

Veille réglementaire et technologique



Projet : Amazon Review Analysis

Auteur : Dyhia TOUAHRI

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. Outils et supports.....	4
2.2. Critères de sélection des sources.....	4
2.3. Organisation de la veille	4
3. Synthèse de la veille technologique et réglementaire	5
3.1. Les tendances observées.....	5
4. Méthode professionnelle de partage	7
4.1. Outils recommandés	7
5. Synthèse des résultats de la veille	8
5.1. Cloud et data platform.....	8
5.2. Modélisation et transformation des données.....	8
5.3. Machine Learning et IA	8
5.4. Cadre réglementaire.....	8
5.5. Responsabilité sociétale et environnementale (RSE)	8
6. Conclusion.....	9
7. Références	10

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. Outils et supports

- Notion : tableaux + synthèses + pages thématiques.
- Google mots-clés : “cloud compliance”, “data governance”, “LLMs”, “RSE numérique”.
- Feedly : flux RSS de sources validées (AWS, CNIL, EDPB, DataGouv, etc.).
- LinkedIn & newsletters professionnelles : observatoires, experts, cabinets.
- Sites officiels : CNIL, RGAA, INSEE, CCPA, EDPB, Garante Privacy, ICO

2.2. Critères de sélection des sources

- Fiabilité : organisme officiel, publication scientifique ou média reconnu.
- Pertinence : lien direct avec les enjeux e-commerce / Big Data / conformité.
- Fraîcheur : publication récente (< 36 mois).
- Neutralité : contenu objectif, sans conflit d'intérêts identifié.

2.3. Organisation de la veille

- Collecte quotidienne via Google Alerts et Feedly.
- Synthèse hebdomadaire dans Notion.
- Classement par thématique : tech, réglementaire, cloud, RSE...
- Évaluation mensuelle : pertinence des articles, taux d'usage interne.

3. Synthèse de la veille technologique et réglementaire

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
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/next-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
RGD (Regulation (EU) 2016/679)	Protection des données personnelles	RGD demeure la référence sur consentement, minimisation, anonymisation/pseudonymisation et transferts transfrontaliers — contraintes essentielles si les données	Texte initial 27 avr. 2016 (applicable depuis 2018)	https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng	RGD, data privacy, anonymisation, consentement, EU compliance

		analysées contiennent des données personnelles (exigence d'anonymisation avant NLP).			
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

4.1. Outils recommandés

- Feedly → Agrégation automatique des sources par mots-clés (NLP, RGPD, AI Act, AWS, etc.)
- Notion → Base de données structurée pour la documentation et les alertes
- Zapier / Make → Automatisation entre Feedly et Notion

- Slack / Teams → Alertes critiques et diffusion rapide
- Scoop.it ou Newsletter interne → Synthèse mensuelle pour le management

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>