

Formation LLM

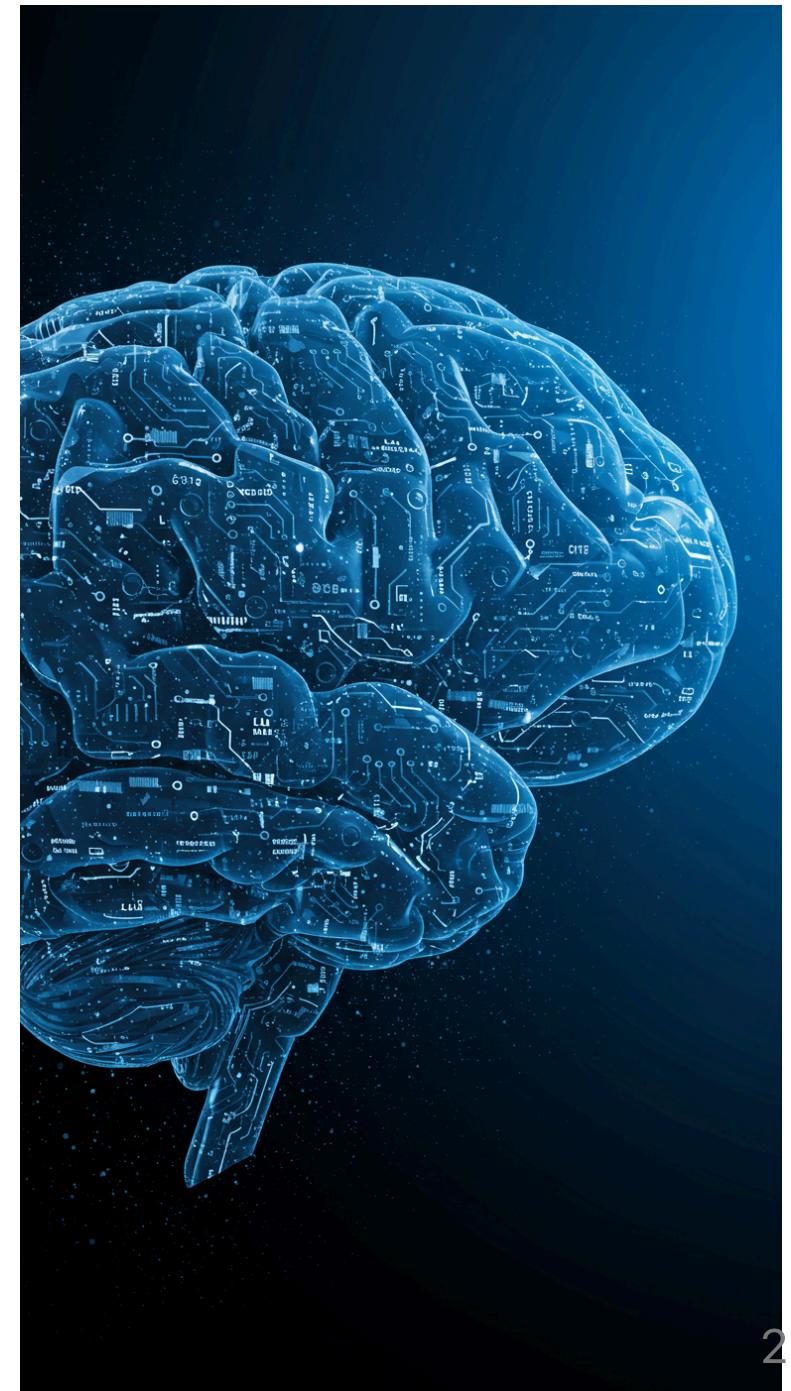
Les Modèles de Langage Expliqués

*Intelligence Artificielle & Réseaux de
Neurones*



1. Historique

L'évolution de l'IA



Étape 1 : Logique Programmée

- Bases de données structurées
- Règles "si... alors..."
- Intelligence codée à la main
- **Limite** : Rigidité

