
Votre collègue de bureau est-il une IA?

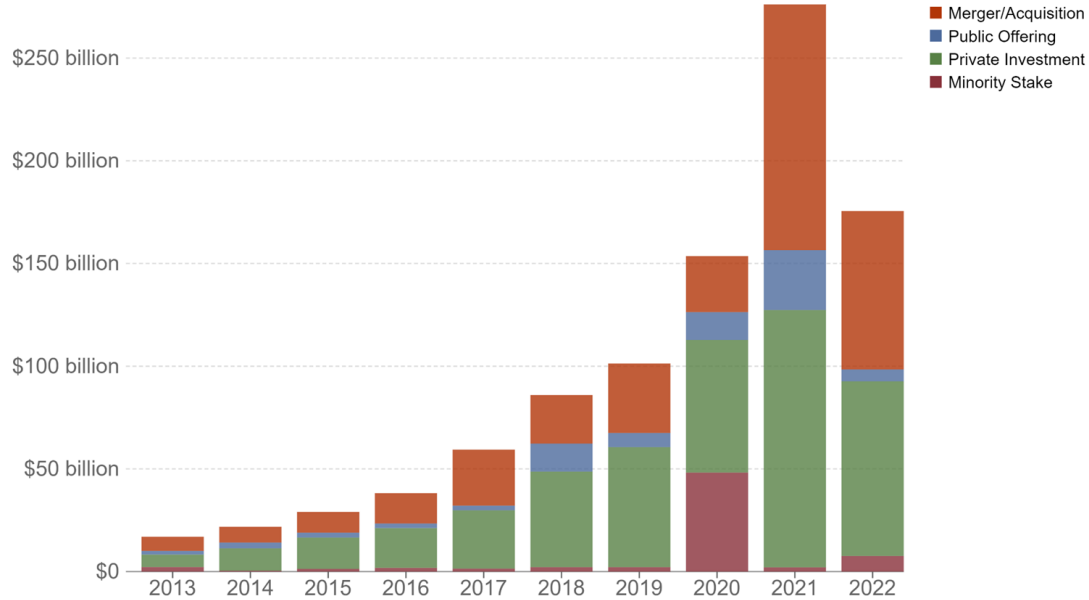
Nicolas Morin | ABES | novembre 2023



L'IA, c'est maintenant

Annual global corporate investment in artificial intelligence, by type

This data is expressed in US dollars, adjusted for inflation.



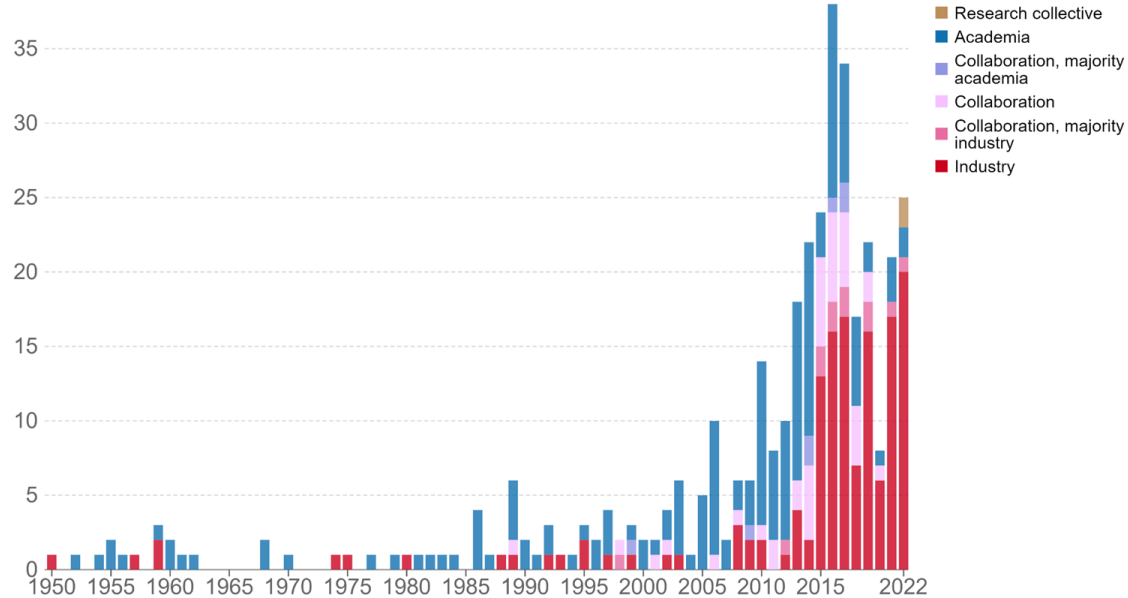
Source: NetBase Quid via AI Index Report (2023)

