

INTELLIGENCE ARTIFICIELLE EN AFRIQUE : COMMENT RATTRAPER LE TRAIN EN MARCHE ?

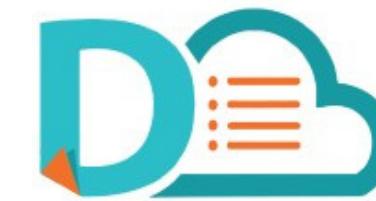
Enjeux et perspectives de l'IA en Afrique



ABDOU MOHAMED Naira

Lead Data Scientist - ToumAI Analytics
Doctorant en Systèmes Intelligents - Institut
National de Statistique et d'Economie
Appliquée (INSEA)

Journée **Data X AI**



جامعة محمد الأول بوجدة
UNIVERSITE MOHAMMED PREMIER OUJDA
+٥٣٨٠٦٤٢٩٨٢٠٥٧٥٠٨

02

Agenda

Première Partie : Introduction générale à l'Intelligence Artificielle

Deuxième Partie : L'Intelligence Artificielle et l'Afrique

A propos de moi

L'IA : D'où ça vient?

L'IA : Comment ça marche?

Un biais dans la matrice

Vers un goulot d'étranglement?

L'Afrique en Chiffre

L'IA en Afrique : Où en Sommes-nous?

Rattraper le train en marche : Comment?

Vers une IA plus durable

03

A propos de moi

- Origine : Comorienne.
- Formation : Faculté des Sciences Dhar el Mehraz de Fès, Université Ibn Tofail de Kénitra, Institut National de Statistique et d'Economie Appliquée de Rabat.
- Fonction actuelle : Lead Data Scientist chez ToumAI Analytics.
- Passions : Ecriture, Lecture, Musique, Cinéma.

[Retourner à l'Agenda](#)



04

A propos de moi



SpeechBrain

Models 5

Sort: Recently updated

- speechbrain/asr-wav2vec2-dvoice-swahili
Automatic Speech Recognition • Updated 5 days ago • 34 • 2
- speechbrain/asr-wav2vec2-dvoice-wolof
Automatic Speech Recognition • Updated 5 days ago • 3 • 3
- speechbrain/asr-wav2vec2-dvoice-amharic
Automatic Speech Recognition • Updated 5 days ago • 18 • 6
- speechbrain/asr-wav2vec2-dvoice-darija
Automatic Speech Recognition • Updated 5 days ago • 9 • 2
- speechbrain/asr-wav2vec2-dvoice-fongbe
Automatic Speech Recognition • Updated 5 days ago • 2

[Retourner à l'Agenda](#)

Datasets Creation and Empirical Evaluations of Cross-Lingual Learning on Extremely Low-Resource Languages: A Focus on Comorian Dialects

Abdou Mohamed Naira^{1,2}, Abdessalam Bahafid^{1,2}, Zakarya Erraji¹, Imade Benelallam^{1,2}

¹ INSEA, Rabat, Morocco ² ToumAI Analytics, Rabat, Morocco

{nabdoumohamed, i.benelallam, a.bahafid, zerraji}@insea.ac.ma
{naira, imade, abahafid}@toum.ai

Abstract

In this era of extensive digitalization, there are a profusion of Intelligent Systems that attempt to understand how languages are structured for the aim of providing solutions in various tasks like Text Summarization, Sentiment Analysis, Speech Recognition, etc. But for multiple reasons going from lack of data to the nonexistence of initiatives, these applications are in an embryonic stage in certain languages and dialects, especially those spoken in the African continent, like Comorian dialects. Today, thanks to the improvement of Pre-trained Large Language Models, a spacious way is open to enable these kind of technologies on these languages. In this study, we are pioneering the representation of Comorian dialects in the field of Natural Language Processing (NLP) by constructing datasets (Lexicons, Speech Recognition and Raw Text datasets) that could be used on different tasks. We also

2021; Chamanga, 2022) and according to the ORELC¹ lexicon (See Fig. 1) in which we can observe that in a dictionary of 7,386 entries, 15.38% of the words are shared by all the dialects, 6.47% by three and 16.10% by two dialects. Indeed, these dialects can be divided into two groups: Eastern group (ShiNdzuan and ShiMaore) and Western group (ShiNgazidja and ShiMwali). Moreover, a part of the experiments conducted in (Maurizio and Michele, 2021) has shown through lexical distances calculation that these dialects could be classified into two other different groups, the first one composed of the ShiNgazidja while the second one contains the other three dialects.

The arrival of Transformers (Vaswani et al., 2017) was a real breakthrough in Artificial Intelligence (AI). This architecture allows us to better represent the context within texts which is a major spearhead in Language Understanding. Pre-trained

Models 3

Sort: Recently updated

- SI2M-Lab/DarijaBERT-arabizi
Fill-Mask • Updated Aug 29, 2023 • 70 • 5
- SI2M-Lab/DarijaBERT
Fill-Mask • Updated Aug 29, 2023 • 471 • 18
- SI2M-Lab/DarijaBERT-mix
Updated May 24, 2022 • 20 • 1

05

L'IA : D'où ça vient?

- Une pensée remontant à la Mythologie
- Les siècles des Lumières
- De la fiction à la réalité
- Premières définitions
- L'hiver de l'Intelligence Artificielle

[Retourner à l'Agenda](#)



06

IA : Les prémices



Dans la Mythologie

La légende de Pygmalion et Galatée : une statuette est rendue vivante grâce à Aphrodite.



Les automates anthropomorphes

C'étaient des automates à formes humaines utilisées souvent dans les cours pour divertir les familles royales.

[Retourner à l'Agenda](#)

07

IA : Les prémisses



Vers le siècle des Lumières

Compréhension et propagation des pensées de Descartes et Leibniz pendant le siècle des Lumières.



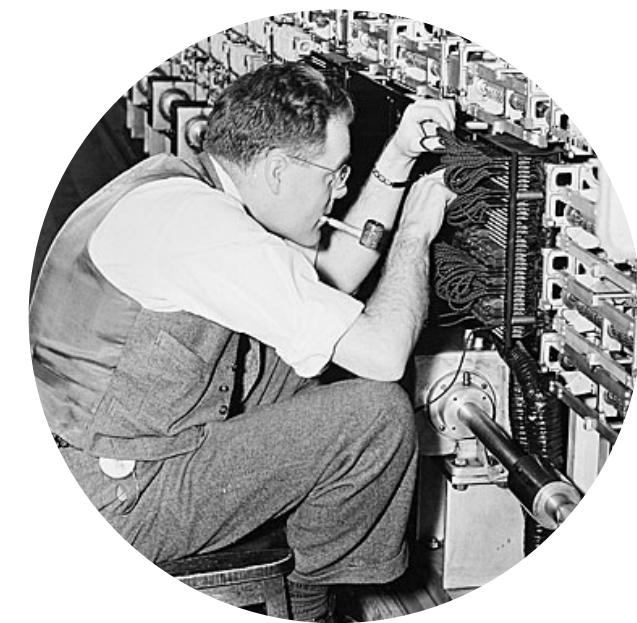
Lorsque la Science Fiction s'en empare

L'idée d'êtres artificiels pensants est devenu un élément central des œuvres de Science Fiction dès le 19ème siècle.

[Retourner à l'Agenda](#)

08

IA : Les prémisses



Alan Turing

L'évolution technologique survenue à la Seconde Guerre Mondiale a initié des projets d'envergures d'où sortiront un certain Alan Turing.



“Intelligence Artificielle” : Naissance du terme

On attribue la paternité du terme “IA” à John McCarthy du MIT, terme qu'il a introduit à la conférence Dartmouth.

[Retourner à l'Agenda](#)

“L’Intelligence Artificielle est la construction de programmes informatiques qui s’adonnent à des tâches qui sont, pour l’instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l’apprentissage perceptuel, l’organisation de la mémoire et le raisonnement critique.”

Marvin Minsky

[Retourner à l’Agenda](#)

IA : Les prémisses



Augmentation de la “hype”

La pop culture s’empare de l’IA et des films, des romans à succès avec comme sujets centraux la pensée artificielle commencent à sortir.



L’hiver de l’Intelligence Artificielle

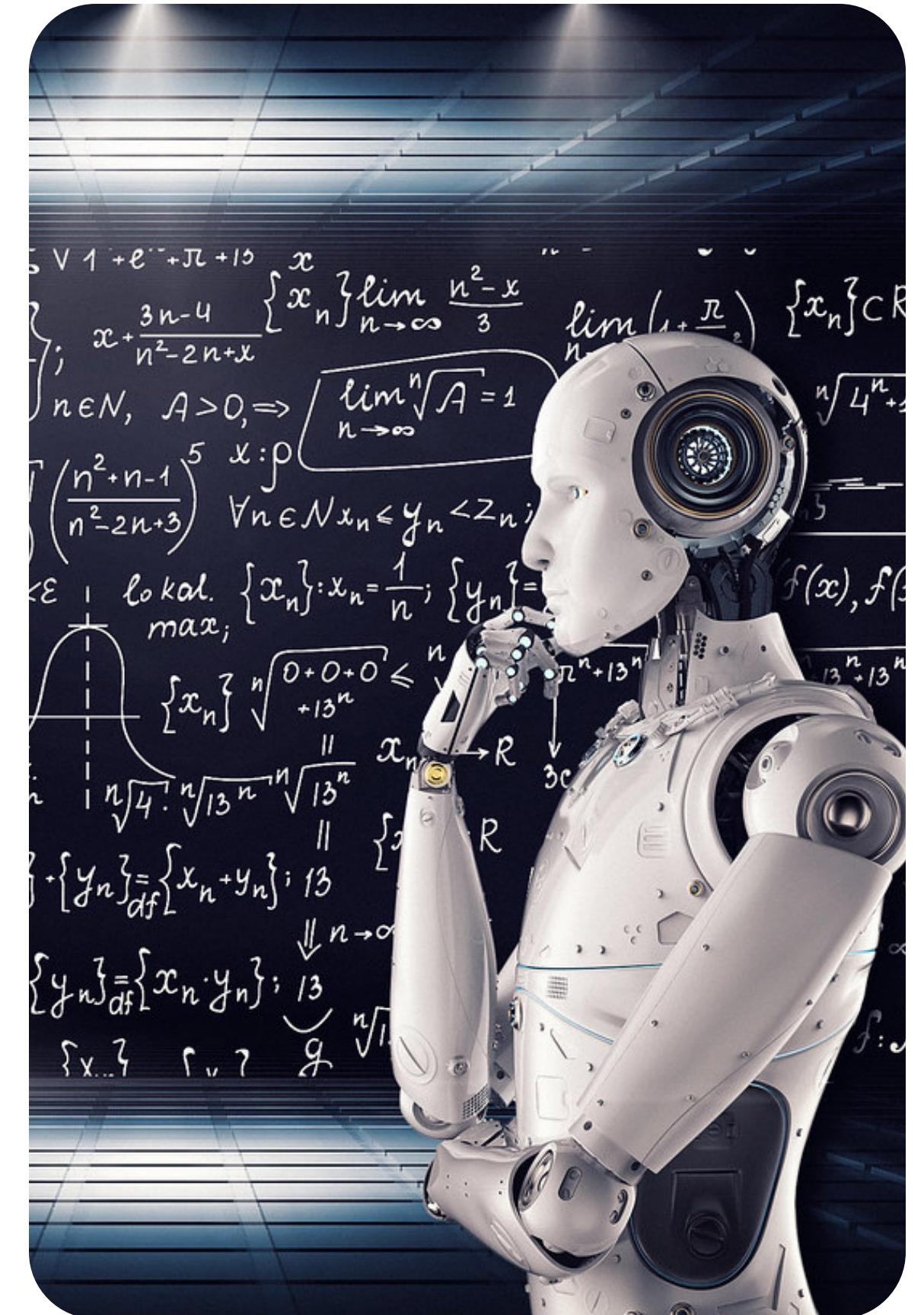
Des gros budgets investis dans les centres de recherches, des attentes trop grandes et de fables résultats ont conduit à un arrêt de travaux liés au développement de l’IA pendant un peu près de deux décennies.

