Parlons de l'IA Générative pour les curieux de la Data

BYMADA

Sommaire

Salut, ici BYMADA.

Ce fut un long processus de réflexion, mais BYMADA a été lancé en 2021 par Laura pour aider des PME et TPE à vraiment utiliser leurs données pour mieux naviguer dans la digitalisation et l'optimisation des marques, et ainsi leur permettre de gagner en agilité (surtout en ces temps demandant de se réinventer en permanence).
Pour la petite histoire, le nom BYMADA vient de la concaténation de « BY MArketing and DAta ».
C'est énoncer dès le départ qu'une bonne décision est le fruit d'un travail commun entre le business et des données de bonne qualité.

Depuis 2 ans maintenant, les outils IA prennent de plus en plus de place dans quotidien et en Décembre 2022, OpenIAI avec ChatGPT a bouleversé le monde de la data.

Dans notre mission de démocratiser la donnée, ce livret regroupe nos meilleurs articles de blog sur l'IA, nos explications et convinctions pour un usage éthique et respectueux envers les clients, les employés, et notre société.

Bonne lecture et à très bientôt!
Laura & Thomas

INTRODUCTION

IA Ethique – Qu'est-ce que cela veut dire?

Cette introduction sert de réflexion aux sujets actuels autour l'IA. Les sujets d'IA éthique et utilisation responsable, commencent à s'accélérer avec l'approbation de plusieurs réglementations dans le monde (+50 pays réfléchissent ou font voter à une réglementation locale sur l'IA).
Nous voulions donc faire un récapitulatif des enjeux actuels. Mais, avant de commencer, rappelons-nous des principales technologies sur le marché:

📈 **Advanced Analytics**: Marché mature avec des solutions performantes (et qui vont aller plus loin, avec de nouvelles opportunités dans le cloud).

📱 **Generative AI** (ChatGPT, DALL E, BARD etc.): LA nouveauté 2023, qui s'installe dans notre quotidien très rapidement, à l'image des smartphones dans les années 2000.

🧠 **Machine Learning**: utilise des algorithmes utilisant des données pour réaliser des prédictions ou décisions, où les performances s'améliorent lorsque exposé à plus de données à travers le temps. C'est en cours d'adoption dans les entreprises (car nécessite une compréhension technique et les compétences) avec des Use-Cases où le ROI commence à prendre forme.

Qu'est-ce que l'IA éthique ?
L'IA éthique fait référence au développement et au déploiement de systèmes d'intelligence artificielle (IA) qui mettent l'accent sur l'équité, la transparence, la responsabilité et le respect des valeurs humaines.
L'objectif est de promouvoir une utilisation sûre et responsable de l'IA, de réduire les nouveaux risques liés à l'IA et de prévenir les préjudices. Une grande partie du travail dans ce domaine se concentre sur quatre principaux aspects :

- **Biais** – risque que le système discrimine injustement des individus ou des groupes.
- **Explicabilité** – risque que le système ou ses décisions ne soient pas compréhensibles pour les utilisateurs et les développeurs.
- **Robustesse** – risque que l'algorithme échoue dans des circonstances inattendues ou sous l'attaque.
- **Confidentialité** – risque que le système ne protège pas adéquatement les données personnelles.

Pourquoi l'IA éthique est-elle importante ?
Bien que l'IA et l'automatisation puissent présenter d'importants avantages, tels qu'une plus grande efficacité, une plus grande innovation, une personnalisation des services et une réduction de la charge de travail pour les travailleurs humains, l'utilisation de l'IA peut également présenter de nouveaux risques qui doivent être abordés: Le traitement des données personnelles et dans quels usages: vente, recrutement, les droits d'auteurs vis-à-vis des métiers créatif etc.

Comment les entreprises sont-elles concernées ?
À mesure que l'adoption de l'IA devient plus répandue (ex: ChatGPT, extensions de logiciels vendeurs), la sensibilisation du public aux risques augmente et l'attention réglementaire s'intensifie, les entreprises sont sous pression croissante pour s'assurer qu'elles conçoivent et déploient l'IA de manière éthique.
Bientôt, les entreprises seront légalement tenues d'incorporer l'IA éthique dans leur travail.
Le projet de loi de l'UE sur l'IA (AI Act voté en juin 2023) est la proposition de loi de l'UE visant à réglementer le développement et l'utilisation des systèmes d'IA à « haut risque ». Il s'agit de la première loi dans le monde régissant le développement et l'utilisation de l'IA de manière exhaustive. La loi sur l'IA de l'UE est destinée à devenir le « RGPD de l'IA », avec des sanctions lourdes en cas de non-conformité, une portée extraterritoriale et un ensemble étendu d'exigences obligatoires pour les organisations qui développent et déploient l'IA.

En tant qu'entreprise, quelles actions entreprendre ?
✅ S'aligner sur la vision et les croyances de l'entreprise: il est essentiel d'être clair sur la vision de l'entreprise, les valeurs qu'elle soutient, et comment un éventuel cas d'utilisation des données s'aligne avec ces valeurs. Cela peut orienter les décisions concernant l'utilisation des données (et modèles d'IA des vendeurs à utiliser ou non).

✅ Déterminer l'Ownership des données et l'atténuation des risques (et cela devient impératif)
Définit les rôles pour l'utilisation éthique des données et la propriété des données (Data Owners) devient indispensable. Ainsi, si un algorithme doit être modifié, par exemple, ou si l'accès aux données d'un système doit être ajusté, il est clair qui doit apporter ces modifications.
Les organisations doivent également être conscientes des risques existants liés aux données, tels que l'utilisation des informations de contact personnelles des clients.

✅ Faire évoluer la culture et les compétences.

✅ Evaluer les usages actuels de l'IA dans l'entreprise.

✅ Créer un comité d'éthique des données
Idéalement, un comité d'éthique des données serait un comité pluridisciplinaire composé de représentants des domaines de l'entreprise, de la conformité et du juridique, des opérations, de l'audit, de l'informatique et de la haute direction. La représentation de l'IT est essentielle en raison des responsabilités et de ses connaissances techniques. En effet, ce département est responsable de plusieurs aspects de la gestion et de la protection des données.
Cependant, il incombe toujours aux Métiers et aux Data Owners de veiller à ce que leurs fonctions respectent les politiques adoptées et de surveiller en permanence les nouveaux cas d'utilisation qui pourraient nécessiter une évaluation des risques liés aux données.

✅ Se former: votre responsabilité est engagée et le sujet n'est pas simple.

la fin en quelques mots

Pour aller plus loin je vous invite à vous abonner à la newsletter de la CNIL, mais aussi à vous former. J'écrirai un nouvel article sur l'AI Act d'ici la fin de l'année. D'ici là, analysez vos processus où vous utilisez de l'IA et à la manière dont vous transmettez des informations confidentielles ou des données personnelles.

CHAPITRE I

IA Générative, LLM explosent le game. Ce qu'il faut savoir

Remontant aux premières explorations de la linguistique informatique dans les années 1950-1960, la technologie a parcouru un long chemin. L'arrivée du T9 dans les années 1990 a révolutionné la saisie de texte sur les téléphones mobiles, posant les bases de la prédiction de texte. Puis, au tournant du millénaire, Google Traduction a transformé notre manière de franchir les barrières linguistiques, en appliquant des algorithmes statistiques à la traduction de langues. Cependant, c'est avec l'avènement du Deep Learning dans les années 2010 que les LLMs ont vraiment pris leur envol.

Les LLMs, ou « Large Language Models », sont des systèmes d'intelligence artificielle (IA) avancés conçus pour comprendre, générer et interagir avec le texte en langue naturelle. Ils sont formés sur de vastes ensembles de données textuelles et peuvent effectuer une variété de tâches liées au langage, telles que la traduction, la rédaction de textes, la réponse à des questions et la synthèse d'informations.

Voici quelques examples de LLMs connus:

✓ GPT (Generative Pre-trained Transformer) Développé par OpenAI, il excelle dans diverses tâches telles que la rédaction de texte, la traduction, et la réponse aux questions, grâce à son apprentissage sur d'énormes volumes de données textuelles.

✓ BERT (Bidirectional Encoder Representations from Transformers) Développé par Google, est utilisé pour améliorer la compréhension du langage naturel, en particulier pour les moteurs de recherche.

✓ LaMDA : Créé par Google également, utilisé pour alimenter le bot de conversation Bard, est connu pour ses compétences conversationnelles avancées

✓ RoBERTa: Développé par Meta, il est une version améliorée de BERT qui a été optimisée pour offrir de meilleures performances.

✓ Falcon : Développé par le Technology Innovation Institute, Falcon est un LLM open source qui s'est classé parmi les LLMs pré-entraînés les plus performants sur la plateforme Hugging Face.

