

# **Atelier sur LLM - Revolution IA 2024**

Simon-Pierre Boucher

Invalid Date

## **Table of contents**

# Preface

Bienvenue à notre atelier dédié aux Large Language Models (LLM), une exploration fascinante à l'intersection de l'intelligence artificielle, du traitement automatique du langage naturel et de l'apprentissage automatique. Dans un monde numérique en constante évolution, les LLM s'avèrent être des outils puissants, capables de comprendre, d'interpréter et de générer du langage humain de manière cohérente et contextuellement pertinente.

## Qu'est-ce qu'un Large Language Model ?

Un Large Language Model est un type de modèle d'intelligence artificielle qui a été entraîné sur d'énormes ensembles de données textuelles. Grâce à cet entraînement, ces modèles sont capables de réaliser une multitude de tâches liées au langage, telles que la traduction automatique, la génération de texte, la compréhension de questions et de réponses, et bien plus encore. Les LLM, tels que GPT (Generative Pre-trained Transformer) d'OpenAI, ont révolutionné la manière dont nous interagissons avec la technologie, ouvrant la voie à de nouvelles applications et services innovants.

## Pourquoi cet atelier est-il pertinent ?

L'essor des LLM transforme divers secteurs, allant de l'éducation à la santé, en passant par le service client et le divertissement. Comprendre le fonctionnement de ces modèles, leurs capacités, leurs limites, et surtout, leur potentiel pour résoudre des problèmes complexes, est devenu indispensable pour les professionnels, les chercheurs et les passionnés de technologie.

Cet atelier a été conçu pour vous fournir une compréhension approfondie des LLM, à travers des présentations théoriques, des études de cas, et des sessions pratiques. Vous découvrirez comment ces modèles sont construits, comment ils apprennent à partir des données et comment ils peuvent être appliqués pour créer des solutions innovantes.

## Objectifs de l'atelier

- **Comprendre les Fondements des LLM** : Acquérir une connaissance solide des principes sous-jacents des LLM, y compris leur architecture et leur processus d'entraînement.
- **Explorer les Applications des LLM** : Découvrir comment les LLM sont utilisés dans divers domaines pour améliorer l'efficacité, la créativité et la prise de décision.
- **Développer des Compétences Pratiques** : Appliquer vos connaissances à travers des ateliers pratiques, vous permettant de mettre en œuvre des LLM pour vos propres projets ou recherches.

Nous sommes à l'aube d'une nouvelle ère de l'intelligence artificielle, où les LLM jouent un rôle central. Rejoignez-nous pour cet atelier afin d'explorer le potentiel de ces technologies transformatrices et de vous préparer à contribuer à l'avenir de l'IA.

