



Par Rim AMEUR,
Architecte d'Entreprise, Partenor Digital

Vers une adoption réussie de l'IA Générative

Une approche pragmatique par les niveaux de maturité

Il est évident que l'IA générative est en train de redéfinir l'entreprise sur tous les plans. Son adoption stratégique peut conduire à un avantage concurrentiel significatif, marquant une ère nouvelle où la créativité et l'intelligence artificielle fusionnent pour débloquer des opportunités sans précédent.

Cependant, malgré ces avantages potentiels, de nombreuses entreprises hésitent à adopter cette technologie en raison des investissements requis et des risques associés, tels que la confidentialité des données et le manque de précision et la pertinence des résultats obtenus [Fatimazahra Moraux, Déc 13, 2023]¹. Dans cette optique, les architectes d'entreprise jouent un rôle essentiel en assurant l'alignement entre les objectifs stratégiques de l'entreprise et les capacités offertes par l'IA générative. Leur responsabilité consiste à établir les bases nécessaires pour intégrer cette technologie de manière sécurisée et éthique, tout en maximisant son potentiel de création de valeur.

Ainsi, nous proposons une approche pragmatique de l'intégration de l'IA Générative dans les entreprises en mettant en lumière les stratégies pour appréhender et désagréger sa complexité, favorisant ainsi une mise en œuvre réussie. Cette approche repose sur une analyse des niveaux de maturité de l'entreprise et les patterns d'architecture en matière d'IA Générative.

Définition des niveaux de maturité d'une entreprise en IA générative

Pour définir les niveaux de maturité d'une entreprise, nous allons établir un pattern d'architecture «type», comprenant les différentes fonctionnalités requises pour une solution d'IA générative.

Pattern d'architecture de l'IA Générative

Ce pattern d'architecture (voir figure 1) comporte une interface utilisateur qui recueille les inputs via des prompts. Ces inputs sont ensuite traités par un module central qui orchestre la réponse à travers diverses composantes de l'IA.

Les LLM pré-entraînés servent de base pour générer du contenu, qui peut être affiné par fine-tuning pour mieux correspondre aux besoins spécifiques de l'entreprise.

¹ [Fatimazahra MORAUX, « Les Décideurs » ; Déc 13, 2023] : [L'IA Générative](#)

En parallèle, la personnalisation et l'intégration du contexte sont réalisées grâce au RAG («Retrieval Augmented Generation»), en utilisant des données d'entreprise spécifiquement préparées et vectorisées.

Tout ceci est piloté par un composant de monitoring qui veille à la sécurité, à l'alignement éthique et à la réduction des biais potentiels dans les réponses de l'IA, réduisant ainsi les risques liés à l'IA générative.

Cela conduit à la création d'applications personnalisées.

Les capacités principales de l'IA Générative

Cette architecture permet d'identifier les principales capacités de l'IA générative au sein d'une entreprise. Une capacité fait référence à la fonctionnalité mais aussi aux compétences, ressources et aptitudes technologiques de l'entreprise en ce qui concerne l'intégration et l'utilisation de l'IA générative.

Ces capacités serviront à définir les différents niveaux de maturité de l'entreprise en IA générative ainsi que les patterns d'architecture correspondants, en fonction de l'activation ou non de ces capacités (voir figure 2). On obtiendra autant de patterns d'architecture que de niveaux de maturité.