On peut effectivement dire que ces dernières années ont été marquées par une sorte de « course à la technologie » entre les plus grandes entreprises de tech dans le domaine. Cette compétition a été particulièrement visible avec la montée en puissance de GPT-3.5 d'OpenAI, qui a connu un succès notable tant auprès des entreprises que du grand public (OpenAI revendique aujourd'hui 180 millions d'utilisateurs par mois).
Les LLMs jouent un rôle essentiel dans les systèmes d' IA générative. Par exemple des modèles de création d'images, comme DALL-E ou Midjourney, en ont besoin pour servir de intermédiaire, traduisant les descriptions textuelles des utilisateurs en instructions compréhensibles pour les algorithmes de génération d'images. Cette capacité rend l'IA génératrice d'images plus accessible et augmente la précision et la pertinence des images produites.

A quoi ça sert ?
On remarque une tendance à l'intégration de LLM spécialisés dans des outils et plateformes spécifiques.
Voici quelques exemples en entreprise:

👉 **Productivité des employés**: Microsoft Copilot, est un outil de productivité, lancé en février 2023, qui intègre les LLMs avec les applications Microsoft 365, comme Word, Excel, et PowerPoint, pour améliorer la créativité, la productivité et les compétences des utilisateurs. Il ne s'agit pas d'un LLM indépendant, mais plutôt d'une plateforme qui l'utilise pour offrir des fonctionnalités intelligentes et contextuelles dans diverses applications Microsoft. Le test que j'ai réalisé qui m'a le plus marqué est la rédactions des comptes-rendus… Merci Copilot!

👉 Pour l'**optimisation des processus** internes, ils automatisent et rationalisent des tâches complexes, allant de la planification de réunions à la gestion de projets, en passant par la gestion de la documentation.

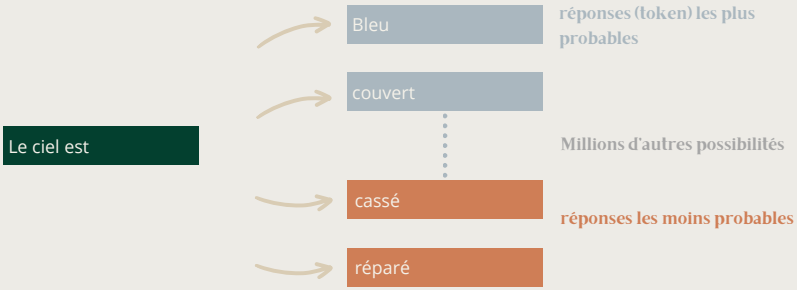
👉 Dans le domaine de l'**analyse de données** et de la génération de rapports, les LLMs peuvent traiter rapidement de grands ensembles de données pour en extraire des insights pertinents, générer des rapports automatisés et fournir des résumés détaillés, aidant ainsi les entreprises à prendre des décisions éclairées basées sur des données. On peut cité par exemple l'outil Q&A de Power BI qui permet de poser directement des questions sur notre base de données ou nos rapports.

👉 L'**intégration avec d'autres technologies** est un autre domaine clé où les LLMs montrent leurs valeurs. Ils peuvent se connecter et interagir avec divers systèmes d'entreprise, tels que les CRM et les ERP, pour améliorer la communication entre les différentes plateformes et optimiser les workflows. Par exemple Salesforce Einstein utilise l'IA pour analyser et prédire les tendances dans les données CRM et ERP, offrant des insights personnalisés et des recommandations pour améliorer les interactions client et optimiser les processus financiers.

👉 En matière d'**assistance juridique** et de conformité, ils peuvent analyser et résumer les documents juridiques, aider à la rédaction de contrats et veiller à la conformité réglementaire, réduisant ainsi les risques juridiques (exemple: legartis.ai).

👉 Enfin, dans le secteur du **marketing** et des réseaux sociaux, les LLMs sont capables de générer des idées créatives pour les campagnes, d'analyser les tendances des réseaux sociaux, et de produire du contenu engageant. Nous pouvons citer les deux plus gros éditeurs : Adobe avec Firefly et Canva AI.
Gardons en tête que pour le moment, l'idée n'est pas de réduire le nombre d'employés mais plutôt de leur donner de nouveaux super-pouvoirs pour qu'ils puissent aller plus loin, sans se sentir limité par la technologie ou perdre des heures à rechercher une information.

Alors, comment ça fonctionne ?
Imaginez un magicien qui, en lisant dans vos pensées, devine exactement ce que vous allez dire ensuite. C'est un peu ainsi que fonctionnent les Large Language Models ! Au cœur de ces modèles se trouve l'art de prédire le texte. Tout commence par la transformation des mots en chiffres, ou plus précisément en vecteurs. Cela aide les LLMs à comprendre non seulement la signification des mots mais aussi comment ils se relient les uns aux autres dans le grand puzzle du langage.

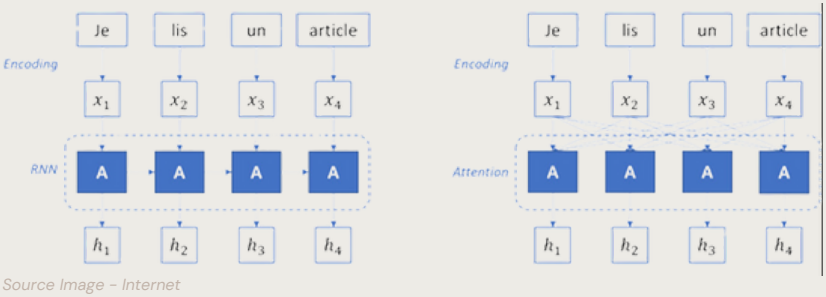


Concrètement un grand modèle de langage utilise un type spécial de Réseau neuronal , appelé l'Architecture Transformer, qui est composé de parties appelées encodeurs et décodeurs. Ces parties ont une fonctionnalité spéciale appelée « auto-attention », qui leur permet de traiter des morceaux de texte simultanément, plutôt qu'un à la fois. Cette méthode est plus avancée que les anciennes méthodes tel que le RNN, qui traitaient le texte séquentiellement.

Mais comment deviennent-ils si doués pour prédire ? En lisant !
Oui, en parcourant un nombre incroyable de textes, des livres aux articles en ligne. C'est cette vaste expérience de lecture qui les rend si habiles à comprendre et à générer du langage, comme un bibliophile qui a lu tous les livres de la bibliothèque.

C'est la phase de pré-entraînement.

Durant cette étape, le modèle est exposé à une quantité massive de données. Cette exposition lui permet d'apprendre les fonctions linguistiques de base et de se familiariser avec une vaste gamme de connaissances et de structures linguistiques. Le nombre élevé de paramètres utilisés par ces modèles, similaire aux souvenirs accumulés par la mémoire humaine durant l'apprentissage, enrichit leur banque de connaissances et améliore leurs performances.
Après le pré-entraînement, les LLMs subissent une phase de fine-tuning, où ils sont ajustés pour des tâches spécifiques. Ici, ils sont formés sur des ensembles de données plus ciblés, ce qui affûte leur capacité à répondre à des besoins spécifiques. Cette étape est comme un cours de spécialisation après un enseignement général. Ainsi, un LLM pré-entraîné peut être adapté pour des applications telles que la traduction, l'assistance juridique, ou même la composition poétique.



Source Image - Internet

CHAPTER I

Quelles sont leurs limites ?

Biais et Préjugés: Imaginez un LLM comme un miroir qui reflète les données avec lesquelles il a été formé. Si ces données contiennent des biais ou des préjugés, le LLM risque de les reproduire, menant parfois à des sorties fausses ou toxiques. L'exemple l'échec de Tay de Microsoft qui en 2016 à générer des propos remplis de préjugés au bout de quelques heures d'existence.

Contexte et de Mémoire: Chaque LLM a une limite de mémoire. Si vous lui donnez trop d'informations à traiter, il peut se « perdre » et ne plus être capable de réaliser correctement la tâche demandée.

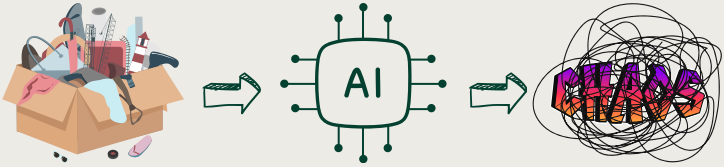
Coûts et Impact Environnemental: Le développement et le fonctionnement des LLMs nécessitent des ressources énormes, tant en termes d'investissement financier qu'en consommation d'énergie (fonctionnement des serveurs).