OurWorldInData.org/artificial-intelligence • CC BY

Note: Data is expressed in constant 2021 US\$. Inflation adjustment is based on the US Consumer Price Index (CPI).

L'IA bascule vers le secteur privé

Affiliation of research teams building notable AI systems, by year of publication



Source: Epoch (2023)

OurWorldInData.org/artificial-intelligence • CC BY

Note: A research collective is a group of AI researchers not organized under an academic or industry affiliation. Systems are defined as "notable" by the authors based on several criteria, such as advancing the state of the art or being of historical importance.

Combien de temps encore avant la révolution?

Le temps d'adoption des révolutions technologiques est long :

- 30 ans pour que l'électrification atteigne 50% d'adoption (~1935 - US)
- 30 ans pour que l'informatique atteigne 50% d'adoption (~2000 - US)



Les liftiers et l'ascenseur automatique

Raisons sociales, raisons psychologiques

“Les opérateurs d'ascenseurs sont un facteur de sécurité et de protection. Il a dit que les agressions, les vols et même les meurtres qui ont eu lieu dans les immeubles d'habitation et dans leurs vestibules auraient pu être évités si les ascenseurs des immeubles étaient occupés par des opérateurs.” (Chicago Evening News, 1952)

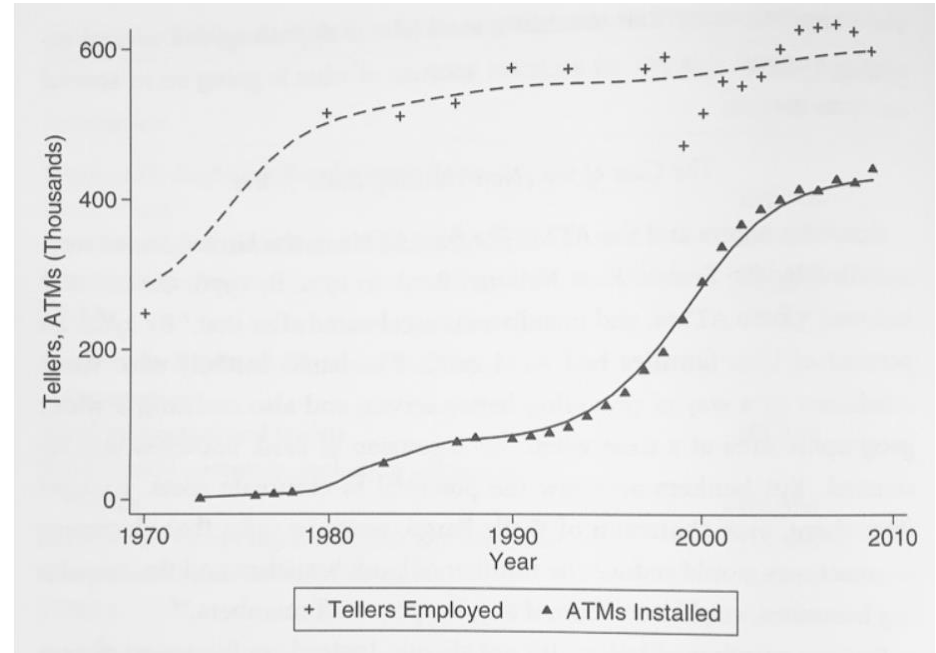
“La porte a appris à penser. Si la porte touche un objet dans son embrasure, elle se retire immédiatement. C'est ce que les fabricants d'ascenseur pensent que la porte pense. Vous savez ce que je pense que la porte pense? Je pense qu'elle pense, cette fois je t'aurais.” (Courier Journal, Louisville, Kentucky, 1957)

“Méfiez vous des voitures qui ont leur propre volonté. Les prédictions selon lesquelles les voitures autonomes alimentées par l'intelligence artificielle créeront un monde plus sûr et plus harmonieux pourraient être fausses.” (Bloomberg, 2019)



