

Veille réglementaire et technologique



Projet : Amazon Review Analysis

Auteur : Amara NAIT SAIDI

Date création : 21 octobre 2025

Mise à jour : novembre 2025

Version 1.0

A propos de ce document

Ce document s'inscrit dans le cadre du projet d'analyse et de valorisation des avis clients. L'objectif de cette veille est d'identifier les technologies cloud, Big Data et réglementaires les plus récentes afin de garantir un traitement des données efficace, conforme et responsable. Elle permet de suivre les innovations (Snowflake, dbt, AWS, Databricks...) et d'anticiper les impacts réglementaires (RGPD, AI Act, CCPA) sur la collecte, le stockage et l'analyse des avis clients.

Enfin, cette veille contribue à orienter les choix techniques vers des solutions durables et éthiques, en cohérence avec les principes de RSE et d'IA responsable.

Table des matières

| | |
|---|----|
| 1. Objectif du dispositif de veille technologique | 4 |
| 2. Méthodologie de veille..... | 4 |
| 2.1. Définitions d'une veille statique et veille dynamique | 4 |
| 2.2. Choix méthodologique : veille statique | 4 |
| 3. Synthèse de la veille technologique et réglementaire | 6 |
| 3.1. Les tendances observées | 6 |
| 4. Méthode professionnelle de partage | 9 |
| 4.1. Outils utilisés pour le partage de cette veille statique | 9 |
| 4.2. Outils recommandés pour une future veille dynamique | 10 |
| 5. Synthèse des résultats de la veille | 10 |
| 5.1. Cloud et data platform | 10 |
| 5.2. Modélisation et transformation des données | 10 |
| 5.3. Machine Learning et IA..... | 10 |
| 5.4. Cadre réglementaire | 10 |
| 5.5. Responsabilité sociétale et environnementale (RSE) | 11 |
| 6. Conclusion | 11 |
| 7. Références | 12 |

1. Objectif du dispositif de veille technologique

Le projet Amazon Reviews vise à analyser et exploiter les avis clients en utilisant des technologies avancées de traitement du langage (NLP, LLMs) tout en respectant les obligations réglementaires (RGPD, CCPA... etc.).

Ce dispositif de veille technologique vise à suivre les évolutions technologiques dans les domaines du numérique, du cloud, du Big Data et du décisionnel afin d'adapter les choix techniques du projet Amazon. Il vise également à identifier les obligations réglementaires liées à la gestion et au traitement des données (RGPD, CNIL, RGAA, RSE).

2. Méthodologie de veille

2.1. Définitions d'une veille statique et veille dynamique

Veille statique (ponctuelle)

La veille statique consiste à effectuer des recherches ponctuelles, à un instant T, pour répondre à un besoin immédiat. Elle se caractérise par une collecte manuelle d'informations non automatisée, sans mécanisme de mise à jour continue. Cette approche présente des limites importantes : les informations deviennent rapidement obsolètes, elle nécessite un effort constant de recherche manuelle, et le risque de passer à côté d'évolutions critiques est élevé. La veille statique convient uniquement aux projets courts avec des technologies stables.

Veille dynamique (continue et automatisée)

À l'inverse, la veille dynamique met en place un système permanent de surveillance technologique et réglementaire. Elle s'appuie sur des flux automatisés (RSS, alertes, API), une mise à jour continue des informations, et des mécanismes d'alerte en temps réel. Cette approche permet d'anticiper les évolutions, de maintenir une conformité réglementaire constante, et d'adapter rapidement les choix techniques aux tendances émergentes.

2.2. Choix méthodologique : veille statique

Justification du choix

Pour ce projet Amazon Review Analysis, nous avons opté pour une veille statique intensive. Ce choix méthodologique repose sur une analyse pragmatique des besoins du projet et des contraintes opérationnelles.

Besoins du projet

Le projet nécessitait en phase de cadrage une photographie précise et exhaustive du paysage technologique et réglementaire à un instant donné. L'objectif était d'établir rapidement un référentiel stable permettant de guider les choix architecturaux initiaux (sélection des technologies cloud, outils de traitement Big Data, frameworks ML) et de garantir la conformité réglementaire dès la conception du système. Cette base documentaire devait être immédiatement exploitable par l'ensemble de l'équipe projet.