Hallucinations: Parfois, un LLM peut créer des réponses complètement fausses ou inappropriées, on appelle cela des « hallucinations » (ex: citer un auteur qui n'existe pas). De plus, s'ils ne sont pas correctement surveillés, ils peuvent poser des risques de sécurité, comme la fuite d'informations personnelles ou la propagation de spams et d'escroqueries.

Consentement et Éthique: Les données utilisées pour entraîner peuvent parfois être collectées sans consentement, soulevant des questions éthiques et légales, notamment en matière de droit d'auteur et de confidentialité.
Concernant ce dernier point, l'année 2023 fut un vrai sujet de préoccupation au niveau des différents gouvernements, notamment en Europe. Le RGPD, entré en vigueur en 2018, axé sur la protection des données personnelles, impose des règles strictes sur la collecte, l'utilisation et la minimisation des données, surtout dans le contexte de l'intelligence artificielle.
Il souligne l'importance de la transparence et de la protection des données, en particulier lorsqu'il s'agit de données sensibles.

CHAPTER II

Qualité des données et LLM – l'essentiel à retenir



Des données en bazar et de mauvaise qualité

Utilisées dans un LLM (ou autre IA d'ailleurs)

Les résultats sont pas fous c'est le CHAOS!

Au cœur des modèles d'IA générative se trouve la qualité des données d'entrée, qui a un impact direct sur l'authenticité, la créativité et l'efficacité du contenu généré.

IA générative & Qualité des données – créer des grands crus

Dans le domaine de l'IA générative, la qualité des données se révèle essentielle. Je vous le démontre en 5 points:

1. Garantir la Précision et la Fiabilité 🎯

Logiquement, le premier challenge est de garantir une précision et une fiabilité des réponses. c'est à dire, avec des données de mauvaise qualité qui contiennent des erreurs, du bruit ou des incohérences, le modèle peut produire des résultats incorrects ou trompeurs, érodant la confiance des utilisateurs.

C'est le fameux "GIGO", Garbage in, Garbage out (Les déchets entrent, les déchets sortent)

2. Combattre les Biais et Promouvoir l'Équité ⚖️

La lutte contre les préjugés dans les modèles d'IA est cruciale. Des données biaisées peuvent perpétuer et amplifier ces biais, menant à des résultats injustes ou discriminatoires.

D'après MIT Technology Review, différents modèles de langage, y compris ChatGPT et GPT-4 d'OpenAI ainsi que LLaMA de Meta, présentaient des biais politiques distincts. GPT-4 étaient considérés comme ayant une tendance politique plus à gauche et libertaire, tandis que LLaMA de Meta était plus à droite et autoritaire.
L'étude montre également que les systèmes de génération d'images tels que DALL-E 2 et Stable Diffusion ont également montré des tendances à amplifier les stéréotypes et les préjugés.

3. Faciliter l'entrainement des algorithmes 🌐
Tout d'abord, la généralisation est la capacité d'un modèle d'IA à produire des résultats pertinents et nouveaux. Une mauvaise qualité des données peut entraver cette capacité, conduisant à des problèmes de surajustement ou de sous-ajustement. Le surajustement se produit lorsque le modèle mémorise les données d'entraînement et ne parvient pas à fonctionner correctement sur des données invisibles, tandis que le sous-ajustement fait référence à l'incapacité du modèle à capturer des modèles importants à partir des données. En résumé, l'algorithme n'est pas en capacité d'apprendre comme vous-voulez !

4. Identifier les Intrants Hors Distribution 🔍
Détecter des entrées anormales, évitant ainsi la génération de résultats trompeurs ou absurdes. Cette compétence est essentielle pour la fiabilité et la précision des réponses générées.
Supposons qu'un modèle d'IA est entraîné pour reconnaître et catégoriser des images de fruits. Cependant, si le modèle reçoit une image d'un objet complètement différent il pourrait être incapable de reconnaître qu'il s'agit d'une entrée hors distribution et pourrait tenter de la classer comme un type de fruit. L'aptitude à détecter les entrées hors distribution est nécessaire pour la précision et la sécurité des modèles d'IA.

5. Promouvoir l'Adaptabilité et la Robustesse 🌟
Afin de permettre une meilleure adaptation aux nouvelles tâches et données, renforçant la robustesse et l'applicabilité des modèles d'IA dans divers scénarios. Cela inclut un entraînement sur des données diversifiées, un apprentissage continu et des tests et une validation rigoureux.

Finalement, on pourra conclure, que si vous voulez un bon usage de votre investissement, vous devez gagner la confiance de vos utilisateurs envers les applications d'IA qui dépendent grandement de la qualité des données.

CHAPTER II

4 actions que vous pouvez entreprendre

📊 **Établir une Stratégie de Gouvernance des Données**
Afin de réellement maîtriser les données, la mise en place d'une gouvernance de données est indispensable. Définir les Data Owners et Stewards permet d'assurer la qualité des données tout au long de son cycle de vie. C'est une condition sinequanone pour le bon fonctionnement des IA en entreprise.

💾 **Mettre en place une stratégie de stockage de données**
Pour s'adapter aux besoins imposés par les volumes croissants de données, il est crucial d'évaluer et d'améliorer l'infrastructure de stockage existante pour garantir une gestion efficace. Dans cette optique, l'adoption de solutions de stockage cloud ou hybrides devient un choix stratégique, offrant une flexibilité et une accessibilité accrues.

✅ **Développer une plan long terme pour assurer la qualité des données**
Dans une IA, les données passent par plusieurs étapes pour entraîner le modèle. Pour assurer une qualité de données fiable, il est essentiel de mettre en place des processus de nettoyage (à la source), de validation et d'enrichissement, avec une gouvernance associée. Ces étapes garantissent que les données restent précises, pertinentes et à jour.
Parallèlement, l'utilisation d'outils d'analyse de données et la réalisation d'audits aident dans l'identification et la correction des erreurs de qualité des données (PowerBI, Tableau ou des outils spécialisés comme SODA ou Attacama).

🤖 **Intégrer l'IA dans le développement et la gestion de la qualité des données**
Avec des volumes de données toujours plus croissants, il devient impossible de corriger l'ensemble des données manuellement. Avec les capacités actuelles, utiliser l'IA (un combo Machine Learning/Generative AI) pour détecter, corriger les problèmes de qualité et proposer de nouvelles règles de qualité, devient un game-changer.
Vous pouvez ainsi utiliser des solutions sur le marché ou développer vous-mêmes des algorithmes de nettoyage des données.
Cependant, même si l'IA peut automatiser de nombreuses tâches, elle ne doit pas remplacer entièrement la surveillance humaine.

CHAPTER III

European IA Act, c’est acté !

Le « Artificial Intelligence Act », initié en 2021, a été approuvé le 21 Mai 2024 par les pays membres de l’Union européenne.
Le texte vise à assurer que les SIA (Systèmes d’Intelligence Artificielle) soient sûrs, respectent les droits fondamentaux et adhèrent aux valeurs européennes. C’est à dire, réguler le développement et l’utilisation d’IA pour un usage responsable et éthique.

Cela commence par la définition des acteurs et de leurs rôles

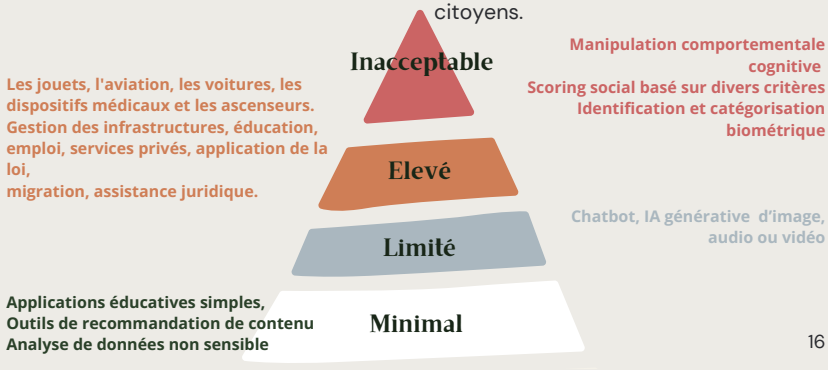
- **Fournisseurs** : Ceux qui conçoivent, développent ou mettent sur le marché des services d’IA. Ils doivent s’assurer que leur système respecte les exigences de documentation technique, d’évaluation des performances et de consommation d’énergie.
- **Importateur** : L’importateur est celui qui introduit des systèmes d’IA provenant de pays tiers sur le marché de l’UE. Il doit veiller à ce que les produits importés soient conformes aux normes européennes.