[Retourner à l’Agenda](#)

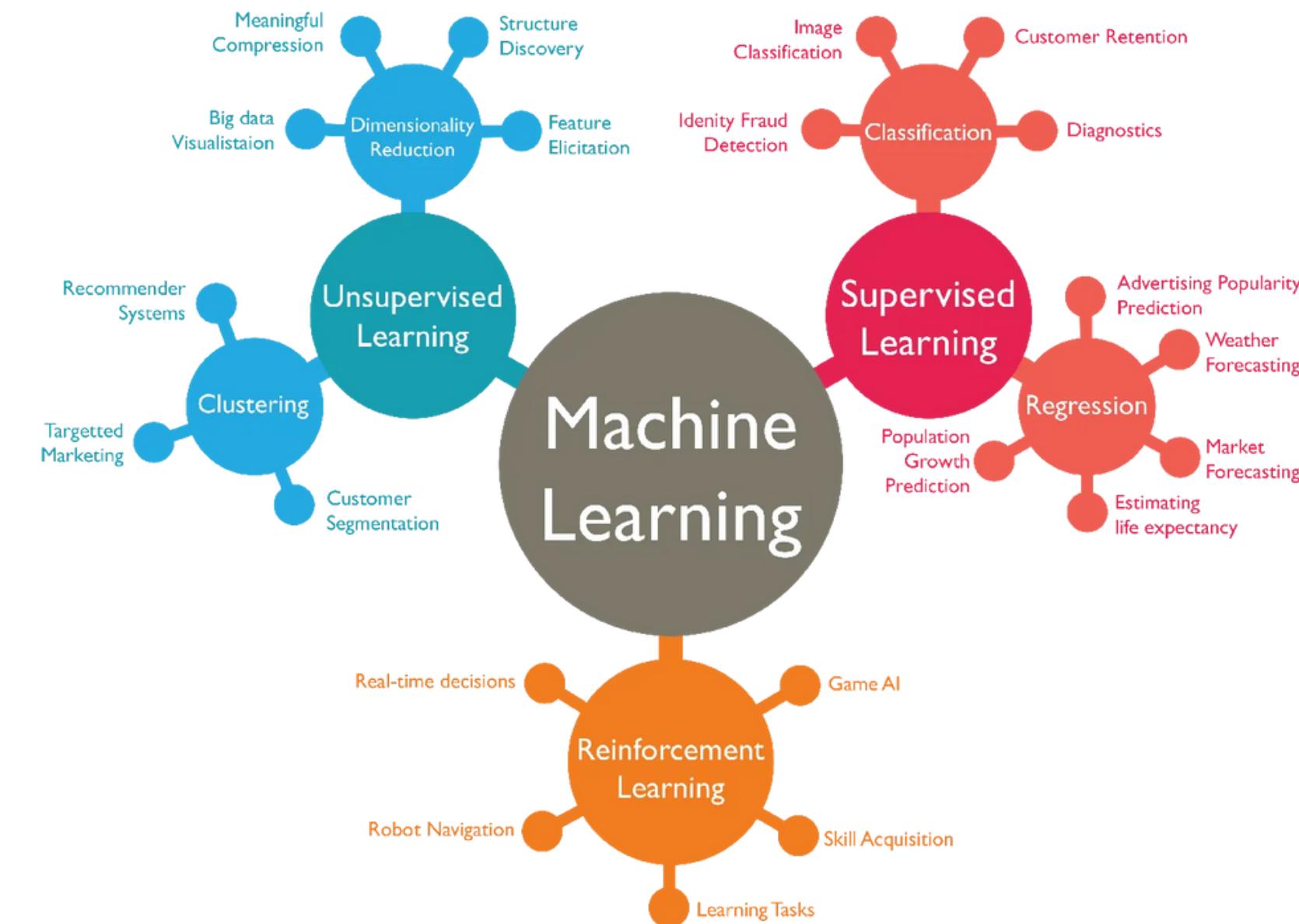
Comment ça marche?

- Introduction au Machine Learning
- Les Réseaux de Neurones Profonds
- Comment la machine apprend?
- Attention is All You Need
- Les Large Language Models
- Intelligence Artificielle Générative

[Retourner à l'Agenda](#)



Machine Learning



Machine Learning

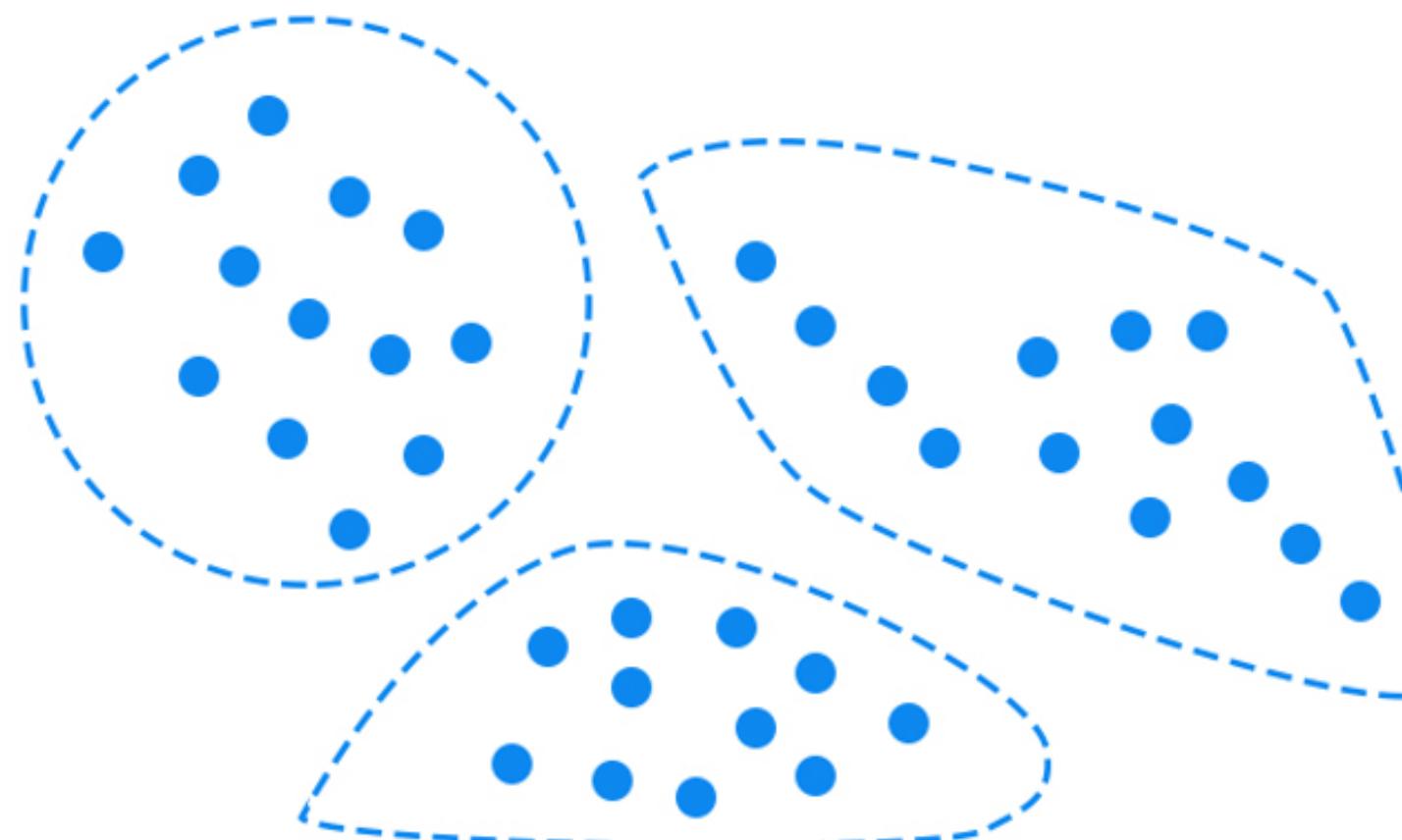
Apprentissage supervisé

- Le Data Scientist enseigne à l'algorithme les conclusions à tirer.
- Cette approche nécessite des données annotées.