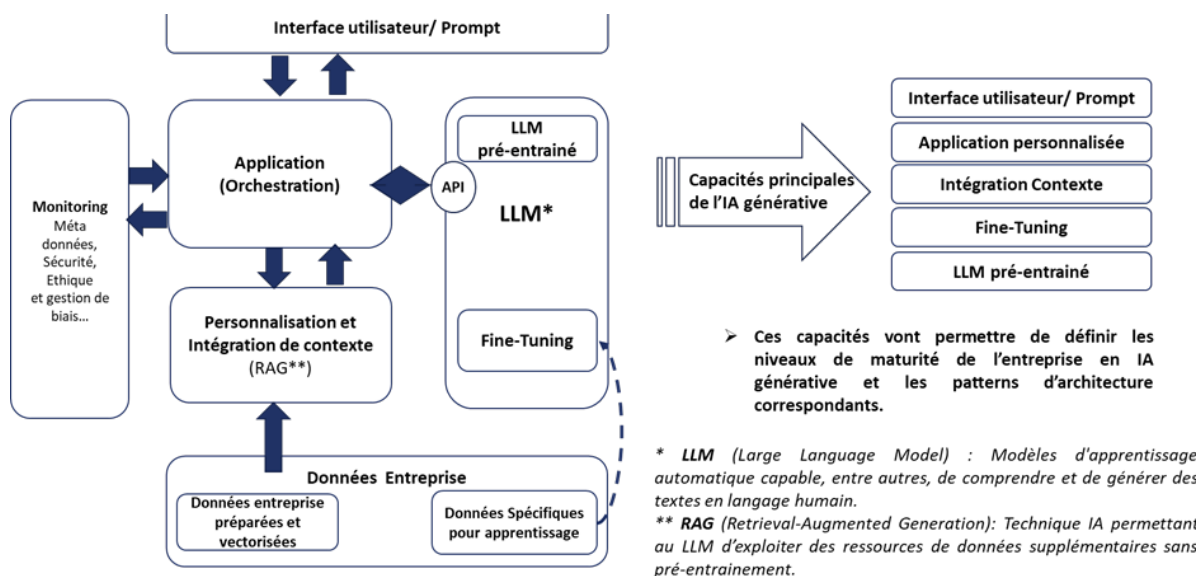


Figure 1. IA générative : Pattern d'Architecture

Les niveaux de Maturité d'une entreprise en IA générative

Sur la base des éléments ci-dessus, nous identifions les cinq (5) niveaux de maturité suivants :

- 1. Au niveau exploratoire**, l'entreprise se familiarise avec les modèles de langage pré-entraînés disponibles publiquement, tels que ChatGPT, Gemini, et plus récemment Copilot en les utilisant pour des applications simples basées sur des prompts. Cette phase permet à l'entreprise de comprendre les fonctionnalités et les limitations de ces modèles. Dans ce contexte, on peut se faire accompagner par un « Prompt Engineer » qui peut aider à formuler des requêtes précises et efficaces pour obtenir les résultats souhaités, tout en optimisant les performances et en minimisant les biais.
- 2. En passant au niveau débutant**, l'entreprise commence à développer des applications spécifiques en utilisant les API des modèles de langage publics. Cela peut inclure la création de chatbots, d'outils d'analyse de texte adaptés aux besoins de l'entreprise. A ce stade, l'accent est mis sur la personnalisation des solutions en fonction des exigences spécifiques de l'entreprise. On peut appliquer des filtres automatiques en entrée et en sortie pour une meilleure conformité aux réglementations et une meilleure maîtrise des risques.

Par ailleurs, on peut choisir de passer par un cloud privé comme Azure. Ceci offre une gamme de services pour le développement et le déploiement d'applications d'IA, offrant une infrastructure sécurisée et évolutive pour soutenir ces projets [Jérôme VATIN, 2024].

3. Au niveau d'intégration de contexte, les données internes de l'entreprise sont intégrées dans le processus d'IA générative pour contextualiser les réponses générées par les modèles. Cela permet aux modèles de mieux comprendre les requêtes des utilisateurs et de fournir des réponses plus précises et pertinentes via la technologie RAG. L'intégration de données d'entreprise peut nécessiter des efforts supplémentaires en termes d'ingénierie de vectorisation de données, et de pipelines d'intégration. L'usage de Lang Chain peut faciliter la tâche.

4. Lorsque l'entreprise atteint le niveau avancé, elle procède au fine-tuning des modèles pré-entraînés du LLM, avec des données spécifiques de l'entreprise pour obtenir des réponses encore plus personnalisées et adaptées à ses besoins. Ce fine-tuning implique l'ajustement des paramètres du modèle en utilisant des données internes (Les données en entrée et en réponse sont utilisées pour ajuster le modèle), ce qui nécessite une expertise en apprentissage automatique et en traitement du langage naturel (NLP). Cela peut impliquer l'utilisation de plateformes cloud dédiées telles que PyTorch et TensorFlow.