- Distributeur : Cette catégorie inclut les entités qui mettent à disposition des systèmes d'IA sur le marché, sans être impliquées dans leur fabrication ou importation. Les distributeurs doivent s'assurer que les systèmes d'IA qu'ils distribuent respectent les réglementations en vigueur.
- Utilisateur (Déployeur): Les utilisateurs ou déployeurs sont les entités qui utilisent les systèmes d'IA dans leurs opérations quotidiennes. Ils doivent s'assurer que l'utilisation des systèmes est conforme aux exigences légales, en particulier pour les systèmes à risque élevé.
- Opérateur: Les opérateurs sont ceux qui gèrent le fonctionnement quotidien des systèmes d'IA, pouvant être différents des utilisateurs finaux. Ils ont la responsabilité de maintenir la conformité opérationnelle des systèmes d'IA.
- Fabricant : Il s'agit de l'entité qui produit physiquement les composants ou systèmes d'IA.
- Les fabricants doivent garantir que leurs produits sont fabriqués selon les standards de sécurité et de qualité requis.
- Personne affectée : Ce terme désigne toute personne qui est affectée par un système d'IA, par exemple, les individus soumis à une analyse ou une décision prise par un système d'IA. Bien qu'ils n'aient pas de responsabilités directes en matière de réglementation, la législation vise à protéger leurs droits fondamentaux.

Une entité peut changer de catégorie (par exemple, de déployeur à fournisseur) si elle modifie substantiellement un système d'IA ou le commercialise sous sa marque.

L'introduction la catégorisation d'IA basé sur les risques

L'EU AI Act propose une classification des systèmes d'IA basée sur le niveau de risque qu'ils représentent pour la société. Cette approche vise à équilibrer l'innovation technologique avec la protection des droits fondamentaux et la sécurité des citoyens.



● **Risque inacceptable**: Cette catégorie concerne les utilisations d'IA qui sont jugées incompatibles avec les valeurs de l'UE ou qui violent les droits fondamentaux des individus. Ces applications sont interdites ou strictement limitées.
Ils comprennent :

- La manipulation comportementale cognitive des personnes ou de groupes vulnérables spécifiques : par exemple, des jouets activés par la voix qui encouragent un comportement dangereux chez les enfants.
- Le scoring social : la classification des personnes basée sur le comportement, le statut socio-économique ou les caractéristiques personnelles.
- L'identification et la catégorisation biométrique des personnes.
- Les systèmes d'identification biométrique à distance et en temps réel, tels que la reconnaissance faciale.

Certaines exceptions peuvent être autorisées à des fins d'application de la loi.

● **Risque élevé (SIAHR)**: Les systèmes d'IA qui affectent négativement la sécurité ou les droits fondamentaux seront considérés à haut risque et seront divisés en deux catégories :
Les systèmes d'IA utilisés dans des produits relevant de la législation de l'UE sur la sécurité des produits. Cela inclut les jouets, l'aviation, les voitures, les dispositifs médicaux et les ascenseurs.
Les systèmes d'IA relevant de domaines spécifiques qui devront être enregistrés dans une base de données de l'UE :

- La gestion et l'opération des infrastructures critiques
- L'éducation et la formation professionnelle
- L'emploi, la gestion des travailleurs et l'accès à l'auto-emploi
- L'accès et la jouissance des services privés essentiels et des services et avantages publics
- L'application de la loi
- La gestion de la migration, de l'asile et du contrôle des frontières
- L'assistance dans l'interprétation et l'application de la loi.

Tous les systèmes d'IA à haut risque seront évalués avant leur mise sur le marché et tout au long de leur cycle de vie.

Bon à savoir

Pour les Systèmes à Haut Risque : Des obligations s'appliquent, telles que la mise en place de systèmes de gestion des risques, la surveillance post-marché, et la conservation des enregistrements d'activité (logs), jusqu'à qu'ils soient conformes. Les utilisateurs doivent assurer une supervision humaine et réaliser une analyse d'impact sur les droits fondamentaux.

🟡 **Risque limité** : Cette catégorie s'applique aux systèmes d'IA qui nécessitent une certaine transparence dans leur utilisation mais présentent un risque moindre comparé aux catégories précédentes. Par exemple, un chatbot devrait être conçu de manière à ce que les utilisateurs sachent qu'ils interagissent avec une machine et non avec un humain. Bien que les exigences soient moins strictes, les développeurs doivent encore informer les utilisateurs de l'utilisation de l'IA.

🟢 **Risque minimal** : La majorité des applications d'IA tombent dans cette catégorie, où le risque pour les droits et libertés des citoyens est considéré comme négligeable. Ces applications peuvent être largement utilisées sans réglementations supplémentaires strictes. Il s'agit, par exemple, des systèmes de recommandation de films ou de musique. Bien que ces applications soient largement exemptes de contraintes réglementaires, les principes généraux de sécurité et de respect de la vie privée s'appliquent toujours.

L'IA générative, comme ChatGPT, devra se conformer aux exigences de transparence :

- Révéler que le contenu a été généré par l'IA.
- Concevoir le modèle pour empêcher qu'il génère du contenu illégal.
- Publier des résumés des données protégées par le droit d'auteur utilisées pour l'entraînement.

Les modèles d'IA à usage général à fort impact, susceptibles de présenter un risque systémique, comme le modèle d'IA plus avancé GPT-4, devront subir des évaluations approfondies et tout incident grave devra être signalé à la Commission européenne.

Ce que les entreprises doivent faire

✅ Recenser tous les use-cases de l'entreprise contenant de l'IA. Attention ! les systèmes ayant inclus récemment des fonctionnalités et si celle-ci sont utilisées, doivent être répertoriés.

✅ Conduire des audits d'impacts pour classifier les risques des IA selon l'échelle d'IA.

✅ A l'image du GDPR, prévoir de signaler auprès des clients et des employés quand l'IA est utilisée.

✅ Revoir les processus d'achats en incluant des clauses IA (procurement & legal) et mettre en place une gouvernance IA pour collecter et monitorer les use cases.

✅ Commencer à réfléchir à une stratégie IT de tests et déploiements des IA qui sont entrainés.

✅ S'assurer que les données critiques et plus largement que les données de l'entreprises sont managées, gouvernées et sécurisées pour une utilisation responsable des données.

CHAPTER IV

Explorer les limites de la créativité avec le potentiel de l'IA Générative d'images

L'IA Générative d'Image permet de créer des visuels en tout genre ouvrant des perpectives autant pour les entreprises que pour l'art digital. Dans ce chapitre nous explorerons son fonctionnement, son application en entreprise, et les implications juridiques qui l'entourent.

Let's rewind

L'histoire des outils de génération d'images remonte aux débuts des réseaux neuronaux, en particulier avec les Réseaux Neuronaux Convolutifs (CNN) dans les années 80 et 90 (Yann LeCun). Ces réseaux ont marqué une avancée significative dans la capacité des machines à extraire des caractéristiques complexes à partir d'images.

Cependant, c'est en 2014 que la révolution a véritablement commencé avec l'introduction des Réseaux Génératifs Antagonistes (GAN)(Ian Goodfellow). Les GAN ont transformé le paysage de la génération d'images en instaurant un jeu adversatif entre un générateur et un discriminateur.

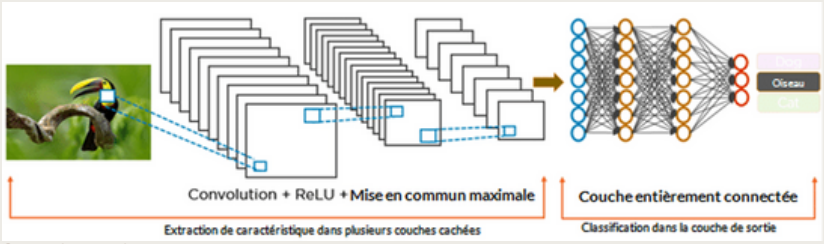
Le modèle générateur a pour but de capturer la distribution des données d'entrée et générer de nouveaux échantillons appartenant au même domaine, alors que le discriminateur a pour but d'estimer les probabilités que des images soient issues de l'ensemble d'entraînement ou bien qu'elles aient été générées par le générateur.

Cette approche a permis de créer des œuvres d'art, de générer des visages de personnes fictives avec une grande réalisme, et d'améliorer la résolution d'images.

Aujourd'hui l'IA générative d'images utilise des algorithmes pour créer des visuels à partir de descriptions textuelles ou de données d'entrée spécifiques. Des outils comme DALL-E et Midjourney illustrent la capacité de ces technologies à transformer des mots en images éloquentes, offrant ainsi un potentiel créatif quasi illimité.

Comment ça marche?