Étape 2 : Machine Learning

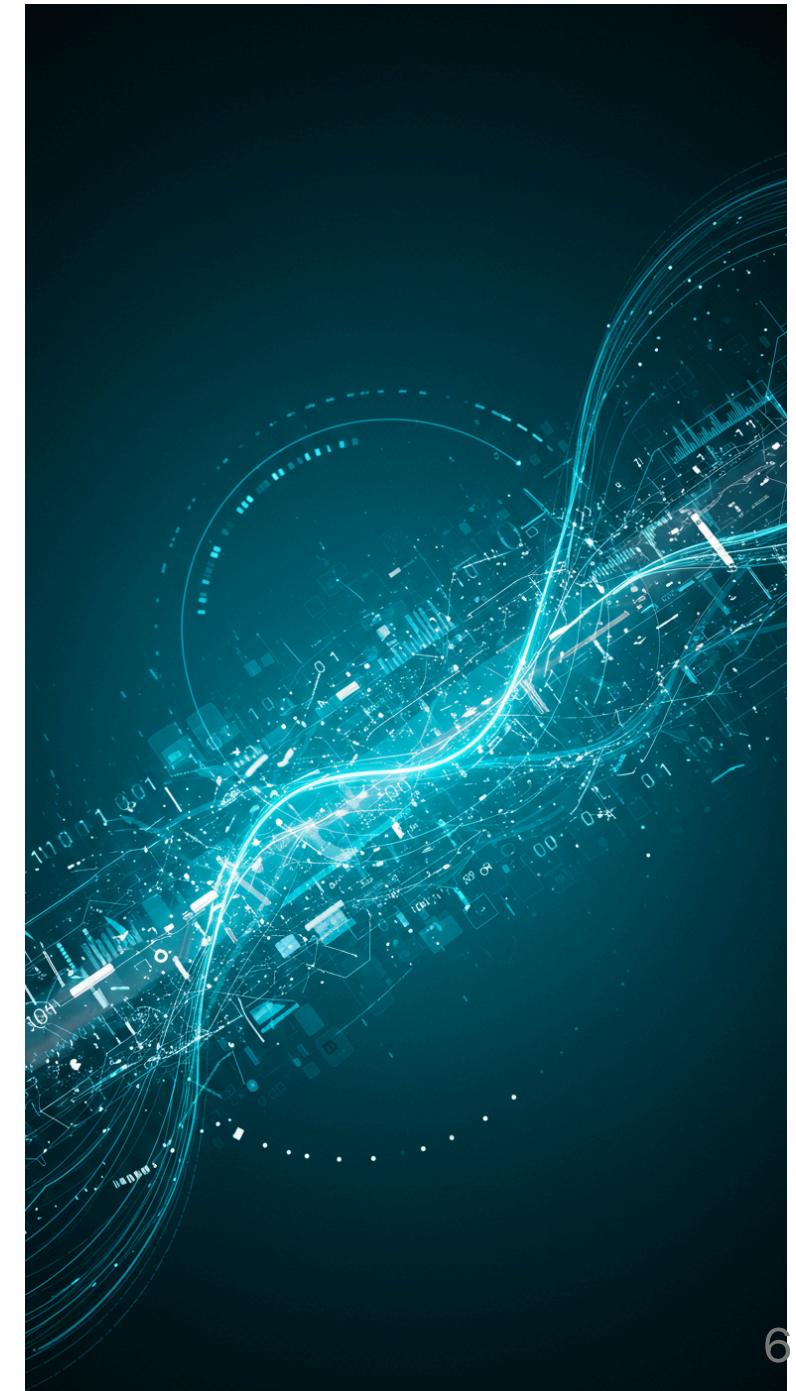
- Détection de patterns
- Apprentissage par l'exemple
- Intelligence par analogie
- **Avancée : Flexibilité**

Étape 3 : Transformers

- Données brutes massives
- Réalité numérique
- Génération de contenu
- **Révolution** : Crédit

2. Concepts Clés

Les briques fondamentales



Le Token

- Plus petite unité de texte
- "Briques de Lego" du langage
- Prédiction du suivant
- Base de la génération

Le Transformer

- Vision globale du contexte
- Mécanisme d'attention
- Liens entre tous les mots
- Compréhension profonde

L'Embedding

- Carte d'identité sémantique
- Vecteur de nombres
- Sens mathématique
- Relations conceptuelles