Une histoire des distributeurs de billets

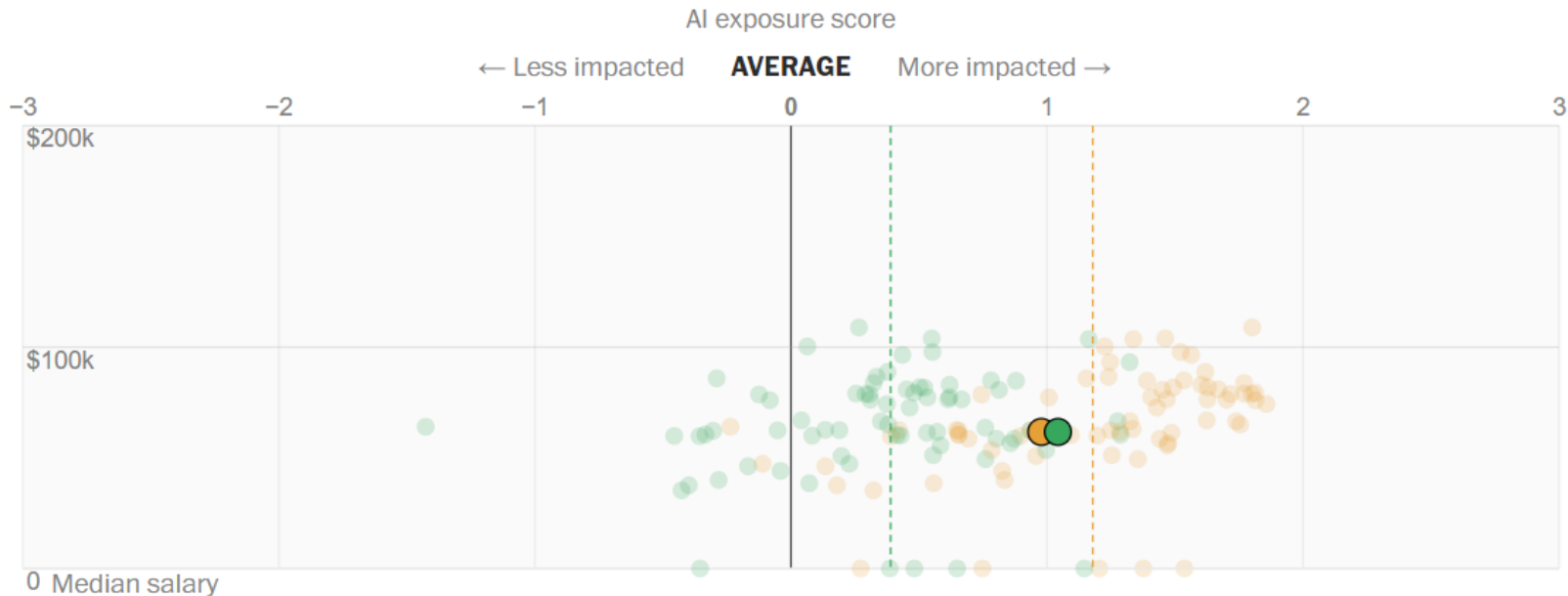
L'effet de l'automatisation n'est pas toujours celui qu'on croit : les distributeurs de billets n'ont pas remplacé les guichetiers de banque



Des métiers “exposés” à l’IA (et à l’automatisation)?

Social services and education

Teacher, librarian, social worker, counselor



Machine Learning Librarian



Daniel van Strien (He/They) · 1st

Machine Learning Librarian
@🤖 | Championing Open Science & ML

Talks about #libraries, #openscience, #machinelearning, and #artificialintelligence

Hugging Face
Glasgow, Scotland, United Kingdom

#libraries #openscience #machinelearning
#artificialintelligence

HuggingFace - créée en 2016 (Paris puis NYC - 2017)
Clément Delangue (CEO)
Revenus 2022 :15m\$
Nombre d'employés 2023 : 160

Start-up franco-américaine développant des outils pour les applications de traitement automatique des langues, et une plate-forme permettant le partage des modèles et des ensembles de données nécessaires à l'apprentissage automatique.