Pertinence de l'approche statique

Une veille statique bien exécutée offre une vision consolidée et cohérente du paysage technologique et réglementaire. Elle permet une analyse comparative approfondie des différentes solutions technologiques disponibles au moment des choix architecturaux critiques. Cette approche constitue une base documentaire stable et partageable immédiatement utilisable par tous les membres de l'équipe, et facilite la prise de décision en fournissant une vue d'ensemble plutôt qu'une accumulation d'informations ponctuelles.

Méthodologie appliquée

Le travail a débuté par l'identification des domaines clés à couvrir : technologies cloud et plateformes data (AWS, Snowflake, Databricks), outils de transformation et de modélisation (dbt, Spark, Kafka), solutions d'IA et de ML (SageMaker, Milvus, LLMs), cadre réglementaire européen et international (RGPD, AI Act, CCPA) ainsi que les enjeux de responsabilité sociétale (RSE numérique, Green AI). Cette étape a aussi permis de définir des critères de sélection des sources : fiabilité, pertinence, fraîcheur des informations et neutralité.

La collecte s'est structurée autour de plusieurs axes. Pour les aspects technologiques, elle s'est appuyée sur les blogs éditeurs, la documentation technique, les notes de version et les annonces de conférences majeures. Pour la réglementation, les informations ont été récoltées depuis les sites officiels des autorités, les textes législatifs et des analyses d'agences de presse reconnues. Les volets RSE et bonnes pratiques ont été complétés par les publications d'organismes spécialisés et les normes ISO. L'ensemble couvre la période allant de 2023 à novembre 2025.

Chaque source a ensuite été analysée selon une grille commune évaluant sa pertinence pour le projet, son actualité, sa fiabilité et ses implications opérationnelles. Cette étape a permis d'établir une synthèse recensant treize sources critiques, leurs points clés, une classification thématique, une évaluation d'impact et des recommandations associées.

Enfin, une phase de consolidation a abouti à la rédaction du document de référence, à la mise en place d'un espace de partage dans Notion et à la définition d'un plan de mise à jour du référentiel.

3. Synthèse de la veille technologique et réglementaire

Notre veille statique a permis de couvrir 13 sources majeures réparties sur l'ensemble des domaines critiques cités plus haut. Cette couverture garantit une vision complète du paysage technologique et réglementaire pertinent pour notre projet.

Ce référentiel a directement alimenté plusieurs décisions architecturales critiques concernant le projet, notamment le choix d’AWS S3, la sélection de Snowflake comme data warehouse et l’adoption des contraintes RGPD et AI Act dès la phase de conception (anonymisation, explicabilité, documentation).

3.1. Les tendances observées

| Technologie / Texte | Sujet | Points clés | Date | Source | Mots-clés utilisés |
|--|---------------------------------------|--|------------|---|--|
| Amazon S3 (Tables, Intelligent-Tiering, features récentes) | Stockage objet & data-lake | S3 renforce ses capacités data-lake (S3 Tables, intégration Iceberg/compaction, tagging pour ABAC et coût), optimisation des coûts via <i>Intelligent-Tiering</i> — utile pour stocker massivement des données brutes et historiser les données. | 2024–2025 | https://aws.amazon.com/blogs/aws/aws-weekly-roundup-amazon-s3-amazon-ec2-and-more-november-10-2025 | data lake, cloud storage, optimisation coût |
| Snowflake (AI & data platform évolutive) | Data warehouse cloud & AI | Snowflake (annonces Summit 2025) pousse l’intégration AI native, fonctions pour ingestion + gouvernance des données et meilleures capacités pour embeddings / ML, facilitant pipelines d’analyse sans transfert massif hors DW. | 03-juin-25 | https://www.snowflake.com/en/blog/announcements-snowflake-summit-2025 | AI, data governance, cloud warehouse, embeddings, LLMs |
| VADER (Valence Aware Dictionary and sEntiment Reasoner) | Analyse de sentiments (NLP classique) | VADER est un lexique + règles conçu pour analyser les sentiments dans des textes courts (réseaux sociaux, applications temps réel). Léger, interprétable, performant sans entraînement ML. Très utile pour des pipelines de sentiment rapide (Kafka, Spark, ETL léger) | 2024–2025 | https://vadersentiment.readthedocs.io/en/latest/ | sentiment analysis, lexicon-based, NLP, real-time |