Machine Learning

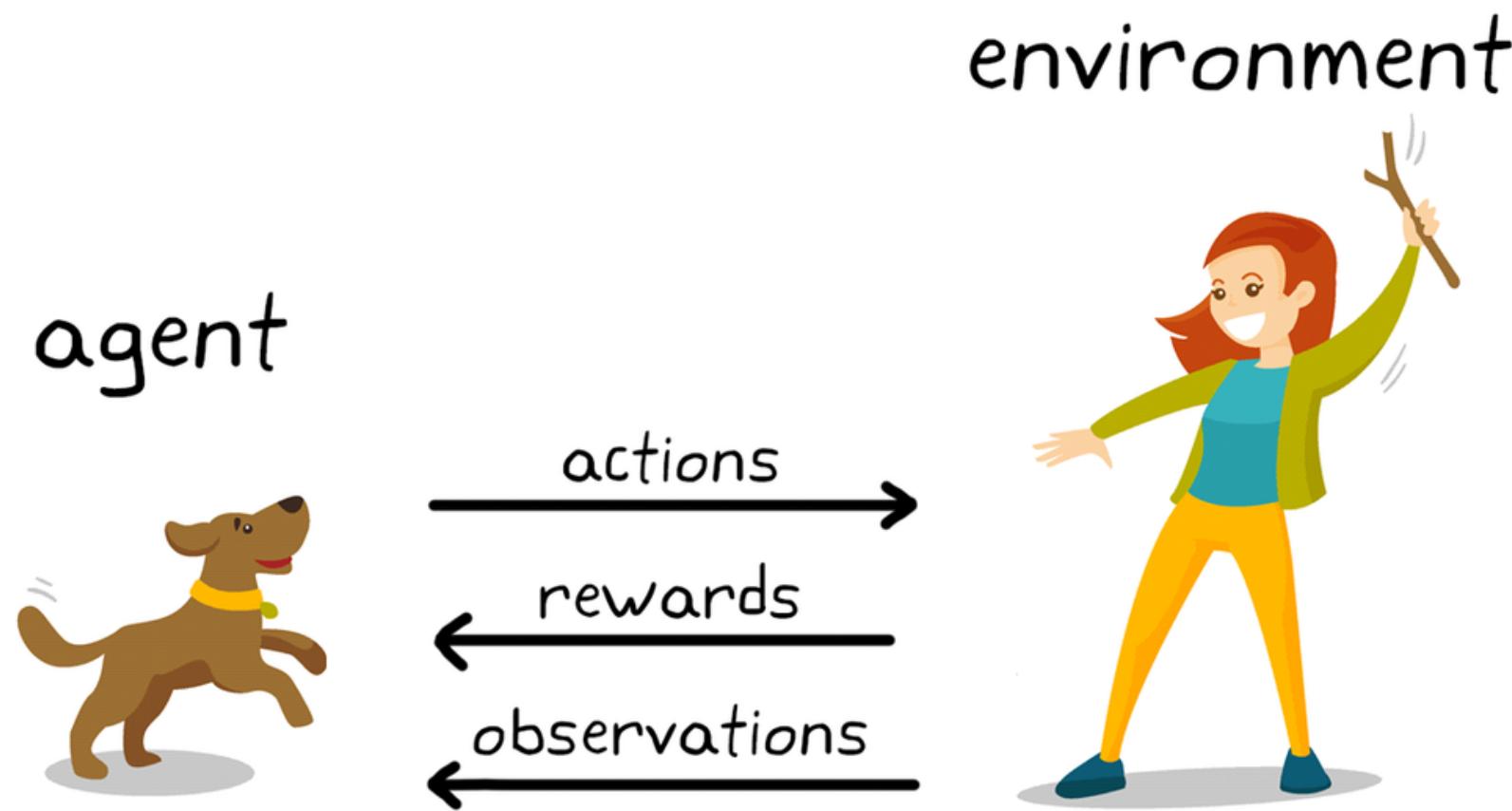
[Retourner à l'Agenda](#)



Apprentissage non supervisé

- L'ordinateur apprend tout seul à identifier des schémas complexes dans les données.
- Pas d'annotation nécessaire.

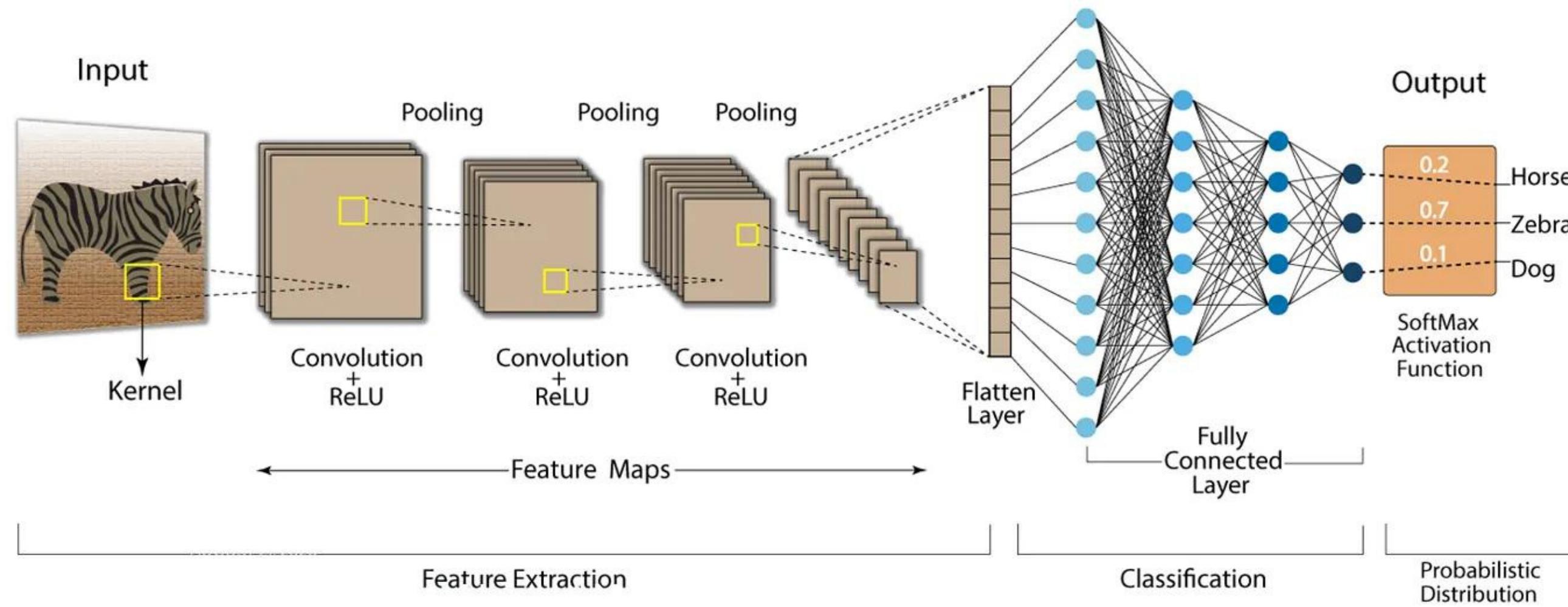
Machine Learning



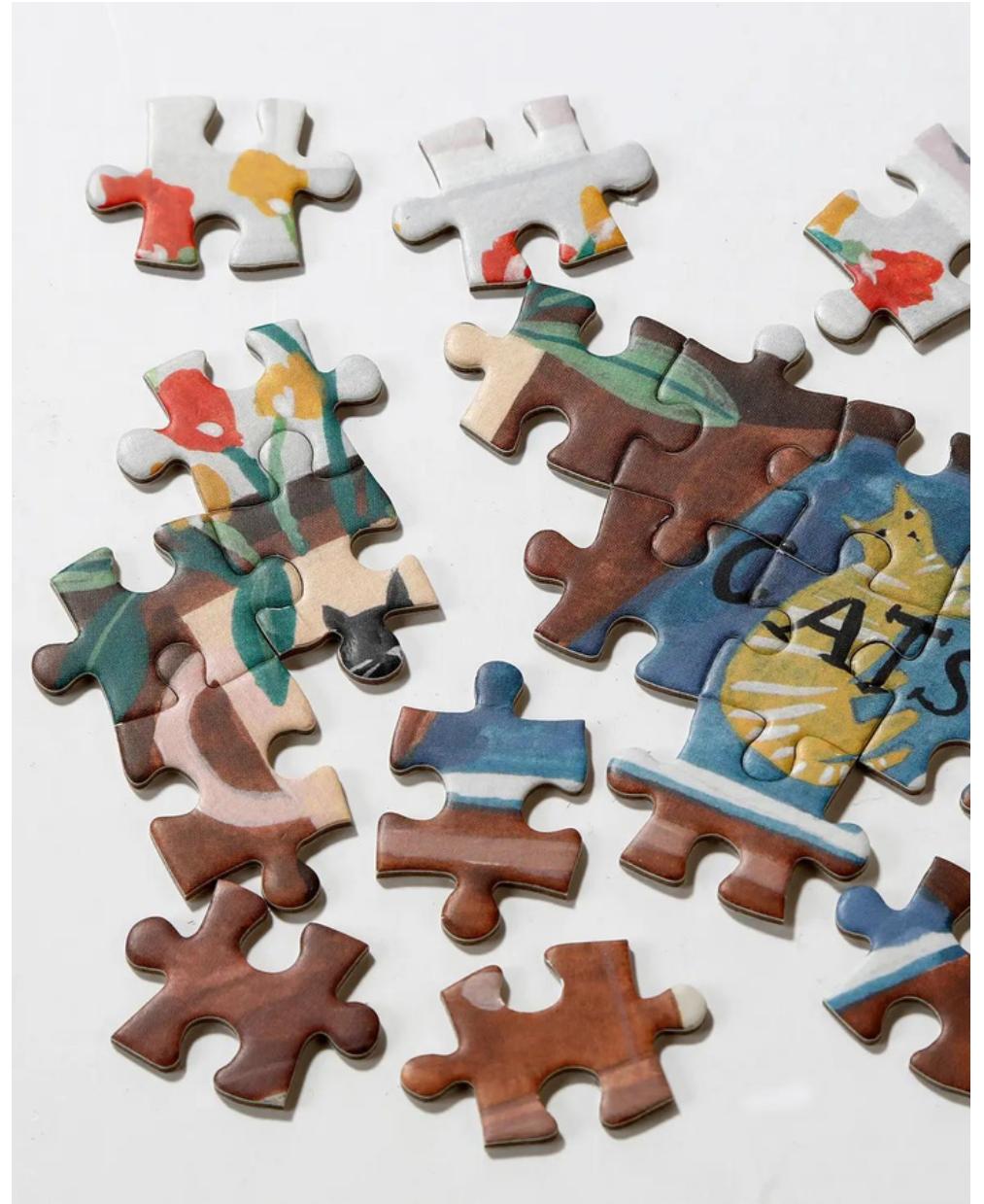
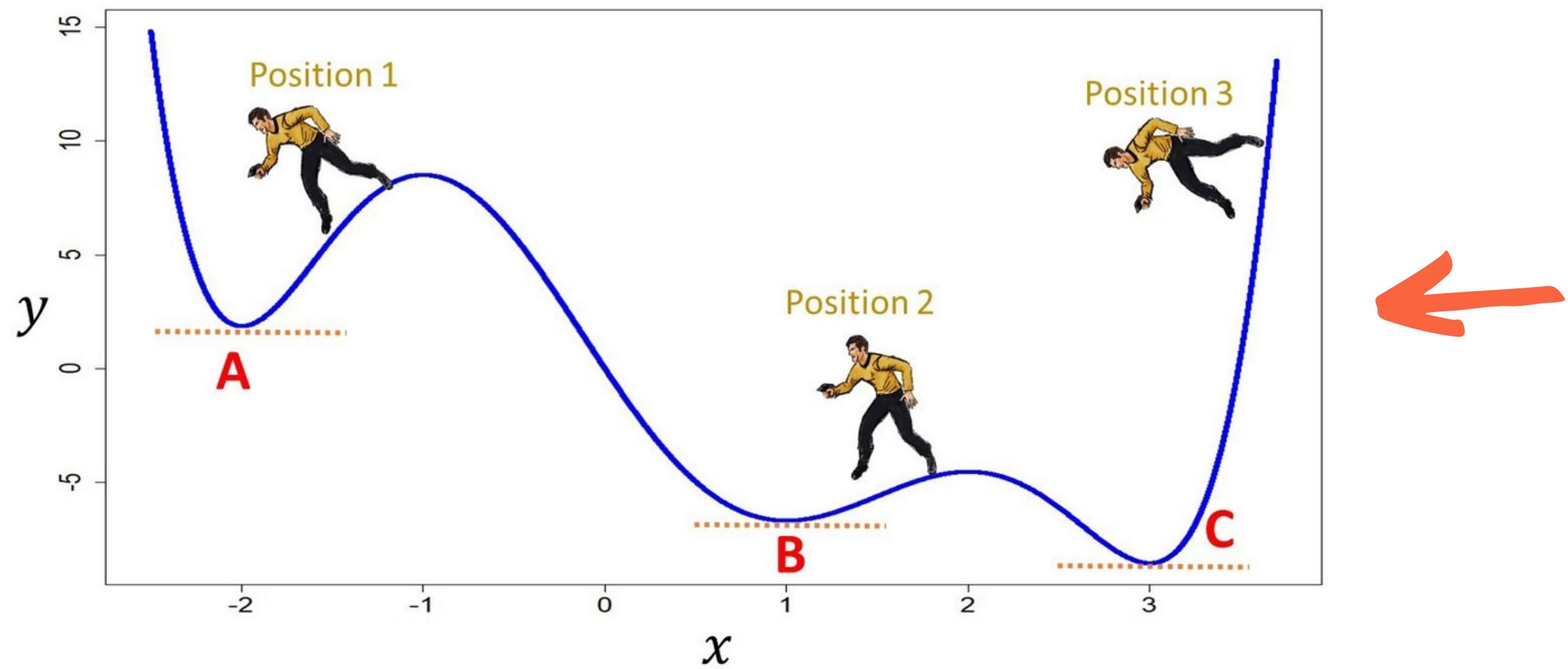
Apprentissage par renforcement