Promesses et illusions de l'IA ?

Aller au-delà d'un discours manichéen, et juger correctement de l'ampleur du changement est difficile.

“Je n'étonnerai personne en évoquant **ma confusion devant l'évolution galopante** d'Internet et plus particulièrement du web.”

.../...

“L'information est de plus en plus massive, disponible et accessible, mais encore faut-il pouvoir la localiser, l'appréhender, l'assimiler **avant de la diffuser** !”

BBF, 2001, 1



Les collections comme données

Chambers, S. (2023) « Position Statements -> Collections as Data: State of the field and future directions ».

- Données de collections numérisées et numériques : des métadonnées au full text
- Données produites par le secteur documentaire en tant que tel : métadonnées en particulier
- Les collections comme *training data* des LLMs et de l'IA en général

Nos collections documentaires sont de la donnée

- Sudoc : 18m de notices
- Science+ : métadonnées pour 600.000 articles à ce stade
- Theses.fr : ~450.000 notices et ~150.000 thèses en Full Text
- Les “agents” : (autorités IdRef) 3,9m personnes et 300.000 collectivités

Les métadonnées sont objet et enjeu stratégique de négociation des contrats de documentation électronique

AI dans une recherche de catalogue : Talpa

Clarivate + LibrayThing

1. Interroge LLM (Claude AI et ChatGPT) pour les livres et autres médias. Elimine les références “hallucinées” en comparant à une bdd bibliographique
2. LLM pour analyser et comprendre les requêtes, auxquelles on répond ensuite à l'aide des données de bibliothèque traditionnelles.

Les données de bibliothèque sont spécifiques et précises : elles permettent la qualité de la réponse : *“Talpa is a magical, AI-powered library search backed by true and authoritative data.”*
(<https://www.talpa.ai/>)



Talpa is a magical, AI-powered library search backed by true and authoritative data.

AI pour les données académiques

AI21 Labs (août 2023)

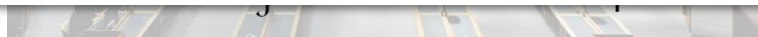
“We’re thrilled to announce the completion of a successful Series C funding round of \$155 million at a valuation of \$1.4 billion, marking a major milestone in our company's journey to bring reliable AI to the enterprise. [...] The new funding round brings the company’s raised capital to a total of \$283 million at a valuation of \$1.4 billion”



AI business insights

Make sense of unstructured text sitting idly in your data lake. Solve topic classification, sentiment analysis, entity extraction and other challenging language understanding tasks in 15 minutes.

[Try Now](#)



Vers un graphe académique mondial... et au-delà

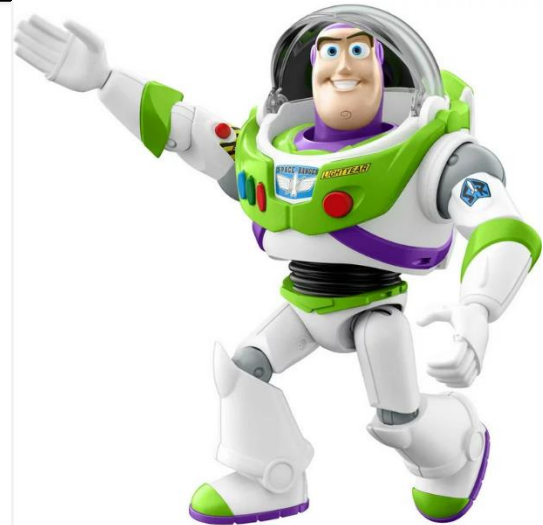


[← BACK TO ALL NEWS](#)

SEPTEMBER 26, 2023

2 MINUTE READ

Clarivate Establishes Academia & Government Innovation Incubator and Acquires Alethea, an AI Student Engagement Solution



Les collections comme données... encore en silos

- Nos données sont encore largement en silos : au sein des bibliothèques, entre bibliothèques, dans le secteur culturel et académique en général, entre acteurs publics et acteurs privés
- Ce n'est pas toujours le secteur documentaire public qui a les "meilleures" données, ni en volume, ni en qualité... le volume étant en soi une qualité.