Les modèles texte-image comportent deux composants clés : un réseau neuronal formé pour associer une image avec un texte qui décrit cette image, et un autre réseau formé pour générer des images à partir de zéro.



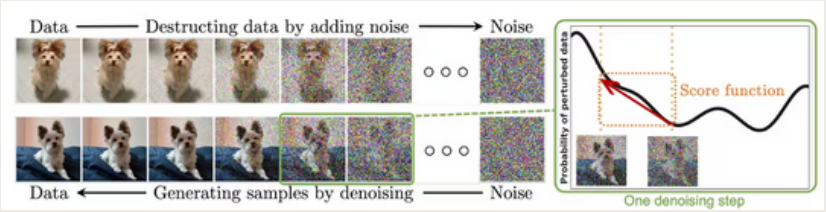
Source Image - Internet

L'idée de base est d'amener le deuxième réseau neuronal à générer une image pour que le premier réseau neuronal accepte comme correspondance avec l'invite.

La grande avancée derrière les nouveaux modèles réside dans la manière dont les images sont générées. La première version de DALL-E utilisait une extension de la technologie derrière le modèle de langage GPT-3 d'OpenAI, produisant des images en prédisant le pixel suivant d'une image comme s'il s'agissait de mots dans une phrase. Cela a fonctionné, mais pas bien.

Au lieu de cela, DALL-E 2 utilise ce qu'on appelle un modèle de diffusion. Les modèles de diffusion sont une technique de génération d'images où un réseau de neurones est formé pour éliminer progressivement le bruit d'une image jusqu'à ce qu'elle devienne claire.

Le processus commence par modifier aléatoirement les pixels d'une image initiale, étape par étape, jusqu'à ce qu'elle se transforme en un ensemble de pixels désordonnés, ressemblant à de la « neige » sur un écran de télévision sans signal. Le réseau neuronal est ensuite entraîné à inverser ce processus et à prédire à quoi ressemblerait la version la moins pixelisée d'une image donnée. Le résultat est que si vous donnez un désordre de pixels à un modèle de diffusion, il essaiera de générer quelque chose d'un peu plus propre. Rebranchez l'image nettoyée et le modèle produira quelque chose de plus propre encore. Faites cela suffisamment de fois et le modèle peut vous emmener de la neige télévisée à une image haute résolution.



Source Image - Internet

Midjourney, d'autre part, a été développé par un laboratoire de recherche indépendant et a rapidement évolué, atteignant la version 5.2. Contrairement à DALL-E, Midjourney utilise un agglomérat de contenus web pour sa formation, exploitant des ensembles de données ouvertes de diverses sources en ligne sans demander de consentement explicite aux artistes ou auteurs des images.

Son processus de génération d'images commence également par un champ de bruit visuel qu'il inverse progressivement pour révéler une image nette et affinée, ce qui est légèrement différent de la méthode de DALL-E 2.

Vous pouvez-ici comparer les images des différents modèles avec le même prompt. Lequel Préférez-vous ?

Test avec le prompt "A realistic photo of a young happy couple"

DALL E

DALL E 2

DALL E 3

MIDJOURNEY

Et les entreprises dans ce game ?

Les entreprises de divers secteurs exploitent aujourd'hui l'IA générative d'images pour leur communication interne, blog, posts de réseaux sociaux, prototypage etc.

Cependant, ces outils ne sont pas sans limites. L'IA générative ne gère pas toujours bien les concepts abstraits et peut s'appuyer sur des données d'entraînement génériques, ce qui pose des défis en termes de personnalisation et d'originalité.

👇 Le cas Adobe

En réponse à Dall-E and Midjourney, Adobe a développé Firefly pour produire des images, vidéos, retouches etc.

Adobe a décidé de s'appuyer exclusivement sur les visuels présents sur Adobe Stock ainsi que sur d'autres contenus 100 % libres de droit, afin que ses utilisateurs puissent utiliser les images générées à des fins commerciales, sans craindre un quelconque problème de propriété intellectuelle. Enfin, la puissance de Firefly est que c'est directement intégré dans des applications comme Photoshop, Illustrator et Adobe Express, offrant des fonctionnalités telles que le remplissage génératif et l'expansion d'images, ainsi que la génération de graphiques vectoriels personnalisables.

CHAPTER IV

La question de la protection de l'image, des marques et créateurs

La réglementation de l'utilisation des outils d'IA de génération d'image dans l'Union Européenne sera principalement encadrée par le règlement sur l'intelligence artificielle (AI Act) approuvé le 2 février 2024.

Commençons par parler de la question des droit d'auteur, deux questions se pose, à qui appartient l'œuvre créée? et peut-on se servir d'une œuvre protégée lors de l'entrainement des modèles?

Le domaine juridique s'adapte pour répondre aux défis posés par l'IA générative d'images. Les lois actuelles sur le droit d'auteur ne reconnaissent pas les œuvres créées sans intervention humaine significative comme éligibles à la protection du droit d'auteur.

⚖️ **En droit français, seule une personne physique peut avoir la qualité d'auteur.**

Une machine ou un logiciel ne peut donc pas se voir reconnaître comme l'auteur d'une œuvre au sens du Code de la propriété intellectuelle. (La situation est différente au Royaume-Uni par exemple.) Ainsi, dans la mesure où l'utilisateur n'a pas la maîtrise des résultats qui seront générés par l'IA, il est difficile de considérer que ceux-ci sont empreints de sa personnalité. En revanche, si le résultat généré par IA est retravaillé par l'utilisateur, celui-ci peut être protégé par un droit d'auteur s'il répond à l'exigence d'originalité du Code de la propriété intellectuelle.

⚖️ **Pour établir une violation des droits d'auteur afférents à une œuvre qui aurait contribué à nourrir une IA, il faut démontrer que le résultat généré grâce à l'IA reproduit en intégralité ou au moins en partie les caractéristiques originales de cette œuvre. Or, cela ne sera pas le cas dans la majeure partie des cas.**

Pour répondre a cette problématique L'IA Act prévoit notamment d'imposer aux plateformes d'IA spécifiquement destinées à générer du contenu tel que du texte complexe, des images, du son et des vidéos, à communiquer au public un « résumé suffisamment détaillé » de l'utilisation des données d'entraînement protégées par le droit d'auteur. Cette disposition (article 28b, 4. c) du projet) semble imposer aux IA de citer leurs sources, reste à voir si cela est faisable en pratique.

Pour une entreprise qui utilise sa propre IA, le premier réflexe à adopter est de sélectionner les données/images qui sont utilisées pour nourrir l'IA et s'assurer que l'auteur de celles-ci ne s'est pas opposé à leur reproduction/extraction aux fins de la fouille des textes et de données. Pour une entreprise qui utilise des IA développées par des tiers comme OpenAI, il est important de prendre connaissance des conditions générales d'utilisation de ces plateformes, et notamment des dispositions relatives à la titularité des droits de propriété intellectuelle sur les contenus générés et sur la responsabilité de la plateforme, ainsi que les éventuelles

licences octroyées à la plateforme pour la réutilisation des œuvres générées pour nourrir l'IA au bénéfice d'autres utilisateurs.
Le troisième point où la loi prévoit des restrictions, est par rapport aux faux contenus et à l'utilisation d'image de personne publique.

En France, certains mécanismes juridiques peuvent déjà être utilisés par un individu souhaitant empêcher et réprimer le détournement de son image sans son autorisation. Ainsi en est-il, sur le plan civil, de l'article 9 du Code civil, qui consacre le droit à la vie privée. Sur ce fondement les personnes apparaissant sur des Deepfakes pourraient invoquer une atteinte à leur image.
Sur le plan pénal, la reprise des attributs de la personnalité d'une personne sans son consentement est susceptible**, sous certaines conditions, de constituer l'infraction d'atteinte** volontaire à la vie privée d'autrui au sens de l'article 226-1 du Code pénal.
Toutefois, dans l'objectif de créer un environnement numérique plus respectueux des droits et de la dignité de chaque individu et de lutter contre le phénomène de désinformation, l'état souhaite renforcer les mécanismes existants, pour y inclure expressément les Deepfakes (l'affaire Taylor Swift est le cas plus récent à ampleur global). Un projet de loi visant à sécuriser et réguler l'espace numérique (projet de loi SREN)

a introduit des amendements spécifiques pour lutter contre ces Deepfakes. Ce projet de loi a été adopté par les députés le 17 octobre 2023 lors de sa première lecture.

- Le premier amendement punit le fait de diffuser un contenu sur une personne généré par une intelligence artificielle (IA) sans son consentement, et sans mentionner qu'il s'agit d'un faux.
- Le second amendement créé un nouveau délit de publication d'hypertrucage (deepfake) à caractère sexuel, c'est-à-dire la diffusion de vidéos pornographiques créées par l'IA et représentant une personne sans son consentement.

