

The CNRS logo is a white circle containing the lowercase letters 'cnrs' in a dark blue, sans-serif font. It is positioned in the upper right quadrant of the slide, overlapping a network of yellow and purple nodes and lines. The background is a dark blue gradient with a network of yellow and purple nodes and lines, and a bright yellow glow on the left side. There are also small purple plus signs in the top right and bottom right corners.

cnrs

Extraire de l'information dans la littérature scientifique avec TDM Factory

ANF IA-TDM 2025

Lucas ANKI et Léo GAILLARD

→ Institut de l'information scientifique et technique

Programme

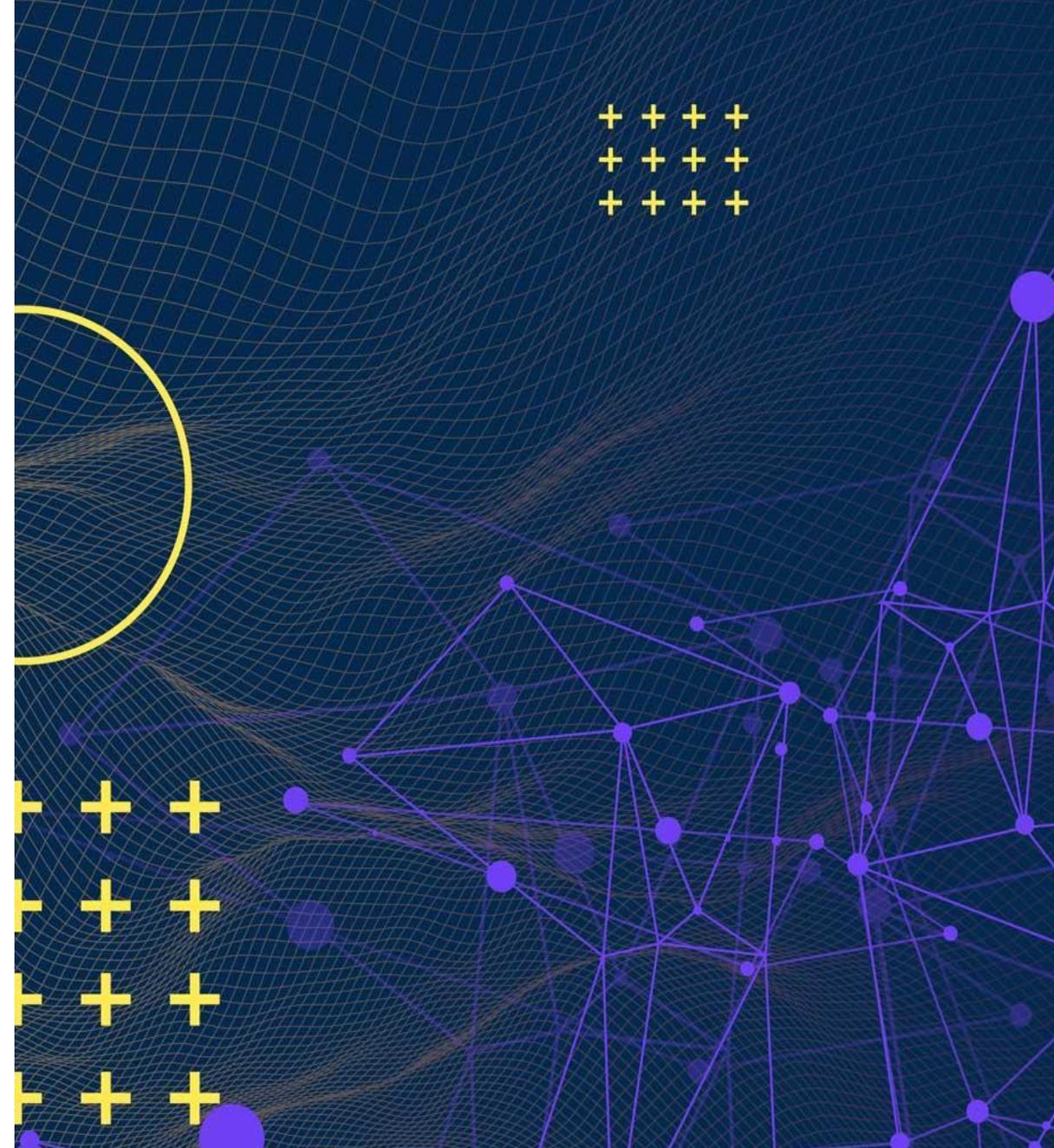
1. Web services : utilisation de petits LLM adaptés

- a. Qu'est-ce qu'un web service ?
- b. Fine tuning de petits LLM

2. TDM Factory : extraire de l'information dans la littérature scientifique avec de l'intelligence artificielle

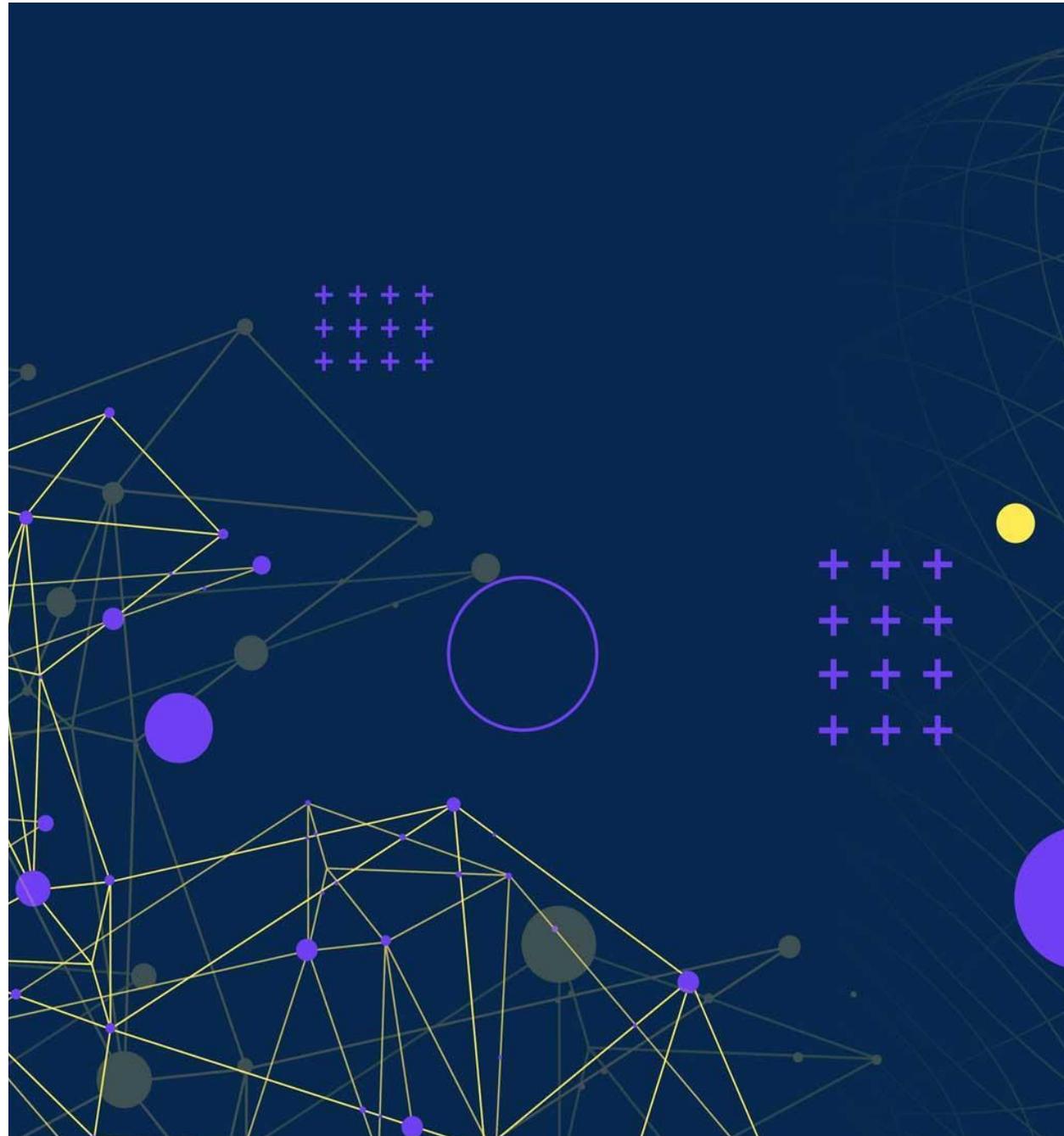
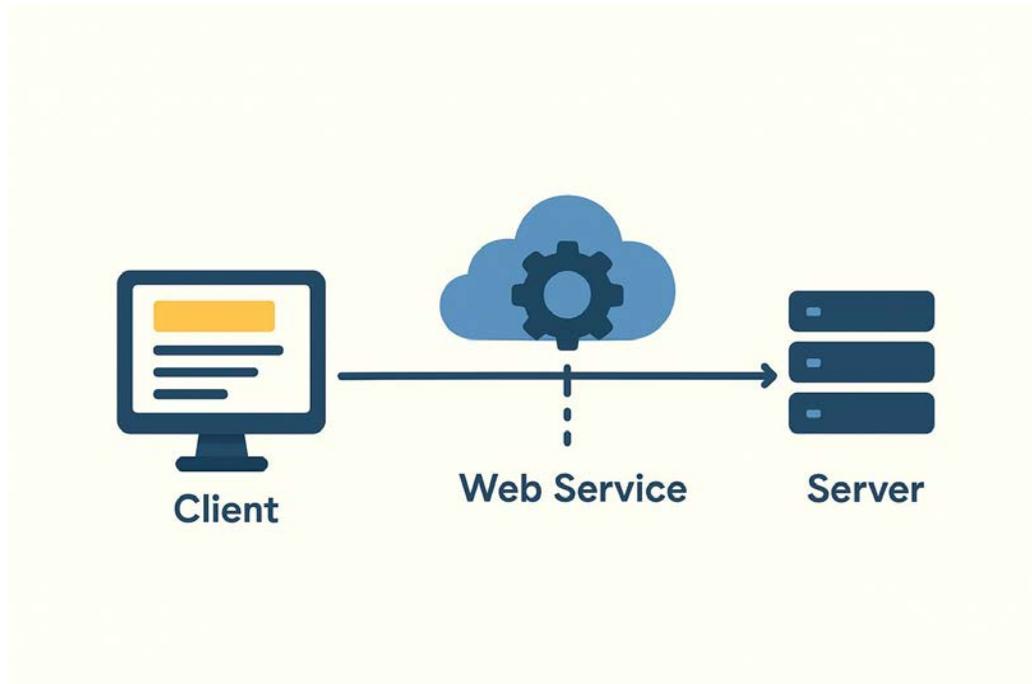
- a. Présentation de l'interface TDM Factory
- b. Démonstrations : exemples de web services
 - i. astroTagger
 - ii. aiAbstractCheck
 - iii. textSummarize
 - iv. bibCheck

3. Istex TDM : présentation du catalogue de l'ensemble des web services



1. Web services : utilisation de petits LLM adaptés

a. Qu'est-ce qu'un web service ?



1. a. Qu'est-ce qu'un web service ?

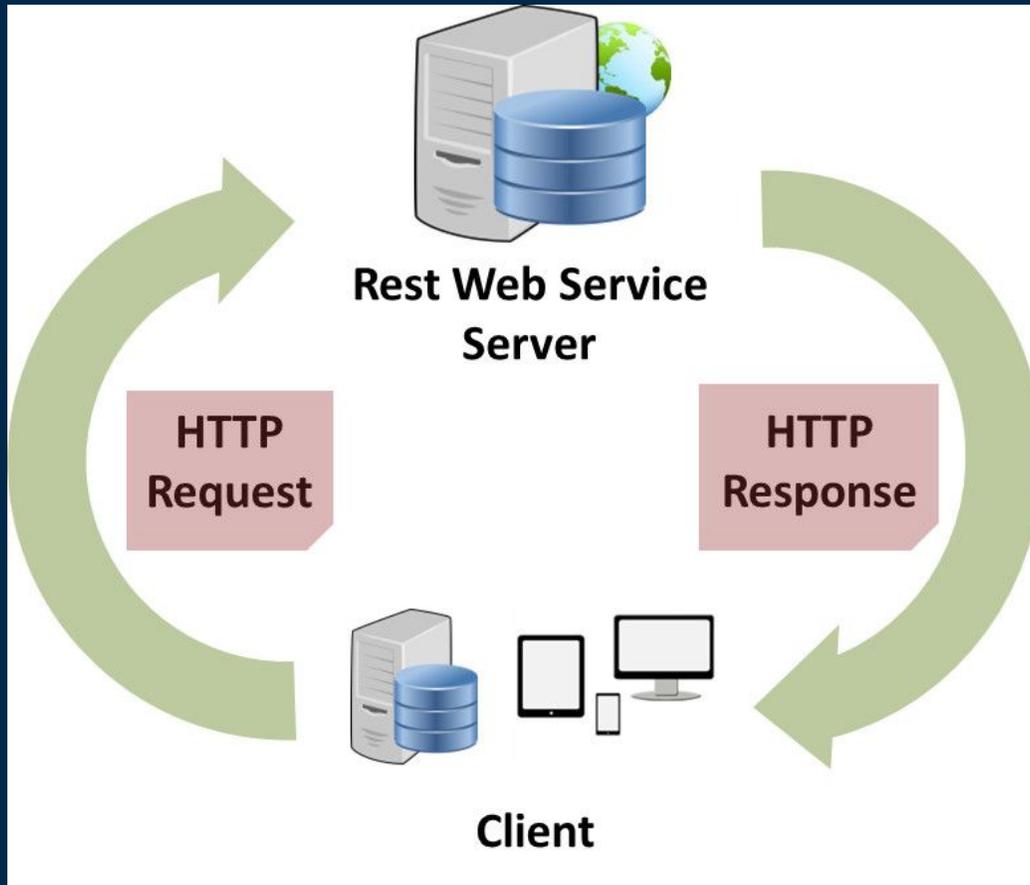
Qu'est ce qu'un web service

Un web service est une application accessible via Internet qui permet la communication entre différentes applications, quel que soit leur langage de programmation ou leur plateforme.