- Un agent apprend les actions à prendre à partir d'exemples.
- Une récompense est donnée à l'agent à chaque fois qu'il entreprend la bonne action.

Les Réseaux de Neurones



La Descente du Gradient



“Attention is All You Need”



[Retourner à l'Agenda](#)

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Llion Jones*
Google Research
llion@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

Noam Shazeer*
Google Brain
noam@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaiser@google.com

Niki Parmar*
Google Research
nikip@google.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.0 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature.

[Retourner à l'Agenda](#)

Les Large Language Models

Encoder la Data

Construire des représentations contextualisées des textes en input.

Générer du texte

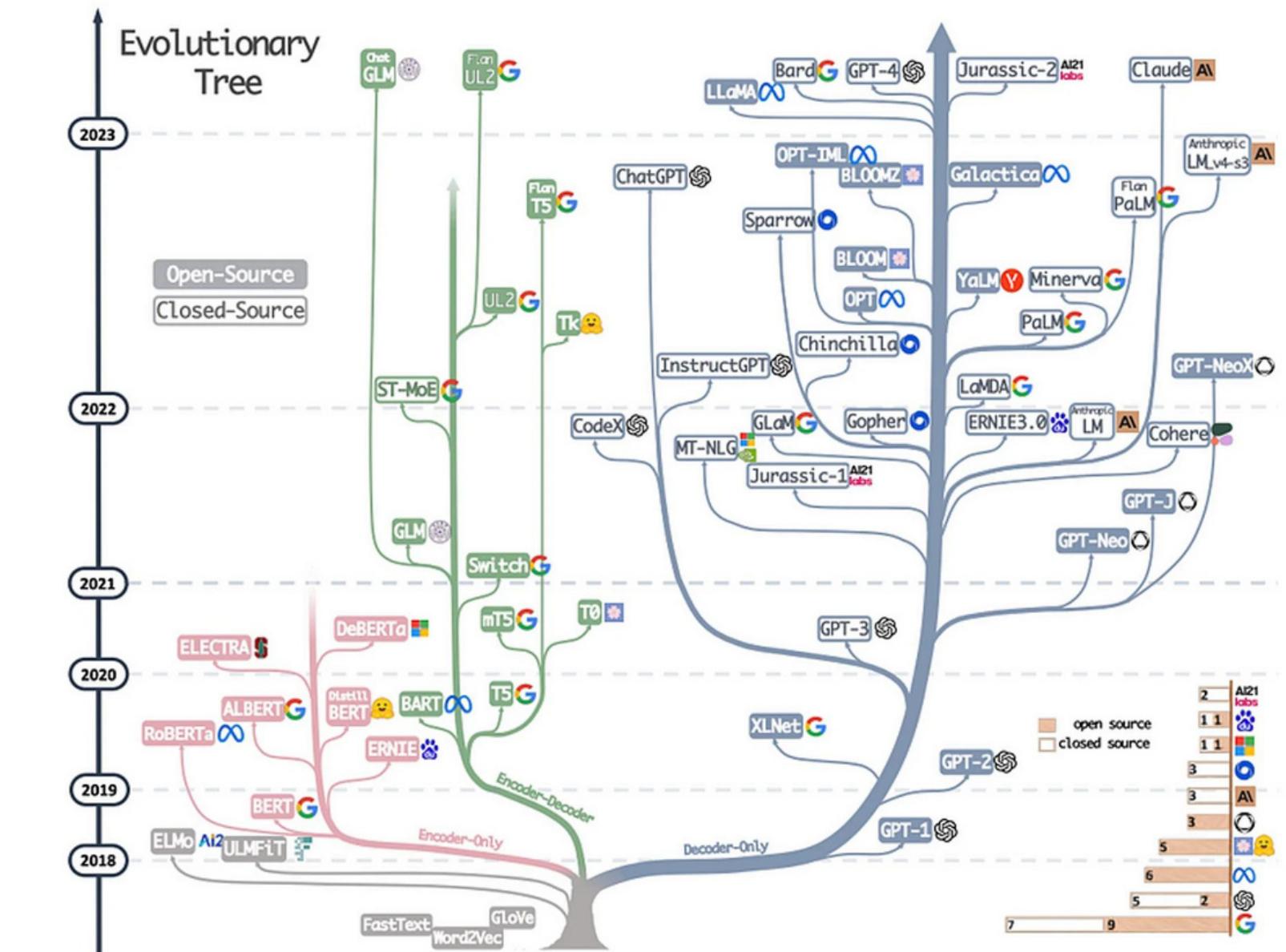
A partir de texte passé dans un prompt d'un chatbot, répondre par du texte

Orienté domaines

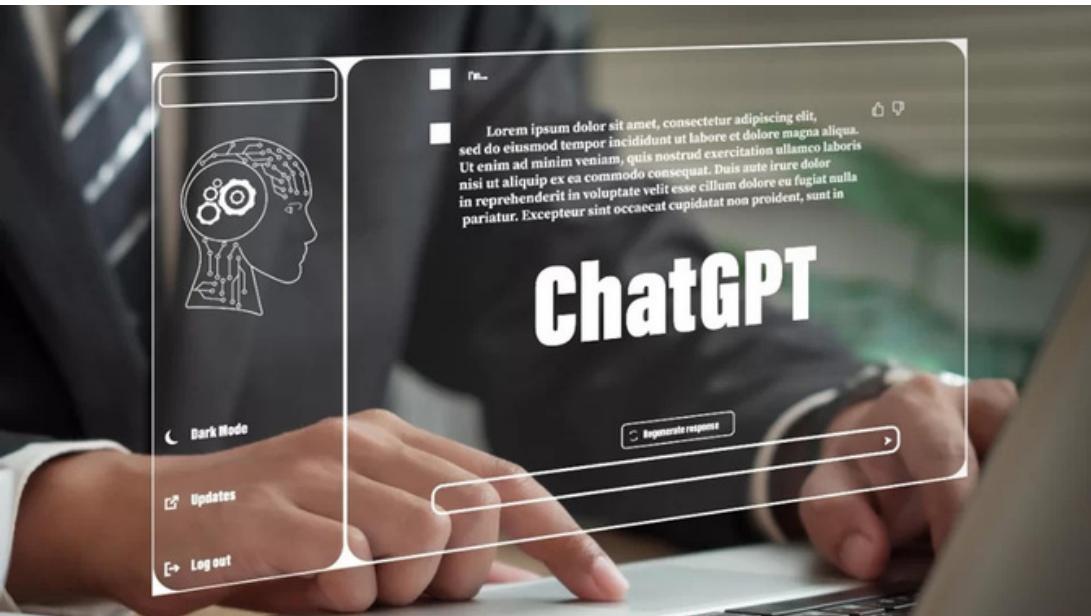
Par exemple Galactica, construit par Facebook pour répondre à des problèmes scientifiques.

Propriétaire vs Communautaire

ChatGPT vs BLOOM



Les IAs Génératives



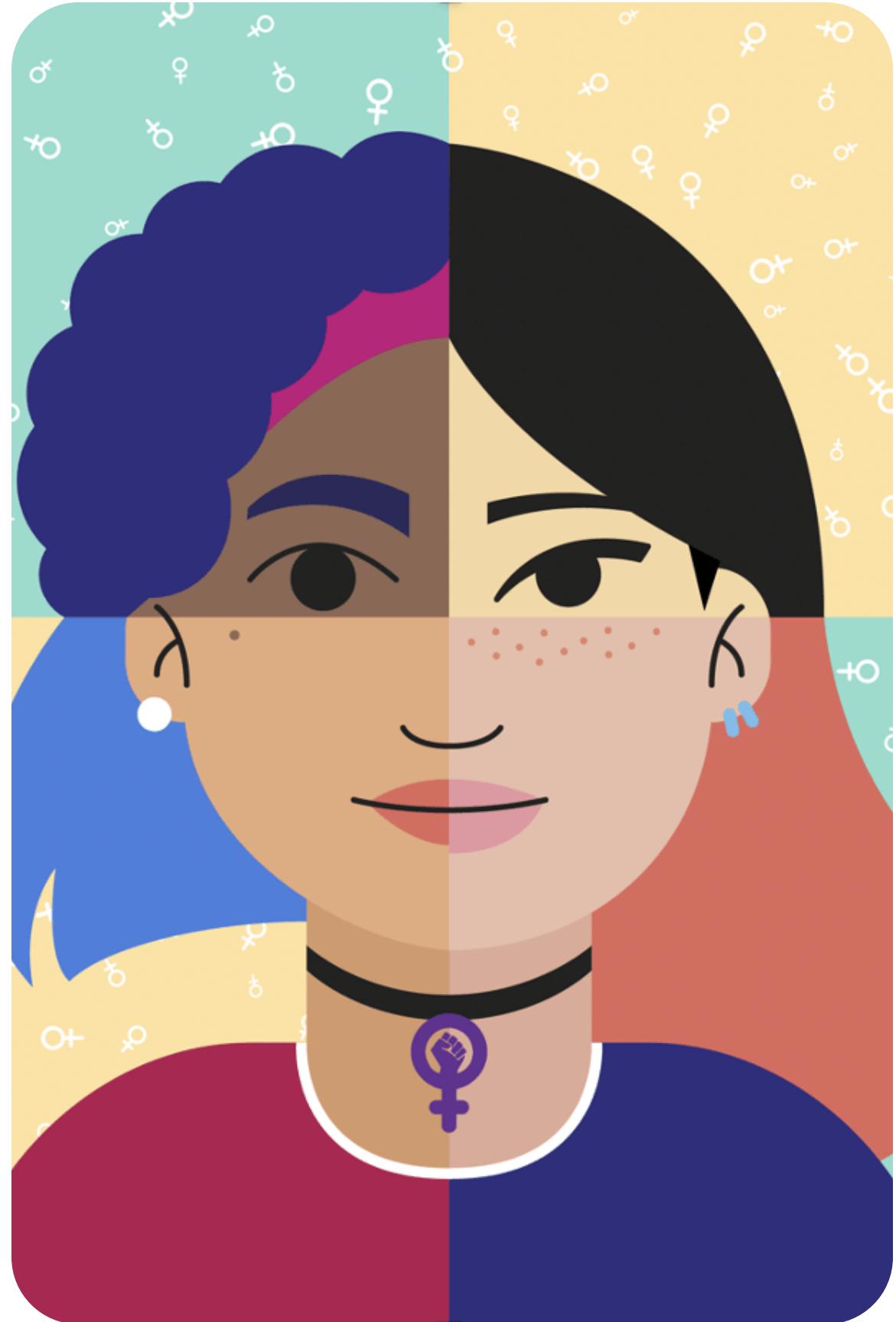
[Retourner à l'Agenda](#)