5. Enfin, au niveau expert, l'entreprise investit dans la construction de nouveaux modèles LLM personnalisés pour répondre à des besoins très spécifiques et complexes. Cela peut impliquer la conception et l'entraînement de modèles sur mesure pour traiter des tâches ou des domaines d'expertise uniques de l'entreprise. Ce niveau nécessite une expertise approfondie en IA et en NLP, ainsi que des ressources importantes en termes de données et de calcul. On peut utiliser des LLM open-source On-Premise, disponibles sur la plateforme « Hugging face ».

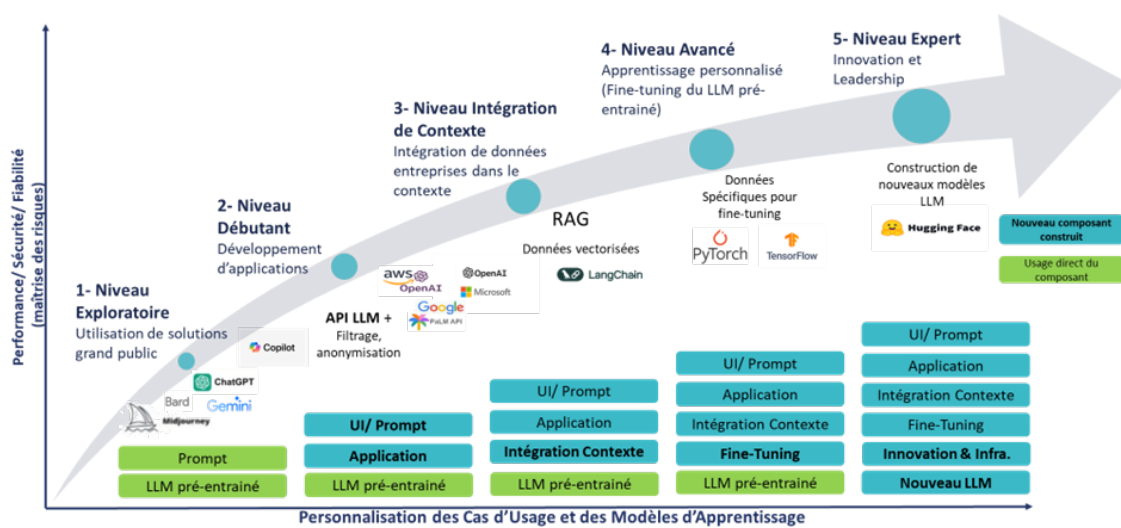


Figure 2. Démarche d'intégration basée sur les niveaux de maturité de l'entreprise en IA Générative et les patterns d'architecture correspondants

Démarche d'adoption basée sur les niveaux de maturité

Cartographie des cas d'usage

La première étape consiste à cartographier les chaînes de valeur de l'entreprise et à identifier des cas d'usage réels et effectifs de l'IA générative, répartis par fonction. Une identification précise des cas d'usage est cruciale pour déterminer où l'IA générative peut apporter une valeur ajoutée significative. Pour illustrer ces propos, nous nous appuyons sur le modèle étendu de la chaîne de valeur de Porter, afin de fournir des exemples de cas d'usage répartis par fonction du modèle (voir figure 3). Ces cas d'usage peuvent varier considérablement d'une entreprise à l'autre en fonction de leurs besoins et objectifs. Par exemple, une entreprise axée sur le service client peut utiliser l'IA générative pour automatiser les réponses aux questions fréquentes (FAQ) et développer une démarche marketing [Clément DOUGE, 2023]², tandis que dans l'industrie pharmaceutique, elle permet aux chercheurs de créer de nouvelles molécules, accélérant ainsi la découverte de médicaments.