Il expose des fonctionnalités ou données à travers des protocoles standards comme HTTP.

Principe de fonctionnement

Avantages



1. a. Qu'est-ce qu'un web service ?

Qu'est ce qu'un web service

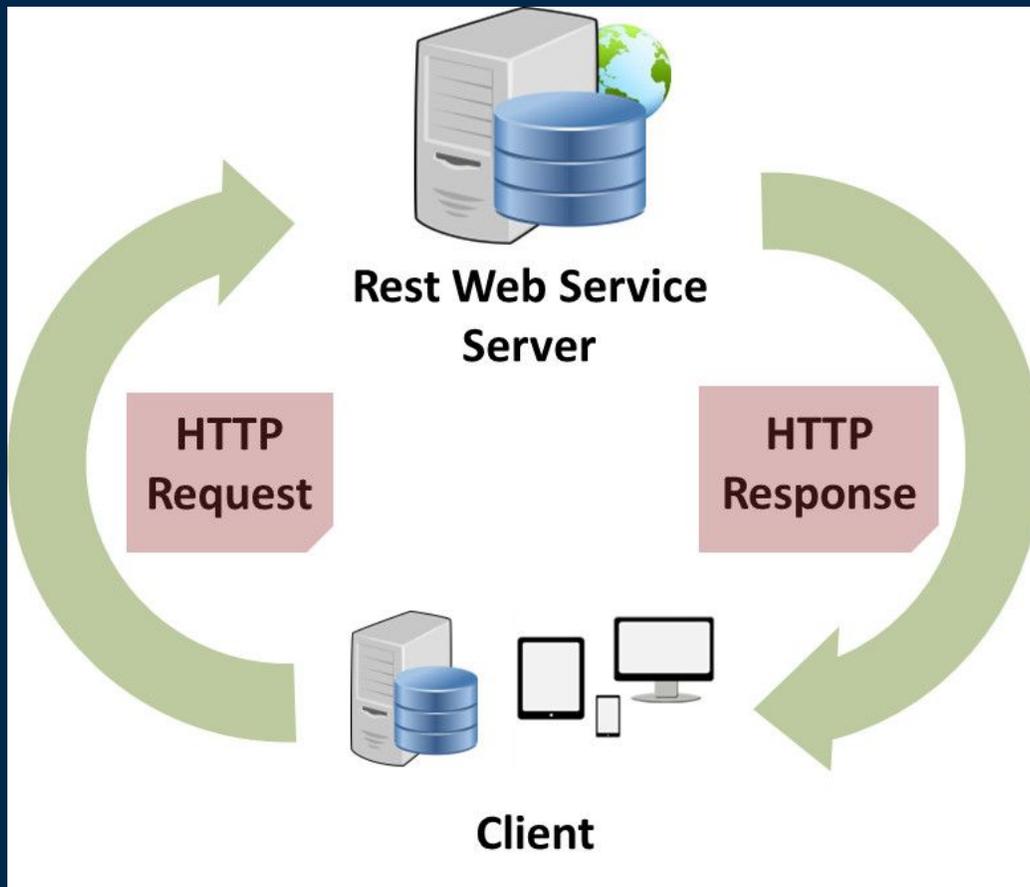
Un web service est une application accessible via Internet qui permet la communication entre différentes applications, quel que soit leur langage de programmation ou leur plateforme.

Il expose des fonctionnalités ou données à travers des protocoles standards comme HTTP.

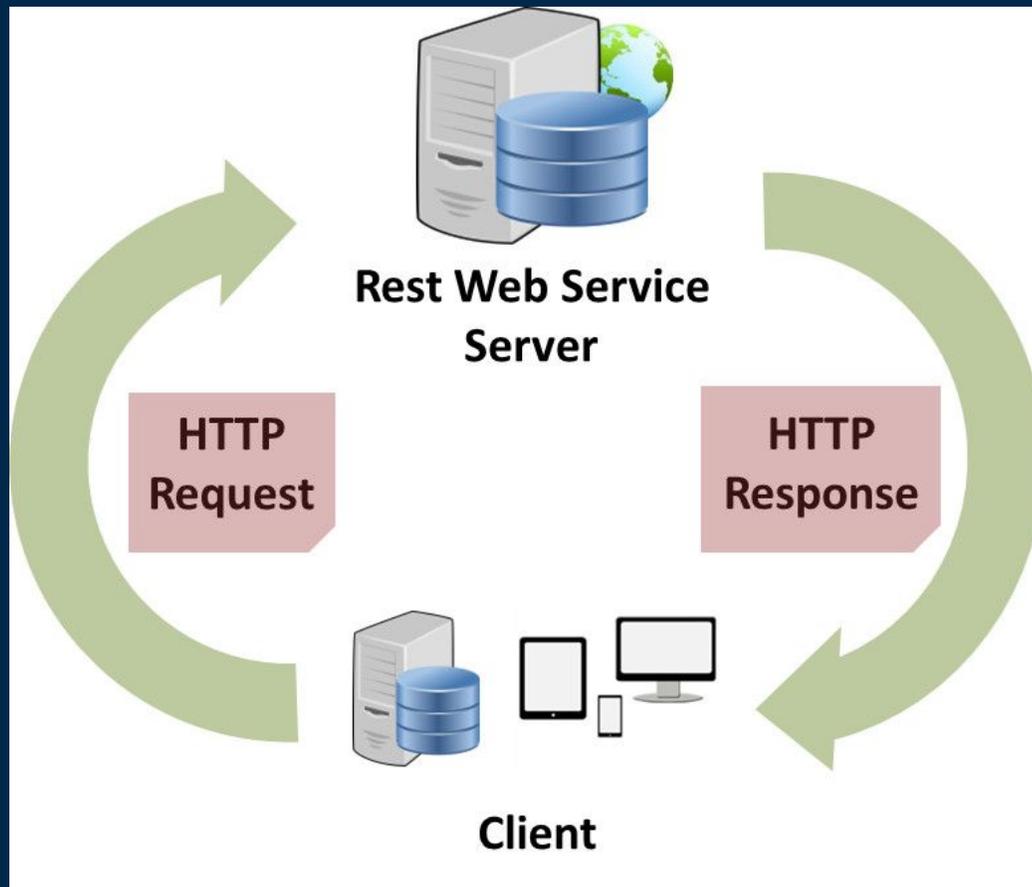
Principe de fonctionnement

- Le client envoie une requête HTTP au serveur du web service.
- Le serveur traite la demande, exécute l'action demandée et renvoie une réponse HTTP (HTTP Response).
- Les données échangées sont souvent au format JSON ou XML.

Avantages



1. a. Qu'est-ce qu'un web service ?



Qu'est ce qu'un web service

Un web service est une application accessible via Internet qui permet la communication entre différentes applications, quel que soit leur langage de programmation ou leur plateforme.

Il expose des fonctionnalités ou données à travers des protocoles standards comme HTTP.

Principe de fonctionnement

- Le client envoie une requête HTTP au serveur du web service.
- Le serveur traite la demande, exécute l'action demandée et renvoie une réponse HTTP (HTTP Response).
- Les données échangées sont souvent au format JSON ou XML.

Avantages

- Interopérabilité entre systèmes
- Réutilisation des services
- Communication standardisée et automatisée

1. a. Qu'est-ce qu'un web service ?

Les webs services à l'INIST



Un tas de webs services

Des webs services couvrant un tas de domaine, comme par exemple :

- Extraction d'informations (Ex: Entités nommées)
- Classification (Ex: Classification binaire)
- Clustering
- Alignement (Ex: Alignement d'entités nommées avec des notices d'autorités)

Simplicité d'utilisation

Frugalité

1. a. Qu'est-ce qu'un web service ?

Les webs services à l'INIST



Un tas de webs services

Des webs services couvrant un tas de domaine, comme par exemple :

- Extraction d'informations (Ex: Entités nommées)
- Classification (Ex: Classification binaire)
- Clustering
- Alignement (Ex: Alignement d'entités nommées avec des notices d'autorités)

Simplicité d'utilisation

- Format d'entrée divers en fonction des webs services (json, jsonl, csv, pdf, txt, ..)
- Pas, ou très peu, de pré traitement de la part de l'utilisateur
- Accessibilité et simplicité d'utilisation via une interface (TDM factory)
- Résultat récupérable via mail

Frugalité

1. a. Qu'est-ce qu'un web service ?

Les webs services à l'INIST



Un tas de webs services

Des webs services couvrant un tas de domaine, comme par exemple :

- Extraction d'informations (Ex: Entités nommées)
- Classification (Ex: Classification binaire)
- Clustering
- Alignement (Ex: Alignement d'entités nommées avec des notices d'autorités)

Simplicité d'utilisation

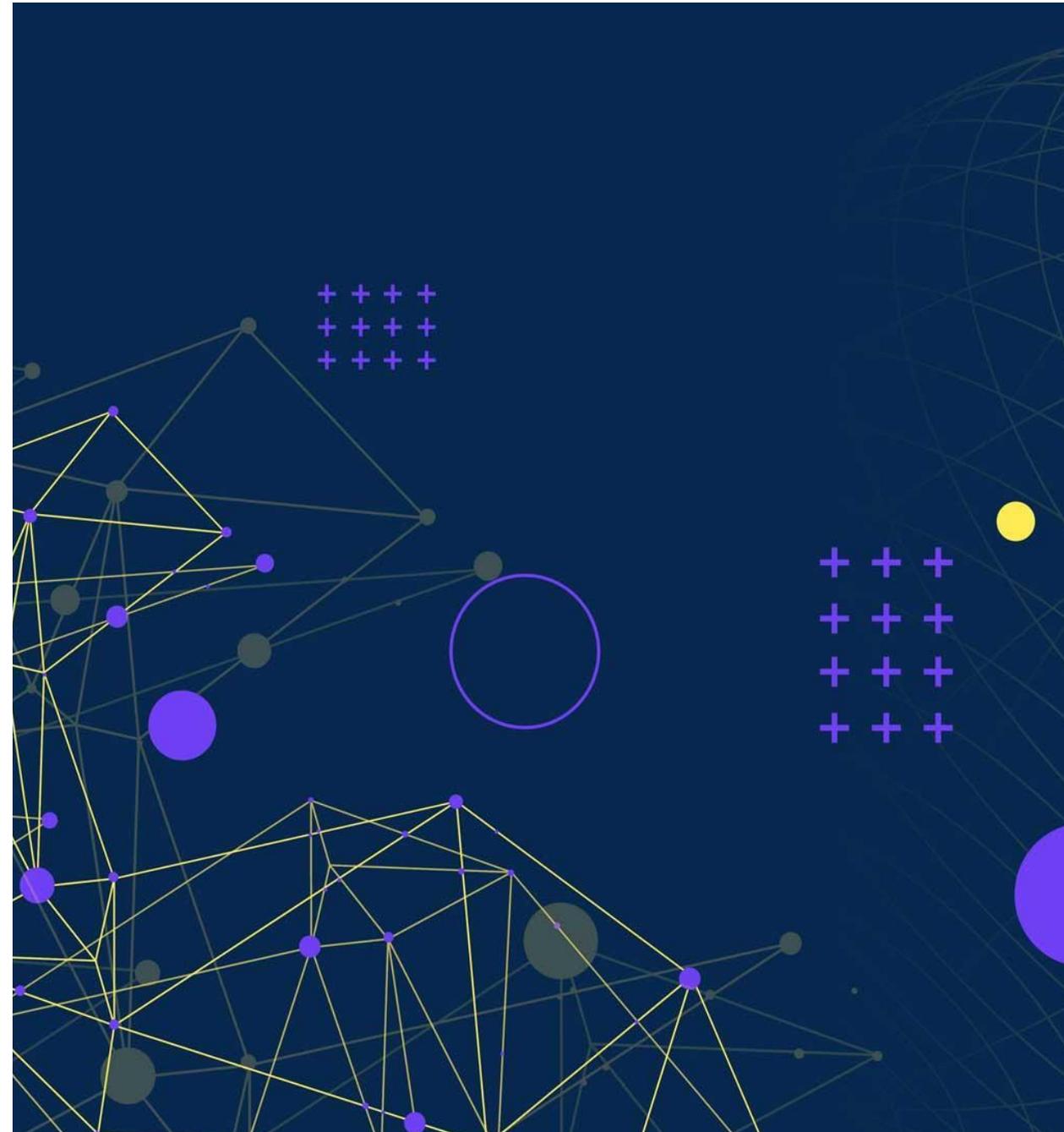
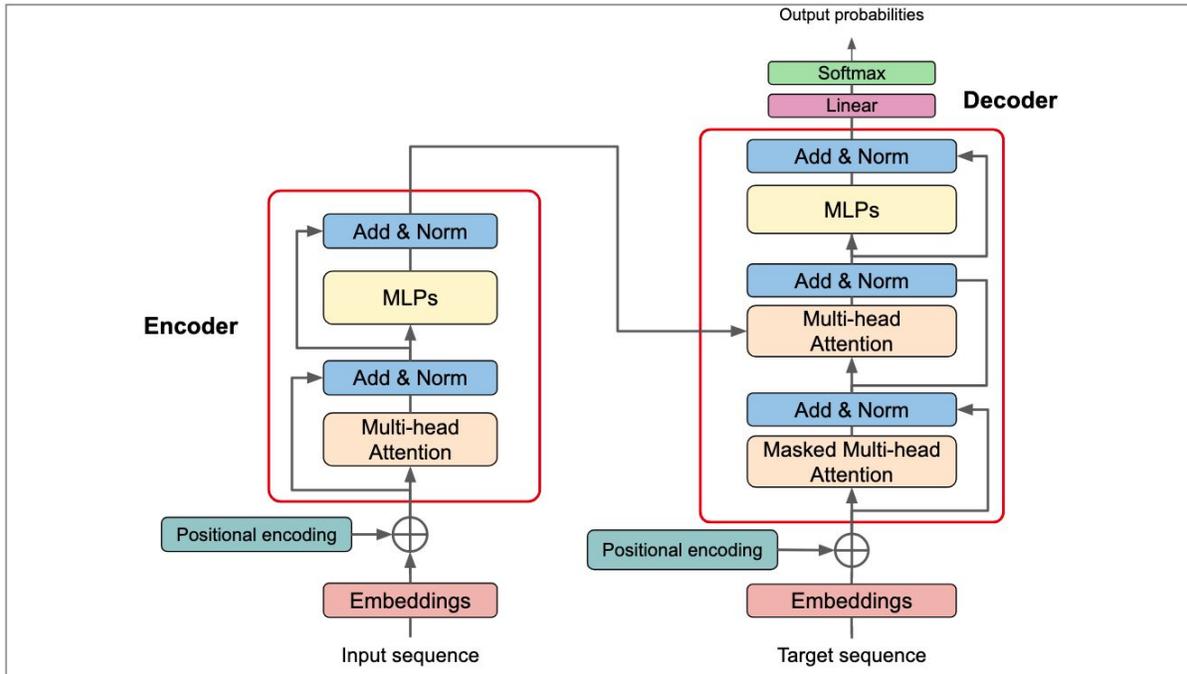
- Format d'entrée divers en fonction des webs services (json, jsonl, csv, pdf, txt, ..)
- Pas, ou très peu, de pré traitement de la part de l'utilisateur
- Accessibilité et simplicité d'utilisation via une interface (TDM factory)
- Résultat récupérable via mail