[More images](#)

Kaitlyn Aurelia Smith – EARS

Label: [Western Vinyl](#) – WV145
Format: [Vinyl](#), LP, Album
Country: [US](#)
Released: [Apr 1, 2016](#)
Genre: [Electronic](#)
Style: [Downtempo](#), [Abstract](#), [Ambient](#)



Tracklist

A1	First Flight	4:38
A2	Wetlands	3:35
A3	Envelop	4:57
A4	When I Try, I'm Full	3:57
B1	Rare Things Grow	3:46
B2	Arthropoda	3:38
B3	Stratus	3:05
B4	Existence In The Unfurling	11:09

Companies, etc.

Copyright © – [Western Vinyl](#)
Copyright © – [Kaitlyn Aurelia Smith](#)

Credits

Artwork – [Salihah Moore Kirby](#)
Composed By, Arranged By, Performer, Producer, Recorded By, Mixed By – [Kaitlyn Aurelia Smith](#)
Mastered By – [Simon Heyworth](#)
Voice, Flute, Clarinet, Alto Saxophone, Tenor Saxophone, Bass Clarinet, Synthesizer [Buchla Music Easel, EMS Synthi, ARP 2600, Oscar, Korg Mono/Poly, Electrocomp 101, Moog Werkstatt] – [Kaitlyn Aurelia Smith](#)
Woodwind – [Rob Frye](#)

Notes

Download card included.

Barcode and Other Identifiers

Barcode: 7 51937 43551 8
Matrix / Runout (Side A, Etched): WV 145-A-RE1 WMM/CJIII
Matrix / Runout (Side B, Etched): WV 145-B-RE1 WMM

Tracklist [\[edit \]](#)


All tracks composed by Kaitlyn Aurelia Smith.

No. Title

- "First Flight"
- "Wetlands"
- "Envelop"
- "When I Try, I'm Full"
- "Rare Things Grow"
- "Arthropoda"
- "Stratus"
- "Existence in the Unfurling"

Personnel [\[edit \]](#)

- Kaitlyn Aurelia Smith - composition, production, Easel, EMS Synthi, ARP 2600, Oscar, Korg
- Rob Frye - woodwind
- Simon Heyworth - mastering
- Salihah Moore Kirby - artwork



Grandes questions éthiques générales commencent à être mieux cernées, par exemple sur les biais de données, la consommation énergétique, etc.

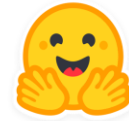
Mais on a besoin de descendre d'un cran pour prendre en compte des responsabilités opérationnelles.

Cf Thomas Padilla. [*Responsible Operations: Data Science, Machine Learning, and AI in Libraries*](#) (OCLC Research, 2019)

1. S'engager : des opérations responsables
2. Description et découverte
3. Méthodes et données partagées
4. Collections exploitables par machine
5. Formation et évolution des personnels
6. Des services de Science des données
7. Maintenir la collaboration interprofessionnelle et interdisciplinaire

Open Source AI: des méthodes et données partagées

L'exemple du projet American Stories



Hugging Face

Melissa Dell (Harvard)

20 millions d'articles complets extraits d'images historiques de périodiques américains de la Library of Congress, sous forme d'un dataset après traitement par un pipeline spécifiquement conçu de Deep Learning :

- détection de mise en page
- classification de lisibilité
- OCR personnalisé
- association de textes d'articles s'étendant sur plusieurs cadres de délimitation, etc.

Abes : études IA ...-2023

1. IA de calcul et diagnostic des liens entre notices bibliographiques et notices d'autorité Personne. IA symbolique (raisonnement logique), pas machine learning !
 - 2012-2016 : projet de recherche ANR Qualinca (Qualité des liens dans les catalogues)
 - 2019 : mise en service de cette IA au sein de l'interface <https://paprika.idref.fr>
Prochaine étape : proposer cette IA comme web service.
2. Indexation RAMEAU automatique.
 - 2021 : stage Master 2 (6 mois)
 - 2023 : prestation de 40 jours
 - 2024 : expérimentation in vivo ?
3. Repérage des auteurs et de leur fonction dans les mentions de responsabilité des notices bibliographiques du Sudoc.
 - 2022 : stage Master 2 (6 mois)
 - 2023 (en cours) : évaluation de son utilisabilité en interne

IA Cluster France 2030

IA MÉDITERRANÉE

Présentation Actualités Soutiens Contact **Appel à signature**

Cluster IA Méditerranée

Bienvenue sur le site du consortium IA Méditerranée ! Un regroupement d'acteurs clés de l'écosystème technologique, académique et entrepreneurial de Montpellier Méditerranée Métropole, engagés dans la recherche et le développement d'intelligences artificielles souveraines et de confiance.