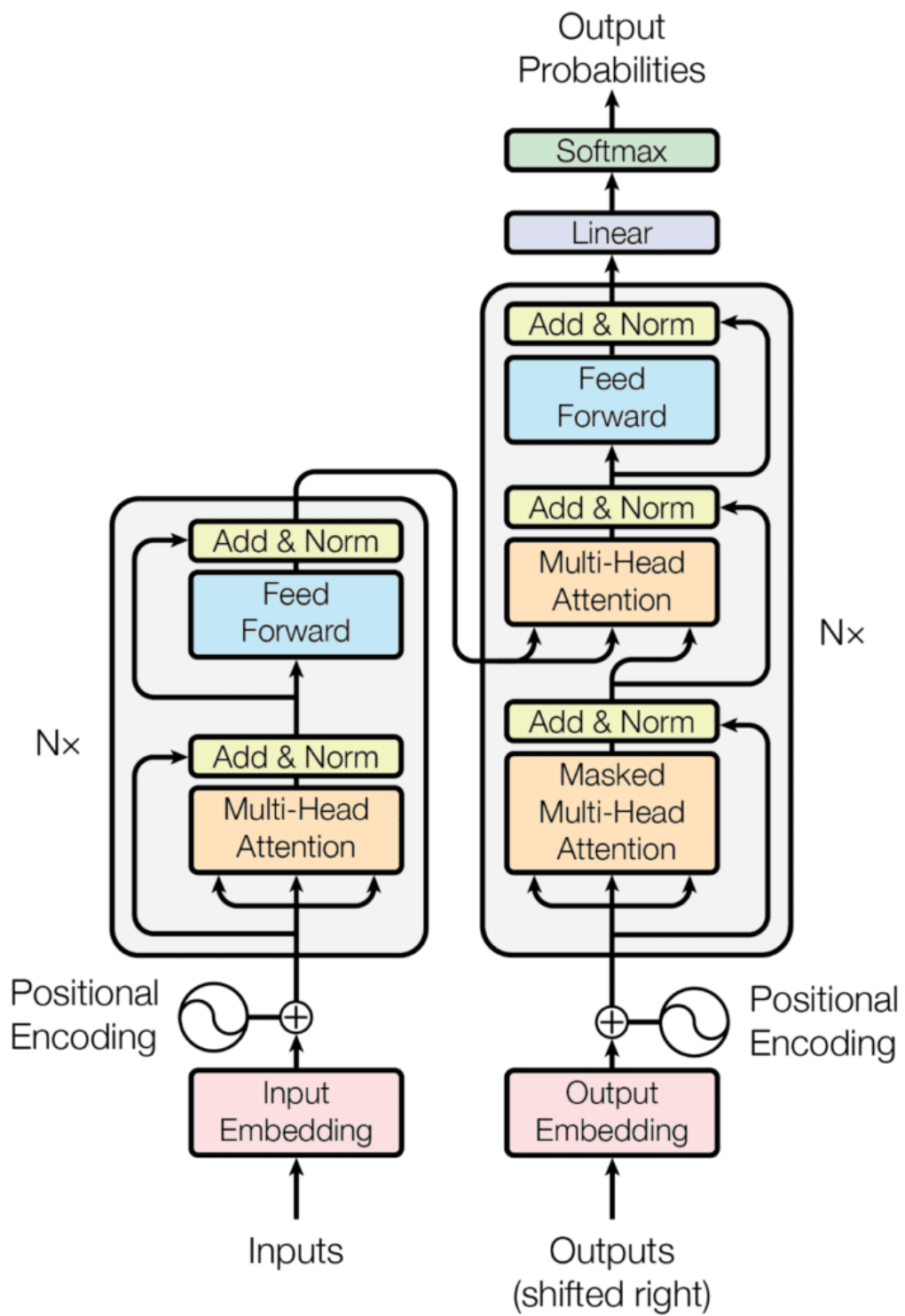
Nous avons hâte de vous accueillir et de partager avec vous cette aventure passionnante !

# 1 Les Transformers : Révolution dans le Traitement Automatique du Langage Naturel

Les Transformers ont révolutionné le champ de l'intelligence artificielle (IA) et du traitement automatique du langage naturel (TALN) depuis leur introduction en 2017 par Vaswani et al. dans le papier intitulé "Attention Is All You Need". Cette architecture novatrice a permis de faire des avancées significatives dans la compréhension et la génération du langage naturel, dépassant les performances des modèles basés sur les réseaux de neurones récurrents (RNN) et les réseaux de neurones convolutifs (CNN) pour de nombreuses tâches.

## 1.1 Qu'est-ce qu'un Transformer ?

Un Transformer est un modèle basé sur le mécanisme d'attention, conçu pour traiter séquentiellement des données avec une efficacité et une flexibilité remarquables. Contrairement aux approches précédentes, qui traitent les séquences mot par mot de manière séquentielle, le Transformer permet un parallélisme complet de traitement, ce qui réduit considérablement les temps de formation.