21

Un biais dans la matrice

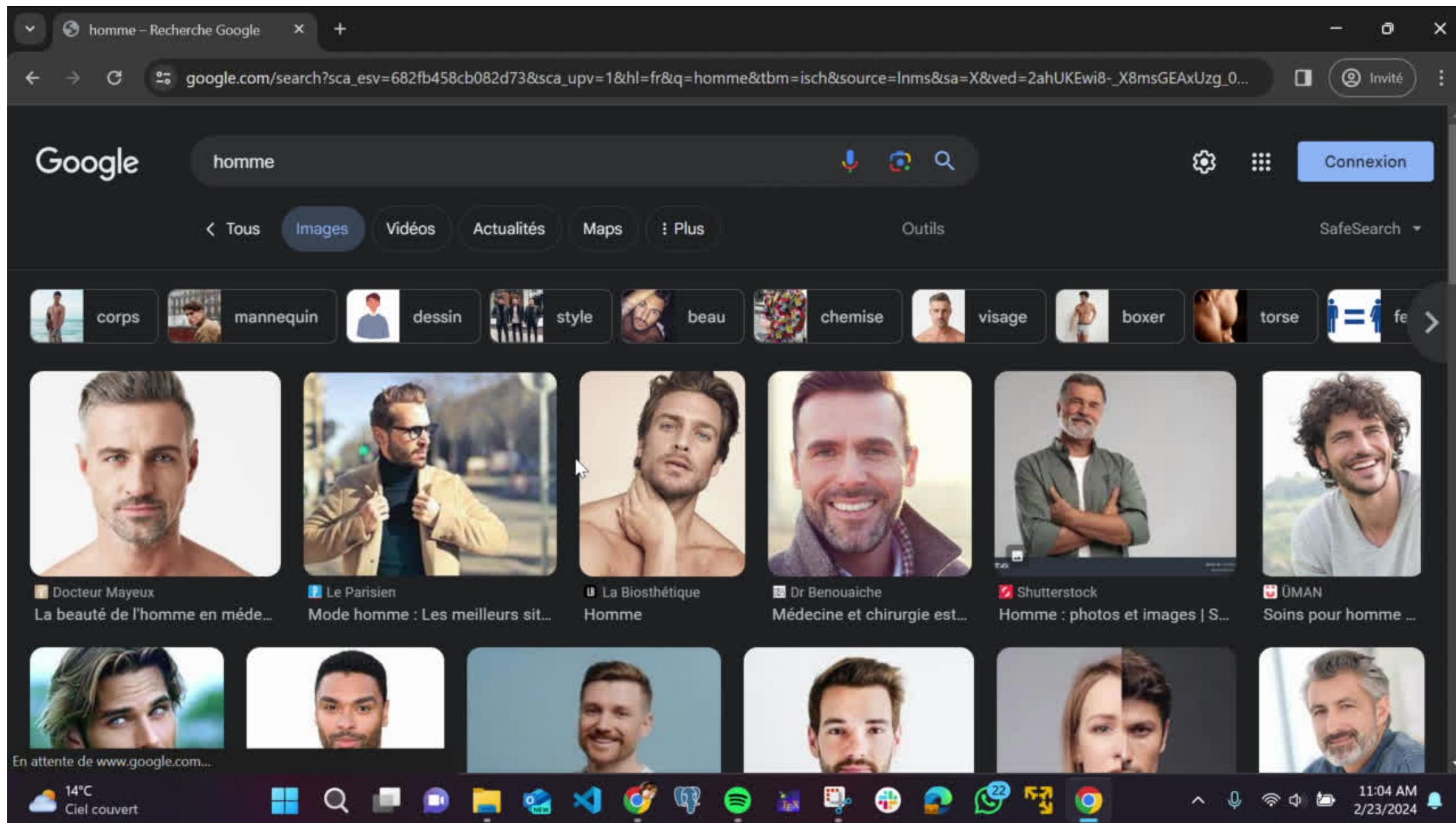
“Le biais dans en IA une situation dans laquelle un système de Machine Learning discrimine un groupe de personnes en particulier.”

[Retourner à l'Agenda](#)



[Retourner à l'Agenda](#)

Le phénomène du biais

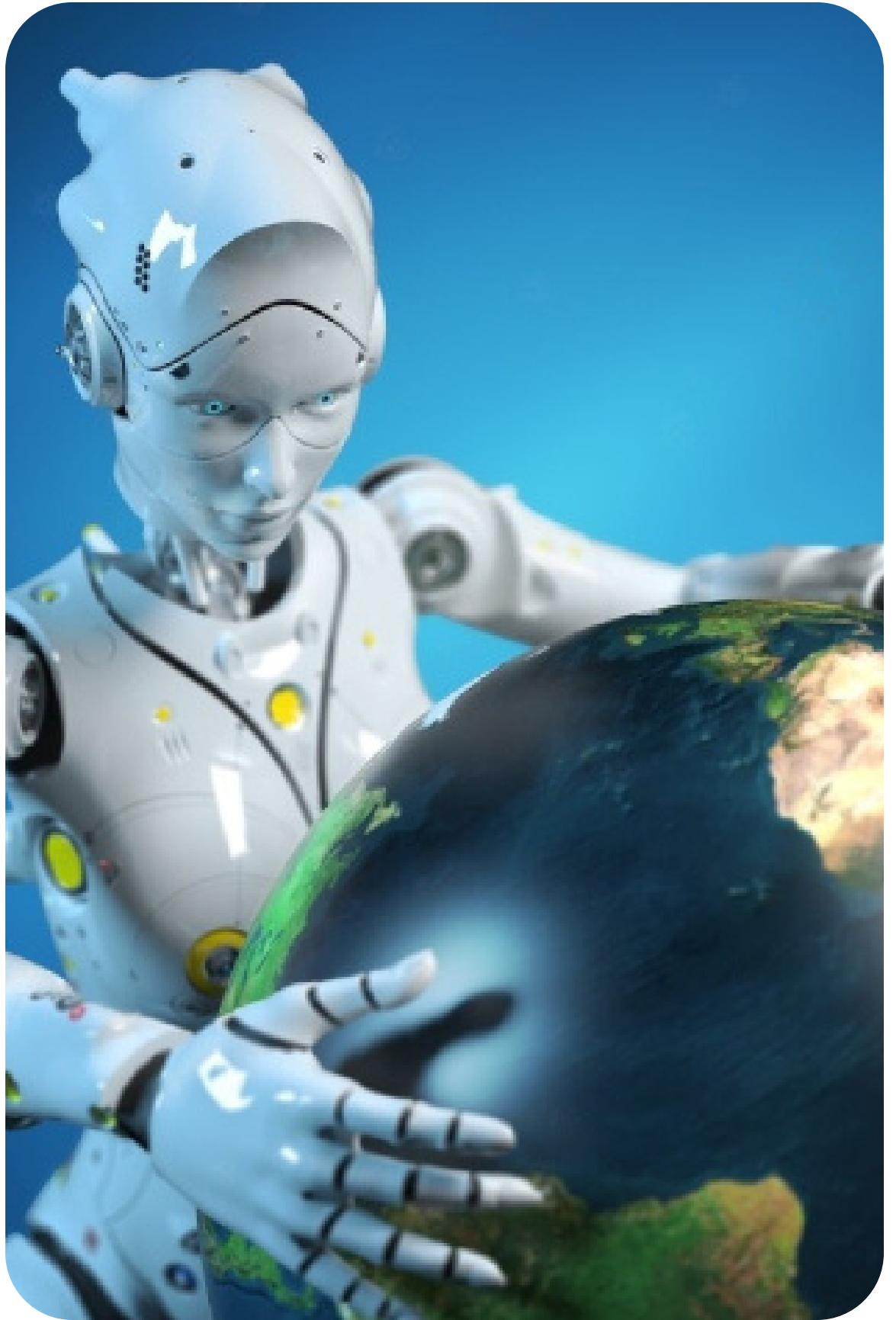


23

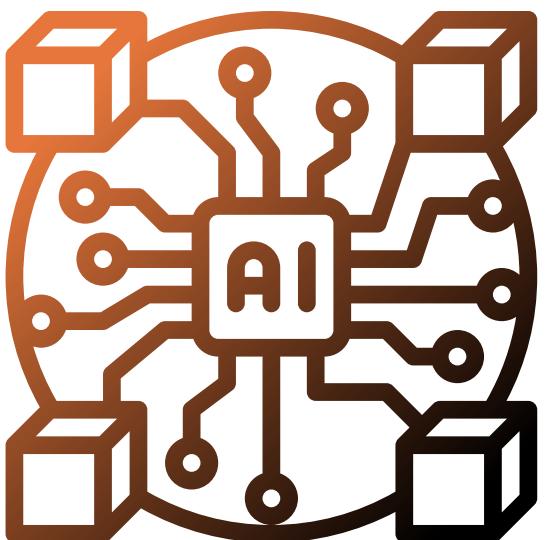
Vers un Goulot d'étranglement?

La profusion d'images générées par des IAs sur Internet pourrait engendrer une sorte de goulot d'étranglement dans le futur.