Le 22 février 2024, Google à retiré de Gemini sa fonctionnalité de génération d'images, après d'importantes imprécisions, biais et problèmes d'inclusivité. La route est longue pour trouver un équilibre entre créativité, droit d'auteurs et droit à l'image. Ces dans de telles problématiques qu'être humain avec un sens éthique, à une valeur très importante dans le business.

CHAPTER V

Garder les IA et LLM sous contrôle avec les guardrails

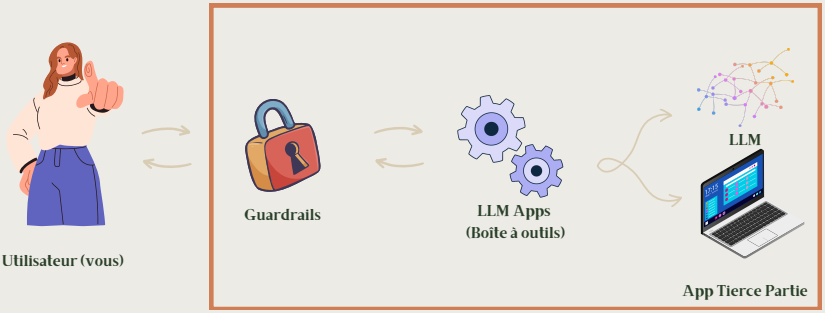
Les récents événements concernant le retrait temporaire de Gemini AI de Google ont mis en lumière l'importance de mettre en place des Guardrails pour éviter les résultats offensants et inacceptables.

🛡️ **Qu'est-ce que les LLM Guardrails ?**
Les LLM tels que ChatGPT d'OpenAI, Llama 2 de Meta et Bard de Google sont dotés de Guardrails (garde-fous) qui limitent leur utilisation et évitent qu'ils ne fournissent des informations dangereuses, tiennent des propos racistes, sexistes ou amplifient la désinformation. Ces Guardrails sont essentiels pour assurer la sécurité et l'intégrité des interactions entre les utilisateurs et les LLM.
Les LLM Guardrails sont des contrôles de sécurité qui surveillent et dictent les interactions entre les utilisateurs et les applications LLM. Ils sont conçus pour assurer le bon fonctionnement de l'IA dans un cadre défini et pour empêcher les utilisateurs de manipuler le modèle à l'aide de prompts malveillants. Les Guardrails permettent également de filtrer les réponses des modèles contenant un vocabulaire toxique ou à caractère sexuel et de corriger certaines hallucinations générées par les LLM.

🔗 **Quelles sont les différentes formes de Guardrails dans les systèmes d'IA ?**
Les Guardrails peuvent prendre diverses formes, selon ses caractéristiques et son utilisation prévue, comme:
👉 de protection de la confidentialité des données,
👉 des procédures pour prévenir les résultats discriminatoires,
👉 des politiques d'audit régulier pour la conformité aux normes éthiques et juridiques,
👉 de transparence pour garantir la compréhension et l'explicabilité des décisions prises par les systèmes d'IA.
Le fonctionnement des Guardrails peut être illustré par l'architecture orientée événements du runtime de Guardrails, déployée en mode asynchrone. Cela signifie que les différentes tâches sont exécutées de manière indépendante et simultanée, plutôt que de manière séquentielle. Les interactions avec une application sont considérées comme

des événements (user_intent ou bot_intent) qui déclenchent des flux prédéfinis. En réponse à la demande d'un utilisateur, une action de type « execute generate_user_intent » est déclenchée, entraînant une recherche de vecteurs selon la méthode K des plus proches voisins sur deux bases de données à l'aide d'un algorithme de classification. Si la demande correspond à une réponse type, le système charge le flux et l'exécute.

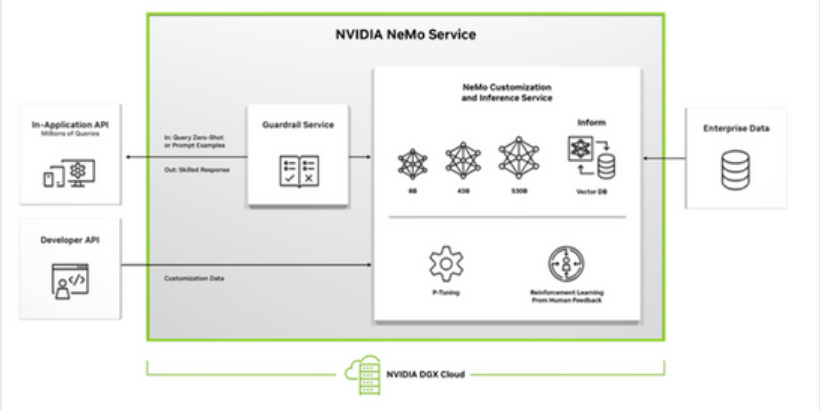
Deux possibilités se présentent alors : la demande est légitime, et la requête appropriée est envoyée au modèle sous-jacent, puis la réponse est retournée à l'utilisateur ; ou la demande est illégitime, et Guardrails déclenche le flux pour la décliner, sans envoyer la demande au modèle sous-jacent.
En somme, les Guardrails fonctionnent en traitant les interactions avec une application comme des événements, qui déclenchent des flux prédéfinis.



🩹 **NeMo Guardrails**
NeMo Guardrails est une boîte à outils open source développée par NVIDIA pour aider les entreprises à construire des systèmes conversationnels LLM sûrs et fiables. Elle s'interpose entre l'utilisateur et une application propulsée par un grand modèle de langage (LLM) et peut être facilement intégrée à d'autres collections d'outils telles que LangChain ou la plateforme d'intégration SaaS Zapier.

NeMo permet aux utilisateurs de définir des flux d'interaction autorisés et non autorisés, ainsi que la manière dont un agent conversationnel engagera une discussion. Pour ce faire, il utilise CoLang, un langage de modélisation qui fournit une interface lisible et extensible permettant aux utilisateurs de soumettre le comportement de leurs bots ou leurs applications en langage naturel à des règles.

La boîte à outils permet également de définir des règles dans trois domaines : le sujet d'une conversation, la sûreté de l'interaction et la sécurité. Par exemple, une entreprise peut utiliser NeMo Guardrails pour déterminer la manière dont la base de connaissances sera interrogée, s'assurer que les demandes ne dérivent pas du sujet principal de l'application, et que le ton des réponses est cohérent avec la politique de communication de l'entreprise.



🤖 Guardrails AI

Guardrails AI est un paquet Python open source qui fournit des cadres de gardrail pour les applications LLM. Il met en œuvre une validation des réponses LLM de type pydantic (bibliothèque Python pour effectuer de la validation de données), y compris la validation sémantique, comme la vérification du biais dans le texte généré, ou la vérification des bogues dans un morceau de code écrit par LLM.

Guardrails AI utilise la spécification RAIL (Reliable AI Markup Language) pour faire respecter des règles spécifiques sur les sorties LLM et fournir une enveloppe légère autour des appels API LLM. RAIL est un dialecte de XML qui permet de spécifier des règles et des mesures correctives.

Chaque spécification RAIL contient trois éléments principaux :

➡ **Produit:** cette composante contient des informations sur la réponse attendue de l'application d'IA. Elle doit inclure les spécifications de la structure des résultats escomptés (par exemple en JSON), le type de chaque champ en réponse, les critères de qualité de la réponse attendue et les mesures correctives à prendre si les critères de qualité ne sont pas remplis.

➡ **Prompt** : ce composant contient les instructions de pré-prompt de haut niveau qui sont envoyées à une application LLM.

➡ **Script**: ce composant optionnel peut être utilisé pour implémenter du code personnalisé pour le schéma, ce qui est particulièrement utile pour mettre en œuvre des valideurs personnalisés et des mesures correctives personnalisées.

Par exemple, une spécification RAIL peut être utilisée pour générer du code SQL sans bogues à partir d'une description en langage naturel du problème. Si les critères de sortie échouent en raison d'un bogue, le LLM ne se contente pas de réitérer l'invite, mais génère une réponse améliorée.
Guardrails AI permet également de prendre des mesures correctives et de faire respecter la structure et les garanties de type. Il peut être utilisé avec LangChain en créant un "Guardrails Output Parser".