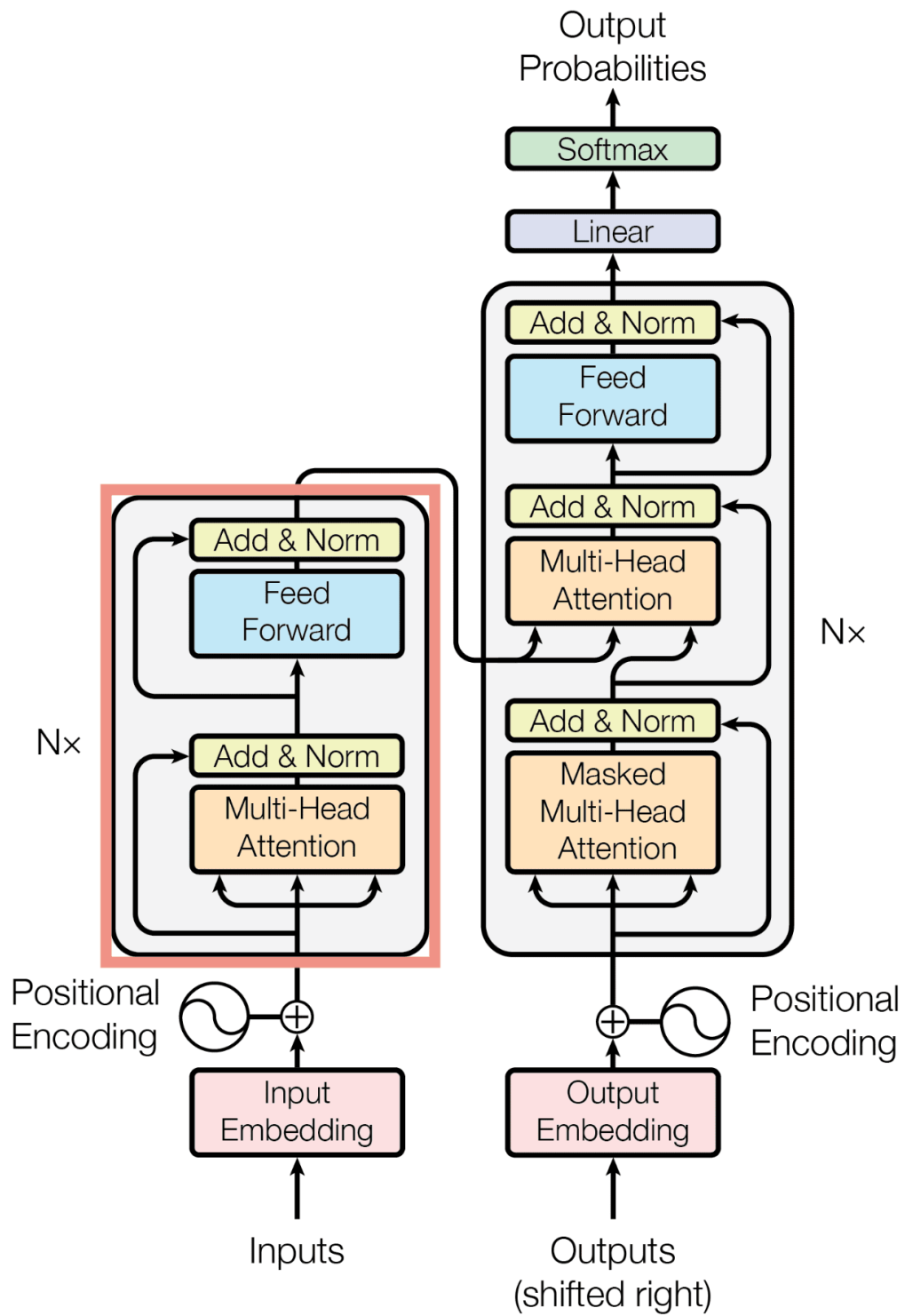
### 1.1.1 Le Mécanisme d'Attention

Le cœur du Transformer est le mécanisme d'attention, spécifiquement l'attention multi-têtes. Ce mécanisme permet au modèle de se concentrer sur différentes parties d'une séquence d'entrée lors de la prédiction d'une partie d'une séquence de sortie, améliorant ainsi sa capacité à comprendre les relations complexes et lointaines dans les données textuelles.