| | | | | | |
|--|---|---|---------------|---|---|
| Zero-shot Classification (transformers / LLM) | Classification thématique & NLP avancé | Les modèles zero-shot, basés sur des architectures type LLM/transformers, permettent d'étiqueter du texte sans entraînement supervisé. Très utile pour catégoriser contenus, détecter intentions, analyser sujets/émotions à grande échelle (pipeline ML, embeddings, lakehouse). | 2024–2025 | http://huggingface.co/tasks/zero-shot-classification | zero-shot, NLP, LLM, classification, embeddings, topic modeling |
| dbt (Data Build Tool) | Modélisation ELT et gouvernance SQL | dbt continue d'évoluer (release notes 2025) : meilleures performances, intégrations cloud, gestion du lineage et tests data — excellent pour stabiliser transformations et réutilisabilité de modèles analytiques. | nov-25 | https://docs.getdbt.com/docs/dbt-versions/dbt-cloud-release-notes | data lineage, SQL pipeline, data quality |
| Apache Spark (moteur de calcul distribué) | Traitement Big Data / ETL distribué | Spark 3.x améliore performance, intégration ML et compatibilité avec moteurs modernes (optimisations pour ETL massif) — choix classique pour transformations NLP à grande échelle. | 2023–2025 | https://spark.apache.org/news | ETL, big data, distributed computing, NLP pipeline |
| Apache Kafka / Amazon MSK | Bus d'événements et ingestion temps réel | Kafka (versions 2024–2025) et offres managées (MSK) permettent ingestion continue des données et logs, résilience et faible latence pour pipelines temps-réel (sentiment en near-real time). | 2024–mai 2025 | RN : https://kafka.apache.org/downloads | streaming, event bus, MSK, ingestion temps réel |
| Databricks (Unity Catalog, Lakehouse features) | Lakehouse / gouvernance & productivisation ML | Databricks renforce Unity Catalog (gouvernance, métriques, interop multi-format) et fonctionnalités Data+AI 2025 — facilite catalogage, lineage et sécurité des features/embeddings. | juin-25 | https://www.databricks.com/blog/whats-new-databricks-unity-catalog-data-ai-summit-2025 | Unity Catalog, data governance, lakehouse, MLops |

| | | | | | |
|---|---|--|---|---|--|
| Amazon SageMaker (nouvelle génération / Unified Studio) | Plateforme ML managée | SageMaker Unified Studio & Lakehouse (annoncé re:Invent 2024 / updates 2024–2025) unifie entraînement, fine-tuning et déploiement (HyperPod, outils GPU distribués) — utile pour garder entraînements/inférences sur AWS pour contrôle des données. | 03 déc. 2024 | https://aws.amazon.com/about-aws/whats-new/2024/12/new-generation-amazon-sagemaker | machine learning, AWS, model training, LLM deployment |
| Milvus (vector DB / embeddings store) | Stockage vecteurs pour recherche sémantique | Milvus (v2.x) est une solution open-source mature pour indexation/lookup d’embeddings, scalabilité pour similarity search — pertinent pour recherche thématique et clustering (semantic search). | 2024–2025 | https://milvus.io/docs/v2.4.x/overview.md | embeddings, semantic search, similarity |
| AI Act (Règlement UE) | Régulation IA (transparence, obligations) | L’AI Act impose obligations de transparence (déclarations si contenu IA), restrictions sur manipulation — impact direct sur synthèses de données générées par LLM (étiquetage, XAI, audits). Débats et implémentations en 2025. | 2024–2025 (mesures d’application 2025). | https://www.reuters.com/technology/tech-lobby-group-urges-eu-leaders-pause-ai-act-2025-06-25 | IA Acte, réglementation IA, transparence, conformité UE |
| RGPD (Règlement (EU) 2016/679) | Protection des données personnelles | RGPD demeure la référence sur consentement, minimisation, anonymisation/pseudonymisation et transferts transfrontaliers — contraintes essentielles si les données analysées contiennent des données personnelles (exigence d’anonymisation avant NLP). | Texte initial 27 avr. 2016 (applicable depuis 2018) | https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng | RGPD, data privacy, anonymisation, consentement, EU compliance |

