

Systemes Experts et Modeles de Langage (LLM)

1. Principe de Fonctionnement :

Un **systeme expert** est basé sur une base de connaissances, souvent construite manuellement, qui contient des regles et des faits. Il utilise un moteur d'inférence pour appliquer des regles logiques aux faits afin de faire des recommandations ou prendre des décisions. Ces systemes sont généralement conçus pour des domaines spécifiques (comme la médecine, l'ingénierie, la finance, etc.) et sont souvent utilisés pour des tâches de diagnostic ou de conseil.

Un **LLM**, comme GPT-4, est un modele statistique formé sur de grandes quantités de données textuelles. Il ne repose pas sur des regles explicites, mais sur des relations et des probabilités observées dans les données. Il génère du texte ou répond à des questions en se basant sur des patterns appris à partir de ces données, sans avoir une structure fixe de connaissances ou une base de regles rigide.

2. Type de Connaissances :

Les connaissances sont explicitement codées dans le systeme sous forme de regles et de faits. Un expert humain doit généralement définir et intégrer ces connaissances.

Les connaissances ne sont pas codées manuellement. Le modele apprend des données textuelles et peut générer des informations ou répondre à des questions en fonction des relations observées dans ces données, mais il ne possède pas une base de connaissances structurée de maniere explicite.

3. Flexibilité :

Il est conçu pour des domaines spécifiques et sa capacité à traiter de nouvelles informations ou situations dépend directement de la qualité et de la mise à jour de ses regles et de sa base de connaissances. Il est donc relativement rigide dans son domaine d'application.

Un LLM est beaucoup plus flexible, capable de traiter une large gamme de sujets et de générer du texte dans divers contextes. Sa flexibilité vient de son entraînement sur de vastes corpus de données, ce qui lui permet d'aborder un éventail plus large de tâches, mais il peut manquer de profondeur spécialisée dans un domaine précis comparé à un systeme expert bien conçu.

4. Adaptabilité :

Un systeme expert nécessite des mises à jour manuelles pour ajouter de nouvelles connaissances ou regles. Cela peut être un processus long et coûteux.

Un LLM peut être mis à jour de maniere plus simple en ré-entraînant le modele sur de nouvelles données, mais il ne peut pas "apprendre" en temps réel ou interagir directement avec l'environnement de maniere continue. De plus, sa capacité à assimiler de nouvelles connaissances peut être limitée sans un réentraînement complet.

5. Explicabilité :

Les décisions prises par un systeme expert sont généralement explicites et peuvent être suivies grâce aux regles qui ont été appliquées. Cela permet une certaine transparence dans le processus décisionnel.

Les décisions ou les réponses générées par un LLM ne sont pas aussi explicites. Les résultats peuvent sembler être le fruit de raisonnements complexes, mais il est souvent difficile d'expliquer pourquoi le modele a produit une réponse particulière, car il est basé sur des corrélations apprises à partir des données.

6. Performance et Précision :

La performance du systeme expert dépend de la qualité de la base de connaissances et des regles intégrées. Si le systeme a accès à des connaissances très spécialisées et à jour, il peut être extrêmement précis dans ce domaine.

Un LLM peut parfois générer des réponses inexacts ou incohérentes, surtout pour des questions très spécialisées, car il n'a pas une connaissance rigoureuse et structurée comme un systeme expert. Cependant, il peut souvent fournir des réponses suffisamment bonnes pour des tâches générales ou pour des sujets où la connaissance de base est large.

7. Applications :

Il est utilisé dans des domaines où des connaissances spécialisées sont cruciales, comme la médecine (diagnostic), l'ingénierie (conception de systemes), ou la finance (conseils d'investissement).

Les LLM sont utilisés dans une variété d'applications telles que la génération de texte, la traduction automatique, la rédaction de contenu, les chatbots, l'assistance virtuelle, etc. Ils sont plus adaptés à des tâches générales de traitement du langage naturel.