[Découvrir le projet](#) [Nous contacter](#)

UNIVERSITÉ DE MONTPELLIER

L'INSTITUT agro Montpellier

Inria

INES

abes

amue

LA TECH TECH MÉDITERRANÉE

Digital113

CNRS

CHU

ICM Institut régional du Cancer Montpellier - Languedoc

MEDVALLÉE MONTPELLIÈRE

Montpellier Méditerranée métropole

La Région Occitanie Pyrénées-Méditerranée

CITÉ DE L'ÉCONOMIE ET DES MÉTIERS DE GRASSE

L'IA dans le projet d'établissement de l'Abes

Projet 2024-2028 :

- Appliquer l'IA sur les données gérées par l'Abes, dans une démarche d'ouverture à la fois des données source, des données résultant des traitements, et des modèles utilisés pour en permettre la plus grande réutilisation
- Proposer des services d'assistance par l'IA aux utilisateurs des logiciels proposés par l'agence

L'IA dans le projet d'établissement de l'Abes

Projet 2024-2028, quatre types de livrables et d'actions :

1. Transformer et améliorer les données gérées par l'Abes ;
2. Equiper les outils mis à la disposition des utilisateurs professionnels de fonctionnalités d'aide à la décision voire d'automatisation ;
3. Partager ces fonctionnalités sous la forme d'API, de modèles ouverts et de code ouvert, pour permettre à nos partenaires de les utiliser sur leurs propres données et dans leur propre contexte ;
4. Proposer nos données comme intrants des IA de tiers, y compris pour nourrir les grands modèles (LLM).

Cas d'usage : quelques exemples possibles

- FAQ en amont d'abesstp par ex., pour les réponses de type tickets (*chatbot*)
- Aide à l'indexation : en arrière-fond à l'ingestion d'une notice par le système (Create / Update); suggestion d'indexation dans une interface de catalogage (web)
- Traduction systématique de résumés en ENG quand il y a un résumé en FRE
- Pour une bibliothèque numérique / numérisation, génération de texte alt descriptif pour images. Utilisation via API, plugin omeka-s par exemple.
- Catalogage d'une notice candidate à partir du scan de la page de titre d'un livre
- .../...

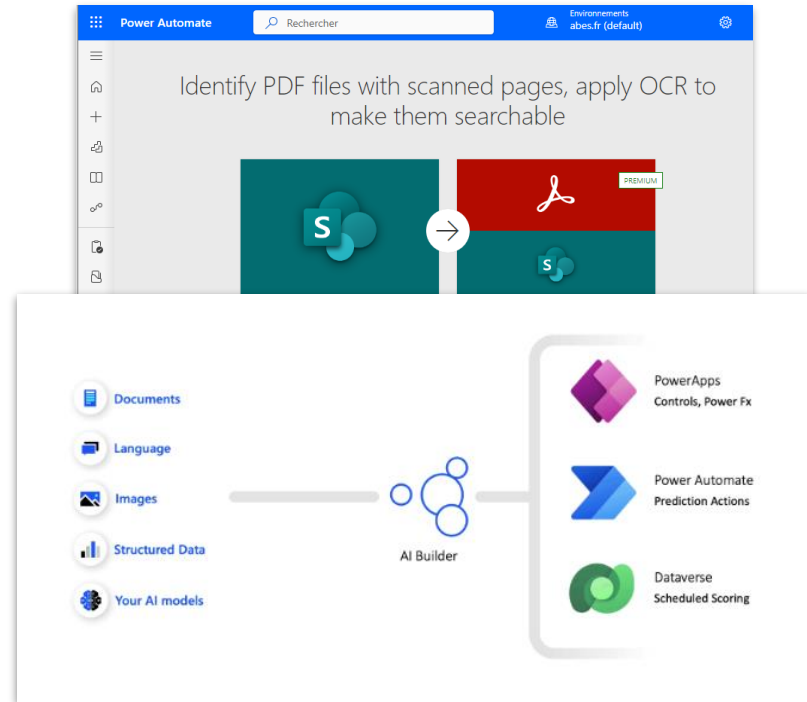
Cas d'usage : catalogage des imprimés

Hannes Lowagie - Metadata Librarian KBR

- Scan des pages de titre, colophon et 4e de couv.
- recherche ISNI + indexation automatique
- génération d'une notice marc21XML