| | | | | | |
|---|--|--|--|--|---|
| CCPA (California Consumer Privacy Act) | Réglementation américaine sur la protection des données personnelles | L'équivalent californien du RGPD : impose le droit à l'information, à la suppression et à l'opposition à la vente des données personnelles. Depuis 2023, le CPRA (California Privacy Rights Act) renforce le CCPA (obligation de transparence accrue et nouvelles catégories de données sensibles). Pertinent pour les traitements de données provenant d'utilisateurs US (collecte, stockage sur S3/Snowflake, partage IA). | Adopté en 2018 – renforcé en 2023 (CPRA) | https://oag.ca.gov/privacy/ccpa | CCPA, CPRA, data privacy US, consentement, droit à la suppression |
| RSE Numérique (Responsabilité Sociétale & Environnementale) | Réglementation / bonnes pratiques | La RSE numérique (CSRD, Green AI, ISO 26000) encourage une IA éthique et sobre : évaluation de l'empreinte carbone des modèles, choix cloud responsables, et inclusion sociale dans les projets IA. Tendance forte en 2025. | 2024–2025 | - https://greenit.eco/nos-etudes-et-essais/impacts-environnementaux-de-lia-dans-le-monde-en-2025 - https://numeum.fr/numerique-responsable | RSE numérique, CSRD, Green AI, IA responsable, durabilité |

4. Méthode professionnelle de partage

Notre méthodologie de partage s'appuie sur un dispositif combinant plusieurs supports complémentaires. Notion sert de base centrale, structurée et accessible, pour organiser la documentation et la mettre à disposition de l'ensemble de l'équipe. Les échanges courants et les informations ponctuelles sont relayés via Slack ou Teams, permettant une communication rapide et ciblée lorsque des résultats, alertes ou mises à jour doivent être diffusés. Enfin, un document de synthèse consolidé est produit afin d'offrir au management et aux parties prenantes une vision claire, stable et exploitable de l'ensemble du référentiel technologique.

4.1. Outils utilisés pour le partage de cette veille statique

1. Notion : base de données structurée pour la documentation et le partage avec l'équipe.
2. Slack / Teams : communication des résultats et alertes ponctuelles à l'équipe.
3. Document de synthèse : référentiel consolidé pour le management et les parties prenantes.

4.2. Outils recommandés pour une future veille dynamique

En cas d'évolution vers une veille continue, les outils suivants seraient pertinents :

- Feedly → Agrégation automatique des sources par flux RSS
- Zapier / Make → Automatisation entre Feedly et Notion
- Google Alerts → Surveillance automatique de mots-clés critiques
- Newsletter interne → Diffusion automatisée des synthèses mensuelles

Cette infrastructure n'est pas déployée actuellement mais constitue une feuille de route pour une surveillance continue future.

5. Synthèse des résultats de la veille

La veille a permis d'identifier les grandes tendances actuelles du cloud, du Big Data et de la gouvernance des données à l'horizon 2025.

Les outils analysés (AWS S3, Snowflake, dbt, Databricks, Kafka, etc.) confirment la convergence entre infrastructures cloud, gouvernance, IA et durabilité.

5.1. Cloud et data platform

Les solutions cloud (AWS S3, Snowflake, Databricks) évoluent vers des plateformes unifiées, intégrant nativement la gouvernance, la sécurité et l'intelligence artificielle.

Les nouveautés 2024–2025 (S3 Tables, Snowflake AI Native, Unity Catalog) renforcent la traçabilité et la maîtrise des coûts, tout en répondant aux besoins de scalabilité et d'automatisation des pipelines.

