Défaut)	Google	Gemini 1.5 Flash	Très faible / Gratuit	Classification rapide, Sentiment
2	Mistral AI	Mistral Small	Faible	Traitement souverain (Europe)
3	OpenAI	GPT-4o-mini	Moyen	Analyse complexe
4	OpenAI	GPT-4o	Élevé	Cas très ambigus ou VIP

*Note : Les prix sont dynamiques et surveillés en temps réel par notre module d'administration.*

**Infrastructure Hybride** : Actuelle (On-Premise) : ~0 € / mois, souveraineté des données. Cible (Cloud AWS/Azure) : ~500 € / mois, scalabilité automatique.

**ROI** : Réduction de 30-40% du temps de traitement par agent, réduction des erreurs de routage, détection précoce des crises. Période de retour estimée à 12-18 mois.

### 2.2. Impacts Organisationnels

**Réduction du Temps de Réponse** : Les tweets High Priority sont traités en priorité. Élimination des transferts entre services. Réduction estimée de 20-30% du temps de réponse moyen pour les cas prioritaires.

**Transformation des Rôles** : Les agents évoluent vers un rôle de **résolution complexe** (focus sur les problèmes techniques ou relationnels complexes) plutôt que de tri et catégorisation manuels. Nécessité de formation accrue sur les outils techniques.

**Intégration Multi-Canal** : Extension possible à Facebook, Forums, Emails, Chat. Vision unifiée, routage intelligent multi-canal, analyse comparative.

**Préservation de l'Humanité** : L'outil assiste, ne remplace pas l'humain. Toutes les réponses doivent être validées et personnalisées par un agent. L'IA fournit le contexte, l'humain apporte l'empathie.

## 2.3. Impacts Sociaux

**Amélioration des Conditions de Travail** : Réduction de la charge mentale (priorisation automatique évite la sensation de submergement). Focus sur des tâches gratifiantes (résolution de problèmes) plutôt que répétitives. Développement de compétences techniques et relationnelles avancées.

**Enjeux d'Emploi** : Les postes évoluent vers des rôles plus qualifiés plutôt que d'être supprimés. Nécessité de profils spécialisés (analystes data, gestionnaires de l'IA). Accompagnement nécessaire : formation continue, reconversion, dialogue social.

## 3. Risques Éthiques et Réglementaires

### 3.1. Traitement Automatique de Données Sensibles

**Données en Open Data** : Les utilisateurs Twitter ne consentent pas explicitement au traitement automatisé de leurs tweets par des entreprises. Précautions : information claire sur l'utilisation des données, conformité avec les règles de l'API Twitter, anonymisation optionnelle.

**Données Personnelles Identifiables** : Risques d'identification indirecte, utilisation de données de localisation (nécessaire pour le contexte métier mais sensible), croisement avec d'autres données Free. Mesures : collecte uniquement des données strictement nécessaires, limitation de la durée de stockage, contrôle d'accès basé sur les rôles (RBAC).

### 3.2. Erreurs de Classification

**Types d'Erreurs** : Faux positifs/négatifs, erreurs de sentiment (ironie mal interprétée), erreurs de thème (mauvais routage). Impacts : réponse inadaptée, érosion de la confiance, risques juridiques.

**Mesures de Contrôle** : Validation manuelle d'un échantillon, correction des erreurs par les agents pour améliorer les modèles, score de confiance pour chaque classification, mécanisme d'escalade pour les cas ambigus.

### 3.3. Déshumanisation du Service

**Risques** : Réponses générées automatiquement sans validation, réponses techniques sans dimension émotionnelle, perception robotisée du service. **Précautions** : validation et personnalisation humaine de toutes les réponses, formation à l'utilisation de l'IA comme outil d'aide, suivi de la satisfaction client.

### 3.4. Stratégie RGPD et Souveraineté

**Dilemme de l'Anonymisation** : L'anonymisation totale pose un problème majeur : **la perte du contexte géographique**. Pour Free, savoir qu'un utilisateur se plaint de "coupure réseau" est inutile sans savoir qu'il se situe à "Bordeaux".

#### **Solutions Retenues (Accords & Local Hosting):**

- **Option A : Accords "Entreprise" (Zero Data Retention)** : Le fournisseur s'engage contractuellement à ne pas utiliser nos données pour entraîner ses modèles et à ne rien stocker (Stateless). Avantage : pas besoin d'anonymiser, conservation de la précision géographique.
- **Option B : Modèles Locaux (On-Premise)** : Souveraineté totale, conformité RGPD par défaut, données ne quittent jamais l'infrastructure Free. Limite : coûts d'infrastructure plus élevés.
- **Option C : API Publique (Cas dégradé)** : Application d'un script de "PII Scrubbing" avant envoi. Limite : perte de précision géographique.

**RBAC** : Rôles granulaires (Administrateurs, Analystes, Agents) avec permissions différenciées. Seuls les administrateurs peuvent configurer les modèles LLM et visualiser les logs techniques.

## 5. Scénarios d'Évolution

### 4.1. Court Terme (1 an)

**Intégration au Back-Office** : Intégration avec les systèmes CRM existants, partage bidirectionnel des données, accès depuis l'interface de travail habituelle. Bénéfices : vision 360° client, réduction des allers-retours entre systèmes.