Appuyé sur des solutions Microsoft

“standard” : Power Automate + AI Builder



L'IA dans le projet d'établissement de l'Abes

L'Abes mènera ces travaux de R&D dans le cadre de partenariats :

- Organisations ayant des démarches similaires de R&D utilisant l'IA dans le secteur documentaire
- Partenaires permettant de valider les cas d'usages, en particulier les établissements utilisateurs des produits et services de l'Abes, qui peuvent contribuer en particulier à l'analyse des besoins, à l'analyse ou encore à l'évaluation des projets dont ils sont les parties prenantes
- Partenaires privés apportant leurs compétences, leur force de travail et leurs capacités industrielles

Résumons-nous

- L'IA est un futur inégalement distribué, et encore partiellement indéterminé : nous avons des “marges de manoeuvres” à son égard
- Pour le secteur documentaire (de l'ESR), il y a collaborations et concurrences réelles entre secteur public et secteur privé à ce sujet
- Nous pouvons et devons :
 - Travailler sur nos collections comme données
 - Utiliser l'IA pour le secteur documentaire quand c'est pertinent
 - Accompagner dans la (longue) durée les évolutions de notre secteur, y compris les métiers
 - Être modestes et avoir un spectre large d'analyse : l'avenir de l'IA ne dépend pas des bibliothèques
- Il sera d'autant plus facile de nous positionner que nous nous appuierons sur des valeurs fondamentales : ouverture, transparence, adaptation aux besoins des communautés que nous desservons

Enfin, mon voisin de bureau est-il une IA?

En faveur de l'anthropomorphisme. Des avantages de considérer l'IA comme une personne.

D'abord c'est inévitable : nous avons de grandes facilités à anthropomorphiser notre environnement. Bénéfices prouvés par ex. dans le domaine du bien être animal.

Si une IA est perçue comme ayant des caractéristiques humaines il est plus difficile de lui demander des choses éthiquement discutables.

- David Scott Bernstein. *Prompt Engineering for Everyone* (2023).
- Sven Nyholm, *Humans and Robots: Ethics, Agency, and Anthropomorphism*, Rowman & Littlefield (2020)
- Sven Nyholm et al. *This is Technology Ethics* (2023). Chap. 5. Artificial intelligence, value alignment, and the control problem.



Merci de votre attention

Aucune IA n'a été utilisée pour la réalisation de cette présentation



Sources

- Michael Webb. The Impact of Artificial Intelligence on the Labor Market. Stanford University, January 2020
- US Bureau of Labor Statistics
- Felten et al. Occupational Heterogeneity in Exposure to Generative AI (2023)
- <https://ourworldindata.org/artificial-intelligence>
- Michael Webb on whether AI will soon cause job loss, lower incomes, and higher inequality—or the opposite (<https://80000hours.org/podcast/episodes/michael-webb-ai-jobs-labour-market/>)
- <https://www.jasonfeifer.com/episode/how-the-elevator-shaped-our-world/>
- David Scott Bernstein. Prompt Engineering for Everyone (2023).
- Sven Nyholm, Humans and Robots: Ethics, Agency, and Anthropomorphism, Rowman & Littlefield (2020)
- Sven Nyholm et al. This is Technology Ethics (2023). Chap. 5. Artificial intelligence, value alignment, and the control problem.
- Chambers, S. (2023) « Position Statements -> Collections as Data: State of the field and future directions ». Zenodo. doi: 10.5281/zenodo.7897735.
- <https://www.talpa.ai/>
- <https://clarivate.com/news/clarivate-announces-partnership-with-ai21-labs-as-part-of-its-generative-ai-strategy-to-drive-growth/>
- <https://huggingface.co/datasets/dell-research-harvard/AmericanStories>
- https://www.linkedin.com/posts/hanneslowagie_powerautomate-aibuilder-activity-7113592369010069504-XbFO