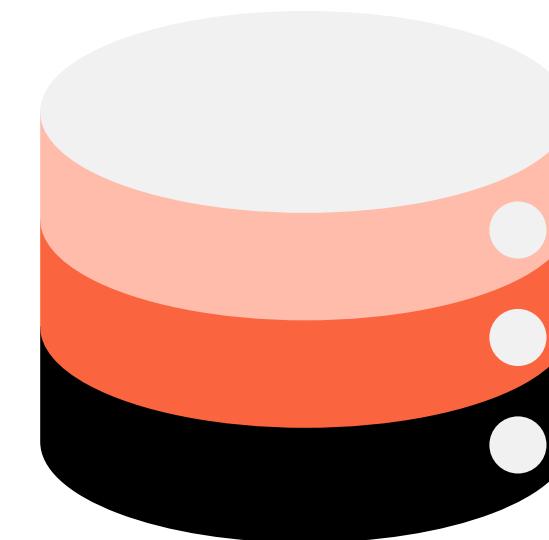
[Retourner à l'Agenda](#)



Vers un Goulot d'étranglement?



IA Générative



Data

25

L'Afrique en Chiffre

“L'Afrique est le berceau de l'humanité.”

[Retourner à l'Agenda](#)



L'Afrique en chiffre

Nombre de pays

54

Population

1,2 Milliards

Alphabétisme

200 Millions d'analphabètes

Age

Espérance de vie : 52 ans / 50% de la population ont moins de 18 ans

Démographie

52 villes > 1 millions d'hab/ En 2050, 50% seront des urbains en 20230

Langues

Environ 3000 langues et dialectes sont parlées sur le continent

Superficie

30 millions de kilomètre carré

Agriculture

60% des réserves de terres cultivables mondiales

Internet

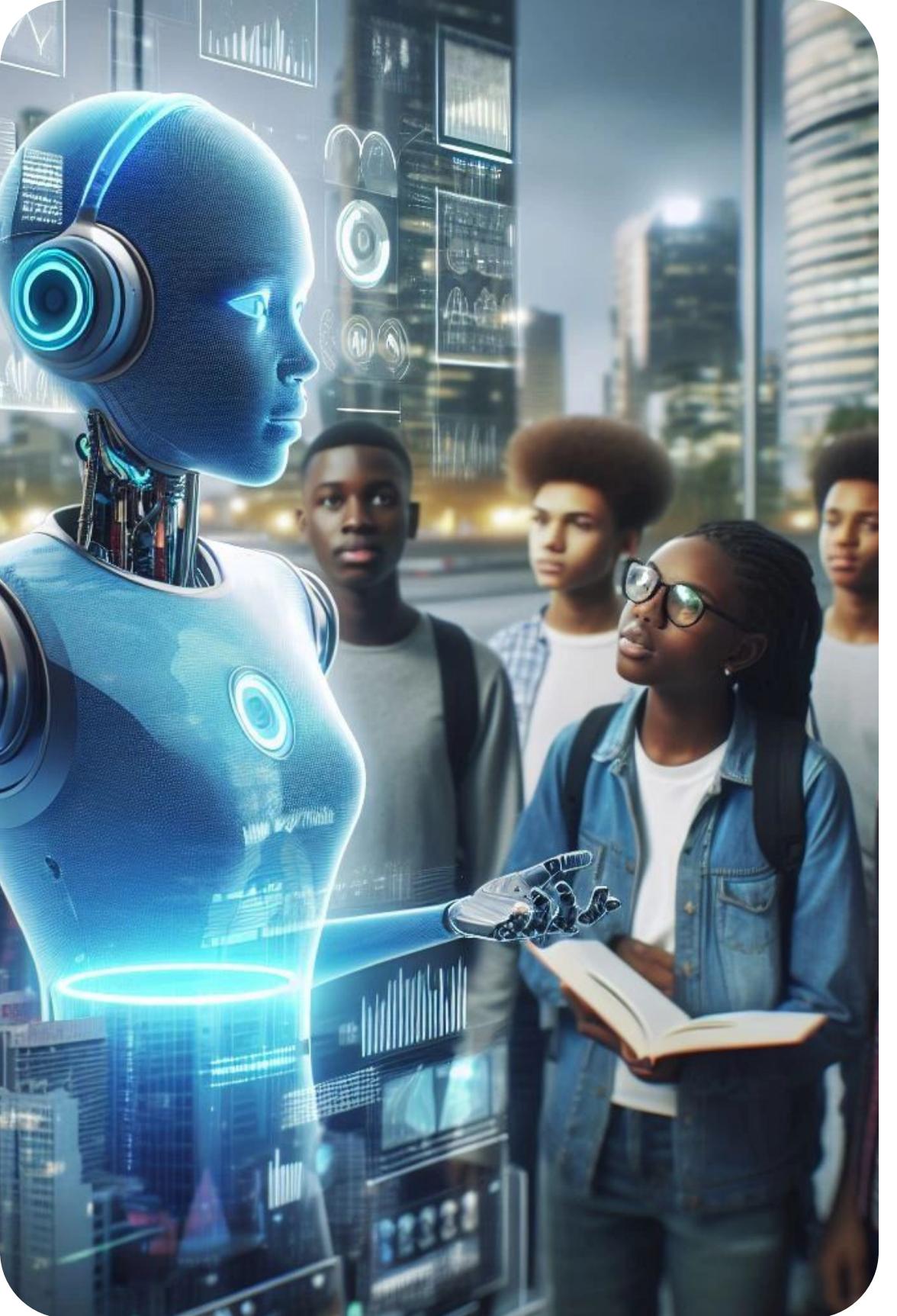
Un taux de pénétration de 43%

27

L'IA en Afrique

- Les raisons du retard
- Les pays qui se démarquent
- Quelques initiatives salutaires

[Retourner à l'Agenda](#)



Les raisons du retard

Priorité

Plusieurs autres actions sont plus prioritaires à entreprendre comme la lutte contre les épidémies, la famine, etc.

Budget

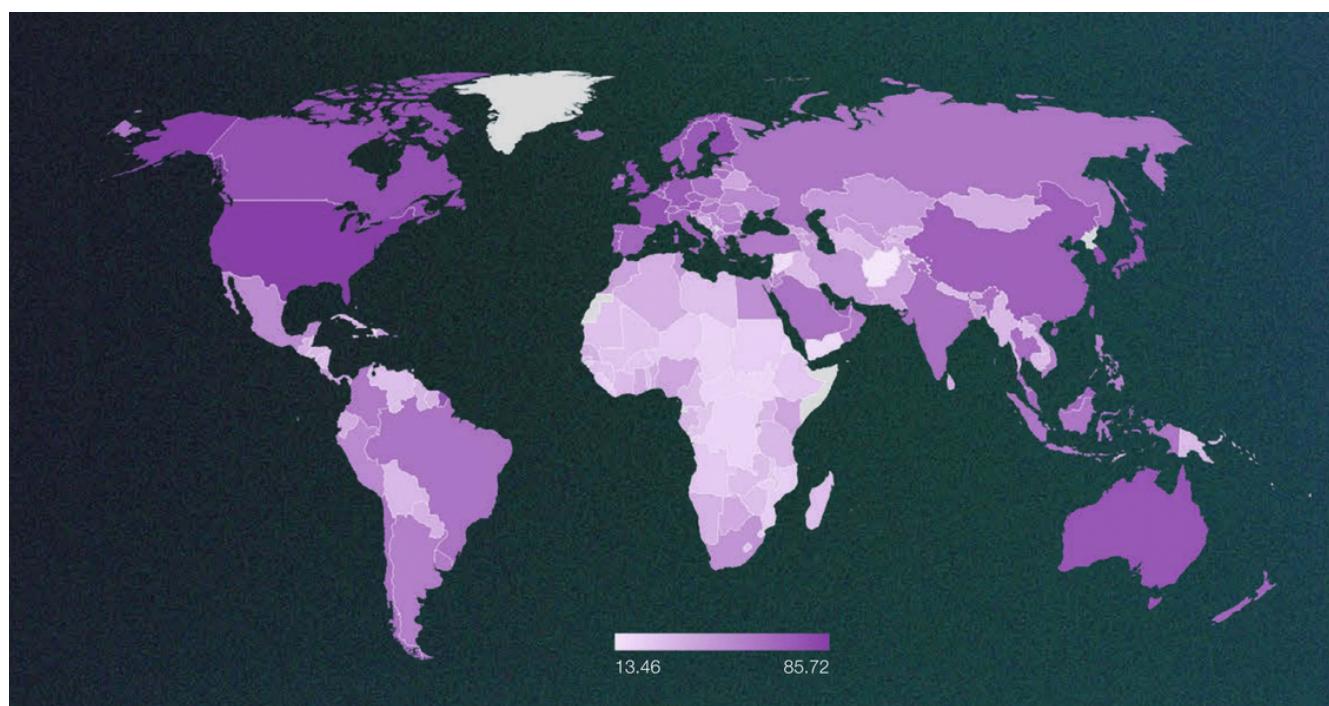
Pour développer des IAs, des budgets souvent faramineux sont nécessaires.

Volonté

Un manque de volonté dû probablement à la non compréhension de l'apport que pourrait apporter l'IA dans la facilitation des quotidiens de tous.

[Retourner à l'Agenda](#)