Frugalité

- Plein de petits web services pour différents problèmes
- Méthode de traitement adapté à la complexité du problème : Pas d'usine à gaz si pas nécessaire
- Web services conçus pour pouvoir tourner sur CPU (Moins d'utilisation GPU, plus accessible)

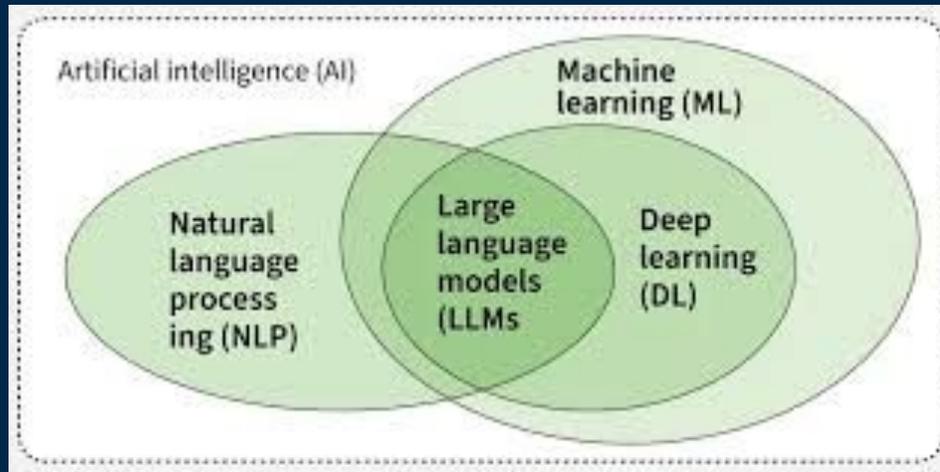
1. Web services : utilisation de petits LLM adaptés

b. Fine tuning de petit LLM



1. b. Utilisation de petits LLM adaptés

La naissance du LLM moderne

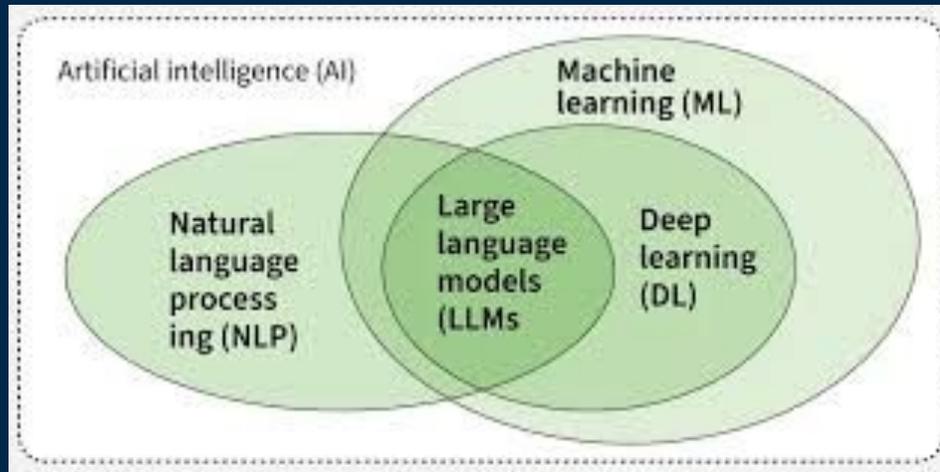


La route vers le LLM moderne

- 1957 – Perceptron (Frank Rosenblatt)
 - Premier modèle de réseau de neurones artificiels, capable d'apprendre des motifs simples.

1. b. Utilisation de petits LLM adaptés

La naissance du LLM moderne

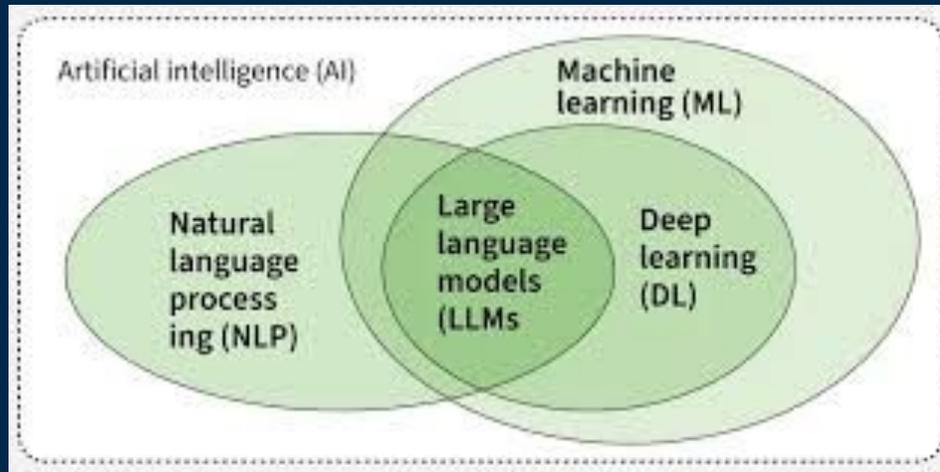


La route vers le LLM moderne

- 1957 – Perceptron (Frank Rosenblatt)
 - Premier modèle de réseau de neurones artificiels, capable d'apprendre des motifs simples.
- 1986 – Backpropagation (Rumelhart, Hinton, Williams)
 - Début des réseaux de neurones profonds.
- 1998 – LeNet (Yann LeCun)
 - Premier réseau de neurones convolutif (CNN) utilisé pour la reconnaissance de chiffres manuscrits (MNIST).

1. b. Utilisation de petits LLM adaptés

La naissance du LLM moderne

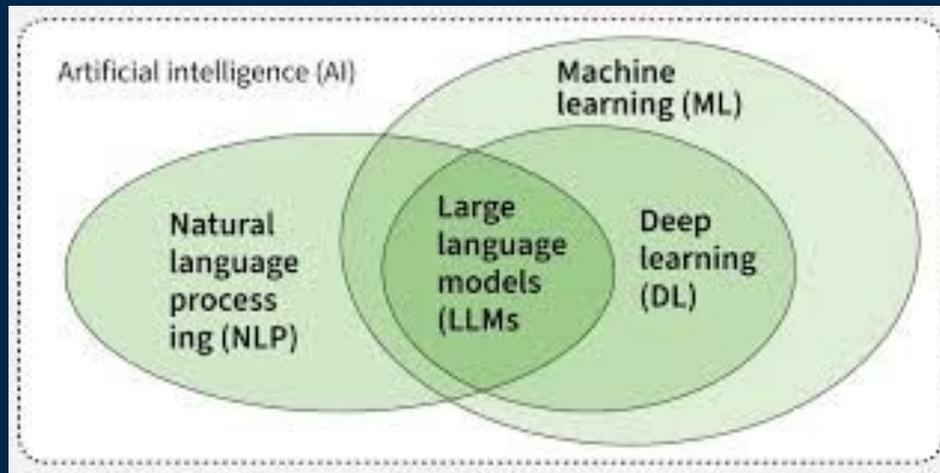


La route vers le LLM moderne

- 1957 – Perceptron (Frank Rosenblatt)
 - Premier modèle de réseau de neurones artificiels, capable d'apprendre des motifs simples.
- 1986 – Backpropagation (Rumelhart, Hinton, Williams)
 - Début des réseaux de neurones profonds.
- 1998 – LeNet (Yann LeCun)
 - Premier réseau de neurones convolutif (CNN) utilisé pour la reconnaissance de chiffres manuscrits (MNIST).
- 2013 – Word2Vec (Google)
 - Introduction des représentations vectorielles de mots (embeddings).
- 2015 – Mécanisme d'attention (Bahdanau et al.)
 - Introduction du mécanisme d'attention, permettant au modèle de “se concentrer” sur les parties importantes d'une séquence.

1. b. Utilisation de petits LLM adaptés

La naissance du LLM moderne

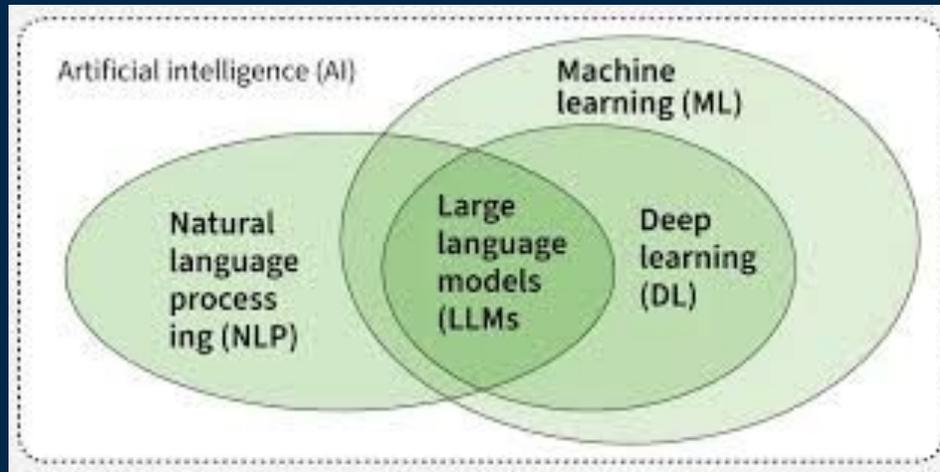


La route vers le LLM moderne

- 1957 – Perceptron (Frank Rosenblatt)
 - Premier modèle de réseau de neurones artificiels, capable d'apprendre des motifs simples.
- 1986 – Backpropagation (Rumelhart, Hinton, Williams)
 - Début des réseaux de neurones profonds.
- 1998 – LeNet (Yann LeCun)
 - Premier réseau de neurones convolutif (CNN) utilisé pour la reconnaissance de chiffres manuscrits (MNIST).
- 2013 – Word2Vec (Google)
 - Introduction des représentations vectorielles de mots (embeddings).
- 2015 – Mécanisme d'attention (Bahdanau et al.)
 - Introduction du mécanisme d'attention, permettant au modèle de "se concentrer" sur les parties importantes d'une séquence.
- 2017 – Le Transformer
 - Architecture basé uniquement sur l'attention, facilement parallélisable et plus performant