¹ [Jérôme VATIN, 2024] : [L'IA générative – Révolution dans le monde de l'entreprise](#)

² [Clément DOUGE, 2023] : [IA Générative x Marketing – Un mariage presque parfait ?](#)

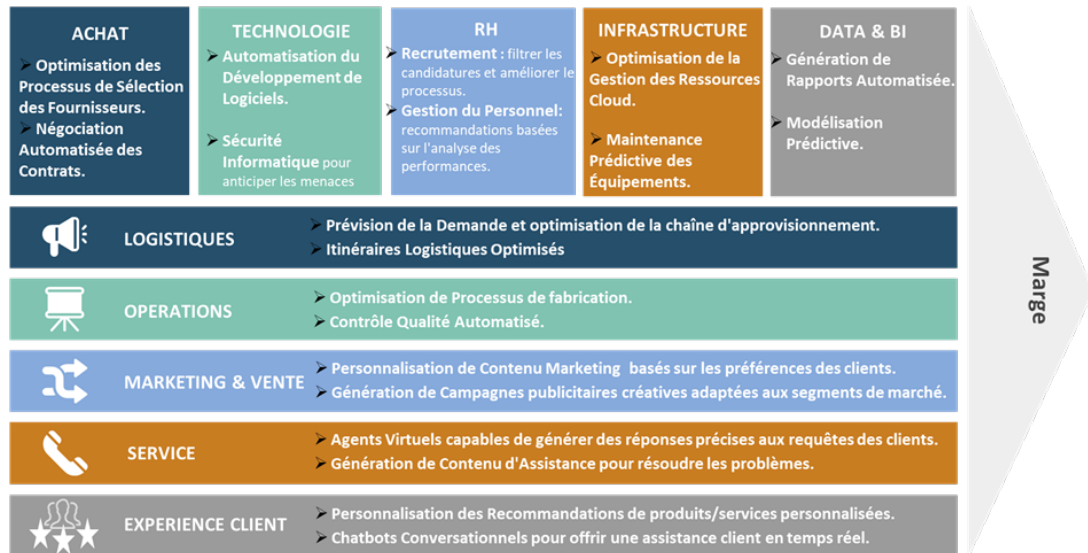


Figure 3. Cartographie des cas d'usage IA générative selon les fonctions de la chaîne de valeurs du modèle de Porter étendu d'une entreprise

Evaluation (maturité et valeur apportée) des cas d'usage

La deuxième étape implique l'évaluation des cas d'usage identifiés lors de la première étape. Cette évaluation approfondie permet de prioriser ces cas d'usage en fonction de leur valeur ajoutée, de leur alignement avec les objectifs stratégiques de l'entreprise, ainsi que de leur faisabilité de mise en œuvre et des risques associés à l'IA générative. Cette analyse de faisabilité et de gestion des risques peut être abordée en tenant compte du niveau de maturité minimal exigé de chaque cas d'usage. En outre, elle peut nécessiter des analyses plus approfondies, telles que des études de marché, des analyses de rentabilité et des évaluations des risques, pour affiner la sélection des cas d'usage les plus pertinents et les plus prometteurs.

Stratégies de Mise en Œuvre de l'IA Générative

Approche incrémentale et évolutive

- **A l'échelle de l'ensemble des cas d'usage**

Dans cette optique, l'identification et la priorisation des cas d'usage jouent un rôle crucial. Au premier stade, l'entreprise met l'accent sur des cas d'usage plus simples et basiques, qui permettent à l'entreprise de se familiariser avec les concepts de l'IA générative et d'acquérir les compétences de base nécessaires. Ces cas d'usage peuvent inclure des applications telles que la génération de contenu textuel simple ou la réponse à des requêtes basiques. À mesure que l'entreprise progresse vers des niveaux de maturité plus avancés, des cas d'usage sont sélectionnés en fonction de leur potentiel à créer de la valeur ajoutée et à contribuer aux objectifs stratégiques de l'entreprise. Des cas d'usage plus complexes et stratégiques, telles que la personnalisation avancée du contenu ou la génération de nouveaux produits, peuvent alors être priorisées.