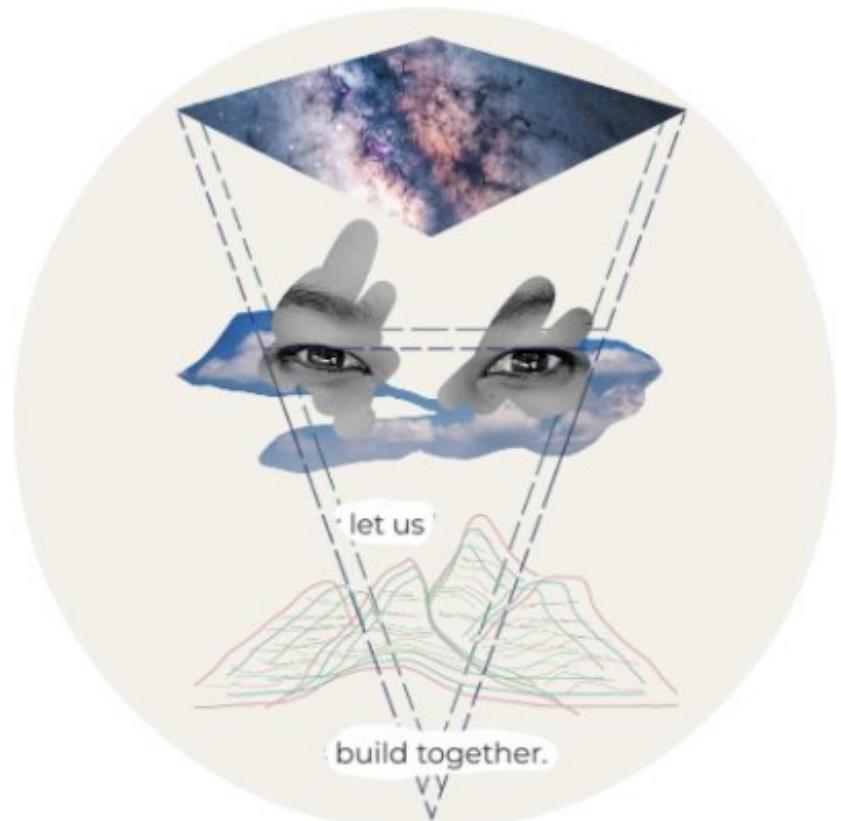
Les pays qui se démarquent



1	United States of America	27	Spain	55	Viet Nam	82	Albania	109	Tajikistan	132	Cambodia	159	Burkina Faso
2	Singapore	28	New Zealand	56	Bahrain	83	Republic of Moldova	110	Paraguay	133	Vanuatu	160	Lesotho
3	United Kingdom	30	Czech Republic	57	Mauritius	84	Armenia	111	Algeria	134	Honduras	161	Eswatini
4	Finland	31	Thailand	58	Romania	85	Fiji	112	Bosnia and Herzegovina	135	El Salvador	162	Ethiopia
5	Canada	32	India	59	Serbia	86	Panama	113	Suriname	136	Côte D'Ivoire	163	Angola
6	Republic of Korea	33	Malta	60	Ukraine	87	Morocco	114	Guatemala	137	Timor-Leste	164	Sierra Leone
7	France	34	Poland	61	Peru	88	Georgia	115	Namibia	138	Cameroon	165	Malawi
8	Australia	35	Chile	62	Mexico	89	Bahamas	116	Nepal	139	Papua New Guinea	166	Comoros
9	Japan	36	Qatar	63	Jordan	90	Kenya	117	Grenada	140	Solomon Islands	167	Niger
10	Netherlands	37	Brazil	64	Slovakia	91	Belarus	118	Cabo Verde	141	Papua New Guinea	168	Guinea Bissau
11	Denmark	38	Lithuania	65	Egypt	92	Pakistan	119	Kyrgyzstan	142	Samoa	169	Sudan
12	Norway	39	Saudi Arabia	66	Croatia	93	Rwanda	120	Saint Vincent and the Grenadines	143	Belize	170	Mozambique
13	Sweden	40	Russian Federation	67	Brunei Darussalam	94	Seychelles	121	Maldives	144	Djibouti	171	Liberia
14	Taiwan	41	Slovenia	68	South Africa	95	Tonga	122	Guyana	145	Togo	172	Chad
15	Germany	42	Hungary	69	Kuwait	96	Barbados	123	Uganda	146	Zambia	173	Burundi
16	Austria	43	Indonesia	70	Tunisia	97	Nigeria	124	Venezuela	147	Eritrea	174	Democratic Republic of the Congo
17	China	44	Bulgaria	71	North Macedonia	98	Botswana	125	Bolivia	148	Zimbabwe	175	Central African Republic
18	Ireland	45	Iceland	72	Kazakhstan	99	Bhutan	126	United Republic of Tanzania	149	Libya	176	Mauritania
19	Estonia	46	Latvia	73	Lebanon	100	Antigua and Barbuda	127	Nicaragua	150	Haiti	177	South Sudan
20	Israel	47	Colombia	74	Azerbaijan	101	Ecuador	128	Togo	151	Gambia	178	Yemen
21	Belgium	48	Cyprus	75	Iran	102	Mongolia	129	Kiribati	152	Myanmar	179	Syrian Arab Republic
22	United Arab Emirates	49	Turkey	76	Montenegro	103	Jamaica	130	Laos People's Democratic Republic	153	Madagascar	180	Afghanistan
23	Switzerland	50	Greece	77	Andorra	104	Ghana	131	Sao Tome and Principe	154	Gabon	181	Guinea
24	Italy	51	Uruguay	78	Costa Rica	105	Sri Lanka	132	Cuba	155	Nicaragua	182	Honduras
25	Portugal	52	Oman	79	Uzbekistan	106	Trinidad and Tobago	133	Mali	156	Turkmenistan	183	Eswatini
26	Luxembourg	53	Argentina	80	Bangladesh	107	Dominican Republic	134	Iraq	157	Saint Lucia	184	El Salvador
		54	Philippines	81	Dominican Republic	108	Benin	135	Congo	158	Cuba	185	Yemen

Quelques initiatives salutaires

[Retourner à l'Agenda](#)

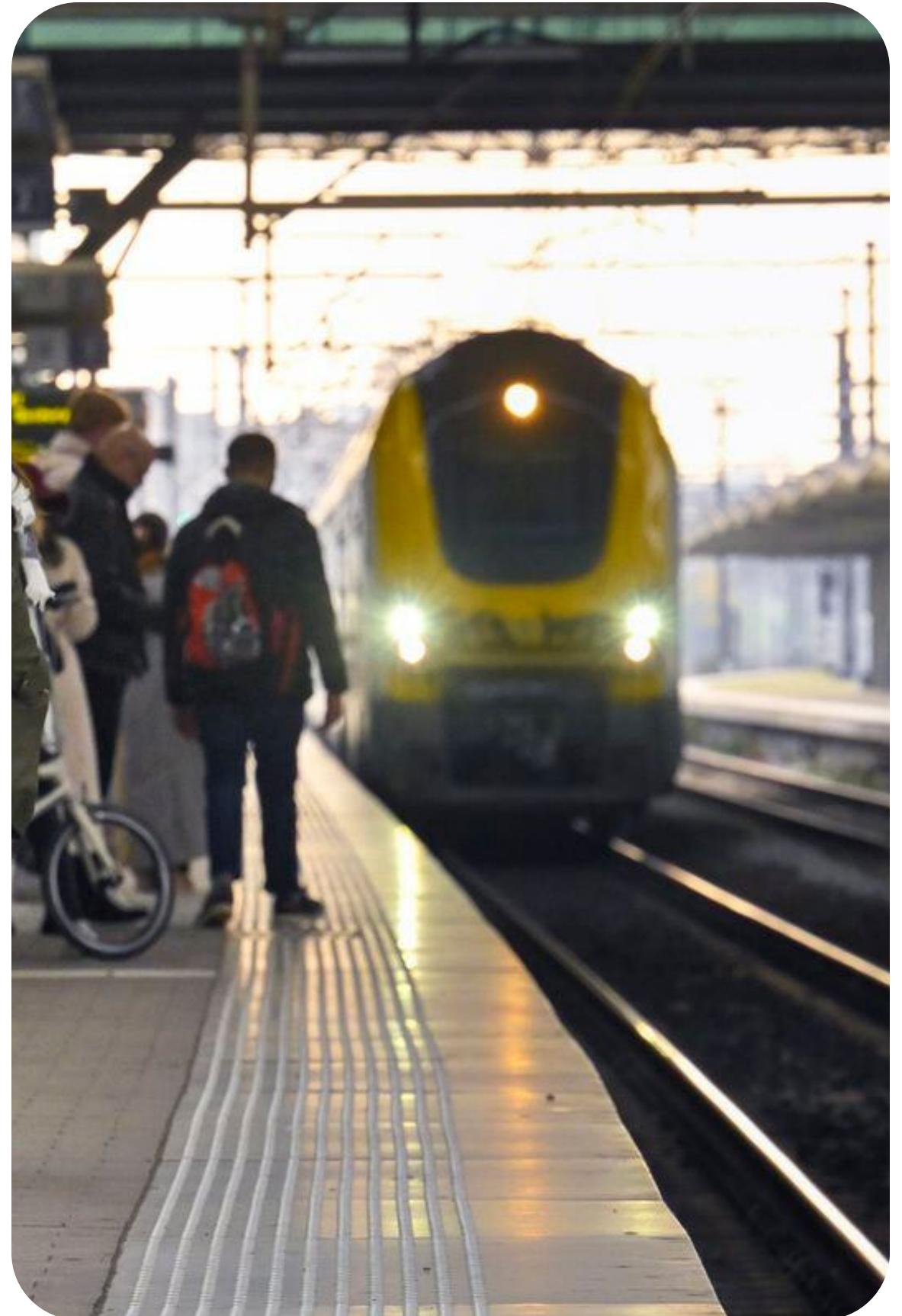


31

Rattraper le train en marche

- Adopter des approches innovantes
- Investir dans l'éducation
- Développer d'autres domaines connexes
- Mettre en place des stratégies nationales

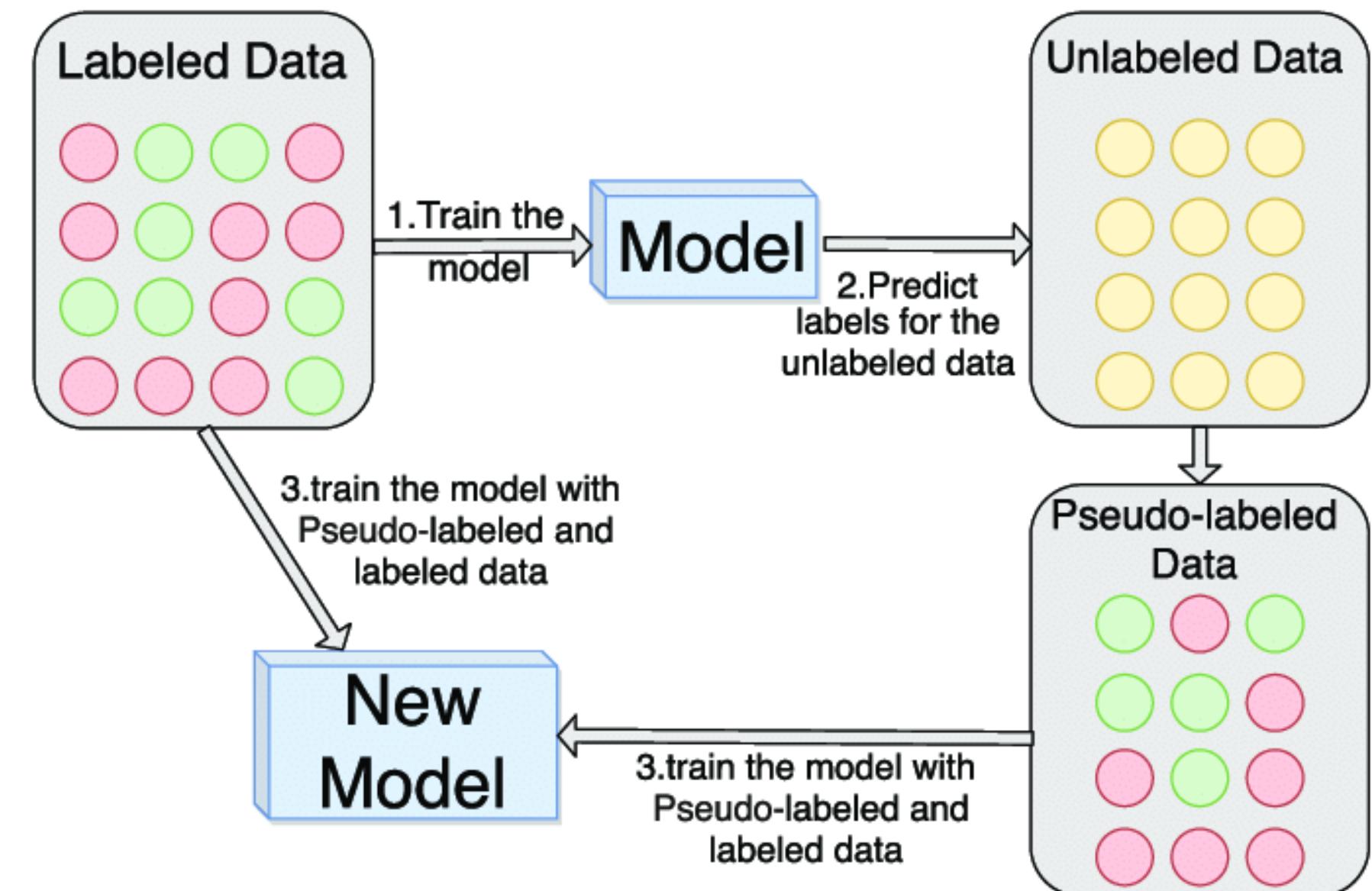
[Retourner à l'Agenda](#)



Approches innovantes

Approches Data-Centric

- Augmentation des données pour régler le problème de manque de données.
- Enrichir les corpus à l'aide de méthodes de pseudo-labeling.

