

L'IA a besoin de contexte. Les données liées le lui apportent.

Axel Kaschte, Ph. D.

3 août 2026 · Traduction française de l'article « [AI needs context. Linked data provides.](#) » publié sur le blog [OCLC Next](#).

L'intelligence artificielle est au cœur des discussions actuelles sur les technologies. Dans les conférences, les titres de l'actualité et les feuilles de route produits, l'IA est partout. Dans le même temps, les données liées, qui ont longtemps suscité un fort intérêt, semblent aujourd'hui moins visibles.

Cette impression peut toutefois être trompeuse sur un point essentiel : l'IA ne remplace pas les données liées. Elle met en évidence à quel point des données structurées et interconnectées sont indispensables.

Il ne s'agit pas de priorités concurrentes. Ce sont des technologies convergentes, qui évoluent ensemble et, dans bien des cas, s'améliorent mutuellement.

Pour comprendre cette relation, il faut commencer par examiner le fonctionnement réel des systèmes d'IA.

L'IA dépend de la qualité de son entraînement

La plupart des systèmes d'IA modernes, notamment ceux reposant sur l'apprentissage automatique et les réseaux neuronaux, fonctionnent en deux phases : l'entraînement et l'inférence.

Lors de l'entraînement, les modèles apprennent à partir de grands volumes de données. Ces exemples permettent au système d'identifier les schémas, de comprendre le fonctionnement du langage et de repérer les relations entre les concepts. Une fois entraîné, le modèle passe à la phase d'inférence, au cours de laquelle il applique ses acquis à de nouvelles situations.

Cette approche est puissante, mais elle soulève aussi des difficultés.

L'entraînement exige de grandes quantités de données, et leur qualité compte tout autant que leur volume. Si les informations sous-jacentes sont incohérentes, incomplètes ou mal structurées, le modèle apprend à partir d'exemples imparfaits. Lorsque des données issues de sources multiples sont combinées, ces incohérences peuvent se multiplier.

Après leur déploiement, les systèmes d'IA peuvent produire des résultats difficiles à vérifier. Les hallucinations, lorsque les systèmes génèrent des réponses assurées, mais incorrectes, restent un défi persistant.

Tout cela conduit à une question centrale : comment fournir à l'IA de meilleures connaissances sur lesquelles s'appuyer ?

Des connaissances structurées améliorent l'apprentissage de l'IA

C'est là que les données liées jouent un rôle essentiel.

Les données liées organisent l'information au moyen de relations structurées. Au lieu de traiter les données comme des notices isolées, elles relient les entités entre elles et explicitent les relations qui les unissent. Elles forment ainsi un réseau de connaissances que les machines peuvent interpréter plus facilement.

Les bibliothèques créent depuis longtemps ce type de structure par l'intermédiaire de métadonnées. Des formats tels que le MARC permettent de décrire avec précision les livres, les auteurs, les sujets et

les éditions. Lorsque ces données évoluent vers des données liées, elles deviennent plus que descriptives : elles sont interconnectées et compréhensibles par les machines à grande échelle.

L'expérience d'OCLC montre que l'utilisation de métadonnées structurées comme données d'entraînement peut faire une réelle différence. Lorsque les relations entre les entités sont déjà définies, les modèles ont moins besoin de faire des déductions à partir de données brutes. La structure est intégrée aux données elles-mêmes.

Cela peut rendre l'entraînement plus efficace et contribuer à réduire les ressources de calcul nécessaires pour obtenir des résultats utiles.

En termes simples : mieux les connaissances sont organisées, plus les systèmes d'IA peuvent apprendre facilement.

Les données liées améliorent également les résultats de l'IA

Les avantages ne s'arrêtent pas à l'entraînement. Les données liées offrent également un moyen de valider les résultats produits par l'IA lors de l'inférence.

