

# Intelligences artificielles : l'Algérie doit renforcer ses données pour mieux se faire connaître

À l'heure où les intelligences artificielles répondent à des millions de requêtes chaque jour, la qualité des données disponibles sur l'Algérie devient un enjeu stratégique. Un document publié par le CARE met en avant les limites de la documentation existante et appelle à la constitution d'un véritable corpus national de connaissances.

Les intelligences artificielles s'installent progressivement dans les usages quotidiens. Pour rechercher une information, comprendre un événement historique, obtenir des données économiques ou découvrir une région, de plus en plus d'utilisateurs se tournent vers des outils capables de produire une réponse en quelques secondes.

Cette évolution pose toutefois une question importante pour l'Algérie : **sur quelles données ces systèmes s'appuient-ils lorsqu'ils répondent aux requêtes concernant le pays ?**

C'est précisément cette problématique que soulève un document publié dans la rubrique « Idées & Débats » du CARE, sous le titre « *De ma vieille Britannica au corpus national : plaidoyer pour une souveraineté du savoir* ». Le texte est signé par le Dr Ali Kahlane.

## **Des données qui ne reflètent pas toujours suffisamment l'Algérie**

Le [document](#) rappelle qu'une intelligence artificielle ne dispose pas d'une connaissance spontanée de l'Algérie. Elle construit ses réponses à partir des informations auxquelles elle a accès.

Or, une partie importante des livres, archives, travaux universitaires, articles de presse et statistiques disponibles sur l'Algérie ont été produits à l'étranger ou restent insuffisamment numérisés et accessibles. Cette situation peut conduire les IA à reproduire des informations exactes, mais également des **angles de lecture extérieurs au contexte algérien**.

Le problème est également linguistique. Le document insiste sur la faiblesse des ressources disponibles pour la darija algérienne et le tamazight. Une faible présence de ces langues dans les données numériques limite mécaniquement la capacité des machines à comprendre certaines réalités linguistiques et culturelles algériennes.

## Wikipédia comme premier levier

Parmi les pistes avancées, le document accorde une attention particulière à Wikipédia.

L'objectif n'est pas de transformer l'encyclopédie en outil officiel, mais de renforcer les contenus consacrés à l'Algérie à partir de **sources algériennes vérifiables**, tout en respectant les règles de neutralité et de fiabilité de la plateforme. Une attention particulière devrait être accordée aux versions arabe et tamazight, aujourd'hui moins développées.

Cette démarche pourrait contribuer à améliorer la présence de données algériennes dans l'écosystème numérique mondial.

## Constituer un corpus national de données pour les intelligences artificielles

Pour le CARE, l'enjeu dépasse toutefois Wikipédia. Le véritable chantier consiste à construire un **corpus national de connaissances**.

Le document relève que l'Algérie dispose déjà de nombreuses ressources dispersées entre la Bibliothèque nationale, les archives, les universités, les centres de recherche, les musées, les administrations, les médias et les maisons d'édition. Des infrastructures existent également, notamment celles développées par le CERIST, comme le SNDL, l'ASJP, le PNST, le Catalogue collectif algérien et BiblioUniv.

Mais la numérisation seule ne suffit pas. Les documents doivent être **classés, datés, identifiés, indexés et rendus accessibles** dans plusieurs langues. Le document recommande notamment de renforcer les données en arabe, tamazight, français et anglais.

## Des recommandations pour mieux alimenter les intelligences artificielles

Plusieurs pistes sont proposées pour alimenter les intelligences artificielles : élargir le dépôt légal au numérique, ouvrir certains fonds publics sous des licences adaptées, travailler avec la diaspora et les institutions étrangères qui détiennent des archives algériennes et constituer des bases de données de référence en darija et en tamazight.

Le document recommande également la création d'un « **corpus-or** » **algérien**, capable de servir de référence pour entraîner de futurs modèles d'intelligence artificielle mais aussi pour vérifier la qualité des réponses fournies par les modèles existants sur les sujets liés à l'Algérie.

## Un enjeu pour des futures intelligences

## artificielles algérienne

L'objectif ne serait pas uniquement de permettre aux intelligences artificielles étrangères de mieux répondre aux questions sur l'Algérie.

La constitution d'un corpus national pourrait également fournir une partie des données nécessaires au développement de futurs modèles algériens capables de comprendre les réalités locales et les différentes langues utilisées dans le pays. Le document présente ainsi la donnée comme une **ressource stratégique**, au même titre que les infrastructures nécessaires au développement de l'intelligence artificielle.

Le message porté par cette réflexion est donc clair : **pour que les intelligences artificielles parlent mieux de l'Algérie, il faut d'abord leur fournir davantage de données fiables, structurées et produites à partir de sources algériennes.**

La question dépasse ainsi la seule technologie. Elle touche à la conservation de la [mémoire nationale](#), à la diffusion du savoir et, plus largement, à la capacité de l'Algérie à documenter son histoire, son économie, sa culture et ses langues dans l'espace numérique.