💼 **Pourquoi les entreprises doivent-elles investir dans les Guardrails des LLM ?**

L'évolution rapide de la technologie de l'IA dépasse souvent les cadres juridiques et réglementaires existants. Les entreprises peuvent être confrontées à des incertitudes concernant les problèmes de conformité lors du déploiement de systèmes d'IA. Les garde-fou des LLM peuvent aider à atténuer ces préoccupations en fournissant des contrôles de sécurité pour surveiller et dicter les interactions des utilisateurs avec les applications LLM. Cependant, la mise en œuvre de garde-fou nécessite une surveillance, une évaluation et un ajustement constants pour s'adapter aux changements dans les lois et les réglementation, tels que l'AI Act en cours d'élaboration par l'Union Européenne.

Les Gardrail sont un élément crucial pour garantir que les systèmes d'IA fonctionnent de manière éthique et responsable. Les entreprises doivent s'adapter aux changements dans les lois et les réglementations et investir dans le développement et la mise en œuvre de Guardrails pour garantir que leurs systèmes d'IA sont conformes aux normes éthiques et juridiques. Leurs perspectives d'avenir sont prometteuses, car ils peuvent aider à préserver la sécurité, l'équité et la transparence des systèmes d'IA tout en permettant l'innovation.

CHAPTER VI

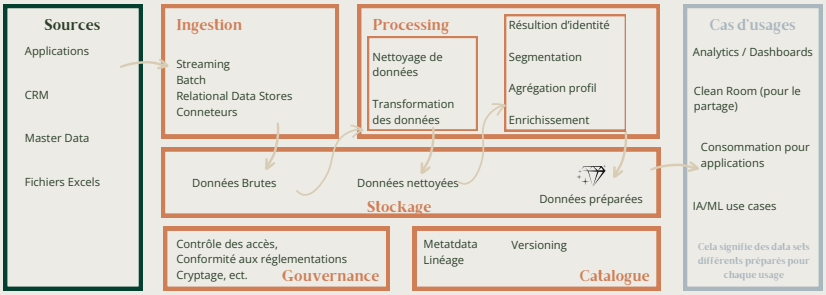
L’enjeu de l’accessibilité des données dans l’IA – Data Platforms

Il y a aujourd’hui un vrai enjeu pour avoir des données de qualité, gouvernée et sécurisée que ce soit pour les flux traditionnels de données, augmenter la capacité de reporting (automatisée) ou créer des dataset (jeux de données) d’entrainement pour des IA (machine learning, predictive analytics ou IA générative).
Dans ce 7e article, je voulais qu’on adresse ensemble le renouveau des enjeux des data platforms.

Modern Data Platform et Data Products
L’un des principaux enjeux des entreprises est de moderniser leurs plateformes actuelles tout en garantissant l’accessibilité et la qualité des données. Aujourd’hui, il y existe une multitude d’approches et différents types de plateformes des données, mais concentrons-nous sur les principes de bases.

Modern Data Platform
Les plateformes de données modernes (Modern Data Platform) sont des infrastructures centralisées qui offrent aux organisations un environnement unifié pour le stockage, le traitement et l’analyse des données. Ils englobent généralement des technologies telles que les data warehouses, les data lakes, les pipelines ETL (Extract-Transform-Load) et les outils d’anal

L’objectif principal d’une modern data platform est d’établir une approche standardisée et intégrée de la gestion des données, permettant d’obtenir des informations précises pour la prise de décision. En centralisant les données dans un référentiel unifié, les modern data platform facilitent l’accès, la collaboration et le contrôle des données.



L'équipe Data Platform est donc responsable pour collecter les besoins business, intégrer de nouvelles sources de données, mettre en place les flux ETL, publier les tables, remédier aux problèmes de qualité de données et souvent en charge de la création des rapports/dashboards.

Voici des exemples de technologies* utilisées dans les plates-formes de données modernes :

1. Data Warehouse: Oracle, Azure (Microsoft), BigQuery (Google)
2. Data Lakes: Talend, Snowflake, MongoDB
3. Outils ETL (Extract-Transform-Load): Talend, Snowflake
4. Outils d'analyse et de visualisation : Tableau, PowerBI

**Aujourd'hui, les éditeurs proposent souvent des solutions tout-en-un*

L'adoption d'une plateforme de données moderne présente certains avantages notables.

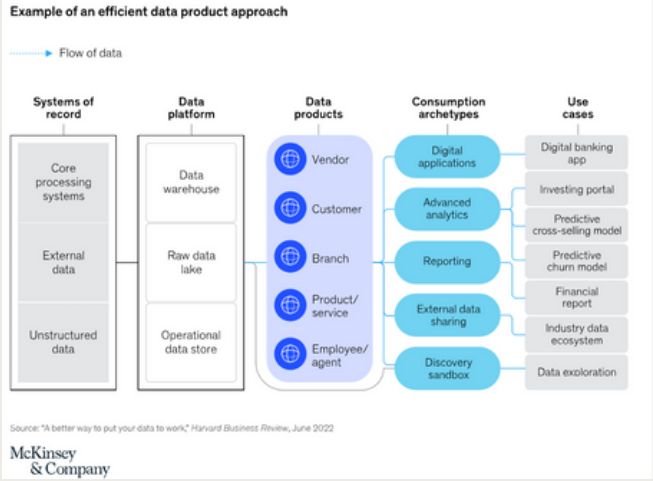
Notamment la centralisation de la gouvernance de données pour garantir la qualité et sécurité, la scalabilité en fonction de l'évolution de l'entreprise et donc une meilleure gestion des coûts.

Gérer sa donnée comme un produit

Né de Mckinsey, le concept de Data Product apparaît comme une solution pour faire face au challenge de qualité de données dans sa distribution à l'échelle. Un data product est set de données de très haute qualité et prêt à l'emploi, qui peut être utilisé dans différents contextes.

Il faut ainsi penser le data product comme une application (data set) avec des fonctionnalités et des attributs définis. Ils construisent un bloc pour les applications en aval, qui agit de manière indépendante.

Les data products, sont regroupés par data domains (voir illustration ci-dessous) facilitant la distribution, sans dupliquer et maintenir des data sets.



L'approche data product, a néanmoins quelques prérequis: comme des data owners et data stewards définis en amont par attribut, la qualité des données gérée à la source et un linéage clair des applications et flux de données.
Ainsi, on repositionne la responsabilité des données entre Métier et IT, garantissant les besoins business pour les processus, et permettant plus d'agilité dans les actions et projets (IA).
Nota bene: Dans d'autres contextes, un data product peut aussi désigner tout outil qui contient de la donnée (dashboard, excel, app etc. – l'équivalent des use-cases dans le schéma ci-dessus).

L'impact de l'IA générative sur les data platforms

Pour générer des résultats fiables, les modèles d'IA nécessitent des quantités importantes de données détectables et de haute qualité. Les data platform reviennent alors au centre des discussions: non-plus que pour de la BI, mais aussi pour la mise à dispositions des données pour les modèles IA.

Pour de nombreuses organisations qui font leurs premiers pas dans cet espace, c'est ici que les choses se compliquent. Ainsi, en pratique, l'adoption de l'IA générative (genAI) nécessite que les équipes aient 5 réflexes:

CHAPTER VI

4 actions que vous pouvez entreprendre

Aligner les stratégies de données et d'IA: Si l'IA générative est sur le point de devenir l'un des cas d'utilisation les plus exigeants pour vos données, votre stratégie de données doit évoluer pour la prendre en charge. Les stratégies doivent être alignées pour garantir que les bonnes données sont collectées, traitées et conservées de la bonne manière pour alimenter vos modèles et applications.

Automatiser autant que possible: La mise en production de modèles d'IA peut être un processus long et complexe. Le MLOps peut vous aider à automatiser les cycles de formation, de déploiement, de surveillance et de feedback pour vous aider à commencer à générer de la valeur rapidement et de manière reproductible.

Créer des processus reproductibles et auditables: Une utilisation responsable et éthique de l'IA exige une transparence totale des processus. En créant des processus reproductibles et auditables, vous pouvez rationaliser la gouvernance de l'IA, garantir que les modèles sont explicables et que vos opérations restent en conformité avec les réglementations émergentes.

Créez des plates-formes de données robustes: Chaque organisation utilisant des outils d'IA générative a besoin d'une data platform dotée des fonctionnalités appropriées pour garantir que les bonnes données de la plus haute qualité sont disponibles, aux bons endroits et au bon moment.

CHAPTER VI

Comment moderniser une data platform? 5 pistes de réflexion:

👉 **Adopter des solutions cloud:** Les data platform cloud coûtent nettement moins cher que les solutions de données on-premise, tout en offrant une durabilité nettement supérieure. Les data platform cloud sont également plus flexibles et plus évolutives et permettent un traitement des données en temps réel.