1. b. Utilisation de petits LLM adaptés

La naissance du LLM moderne

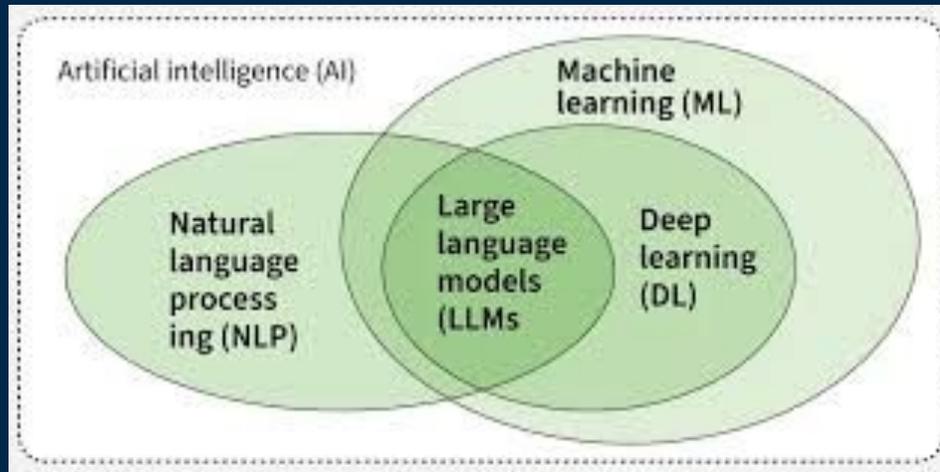


La route vers le LLM moderne

- 1957 – Perceptron (Frank Rosenblatt)
 - Premier modèle de réseau de neurones artificiels, capable d'apprendre des motifs simples.
- 1986 – Backpropagation (Rumelhart, Hinton, Williams)
 - Début des réseaux de neurones profonds.
- 1998 – LeNet (Yann LeCun)
 - Premier réseau de neurones convolutif (CNN) utilisé pour la reconnaissance de chiffres manuscrits (MNIST).
- 2013 – Word2Vec (Google)
 - Introduction des représentations vectorielles de mots (embeddings).
- 2015 – Mécanisme d'attention (Bahdanau et al.)
 - Introduction du mécanisme d'attention, permettant au modèle de "se concentrer" sur les parties importantes d'une séquence.
- 2017 – Le Transformer
 - Architecture basé uniquement sur l'attention, facilement parallélisable et plus performant
- 2018-2020 – Les premiers grands modèle de langue
 - 2018 – BERT (Google)
 - 2018 – GPT (OpenAI)
 - 2019 – GPT-2 (OpenAI)

1. b. Utilisation de petits LLM adaptés

La naissance du LLM moderne

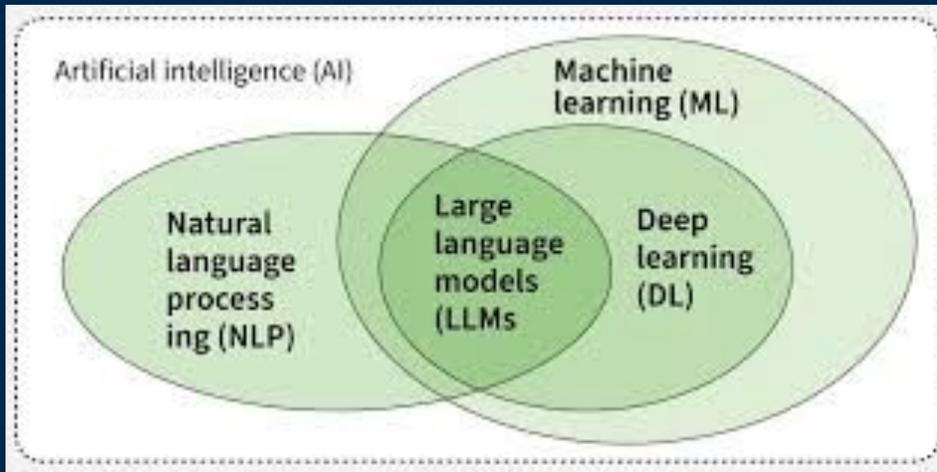


La route vers le LLM moderne

- 1957 – Perceptron (Frank Rosenblatt)
 - Premier modèle de réseau de neurones artificiels, capable d’apprendre des motifs simples.
- 1986 – Backpropagation (Rumelhart, Hinton, Williams)
 - Début des réseaux de neurones profonds.
- 1998 – LeNet (Yann LeCun)
 - Premier réseau de neurones convolutif (CNN) utilisé pour la reconnaissance de chiffres manuscrits (MNIST).
- 2013 – Word2Vec (Google)
 - Introduction des représentations vectorielles de mots (embeddings).
- 2015 – Mécanisme d’attention (Bahdanau et al.)
 - Introduction du mécanisme d’attention, permettant au modèle de “se concentrer” sur les parties importantes d’une séquence.
- 2017 – Le Transformer
 - Architecture basé uniquement sur l’attention, facilement parallélisable et plus performant
- 2018-2020 – Les premiers grands modèle de langue
 - 2018 – BERT (Google)
 - 2018 – GPT (OpenAI)
 - 2019 – GPT-2 (OpenAI)
- 2020–2022 : L’explosion de la taille et des capacités
 - 2020 – GPT-3 (175 milliards de paramètres)
 - 2021 – Megatron-Turing NLG (NVIDIA + Microsoft) (500 milliards de paramètres)

1. b. Utilisation de petits LLM adaptés

La naissance du LLM moderne

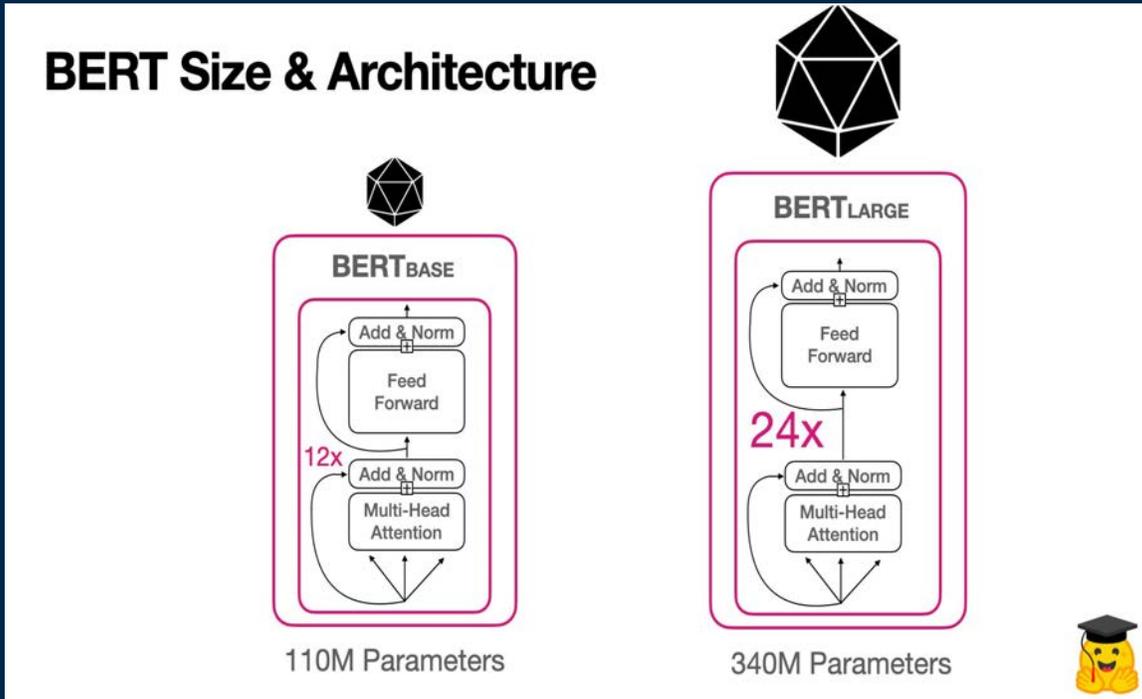


La route vers le LLM moderne

- 1957 – Perceptron (Frank Rosenblatt)
 - Premier modèle de réseau de neurones artificiels, capable d'apprendre des motifs simples.
- 1986 – Backpropagation (Rumelhart, Hinton, Williams)
 - Début des réseaux de neurones profonds.
- 1998 – LeNet (Yann LeCun)
 - Premier réseau de neurones convolutif (CNN) utilisé pour la reconnaissance de chiffres manuscrits (MNIST).
- 2013 – Word2Vec (Google)
 - Introduction des représentations vectorielles de mots (embeddings).
- 2015 – Mécanisme d'attention (Bahdanau et al.)
 - Introduction du mécanisme d'attention, permettant au modèle de "se concentrer" sur les parties importantes d'une séquence.
- 2017 – Le Transformer
 - Architecture basé uniquement sur l'attention, facilement parallélisable et plus performant
- 2018-2020 – Les premiers grands modèle de langue
 - 2018 – BERT (Google)
 - 2018 – GPT (OpenAI)
 - 2019 – GPT-2 (OpenAI)
- 2020–2022 : L'explosion de la taille et des capacités
 - 2020 – GPT-3 (175 milliards de paramètres)
 - 2021 – Megatron-Turing NLG (NVIDIA + Microsoft) (500 milliards de paramètres)
- 2023–2025 : Les LLM avancés et multimodaux
 - 2023 – GPT-4 (OpenAI) Multimodal (texte, image, code)
 - 2024–2025 – LLM multimodaux + agents autonomes. Fusion texte, image, son, vidéo. Intégration dans des systèmes interactifs (agents intelligents, copilotes IA, etc.).

1. b. Utilisation de petits LLM adaptés

BERT



Bert : Un bon compromis

Bert c'est quoi :

BERT (*Bidirectional Encoder Representations from Transformers*) est un modèle de langage basé sur l'architecture Transformer, conçu pour comprendre le langage plutôt que le générer.

Quelques caractéristiques de BERT :

- Bidirectionnel, il lit le texte dans les deux sens : Permet de mieux comprendre le contexte d'un mot.
- Deux versions :
 - BERT-base : 110 millions de paramètres
 - BERT-large : 340 millions de paramètres
- Excelle dans :
 - La classification de texte
 - L'extraction d'entités nommées (NER)
 - Similarité sémantique
 - Résumé et recherche contextuelle

Les avantages de BERT sur un gros LLM :

- Prend moins de mémoire (quelques centaines de Mo contre des dizaines de Go pour un LLM moderne)
- Exécutable sur un CPU en local
- Rapide à l'inférence, beaucoup moins coûteux à interroger qu'un gros LLM
- BERT peut être fine-tuné facilement sur une tâche ciblée **avec peu de données**.
- Open-source, gratuit, et peu coûteux à déployer.

1. b. Utilisation de petits LLM adaptés

L'utilisation des LLM à l'inist



L'utilisation des LLM à l'INIST

Une bonne partie de nos web services utilisant une architecture LLM utilisent une base BERT. Par exemple :

NER :

- AstroTagger
 - Fine tuning de BERT-base avec un petit nombre de données
- GeoTagger
 - Fine tuning de BERT-base avec un petit nombre de données annoté par notre équipe
- DiseaseTag
 - Utilisation du modèle bio-BERT

Classification :

- AiAbstractCheck
 - Fine tuning de BERT-large via des données d'Istex et des données générés par plusieurs gros LLM

Résumé :