- **A l'échelle d'un cas d'usage**

A l'échelle d'un cas d'usage, une approche incrémentale et évolutive, selon les niveaux de maturité, peut être adoptée. Cette approche permet à l'entreprise de commencer pour ce cas d'usage, avec une solution de base, puis d'évoluer vers une solution plus sophistiquée. Par exemple, initialement, une entreprise pourrait utiliser le RAG (Retrieval Augmented Generation) pour créer un système de chatbot capable de répondre aux questions basiques des clients. Le RAG permettrait au chatbot de récupérer des réponses pertinentes à partir d'une base de connaissances existante, offrant ainsi une solution efficace et rapide pour répondre aux demandes initiales. Au fur et à mesure que l'entreprise acquiert de l'expérience et que le volume de requêtes des clients augmente, elle peut décider d'améliorer la performance et la pertinence des réponses en utilisant du fine-tuning sur un modèle de langage pré-entraîné (LLM) et donc de passer du niveau de maturité 3 au niveau 4 pour le même cas d'usage (voir figure 2). Cette approche permet une adaptation agile aux défis émergents tout en optimisant les ressources et les investissements.

Approche Focalisée sur des Cas d'Usage Spécifiques

L'approche ciblée par cas d'usage de l'IA générative, consiste pour une entreprise à se concentrer sur des cas d'usage spécifiques de son activité qui correspondent à ses besoins stratégiques. Plutôt que d'adopter une approche progressive, elle développe directement des solutions sur mesure, même si elles nécessitent un niveau de maturité élevé en IA générative.

Par exemple, dans le cas d'usage du système de recommandation pour une entreprise de commerce en ligne, un niveau de maturité avancé serait nécessaire. Cela impliquerait probablement l'utilisation du fine-tuning des modèles de langage pré-entraînés (LLM) pour adapter les recommandations aux préférences spécifiques des utilisateurs et aux caractéristiques des produits. Bien que cette approche puisse nécessiter des investissements initiaux plus importants en termes de ressources et d'expertise, elle permet à l'entreprise de se démarquer de la concurrence en offrant une expérience client hautement personnalisée et différenciée.

Approche Hybride : Combiner Progression et Focalisation

Cette approche combine les deux précédentes en adoptant une approche incrémentale tout en se concentrant sur des cas d'usage spécifiques. En effet, les niveaux de maturité d'une entreprise peuvent être adressés en parallèle, adaptant l'adoption de l'IA générative aux exigences spécifiques de chaque cas d'usage. Cela signifie qu'une entreprise peut simultanément expérimenter avec des applications de cas d'usage basiques de l'IA générative dans certains domaines (comme GitHub Copilot pour améliorer la productivité des développeurs en générant automatiquement du code), tout en développant des solutions plus avancées et personnalisées là où les besoins et le potentiel de valeur ajoutée sont plus élevés, comme la personnalisation des interactions client à travers des chatbots intelligents.

Pour réussir l'adoption de l'IA générative dans les entreprises, il est essentiel de maintenir un équilibre entre la valeur attendue et la capacité nécessaire, en adoptant une approche axée sur les niveaux de maturité et les patterns d'architecture correspondants mais aussi sur les cas d'usage spécifiques. Ce qui constitue la clé pour une adoption durable de l'IA générative, permettant à l'entreprise de rester compétitive et à la pointe de l'innovation.

Ainsi, les entreprises peuvent progressivement monter en maturité, en s'appuyant sur des succès initiaux pour explorer des applications plus complexes et innovantes de l'IA générative. Cette perspective stratégique encouragerait les entreprises à élaborer un schéma directeur et une feuille de route adaptative d'intégration de l'IA générative, en choisissant la voie qui maximise la valeur tout en optimisant les investissements et les ressources. Cela nécessite également une culture organisationnelle collaborative et réactive, prête à s'adapter aux changements, ainsi qu'une gouvernance des risques flexible, adaptée aux niveaux de maturité et aux patterns d'architecture correspondants.

Par ailleurs, les entreprises doivent rester à l'affût des dernières avancées technologiques et ajuster leurs stratégies en fonction des retours d'expérience et des évolutions des outils et solutions du marché. En se référant aux niveaux de maturité, on peut ainsi dresser un panorama des solutions disponibles.