3. La Donnée

Le carburant des LLM



Processus de Création

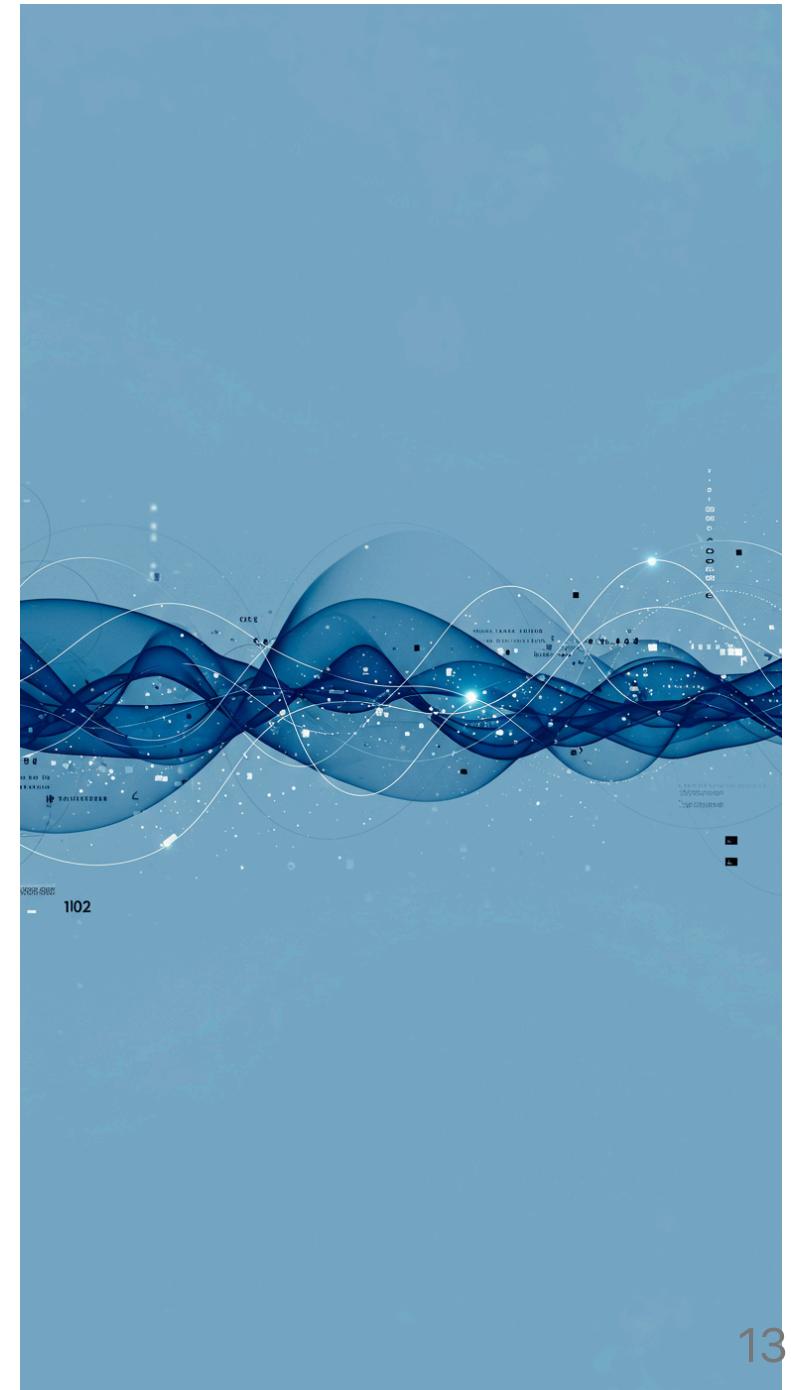
- **Collecte** : Milliards de textes
- **Nettoyage** : Filtrage qualité
- **Transformation** : Langage machine
- **Métadonnées** : Étiquetage
- **Intégration** : Compression

Qualité = Résultats

- Imitateurs statistiques
- GitHub : Code hétérogène
- Biais reproduits
- Filtrage complexe
- Amélioration continue

4. Capacités & Limites

Pouvoirs et contraintes



Limitations du LLM "Nu"

- Connaissances figées
- Pas d'internet intégré
- Pas de notion du temps
- Pas d'actions concrètes
- Hallucinations possibles

Super-Pouvoirs Ajoutés

- **RAG** : Données externes
- **Outils** : Recherche web
- **Code** : Exécution sandbox
- **APIs** : Fonctions spécialisées
- **Plugins** : Capacités infinies

Le Principe RAG

- Base vectorielle
- Recherche sémantique
- Contexte enrichi
- Réponses ancrées
- Sources citées

5. En Pratique

L'art du prompting

Règles d'Or

- **Clarté** : Soyez spécifique
- **Contexte** : Donnez le cadre
- **Rôle** : "Agis comme..."
- **Exemples** : Montrez la voie
- **Itération** : Affinez

Paramètres Clés

- **Température** : Créativité
- **Top P** : Sélection probabiliste
- **Longueur** : Limite tokens
- **Stop** : Séquences d'arrêt
- **Sécurité** : Filtres contenus

Outils Avancés

- JSON structuré
- Exécution code
- Appel fonctions
- Recherche Google
- Context URL

Questions ?

Merci pour votre attention