- TextSummarize
 - Utilisation de Bart-Large-CNN

2. TDM Factory : extraire de l'information dans la littérature scientifique avec de l'intelligence artificielle

a. Présentation de TDM Factory



Chargez vos données et découvrez les résultats des services TDM

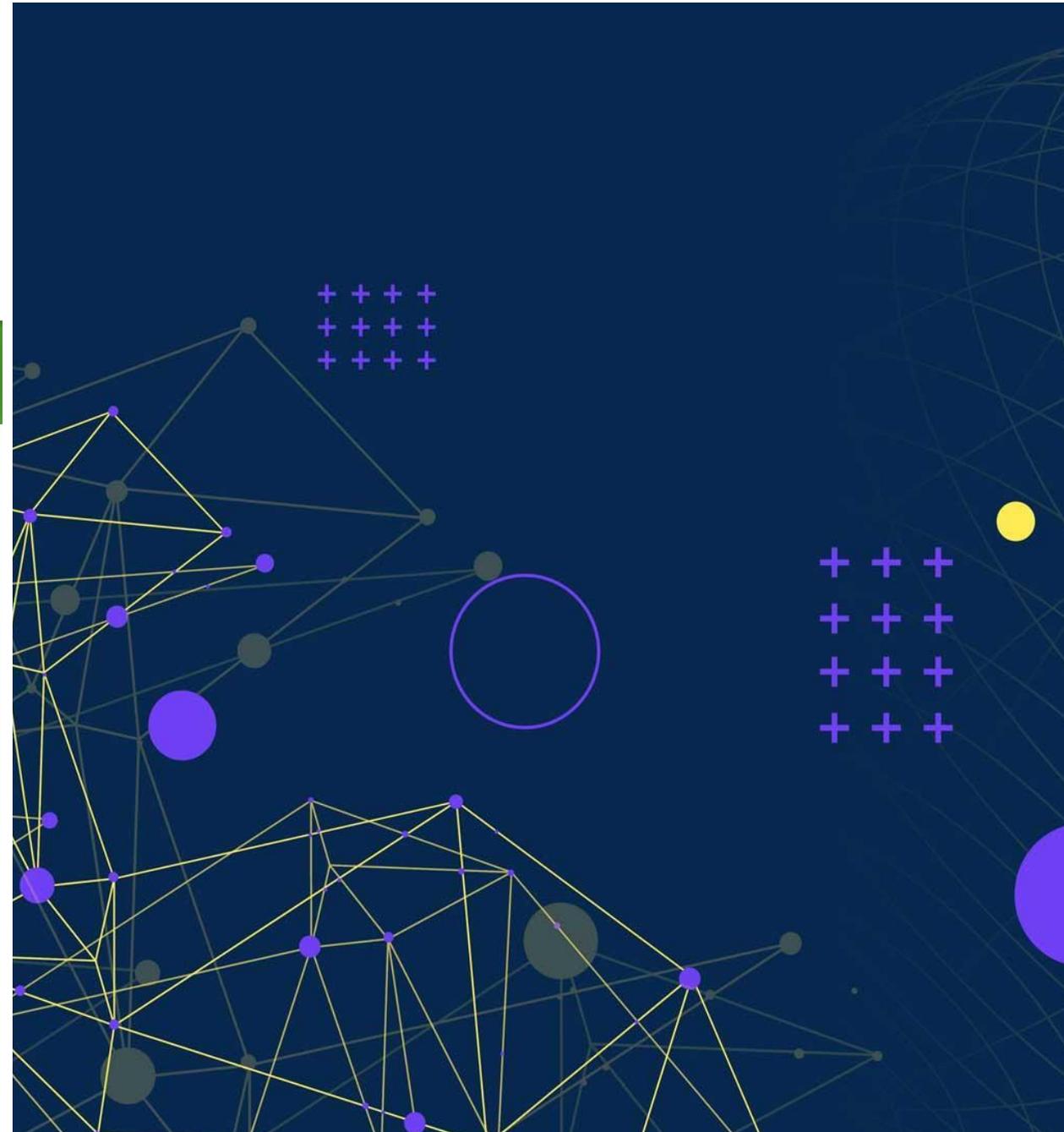


Traiter un article scientifique [Commencer →](#)



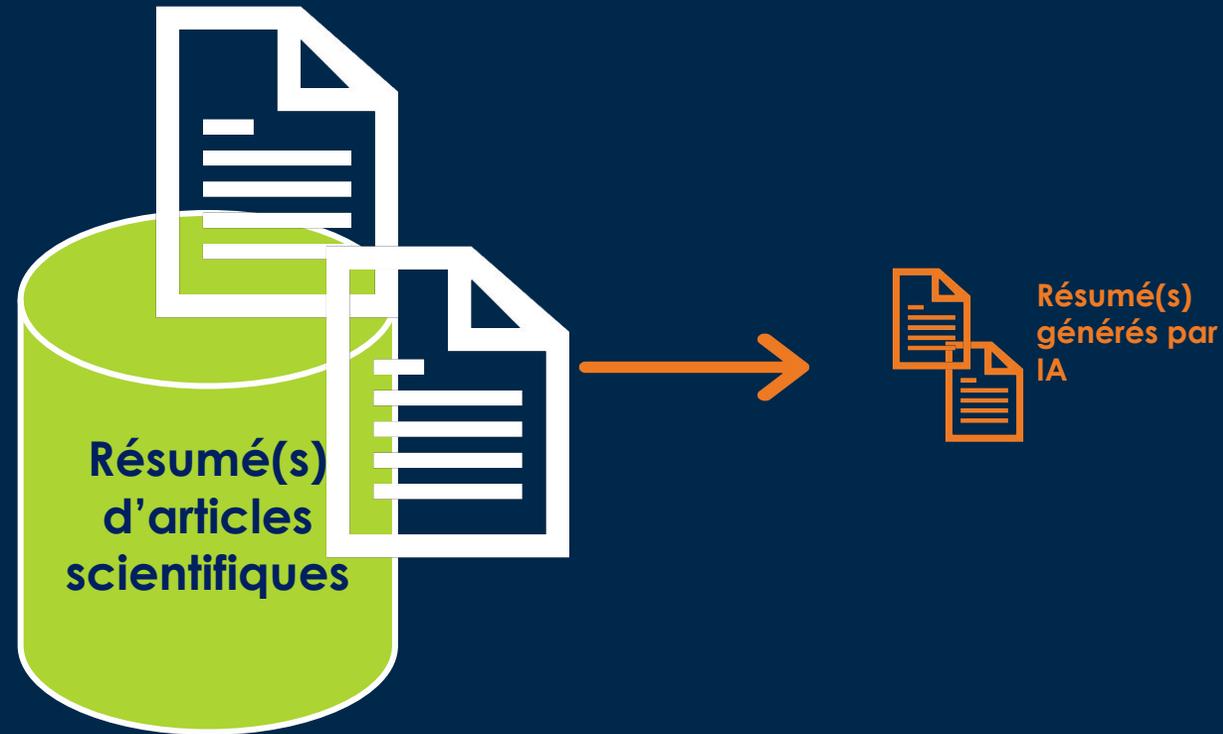
Traiter un corpus d'articles scientifiques [Commencer →](#)

<https://tdm-factory.services.istex.fr/>



textSummarize

Résumé automatique d'articles scientifiques



Objectif

Ce web service permet de résumer un texte scientifique écrit en anglais.

Méthode

Modèle [Bart-large-CNN](#), obtenu après fine-tuning du modèle Bart-large sur le jeu de données [CNN-dailymail](#).

Modèle plutôt frugal ayant obtenu de bons résultats lors de l'évaluation.

Métriques

Sur le jeu de données [dataset scisumm](#) :

- BERT score : 0.87
- ROUGE-1 score : 0.56
- ROUGE-2 score : 0.41
- ROUGE-3 score : 0.37

Résultat de textSummarize

sur l'article présentant BERT

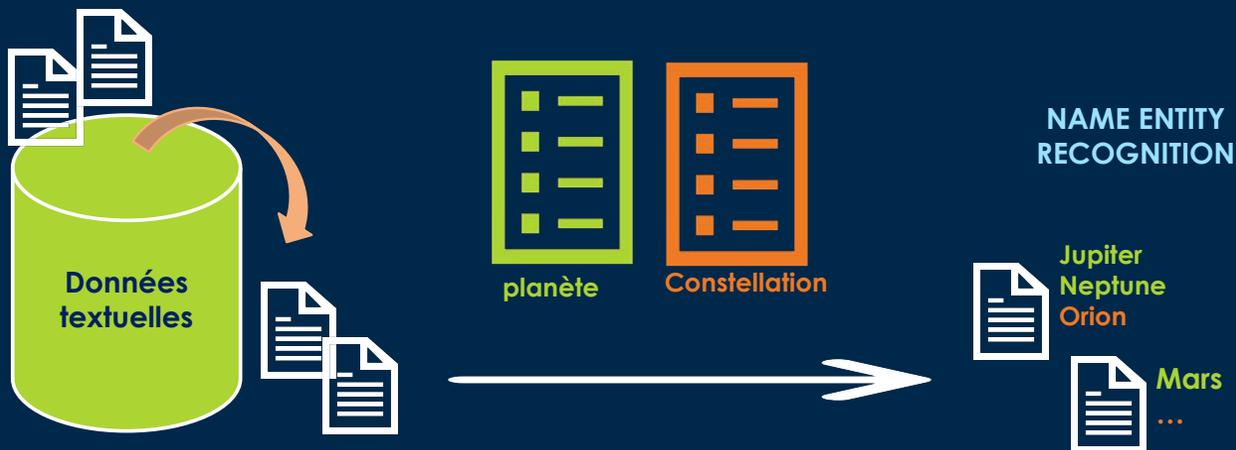
*(BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, 2018, arXiv:1810.04805)*

“ We introduce a new language representation model called BERT, which stands for Bidirectional Encoder Representations from Transformers. BERT is designed to pretrain deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. It obtains new state-of-the-art results on eleven natural language processing tasks.
<AI-generated> „



astroTagger

Extraction d'entités nommées d'Astronomie



Objectif

Détecter des entités nommées en astronomie sur des textes anglais, et les répartir selon 16 classes :

Amas stellaires, Astéroïdes, Constellations, Étoiles, Étoiles binaires (et pulsars), Exoplanètes, Galaxies et amas de galaxies, Nébuleuses et régions apparentées
Objets artificiels, Planètes, Satellites naturels, Statut hypothétique, Supernovas, Sursauts radios, sources radios, autres sursauts, Système solaire, Trous noirs, quasars et apparentés

Méthode

- Création d'un corpus d'Astronomie via la bibliothèque ISTEK
- Annotation manuelle des données
- Prétraitement des données
- Fine tuning d'un modèle Bert avec les données pour une tâche de détection d'entités nommées

Métriques

- Accuracy de 0.91

Résultat de AstroTagger

(sur l'article Photometric Variability of Neptune, 1972–2000
G. W. Lockwood & D. T. Thompson, 06-04-2001)

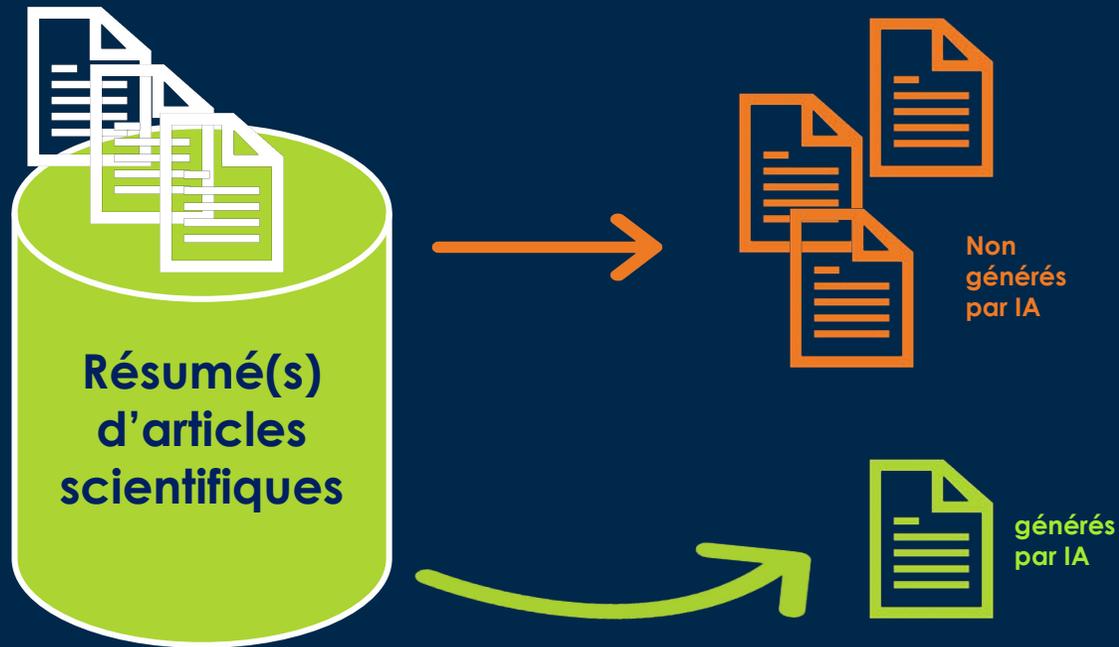
Photoelectric measurements of **Neptune** made each year since 1972 through intermediate band b (472-nm) and y (551-nm) filters are a component of a long-term program of **Solar System** photometry whose lineage derives from an earlier project that began around 1950 at Lowell Observatory. Currently, we also monitor **Uranus** and **Titan** in addition to **Neptune**. Originally, the intent was to study the variations of total solar irradiance, but this task has been carried out more precisely from space since about 1980 ...

Planete	Neptune ,Mars, Uranus ,Earth
Satellite_naturel	Titan ,Galilean satellites,Triton,Moon
Objets_artificiels	Voyager,Hubble Space Telescope,Hubble,HST
Systeme_solaire	Solar System ,Sun



aiAbstractCheck

Détection de résumés générés pas IA



Objectif

Détecter si le résumé d'un texte scientifique en anglais a été généré par intelligence artificielle ou non.