👉 **Arrêter les intégrations point à point (P2P)**: Il est souvent difficile de déplacer rapidement de gros volumes de données. Les interfaces P2P, sont souvent non-scalables et coûteuses dans le temps (en incluant sa maintenance). Dans le cadre de migration dans le cloud, les solutions iPaaS, se révèlent être game-changers (Microsoft, Mulesoft, Oracle, ou même Zapier!). Cela permettra également de rendre vos données disponibles en temps réels pour gagner en agilité globale.

👉 **Garantir la protection, la confidentialité et la gouvernance des données**: Garantir la confidentialité des données est essentiel pour bâtir et maintenir la confiance des clients.

Il est également essentiel de rester conforme aux diverses réglementations sectorielles et gouvernementales, telles que le RGPD ou le CCPA.

Alors que la sécurité des données continue de prendre de l'importance dans le débat culturel et que les réglementations continuent de proliférer, les process évolue pour intégrer des contrôles d'accès et des restrictions de politique de gouvernance plus stricts.

👉 **Traitement automatisé des données avec IA et ML**: L'intelligence artificielle et le machine learning jouent un rôle important dans la modernisation des données. Ils peuvent traiter et analyser rapidement des data lakes/warehouses entiers, puis fournir des informations et des prédictions en temps réel.

À mesure que la technologie progresse, l'IA jouera également un rôle plus important dans la remédiation de la qualité des données, en parcourant les requêtes pour trouver en quelques minutes les erreurs que les analystes humains ont dû chercher pendant des heures.

CHAPTER VII

Les 9 principes de bases pour une IA éthique

Il est l'heure d'arriver à la fin de cet e-book. Alors que nous continuons à explorer et exploiter la puissance de l'intelligence artificielle (IA), il est crucial que les créateurs, entreprises et régulateurs mettent en œuvre et promeuvent l'éthique de l'IA à chaque étape de développement.

L'établissement de principes éthiques peut aider les organisations à protéger les droits et libertés individuels, tout en augmentant le bien-être et le bien commun. Les organisations peuvent appliquer ces principes et les traduire en normes et pratiques, qui peuvent ensuite être gouvernées.
Ces principes IA sont définis par les entreprises pour montrer leur engagement respectueux et en faveur des droits humains, auprès de leurs clients et employés.

Quels sont les 9 principes les plus courants ?

🔎**Bien-être social**: Le développement d'IA doit permettre d'améliorer la société, tout en respectant les droits fondamentaux humains.

🔎 **Surveillance Humaine**: S'assurer que l'incorporation des modèles IA sont éthiques et conforme aux valeurs de l'entreprise, tout en s'assurant de la surveillance humaine lors des prises de décisions IA.

🔎 **Juste et Inclusif**: S'assurer que les données utilisées pour les systèmes IA sont juste et non-biaisés.

🔎 **Transparente**: Les utilisateurs doivent être en mesure de comprendre de manière simple les entrants et les résultats du modèle. Les modèles et données doivent être documentées pour permettre de comprendre toutes les opérations, signification des résultats et limites des modèles.

🔎 **Fiabilité**: Développer ou déployer des systèmes IA précis qui garantissent que les résultats sont fiables et dignes de confiances.

🔎 **Respect de la propriété intellectuelle**: L'utilisation des modèles d'IA doivent s'assurer du respect de la propriété intellectuelle afin de protéger les œuvres et les artistes.

CHAPTER VII

🔎 **Confidentialité & Gouvernance des données**: Donner la possibilité de notification et de consentement, encourager des garanties de confidentialité des données personnelles et assurer une transparence et un contrôle appropriés sur l'utilisation des données.

🔎 **Robustesse technique et sécurisé**: Protéger les systèmes IA d'attaques cyber et physiques

🔎**Durable**: Les modèles d'IA générative consomment des ressources exponentielles en électricité et laissent une emprunte carbone de plus en plus forte. L'usages des modèles IA dit larges, doivent être maîtrisés et l'entreprise doit mettre en place des actions plus fortes pour sauvegarder notre planète.

Pour aller plus loin, je vous invite à lire les principes IA des entreprises suivantes:

- Google
- Microsoft
- Salesforce
- LinkedIn

Que pensez-vous de ces principes ? A mon sens ils doivent aussi être considérés pour les petites entreprises et les entrepreneurs. Chez BYMADA, même si je suis très tech et data, je fais au mieux pour utiliser la créativité humaine et former des plus jeunes.

Après réflexion ces dernières semaines, finalement les plus impactés sont les nouvelles générations qui iront vers des postes qui vont être complètement chamboulés et ne sont pas formé pour.

CHAPTER VIII

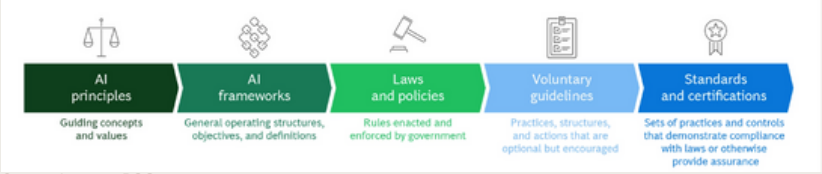
La Gouvernance IA – Une obligation ou nécessité ?

Nous sommes arrivés au terme de cet e-book sur l'IA et l'IA générative, qui conclu d'une certaine manière cette première introduction. 🎉

La Gouvernance IA, pour ne pas être un doublon d'une gouvernance de données ou d'une gouvernance IT a pour objectif de réduire la frontière entre la technologie, les données et l'éthique.

C'est donc le garde fou qui va assurer que les outils et systèmes IA soient sécurisés et éthiques. La gouvernance établi le framework, règles et standards de développement, déploiement et usages pour s'assurer de la sécurité, la justice et le respect des droits humains.

Dans le scope de la gouvernance IA
La gouvernance IA doit adresser les risques comme les biais, la privacy (respect des données personnelles) et mauvais usages tout en promouvant l'innovation et l'acculturation dans l'entreprise. Pour y arriver, une approche structurée pour remédier aux risques est indispensable.
Cette approche doit aussi prévoir la surveillance, l'évaluation et les mises à jour des modèles, pour empêcher que des décisions, impactant négativement des personnes, soient prises. Les risques sont adressés à travers:
👉 une forte politique IA,
👉 les régulations (EU AI Act),
👉 la gouvernance de données,
👉 la maintenance des modèles,
👉 des data sets de haute qualité .



Source Image - BCG

Structure de la gouvernance

La Gouvernance IA a une particularité d'être multidisciplinaires dans son approche. Elle fait intervenir:

✓ La Gouvernance de données: pour assurer l'ownership et la qualité des données

✓ L'IT: pour assurer l'architecture, l'intégration des systèmes ainsi que ses évolutions

✓ Le Juridique : pour la conformité aux règlementations (ainsi que la veille)

✓ Les Achats (ou procurement): pour assurer que les partenaires / solutions externes répondent aux principes de l'entreprise

✓ Le Risque: pour s'assurer que les nouveau projets IA soient avec un risque connu et maîtrisé

✓ Les HR: pour s'assurer que les emplois ne sont pas en périls et que les employés bénéficient du bon niveau d'accompagnement

Ensemble, ils construisent les best-practices et guidelines de l'entreprise concernant l'IA et ses usages.

Certaines grandes entreprises mettent également en place:

- des "Ethics Boards" pour évaluer les projets à risques, s'ils sont contraire ou non aux principes de l'entreprises.
- Des "AI COE": Centre d'Expertise IA qui va accompagner des projets IA avec des compétences expertes (intégration des modèles, entraînement des modèles, recherches de biais dans les data sets, veille innovation, acculturation etc.).

Pour répondre à la question initiale la gouvernance est plus une nécessité qu'une obligation.

Aujourd'hui, il n'y a pas de standards réellement définis car les entreprises sont en "mode découverte", tout en essayant de déterminer les bénéfices qu'ils peuvent réellement avoir. L'IA, sa gouvernance et ses technologies, est un des plus grands terrains d'expérimentation du XXIe siècle.

CONCLUSION

C’est le début d’une aventure incroyable

Il y a encore beaucoup de choses à écrire et à découvrir et ce fut un vrai challenge de choisir les articles pour cet e-book, qui nous espérons vous aura plu! Pour sortir son épingle du jeu et s’assurer de ne pas être avalé par ce rouleau compresseur, il faut impérativement se former.

Si vous souhaitez en savoir plus sur l’IA Responsable, la Gouvernance IA avec des modèles associés, contactez-nous!

Nous avons un module en ligne exclusif d’une durée de 1h pour apprendre ce que vous ne trouverez nulle part ailleurs sur le net - pour un montant de 49e seulement sur le site https://bymada.fr 🔥

A très bientôt pour de nouvelles aventures,
Laura et Thomas de BYMADA.

Parlons de l'IA Générative pour les curieux de la Data

BYMADA