Une approche qui suscite un intérêt croissant est la génération augmentée par récupération d'information (retrieval-augmented generation, ou RAG) à l'aide de graphes de connaissances. Au lieu de s'appuyer uniquement sur les résultats générés par le modèle, les systèmes peuvent les confronter à une source fiable de connaissances structurées.

Dans le cadre des travaux d'OCLC sur les systèmes de recommandation, les suggestions générées par l'IA peuvent être vérifiées dans WorldCat afin de confirmer que les titres existent et que les relations sont exactes avant de présenter les résultats.

Cette approche ajoute une couche essentielle de fiabilité. Elle réduit le risque de résultats hallucinés et fournit une base plus claire et plus traçable pour comprendre comment les résultats sont générés.

Plutôt que de reposer uniquement sur la reconnaissance de schémas, les systèmes d'IA peuvent ancrer leurs réponses dans des

connaissances vérifiables.

Un cycle d'amélioration continue

La relation entre l'IA et les données liées devient encore plus puissante lorsqu'on la considère comme un cycle continu.

L'IA peut analyser les métadonnées existantes et identifier des schémas, des lacunes ou des liens manquants. Ces enseignements peuvent ensuite servir à améliorer les données sous-jacentes et à enrichir le graphe de connaissances.

Ces données améliorées deviennent alors le socle de la phase d'entraînement suivante.

Le processus se poursuit : entraînement, inférence, validation et amélioration.

Chez OCLC, cette approche itérative prend déjà forme, notamment avec notre processus de dédoublonnage par IA des notices de WorldCat. Au début de cette année, nous avons achevé une phase de test ciblant uniquement les livres imprimés en anglais dans WorldCat et fusionné 5 000 000 de notices dupliquées nouvellement détectées. Les livres imprimés en anglais représentent la plus grande catégorie de doublons dans WorldCat et constituent le format qui a fait l'objet des tests et améliorations les plus approfondis dans nos activités de dédoublonnage par apprentissage automatique à ce jour. Nous continuons d'analyser les résultats, mais un point est clair : les résultats produits par l'IA peuvent être vérifiés à partir de données structurées, et ces résultats peuvent contribuer à affiner et à enrichir les données elles-mêmes. Au fil du temps, les données comme les modèles s'améliorent conjointement.

C'est dans ce cadre que la notion de technologies convergentes prend une dimension concrète. L'IA et les données liées n'évoluent pas seulement côte à côte : chacune améliore activement l'efficacité de l'autre.

Pourquoi les données liées restent essentielles

L'attention actuellement portée à l'IA peut donner l'impression que toutes les autres technologies doivent s'adapter ou disparaître. La relation entre l'IA et les données liées suggère pourtant une autre réalité.

L'IA est puissante, mais elle dépend de connaissances structurées et interconnectées pour atteindre tout son potentiel. Les données liées fournissent cette base en organisant l'information pour que les machines puissent la comprendre, la valider et s'en servir pour avancer.

Pour les organisations qui gèrent des métadonnées riches, cela représente un avantage considérable. Ces données ne se contentent pas de décrire les ressources : elles permettent de développer une IA plus précise, plus efficace et plus fiable.

Perspectives

L'IA occupe peut-être aujourd'hui le devant de la scène, mais son efficacité repose sur des données fiables et bien connectées. Les données liées lui apportent ce contexte.

Chez OCLC, cette relation n'est pas théorique. Il s'agit d'un domaine d'investissement actif, dans lequel les données liées et l'IA évoluent ensemble pour s'améliorer au fil du temps.

Plutôt que de se remplacer l'une l'autre, ces technologies convergent. Ensemble, elles transforment la manière dont les connaissances peuvent être organisées, comprises et partagées plus efficacement.

Ce billet s'appuie sur le contenu de la présentation « Building Intelligence: Practical library applications of AI and Linked Data at OCLC », présentée pour la première fois lors de la conférence IFLA WLIC à Astana, au Kazakhstan, en 2025. Vous pouvez regarder l'enregistrement complet de la [présentation ici](#).