Méthode

- Création d'un corpus de résumés d'articles scientifiques via la bibliothèque ISTEEX
- Génération de plusieurs résumés d'articles scientifiques par plusieurs LLM (Mistral-7B, Gwen-3B, LLama-3.2, Deepseek-7B principalement, et quelques données GPT-4).
- Fine tuning d'un modèle BERT pour une tâche de classification binaire (Généré par IA / Non généré par IA)

Métriques

- Validation sur plusieurs jeux de données publiques
- Score F1 moyen de 0.88 sur les deux jeu de données

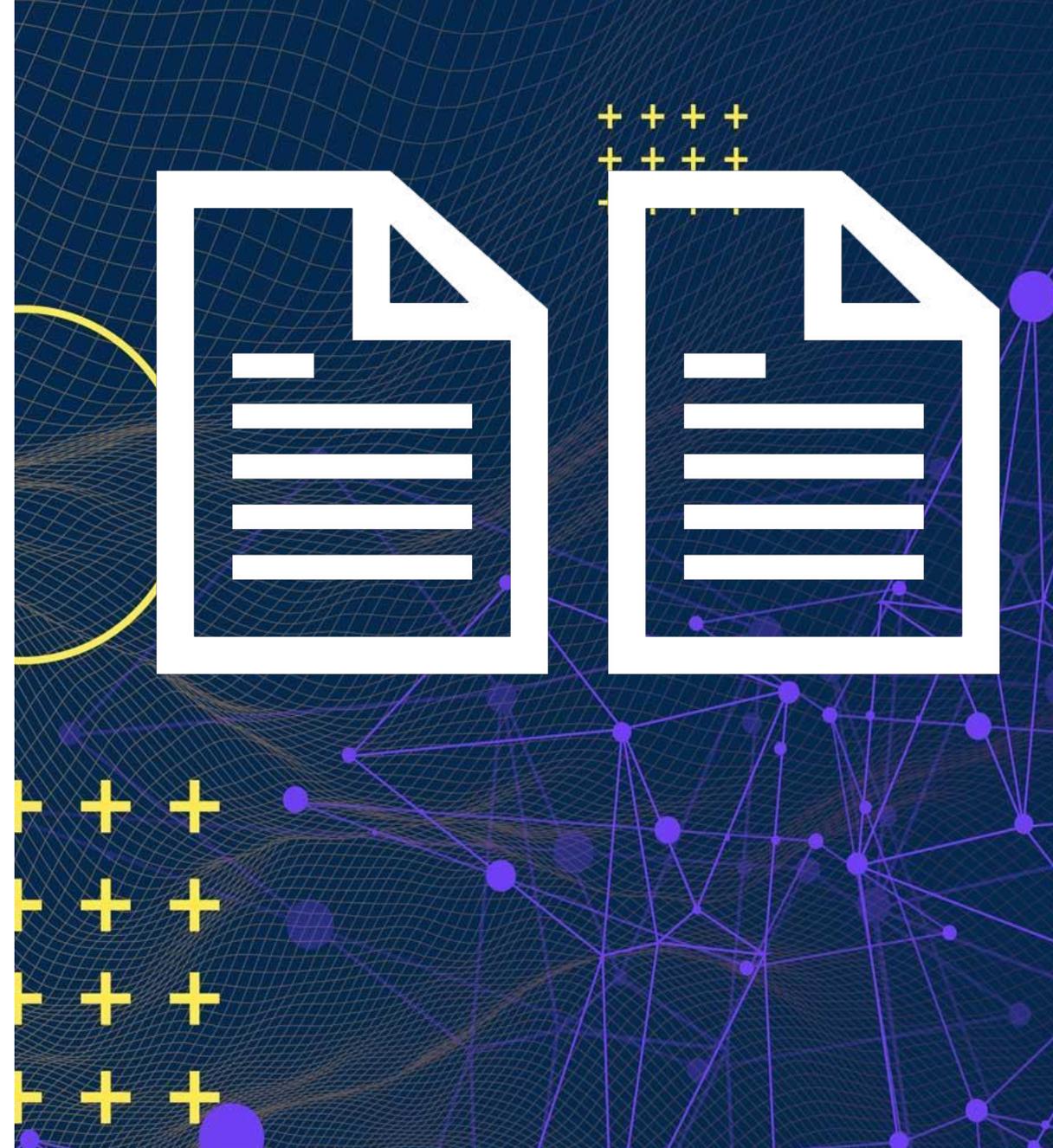
Résultat de AiAbstractCheck

Abstract 1 :

The Conditional Lie-Bäcklund Symmetry (CLBS) method is introduced to analyze systems of evolution equations. It is demonstrated that the ability to reduce a system of evolution equations to a system of ordinary differential equations can be entirely described by the CLBS of the given system. As an application, a specific class of two-component nonlinear diffusion equations is examined. The governing system and its corresponding CLBS are determined. Consequently, exact solutions are derived based on polynomial, exponential, trigonometric, and mixed invariant subspaces through symmetry reductions.

Abstract 2 :

Conditional Lie-Bäcklund symmetry (CLBS) method is developed to study system of evolution equations. It is shown that reducibility of a system of evolution equations to a system of ordinary differential equations can be fully characterized by the CLBS of the considered system. As an application of the approach, a class of two-component nonlinear diffusion equations is studied. The governing system and the admitted CLBS can be identified. As a consequence, exact solutions defined on the polynomial, exponential, trigonometric, and mixed invariant subspaces are constructed due to the corresponding symmetry reductions.



Résultat de AiAbstractCheck

Abstract 1 :

The Conditional Lie-Bäcklund Symmetry (CLBS) method is introduced to analyze systems of evolution equations. It is demonstrated that the ability to reduce a system of evolution equations to a system of ordinary differential equations can be entirely described by the CLBS of the given system. As an application, a specific class of two-component nonlinear diffusion equations is examined. The governing system and its corresponding CLBS are determined. Consequently, exact solutions are derived based on polynomial, exponential, trigonometric, and mixed invariant subspaces through symmetry reductions.

isAiGenerated : **True**
Score : **1.000**

Abstract 2 :

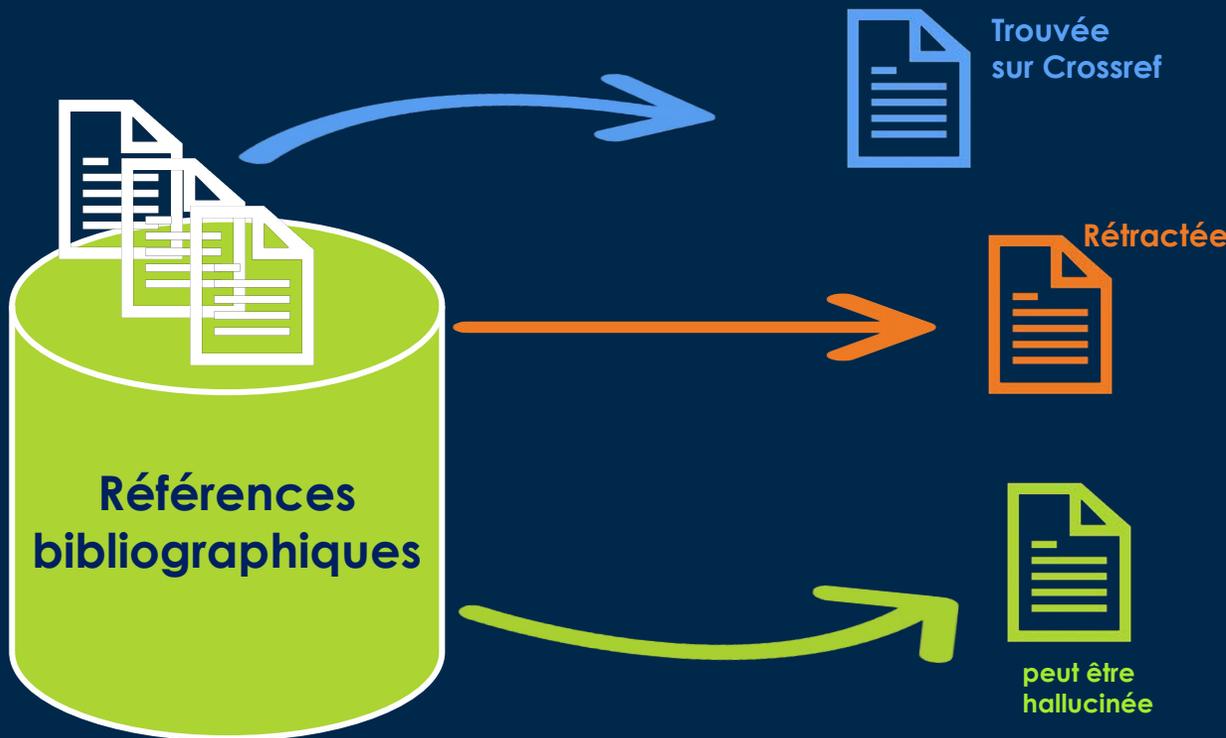
Conditional Lie-Bäcklund symmetry (CLBS) method is developed to study system of evolution equations. It is shown that reducibility of a system of evolution equations to a system of ordinary differential equations can be fully characterized by the CLBS of the considered system. As an application of the approach, a class of two-component nonlinear diffusion equations is studied. The governing system and the admitted CLBS can be identified. As a consequence, exact solutions defined on the polynomial, exponential, trigonometric, and mixed invariant subspaces are constructed due to the corresponding symmetry reductions.

isAiGenerated : **False**
Score : **1.000**



bibCheck

Validation automatique de références bibliographiques



Objectif

Ce web service contrôle une référence bibliographique donnée, en vérifiant sa présence dans [Crossref](#) tout en s'assurant que l'article associé n'est pas rétracté.

Méthode

Utilise l'API [Crossref](#) et le [PPS](#) (*Problematic Paper Screener*, Cabanac, G., Labbé, C., & Magazinov, A. (2022)) pour vérifier la référence bibliographique :

- **found** : la référence a été identifiée sur Crossref et est valide.
- **retracted** : la référence a été identifiée sur Crossref mais est rétractée selon le PPS.
- **not_found** : la référence n'a pas été trouvée sur Crossref mais ne semble pas générée par IA.
- **to_be_verified** : la référence n'a pas pu être identifiée sur Crossref et semble générée par IA.

Usage

Algorithme utilisé à des fins préventives et non curatives. Le but est d'éviter de propager des références rétractées.

Résultat de bibCheck

sur un article récent de climatologie

Vowles, K., & Ekberg, K. (2025). *Beyond the CCCM – The Conditions Facilitating the Swedish Climate Change Reactionary Movement. Environmental Communication, 1–15.*
<https://doi.org/10.1080/17524032.2025.2560392>