### 1.1.2 Architecture du Transformer

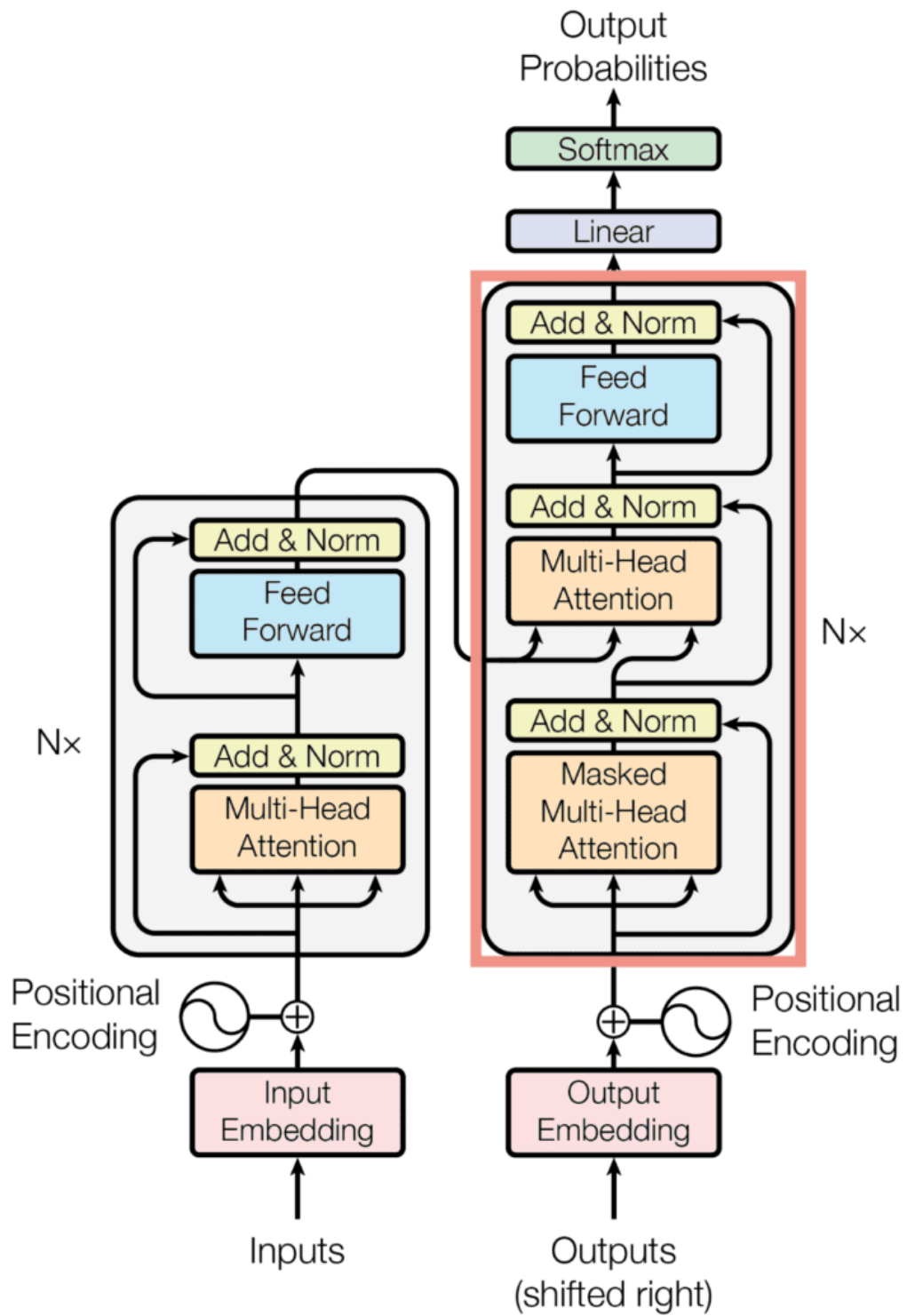
L'architecture du Transformer est constituée de deux parties principales : l'encodeur et le décodeur.

- **L'Encodeur** : Il traite la séquence d'entrée et la transforme en une série de représentations qui contiennent à la fois les informations du mot spécifique et le contexte dans lequel il apparaît. Chaque couche de l'encodeur contient deux sous-couches : une sous-couche d'attention multi-têtes et une sous-couche de réseau de neurones entièrement connecté.