**Usage par les Agents** : Déploiement progressif (Phase 1 : formation pilotes, Phase 2 : déploiement à toutes les équipes, Phase 3 : optimisation continue). Métriques de succès : taux d'adoption >80%, réduction du temps de traitement -20%, amélioration de la satisfaction client +15%.

## 4.2. Moyen Terme (3 ans)

**Extension Multi-Canal** : Extension à Facebook, Forums, Emails, Chat. Architecture unifiée (mêmes modèles LLM, tableau de bord multi-canal, routage intelligent). Bénéfices : compréhension complète de l'expérience client multi-canal, allocation optimale des agents.

**Interface Grand Public** : Dashboard client pour visualiser leur propre sentiment, publication anonymisée de statistiques globales, outils self-service. Bénéfices : transparence, autonomie client, positionnement innovant. Précautions : anonymisation stricte, contrôle client, conformité RGPD.

## 4.3. Long Terme (5 ans)

**Agent Conversationnel Hybride** : L'IA gère les demandes simples et routage les cas complexes vers les humains, assistance en temps réel aux agents, apprentissage continu. Bénéfices : automatisation des tâches simples, focus humain sur le complexe, scalabilité.

**Traitement Vocal** : Transcription automatique des appels, analyse en temps réel du sentiment et de la priorité, suggestions de réponses en temps réel. Bénéfices : analyse unifiée de tous les canaux (texte, vocal), amélioration continue, formation des agents.

**Anticipation des Crises** : Prédiction proactive des problèmes avant qu'ils n'émergent. Technologies : modèles de machine learning pour identifier les signaux faibles, détection de patterns annonciateurs. Exemples : identification de pannes réseau avant qu'elles ne deviennent critiques, anticipation des baisses de satisfaction, ajustement des ressources avant les pics d'activité.

## 4.4. Plan de Continuité d'Activité (PCA) et Fallback

Pour garantir une disponibilité de 99.9%, nous avons prévu une "cascade de repli" en cas de panne des API externes :

1. **Niveau 1 (Nominal)** : Appel aux API Externes (Gemini/Mistral/OpenAI) avec Load Balancing.
2. **Niveau 2 (Panne API)** : Bascule automatique sur **Serveurs GPU Internes**.
  - Action : Démarrage de conteneurs Docker avec **Mistral Open Source** (via Ollama/vLLM).
  - Impact : Surcharge possible des serveurs internes, mais service maintenu.
3. **Niveau 3 (Panne Puissance GPU)** : Mode dégradé "Low Compute".
  - Action : Abandon des LLM génératifs. Utilisation de modèles NLP classiques légers (**BERT** ou **RoBERTa**).
  - Impact : Perte de finesse dans l'analyse (moins de nuances), mais la classification "Positif / Négatif" reste fonctionnelle.

#### 4.5. Optimisation par Tokens (Smart Batching)

Pour optimiser la facturation API (souvent au million de tokens), nous abandonnons le traitement par "nombre de lignes fixes" (ex: 500 tweets).

- **Nouvelle logique** : Calcul préalable du nombre de tokens via un *tokenizer* léger.
- **Constitution des lots** : Remplissage optimal de la fenêtre de contexte (ex: remplir jusqu'à 90% de la limite de contexte du modèle).

**Gain** : Réduction du nombre de requêtes HTTP et utilisation maximale de chaque centime dépensé.

Phase	Infrastructure	Coût Estimé	Avantages / Limites
<b>Actuelle</b>	<b>Serveurs On-Premise (Bare Metal)</b>  Hébergés en interne chez Free.	<b>~0 € / mois</b>  <i>(Hors électricité et amortissement matériel déjà acquis)</i>	<b>+ Coût marginal nul.</b>  - Bande passante limitée (goulot d'étranglement).  - Pas de redondance (SPOF).
<b>Cible (Scale)</b>	<b>Cloud Public (AWS / Azure)</b>  Instances EC2/VM avec Load Balancing.	<b>~500 € / mois</b>  <i>(Variable selon charge CPU/RAM)</i>	<b>+ Scalabilité automatique (Auto-scaling).</b>  <b>+ Load Balancing</b> pour gérer les pics de trafic.  - Coût récurrent (OpEx).



## Conclusion

La solution Dallosh offre un potentiel d'utilisation varié selon les profils (agents, managers, direction, autres services). Les impacts économiques, organisationnels et sociaux sont significatifs : réduction des coûts opérationnels, transformation des rôles vers des tâches à plus forte valeur ajoutée, amélioration des conditions de travail.

Les risques éthiques et réglementaires sont maîtrisables grâce à une stratégie RGPD rigoureuse (accords Zero Data Retention, modèles locaux, RBAC) et des précautions vis-à-vis des erreurs de classification (validation humaine, droit de recours).

Les scénarios d'évolution tracent une feuille de route : intégration back-office à court terme, extension multi-canal à moyen terme, agent conversationnel hybride et anticipation des crises à long terme.

**Recommandations** : Le déploiement doit être progressif, avec une attention particulière à l'accompagnement des équipes et à la préservation de l'humanité du service client. L'outil doit rester un **assistant intelligent** qui amplifie les capacités humaines plutôt que de les remplacer.