	A	B	C	
1	id	value/doi	value/status	value/reference
2	PID1760000124239		not found	Åkerlund, M. (2022). Far right, right here. Umeå University.
3	PID1760000124239		not found	Åkesson, J., Andersson, T., and Kinnunen, M. 2025. DN Debatt. "Vi måste"
4	PID1760000124239	10.1177/09636625221137815	found	Almiron, N., Moreno, J. A., & Farrell, J. (2022). Climate change contrarian
5	PID1760000124239	10.1080/18902138.2014.908627	found	Anshelm, J., & Hultman, M. (2014a). A green Fatwā? Climate change as a
6	PID1760000124239	10.4324/9781315769998	found	Anshelm, J., & Hultman, M. (2014b). Discourses of global climate change
7	PID1760000124239		not found	Atlas Network. (n.d.). Atlas network, proposal. "PROPOSAL FOR FUNDIN
8	PID1760000124239		not found	Benkler, Y., Farris, R., & Roberts, H. (2018). Network propaganda, Oxford
9	PID1760000124239	10.1016/j.gloenvcha.2021.102386	found	Bonneuil, C., Choquet, P.-L., & Franta, B. (2021). 'Early warnings and em
10	PID1760000124239	10.1080/17541328.2015.1057382	found	Boynton, A. (2015). Formulating an anti-environmental opposition: Neoco
11	PID1760000124239	10.1007/s10584-013-1018-7	found	Brulle, R. (2014). Institutionalizing delay: Foundation funding and the crea
12	PID1760000124239	10.1093/oso/9780197762042.003.0001	found	Brulle, R., & Timmons Roberts, J.. (2024). Introduction: The first portrait of
13	PID1760000124239	10.1007/978-3-031-05171-5_12	found	Buns, M. A., & Hinde, D., (2023). Green states in a dirty world: 1975 and t
14	PID1760000124239	10.1007/s10551-011-0950-6	found	Cho, C. H., Martens, M. L., Kim, H., & Rodrigue, M. (2011). 'Astroturfing g
15	PID1760000124239	10.1016/j.jenvp.2025.102579	found	Czarnek, G., Hornsey, M. J., Bucki, S., & Kossowska, M. (2025). As coun
16	PID1760000124239	10.1080/17524032.2024.2363861	found	De Nadal, L. (2024). From denial to the culture wars: A study of climate m
17	PID1760000124239	10.1080/09644016.2021.1979775	found	Dubash, N. K. (2021). Varieties of climate governance: The emergence an
18	PID1760000124239	10.4337/9781789900408.00013	found	Dunlap, R., & Brulle, R. (2020). Sources and amplifiers of climate change
19	PID1760000124239	10.1093/oxfordhb/9780199566600.003.0010	found	Dunlap, R., & McCright, A. (2011). Organized climate change denial. In J.
20	PID1760000124239	10.4324/9781003515494-13	found	Ekberg, K. (2025). Capitalist fictions. In S. Bott, S. Pitteloud, & J. Marina S
21	PID1760000124239	10.4324/9781003181132	found	Ekberg, K., Forchtner, B., Hultman, M., & Jylhä, K. (2023). Climate obstru
22	PID1760000124239	10.3197/096734021x16245313030028	found	Ekberg, K., & Hultman, M. (2021). A question of utter importance: The ear
23	PID1760000124239	10.1017/s096077732200025x	found	Ekberg, K., & Pressfeldt, V. (2022). A road to denial: Climate change and
24	PID1760000124239	10.2478/nor-2021-0007	found	Ekman, M., & Krzyżanowski, M. (2021). A populist turn?: News editorials a
25	PID1760000124239	10.1073/pnas.1509433112	found	Farrell, J. (2016). Corporate funding and ideological polarization about cli
26	PID1760000124239	10.1080/09644016.2022.2048556	found	Forchtner, B., & Lubarda, B. (2022). 'Scepticisms and beyond? A compre
27	PID1760000124239	10.2307/1388959	found	Gale, R. P. (1986). Social movements and the state: The environmental m
28	PID1760000124239		not found	Habul, K., & Forsberg, O. (2018). Jimmie Åkesson om klimatdebatten: "Vä
29	PID1760000124239		not found	Häger, B.. (2016). Den stora konspirationen: Nio raka svar om mörkning.
30	PID1760000124239	10.1088/2515-7620/ad61c6	found	Hall, G., Loy, L., Brulle, R. J., Schell-Smith, K., Hu, M.-M., & Trollback, S.
31	PID1760000124239	10.21525/kriterium.28.h	found	Heidenblad, D. L. (2021). Den gröna vändningen: En ny kunskaphistoria
32	PID1760000124239	10.4324/9781351104043-8	found	Hultman, M., Björk, A., & Viinikka, T.. (2019). Far-Right and climate chang
33	PID1760000124239	10.4324/9781315195223-9	found	Hultman, M., & Pulé, P. M.. (2018). Ecological masculinities: Theoretical f
34	PID1760000124239	10.3390/su122310226	found	Jylhä, K., Strimling, P., & Rydgren, J. (2020). Climate change denial amor
35	PID1760000124239	10.4324/9781351104043-15	found	Kaiser, J. (2019). In the heartland of climate scepticism. In B. Forchtner (E
36	PID1760000124239		not found	Kaun, A.. (2017). Allt är annorlunda nu -Hur ett medielandskap i förändring
37	PID1760000124239		not found	Klenell, J. (2025). Bananrepubliken Sverige: Hur politiker, välfärdskapitalis
38	PID1760000124239		not found	Klimatpolitiska rådet. (2024). Årsrapport 2024. No. 7. Klimatpolitiska rådet
39	PID1760000124239		not found	Knaggård, Å. (2009). Vetenskaplig Osäkerhet i Policyprocessen: En Studi
40	PID1760000124239	10.1177/1461444819887700	found	Larsson, A. O. (2019). Right-wingers on the rise online: Insights from the
41	PID1760000124239	10.1016/j.polgeo.2025.103390	found	Malm, A., Ekberg, K., Englund, C., Grønkaer, J. T., Charlier, M., Medin, C
42	PID1760000124239	10.1177/0263276409356001	found	McCright, A., & Dunlap, R. (2010). Anti-reflexivity: The American conserva
43	PID1760000124239	10.1007/978-3-031-33592-1	found	McKie, R. E. (2023). The climate change counter movement: How the fos
44	PID1760000124239	10.1177/1464884910379704	found	McKnight, D. (2010). A change in the climate? The journalism of opinion a
45	PID1760000124239	10.7765/9781526167804.00009	found	McLean, J.. (2024). United they roll? How Canadian fossil capital subsidiz
46	PID1760000124239	10.1086/230869	found	Meyer, D. S., & Staggenborg, S. (1996). Movements, countermovements.
47	PID1760000124239	10.1093/9780197807705.003.0001	found	Morris, H. E. (2025). Apocalyptic authoritarianism: Climate crisis, media, a
48	PID1760000124239	10.2139/ssrn.2619576	found	Newman, N., Fletcher, R., Kalogeropoulos, A., & Nielsen, R. K. (2019). Re



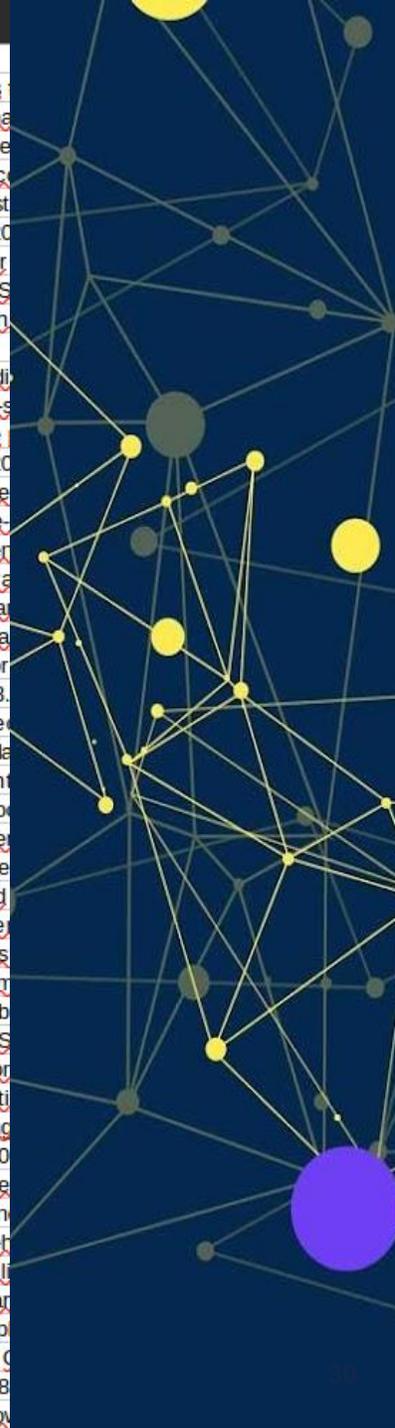
Résultat de bibCheck

sur l'article présentant BERT

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, 2018, arXiv:1810.04805

Spécificité : beaucoup de conférences et aucun DOI dans les références.

	A	B	C	
1	id	value/doi	value/status	value/reference
2	PID1760000985686		not_found	Alan Akbik, Duncan Blythe, and Roland Vollgraf, 2018. Contextual string embeddings
3	PID1760000985686		not_found	Rami Al-Rfou, Dokook Choe, Noah Constant, Mandy Guo, and Llion Jones. 2018. Cha
4	PID1760000985686		to be verified	Rie Kubota Ando and Tong Zhang, 2005. A framework for learning predictive structure
5	PID1760000985686		not_found	Luisa Bentivogli, Bernardo Magnini, Ido Dagan, Hoa Trang Dang, and Danilo Giampic
6	PID1760000985686	10.3115/1610075.1610094	found	John Blitzer, Ryan McDonald, and Fernando Pereira, 2006. Domain adaptation with st
7	PID1760000985686		not_found	Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning, 20
8	PID1760000985686		not_found	Peter F Brown, Peter V Desouza, Robert L Mercer, Vincent J Della Pietra, and Jenifer
9	PID1760000985686	10.18653/v1/s17-2001	found	Daniel Cer, Mona Diab, Eneko Agirre, Inigo Lopez-Gazpio, and Lucia Specia, 2017. S
10	PID1760000985686	10.21437/interspeech.2014-564	found	Ciprian Chelba, Tomas Mikolov, Mike Schuster, Qi Ge, Thorsten Brants, Philipp Koeh
11	PID1760000985686	10.1007/978-981-99-3723-3_16	found	Z. Chen, H. Zhang, X. Zhang, and L. Zhao, 2018. Quora question pairs.
12	PID1760000985686		not_found	Christopher Clark and Matt Gardner, 2018. Simple and effective multi-paragraph readi
13	PID1760000985686	10.18653/v1/d18-1217	found	Kevin Clark, Minh-Thang Luong, Christopher D Man-ning, and Quoc Le, 2018. Semi-s
14	PID1760000985686	10.1145/1390156.1390177	found	Ronan Collobert and Jason Weston, 2008. A unified architecture for natural language.
15	PID1760000985686	10.18653/v1/d17-1070	found	Alexis Conneau, Douwe Kiela, Holger Schwenk, Loic Barrault, and Antoine Bordes, 20
16	PID1760000985686		not_found	Andrew M Dai and Quoc V Le, 2015. Semi-supervised sequence learning. In Advance
17	PID1760000985686	10.1109/cvpr.2009.5206848	found	J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, 2009. ImageNet: A Large-
18	PID1760000985686		not_found	William B Dolan and Chris Brockett, 2005. Automati- cally constructing a corpus of ser
19	PID1760000985686		not_found	William Fedus, Ian Goodfellow, and Andrew M Dai, 2018. Maskgan: Better text genera
20	PID1760000985686		not_found	Dan Hendrycks and Kevin Gimpel, 2016. Bridging nonlinearities and stochastic regula
21	PID1760000985686	10.18653/v1/m16-1162	found	Felix Hill, Kyunghyun Cho, and Anna Korhonen, 2016. Learning distributed representa
22	PID1760000985686	10.18653/v1/p18-1031	found	Jeremy Howard and Sebastian Ruder, 2018. Universal language model fine-tuning for
23	PID1760000985686	10.24963/ijcai.2018/570	found	Minghao Hu, Yuxing Peng, Zhen Huang, Xipeng Qiu, Furu Wei, and Ming Zhou, 2018.
24	PID1760000985686		not_found	Yacine Jernite, Samuel R. Bowman, and David Son- tag, 2017. Discourse-based obje
25	PID1760000985686	10.18653/v1/p17-1147	found	Mandar Joshi, Eunsoo Choi, Daniel S Weld, and Luke Zettlemoyer, 2017. Triviaqa: A la
26	PID1760000985686		not_found	Ryan Kiros, Yukun Zhu, Ruslan R Salakhutdinov, Richard Zemel, Raquel Urtasun, An
27	PID1760000985686		not_found	Quoc Le and Tomas Mikolov, 2014. Distributed rep- resentations of sentences and do
28	PID1760000985686	10.1613/jair.5339	found	Hector J Levesque, Ernest Davis, and Leora Morgen- stern, 2011. The winograd sche
29	PID1760000985686		not_found	Lajanugen Logeswaran and Honglak Lee, 2018. An efficient framework for learning se
30	PID1760000985686		not_found	Bryan McCann, James Bradbury, Caiming Xiong, and Richard Socher, 2017. Learned
31	PID1760000985686	10.18653/v1/k16-1006	found	Oren Melamud, Jacob Goldberger, and Ido Dagan, 2016. context2vec: Learning gene
32	PID1760000985686		not_found	Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Cor- rado, and Jeff Dean, 2013. Dis
33	PID1760000985686	10.18653/v1/d16-1244	found	Ankur P Parikh, Oscar Täckström, Dipanjan Das, and Jakob Uszkoreit, 2016. A decon
34	PID1760000985686	10.3115/v1/d14-1162	found	Jeffrey Pennington, Richard Socher, and Christo- pher D. Manning, 2014. Glove: Glob
35	PID1760000985686	10.18653/v1/p17-1161	found	Matthew Peters, Waleed Ammar, Chandra Bhagavat- ula, and Russell Power, 2017. S
36	PID1760000985686		not_found	Matthew Peters, Mark Neumann, Mohit Iyer, Matt Gardner, Christopher Clark, Kenton
37	PID1760000985686	10.18653/v1/d18-1179	found	Matthew Peters, Mark Neumann, Luke Zettlemoyer, and Wen-tau Yih, 2018b. Dissecti
38	PID1760000985686		not_found	Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever, 2018. Improving
39	PID1760000985686	10.18653/v1/d16-1264	found	Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang, 2016. Squad: 10
40	PID1760000985686		not_found	Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi, 2017. Bidire
41	PID1760000985686	10.18653/v1/d13-1170	found	Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, An
42	PID1760000985686		not_found	Fu Sun, Linyang Li, Xipeng Qiu, and Yang Liu, 2018. U-net: Machine reading compre
43	PID1760000985686	10.1177/107769905303000401	found	Wilson L Taylor, 1953. Cloze procedure: A new tool for measuring readability. Journal
44	PID1760000985686	10.3115/1119176.1119195	found	Erik F Tjong Kim Sang and Fien De Meulder, 2003. Introduction to the conll-2003 shar
45	PID1760000985686		not_found	Joseph Turian, Lev Ratinov, and Yoshua Bengio, 2010. Word representations: A simp
46	PID1760000985686		not_found	Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N O
47	PID1760000985686	10.1145/1390156.1390294	found	Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol, 2008
48	PID1760000985686	10.18653/v1/w18-5446	found	Alex Wang, Amanpreet Singh, Julian Michael, Fe- lix Hill, Omer Levy, and Samuel Bo

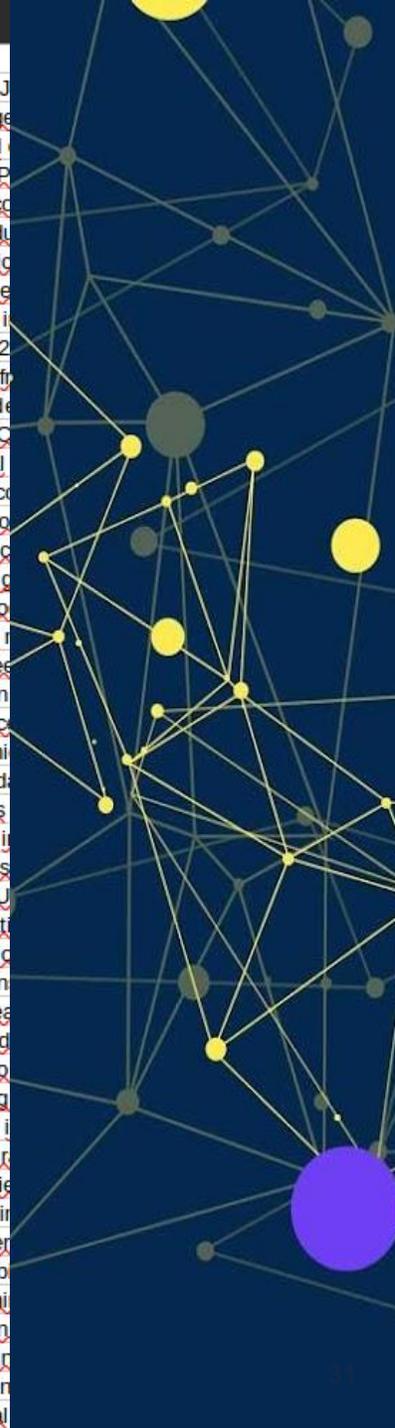


Résultat de bibCheck

sur un article rétracté

Spécificité : beaucoup d'articles rétractés dans les références.