- **Le Décodeur** : Il génère la séquence de sortie, mot par mot, en se basant sur les représentations fournies par l'encodeur et ce qui a déjà été généré. Le décodeur ajoute une troisième sous-couche à celles trouvées dans l'encodeur, qui permet d'appliquer l'attention sur la sortie de l'encodeur.



## 1.2 Avantages des Transformers

Les Transformers offrent plusieurs avantages significatifs par rapport aux architectures précédentes :

- **Traitement Parallèle** : La capacité à traiter l'ensemble de la séquence d'entrée en parallèle conduit à une formation plus rapide des modèles.
- **Gestion des Dépendances à Long Terme** : L'attention multi-têtes permet au modèle de se concentrer sur l'ensemble de la séquence d'entrée pour chaque mot de la séquence de sortie, gérant efficacement les dépendances à long terme.
- **Flexibilité et Adaptabilité** : Les Transformers ont été adaptés avec succès à une grande variété de tâches TALN, y compris la traduction automatique, la synthèse de texte et la compréhension de texte.

## 1.3 Impact des Transformers

L'introduction des Transformers a marqué un tournant dans le domaine de l'IA et du TALN. Des modèles comme BERT (Bidirectional Encoder Representations from Transformers), GPT (Generative Pretrained Transformer), et d'autres ont établi de nouveaux standards de performance sur divers benchmarks TALN. Ces modèles ont non seulement amélioré la qualité de la compréhension et de la génération de texte, mais ils ont également ouvert la voie à des applications innovantes dans le traitement du langage naturel, la recherche d'informations, et bien au-delà.

En somme, les Transformers représentent une avancée majeure dans notre capacité à modéliser et à comprendre le langage humain, conduisant à des progrès significatifs dans de nombreuses applications pratiques de l'IA.

## 2 Acteurs Majeurs dans le Monde de l'IA

L'univers de l'intelligence artificielle (IA) est peuplé de plusieurs acteurs majeurs, chacun contribuant de manière significative à l'avancement et à la démocratisation de la technologie. Voici quelques-uns des plus influents :

### 2.1 TensorFlow & Keras

- **TensorFlow** : Développé par l'équipe Google Brain, TensorFlow est un framework open-source pour le calcul numérique qui facilite la construction, le déploiement et la formation de modèles d'apprentissage profond. Il est particulièrement connu pour sa flexibilité et sa capacité à opérer à grande échelle, ce qui le rend populaire parmi les chercheurs et les développeurs travaillant sur des projets complexes d'IA.
- **Keras** : Intégré dans TensorFlow comme `tf.keras`, Keras est une interface de haut niveau pour les réseaux de neurones, conçue pour l'expérimentation rapide. Facile à utiliser et intuitif, Keras permet de construire et de tester des modèles d'IA avec un minimum de code, le rendant accessible aux novices tout en restant suffisamment puissant pour les chercheurs expérimentés.



## 2.2 Hugging Face

Hugging Face est une entreprise technologique qui a pris une importance considérable dans le monde de l'IA grâce à sa plateforme collaborative et sa bibliothèque de modèles de traitement automatique du langage naturel (TALN), **transformers**. Offrant un accès facile à plus d'une centaine de modèles pré-entraînés et supportant une multitude de langues, Hugging Face démocratise l'accès aux technologies d'IA de pointe. Elle est devenue incontournable pour les chercheurs, les développeurs, et les entreprises souhaitant intégrer des capacités avancées de TALN dans leurs applications.



## 2.3 Mistral

Mistral est une initiative moins connue mais tout aussi importante dans le domaine de l'IA, se concentrant sur le développement et l'optimisation de modèles d'IA pour des performances et une efficacité énergétique accrues. Bien que moins médiatisée que les autres entités mentionnées ici, Mistral joue un rôle crucial dans la recherche de solutions d'IA plus durables et accessibles, en particulier dans des contextes où les ressources de calcul sont limitées.