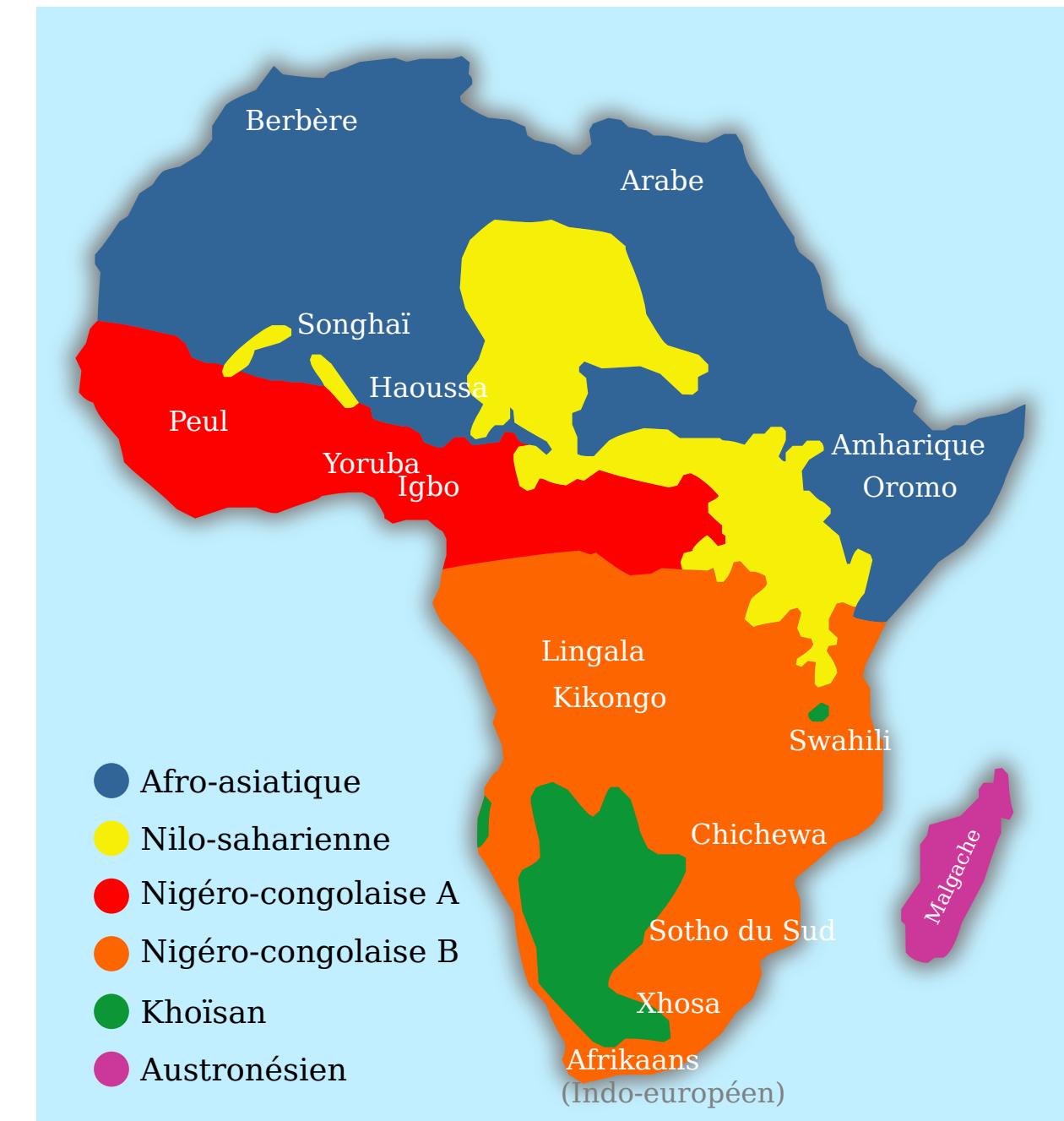


Approches innovantes

[Retourner à l'Agenda](#)

Approches Interlingues

- Utiliser la proximité entre les langues appartenant dans la même famille pour améliorer l'état de l'art des langues peu représentées.
- Recourir aux architectures des Modèles de Langages Multilingues.



34

Investir dans l'éducation



Enseignement secondaire

Intégrer l'enseignement de l'informatique chez les plus jeunes, Proposer des approches multi-domaines, Encourager la mise en place de travaux pratiques, etc.



Investir dans la recherche

Développer les laboratoires de recherche, Adopter des approches R&D, Nouer des partenariats avec d'autres structures à l'international, etc.

[Retourner à l'Agenda](#)

Domaines connexes



Linguistique

Mettre en place des initiatives de normalisation des langues africaines, Valoriser ces langues en incitant les gens à les parler au détriment des langues occidentales, etc.



Sciences Naturelles, Sociales, etc.

L'IA permet aujourd'hui de résoudre des problèmes complexes liés aux Sciences Naturelles, Sociales et d'autres domaines.

[Retourner à l'Agenda](#)

36

Stratégies nationales

“Aujourd’hui, le Sénégal reste le seul pays africain qui a une feuille de route nationale sur l’Intelligence Artificielle.”

[Retourner à l’Agenda](#)



2023

Stratégie nationale et feuille de route du Sénégal sur
l’Intelligence Artificielle
(VERSION RÉSUMÉE)



37

Vers une IA plus durable

Se servir de l'IA pour régler les problèmes impactant
l'Environnement, le Social et la Gouvernance.

[Retourner à l'Agenda](#)



IA durable



Environnement

Comprendre et lutter contre les impacts négatifs des activités humaines sur l'environnement, etc.



Social

Utiliser l'IA pour résoudre les crises sociales observées sur le continent, réduire les inégalités, etc.

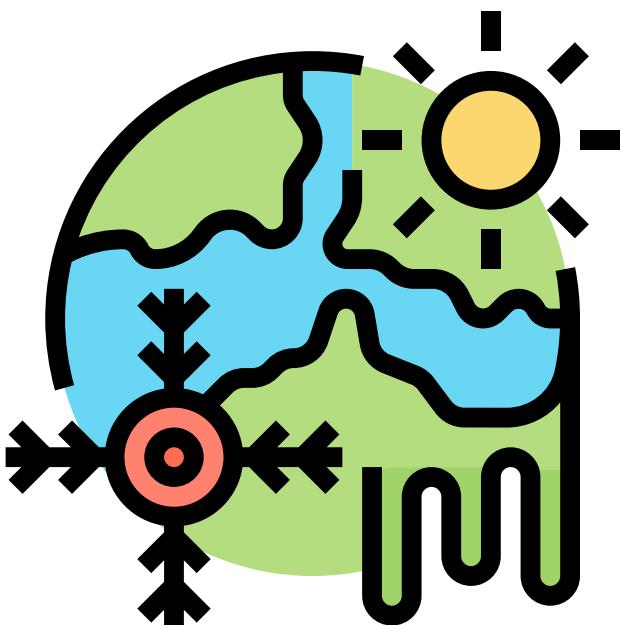


Gouvernance

Encourager la transparence sur la gouvernance des entreprises et des institutions publiques, etc.

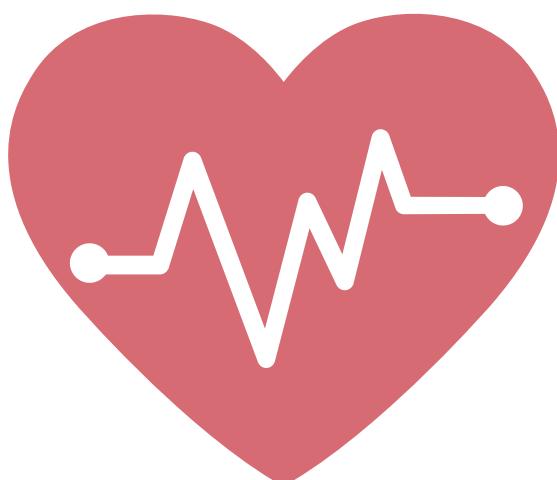
39

IA durable



Mesurer l'impact climatique des activités industrielles

Déetecter et mesurer le type de gaz éjectée par les compagnies d'exploitation pétrolière.



Assistance aux médecins, lutte contre les maladies

Détection des complications prénatales, des cancers, modélisation des protéines.

[Retourner à l'Agenda](#)

40

IA durable



Faciliter l'octroi l'aide humanitaire à l'aide de l'IA

Permettre aux décideurs de retrouver les zones les plus impactées par des aléas naturelles ou socio-politiques, Permettre aux populations impactées de mieux exprimer leurs besoins.



Booster l'apprentissage au niveau des plateformes e-learning

Traduire instantanément en langues locales toute la documentation en ligne.

[Retourner à l'Agenda](#)

Merci!

Des questions?

INTELLIGENCE ARTIFICIELLE EN AFRIQUE : COMMENT RATTRAPER LE TRAIN EN MARCHE ?

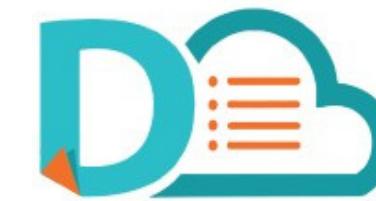
Enjeux et perspectives de l'IA en Afrique



ABDOU MOHAMED Naira

Lead Data Scientist - ToumAI Analytics
Doctorant en Systèmes Intelligents - Institut
National de Statistique et d'Economie
Appliquée (INSEA)

Journée **Data X AI**



جامعة محمد الأول بوجدة
UNIVERSITE MOHAMMED PREMIER OUJDA
+٥٣٨٠٦٤٢٤٩٨٠٦٧٦٥٨