	A	B	C	
1	id	value/doi	value/status	value/reference
2	PID1760001697325	10.7459/es/32.1.02	found	R.A. Rhoads, China's Rising Research Universities: a New Era of Global Ambition, J
3	PID1760001697325	10.1353/chn.2023.a898342	found	M. Svensson, Chinese youth and the communist party of China: cultivating a loyal ge
4	PID1760001697325	10.1155/2022/3904316	retracted	Y. Cheng, Research on dynamic design of ideological and political education based
5	PID1760001697325	10.1155/2021/2648352	found	N. Wang, Ideological and political education recommendation system based on AHP
6	PID1760001697325	10.1155/2022/2845259	retracted	J. Xiao, M. Zhu, An analysis model of psychological resilience of outstanding poor co
7	PID1760001697325	10.1155/2021/3285830	retracted	L. Xu, S.-B. Tsai, The transformation of college students' ideological and political edu
8	PID1760001697325	10.3390/ijerph20032608	found	S. Liu, L. Luo, A study on the impact of ideological and political education of ecologic
9	PID1760001697325		not found	Xuejian Wang, Shi Yan, Characteristics Connotation, Difficulties and response strate
10	PID1760001697325		not found	Han Yi, Study on applied path of teaching civics in ChatGPT perspective, Teaching i
11	PID1760001697325	10.1191/1478088706qp0630a	found	V. Braun, V. Clarke, Using thematic analysis in psychology, Qual. Res. Psychol. 3 (2
12	PID1760001697325	10.47836/pjsrh.29.1.35	found	Mohd Zairul, A thematic review on industrialised building system (IBS) publications fr
13	PID1760001697325	10.1080/23311983.2022.2100129	found	Z. Qiuxia, A.R. Abdul Rahman, H. Wenhong, A thematic review on souvenirs from de
14	PID1760001697325		not found	Daokun Lu, Research Progress, Difficult Focus and Future Direction of Curriculum C
15	PID1760001697325		not found	Xia Yuhan, Changes and Constants in the Modernization of Ideological and Political
16	PID1760001697325	10.1155/2022/9704315	retracted	W. Wu, H. Liu, Analysis of college students' ideological and political dynamics and co
17	PID1760001697325	10.1155/2022/9885274	retracted	Y. Zhang, Analysis of college students' network moral behavior by the history of ide
18	PID1760001697325	10.3389/fpsyg.2021.707973	found	X. He, X. Dong, L. Liu, Y. Zou, Challenges of college students' ideological and politic
19	PID1760001697325	10.1155/2022/7417406	retracted	X. Li, The practical dilemma and path optimization of big data integration into ideolo
20	PID1760001697325	10.1155/2022/2148657	retracted	J. Cheng, Construction of data resource sharing platform for college students' ideolo
21	PID1760001697325	10.3390/ijerph20021569	found	H. Ren, L. Zhao, Demonstration and suggestion on the communication efficiency of r
22	PID1760001697325	10.1155/2022/1283256	retracted	X. Huang, Y. Xie, Y. Li, Exploring online teaching design of curriculum politics by des
23	PID1760001697325	10.1155/2022/5833589	retracted	C. Luo, Optimization algorithm for ideological and political curriculum environment in
24	PID1760001697325	10.1155/2022/3845749	retracted	T. Yang, Z. Duan, Design of personalized ideological and political education resource
25	PID1760001697325	10.1155/2022/8354909	retracted	S. Xu, J. Liu, K. Chen, Y. Yang, Research on the communication path of public opini
26	PID1760001697325		not found	C. Chang, Optimization of ideological and political education teaching by using big d
27	PID1760001697325		not found	Wen-Ting Xu, Internet ideological and political education research retrospect, status
28	PID1760001697325	10.1155/2022/5378694	retracted	J. Wu, Analysis and evaluation of the impact of integrating mental health education i
29	PID1760001697325	10.1155/2022/4206714	found	J. Zhu, Y. Xue, Construction of a mental health education model for college students
30	PID1760001697325	10.3389/fpsyg.2022.908154	found	S. Shen, J. Fan, Emotion analysis of ideological and political education using a GRU
31	PID1760001697325	10.1155/2022/2294908	retracted	K. An, Exploration of intelligent teaching methods for ideological and political educat
32	PID1760001697325	10.1155/2022/7465860	retracted	X. Hu, D. Sturdivant, Dynamical alert of thought and politics teaching based on the lo
33	PID1760001697325		to be verified	J. Qi, X. Jia, Text score analysis under the IPE environment based on improved tran
34	PID1760001697325	10.1155/2022/4754972	retracted	L. Qiao, Teaching design of online ideological and political course based on deep lea
35	PID1760001697325	10.1155/2022/3119014	retracted	S. Tao, Parameter optimization of educational network ecosystem based on BERT d
36	PID1760001697325	10.1155/2022/4139323	retracted	W. Zhao, J. Liu, Application of knowledge map based on BiLSTM-CRF algorithm mo
37	PID1760001697325	10.1155/2022/3496676	retracted	C. Cao, Artificial intelligence and internet-of-things technology application on ideolog
38	PID1760001697325	10.1155/2021/9914790	retracted	X. Zhang, X. Gao, H. Yi, Z. Li, Design of an intelligent virtual classroom platform for i
39	PID1760001697325	10.1155/2022/3934381	found	L. Huang, The establishment of college student employment guidance system integr
40	PID1760001697325	10.1155/2022/5885065	retracted	Q. Liu, Research on the innovation of ideological and political education in universiti
41	PID1760001697325	10.1155/2022/5230215	retracted	Q. Sun, Ideological and political education reform using mobile phones as a carrier in
42	PID1760001697325	10.1155/2022/1415142	retracted	Y. Zhou, Beliefs and practice evaluation based on artificial intelligence models under
43	PID1760001697325	10.1155/2022/5908754	retracted	D. Chu, D. Yu, Analysis of the innovation path of marxism popularization based on b
44	PID1760001697325	10.1155/2022/4305886	retracted	X. Luo, Analysis the innovation path on psychological ideological with political teachi
45	PID1760001697325	10.1155/2022/1036168	retracted	D. Qiu, Innovation and reform of ideological and political course mode in colleges an
46	PID1760001697325	10.1155/2022/7439862	retracted	Y. Zhao, Innovation of network ideological and political education in universities facir
47	PID1760001697325	10.1155/2022/2564447	retracted	D. Yin, Innovation of network political and ideological education in colleges based on
48	PID1760001697325	10.1155/2021/4263981	found	J. Hong, Research on the development of innovation path of ideological and political



3. Istex TDM : présentation du catalogue de l'ensemble des web services

ISTEX TDM
Les services Istex pour la fouille de textes

Rechercher un web service

Tapez ici votre recherche, p.ex. : Classification RECHERCHER

<p>Texte Intégral</p> <p>datatableExtract DETECTION ET EXTRACTION DE TABLEAUX DANS UN ARTICLE SCIENTIFIQUE</p>	<p>Citations - Résumés</p> <p>textSimilarity CALCUL DE SIMILARITE ENTRE DES METADONNEES</p>
<p>Eléments catalographiques</p> <p>dataHomogenise HOMOGENISATION AUTOMATIQUE DE MOTS-CLÉS</p>	<p>Résumés</p> <p>aiAbstractCheck DETECTION DE RESUME SCIENTIFIQUE GENERE PAR IA</p>
<p>Texte Intégral</p> <p>textSummarize RESUME AUTOMATIQUE D'UN ARTICLE SCIENTIFIQUE</p>	<p>Citations</p> <p>topRefExtract EXTRACTION DES REFERENCES PHARES D'UN CORPUS</p>

Trouvez un web service correspondant à vos besoins

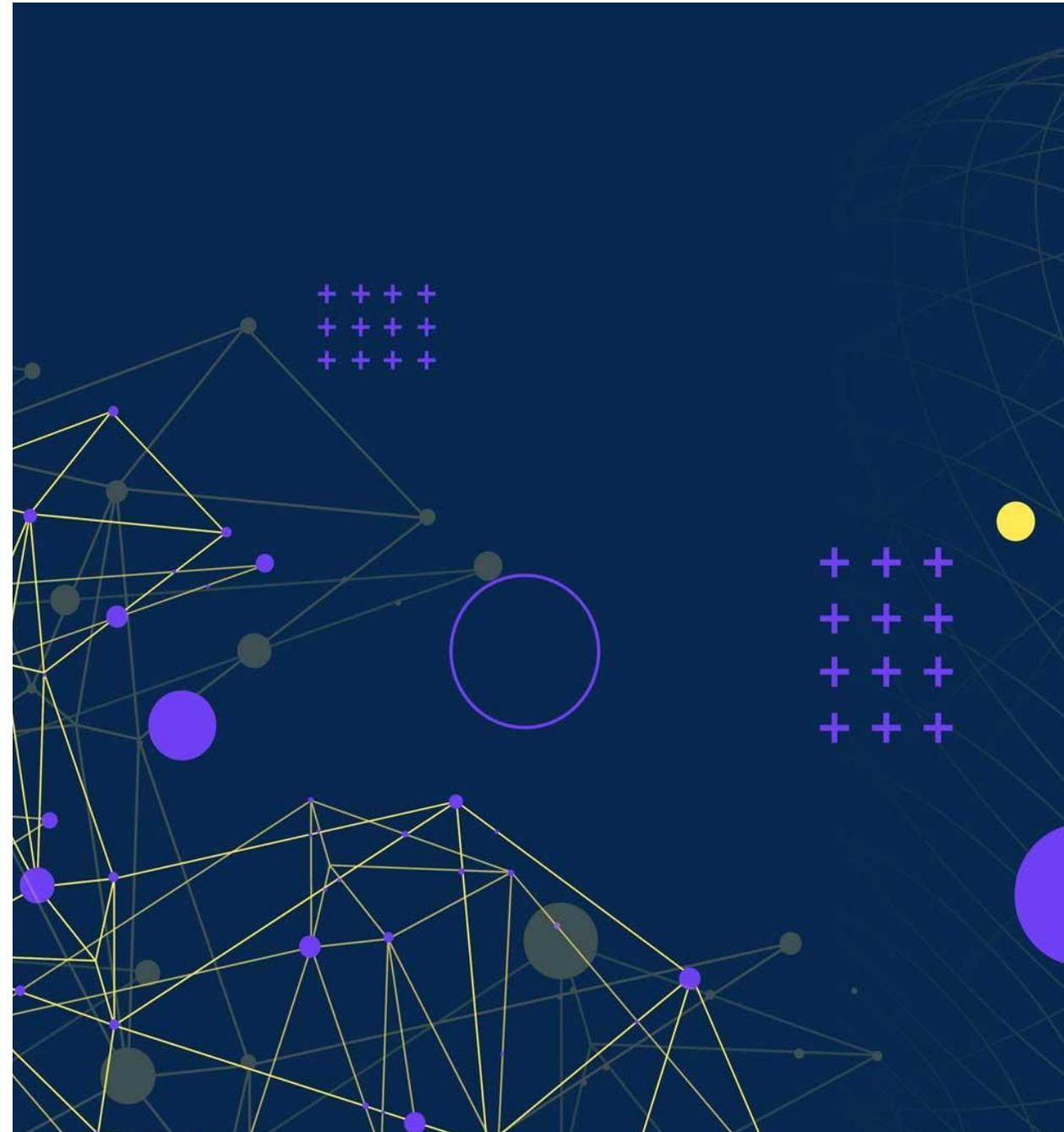
Nous développons et mettons à votre disposition des web services de TDM (Text and Data Mining) faciles à mettre en œuvre, couplés à un outil de création de tableaux de bord dynamiques.

Actuellement **44** web services sont disponibles

[COMMENT LES UTILISER ?](#)

[VOIR LA DOCUMENTATION](#)

<https://services.istex.fr/>



Merci

Des questions ?