5.2. Modélisation et transformation des données

dbt s'impose comme un standard pour le data lineage, la documentation et les tests automatiques.

Combiné à Spark ou à Kafka, il permet une industrialisation maîtrisée des traitements ELT/ETL et une meilleure qualité des jeux de données, essentielle à toute démarche d'IA responsable.

5.3. Machine Learning et IA

Les outils de type SageMaker et Milvus montrent la tendance vers des environnements intégrés, modulaires et sécurisés.

SageMaker (Unified Studio) centralise les phases de MLOps, tandis que Milvus offre un stockage vectoriel performant pour la recherche sémantique et les applications de LLM.

L'objectif commun est de rapprocher la donnée de l'intelligence artificielle, sans la déplacer inutilement.

5.4. Cadre réglementaire

La réglementation (RGPD, AI Act, CCPA) continue d'évoluer pour encadrer l'usage des données et des modèles d'IA.

Les entreprises doivent anticiper les obligations de transparence, d'auditabilité et de protection des données personnelles, notamment dans le contexte du traitement de données

clients à grande échelle.

L'AI Act, en cours de déploiement, impose de nouvelles exigences de documentation, d'explicabilité et de gestion du risque algorithmique.

5.5. Responsabilité sociétale et environnementale (RSE)

La dimension RSE émerge comme un pilier de la gouvernance data.

Les initiatives “Green AI” et “Numérique Responsable” (Numeum, GreenIT.eco) insistent sur la sobriété énergétique, la durabilité des infrastructures cloud, et l'intégration d'indicateurs environnementaux dans les projets data.

Cette tendance traduit une mutation de la data en un actif éthique et soutenable, non seulement performant.

6. Conclusion

Cette veille met en évidence la maturité croissante de l'écosystème data & cloud, désormais structuré autour de trois axes stratégiques :

- L'intégration technologique : les frontières entre stockage, gouvernance et IA s'effacent au profit de plateformes unifiées (Snowflake, Databricks, AWS).
- La conformité et la transparence : la régulation (RGPD, AI Act, CCPA) devient un facteur d'innovation autant que de contrainte, guidant les pratiques de gouvernance et de documentation.
- La durabilité et la responsabilité : la RSE s'impose comme un levier d'optimisation éthique et environnementale, soutenue par des initiatives sectorielles (Green AI, Numeum, GreenIT.eco).

En somme, les tendances 2024–2025 confirment que la valeur de la donnée ne repose plus uniquement sur la puissance technologique, mais sur sa qualité, sa conformité et sa soutenabilité.

Les entreprises qui parviennent à articuler ces trois dimensions — performance, gouvernance et responsabilité — disposeront d'un avantage compétitif durable.

7. Références

- [Technology Watch Template](#)
- [Source evaluation methodology](#)
- [A great example of newsletter that keeps everything concise and may serve as a template for your summary page](#)
- [RGAA](#)
- [CNIL](#)
- [La responsabilité sociétale des entreprises \(RSE\) Insee.](#)
- [RSE Economie.gouv](#)
- [CCPA California Consumer Privacy Act](#)
- <https://aws.amazon.com/blogs/aws/aws-weekly-roundup-amazon-s3-amazon-ec2-and-more-november-10-2025>
- <https://www.snowflake.com/en/blog/announcements-snowflake-summit-2025/>
- <https://docs.getdbt.com/docs/dbt-versions/dbt-cloud-release-notes>
- <https://spark.apache.org/news>
- <https://kafka.apache.org/downloads>
- <https://www.databricks.com/blog/whats-new-databricks-unity-catalog-data-ai-summit-2025>
- <https://aws.amazon.com/about-aws/whats-new/2024/12/next-generation-amazon-sagemaker>
- <https://milvus.io/docs/v2.4.x/overview.md>
- <https://www.reuters.com/technology/tech-lobby-group-urges-eu-leaders-pause-ai-act-2025-06-25>
- <https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>
- <https://oag.ca.gov/privacy/ccpa>
- <https://greenit.eco/nos-etudes-et-essais/impacts-environnementaux-de-lia-dans-le-monde-en-2025>
- <https://numeum.fr/numerique-responsable>