

l'Informatique Approximative

David Soulayrol
david@ti-nuage.fr

Association Ti Nuage
<https://ti-nuage.fr>

Lannion, 2024



Ti Nuage est une association loi 1901 située à Vieux-Marché. Ses buts :

- la fourniture de services numériques libres et éthiques ;
- l'accompagnement aux usages du numérique.

Elle propose :

Des services libres sur <https://www.ti-nuage.fr/services.html>

Des rendez-vous tous les premiers jeudis du mois.



└ Présentations

- Présentation du conférencier.
- Présentation de l'auditoire.

- la fourniture de services numériques libres et éthiques ;
- l'accompagnement aux usages du numérique.

Elle propose :

Des services libres sur <https://www.ti-nuage.fr/services.html>

Des rendez-vous tous les premiers jeudis du mois.

2025-02-15

└ Introduction

Introduction

1 Aperçu historique rapide

2 Un éventail de techniques

3 Connexions avec les activités humaines et l'environnement

4 L'IA (généraliste) est un faux sujet

Préhistoire

Les automates Les golems Le canard de Vaucanson Le joueur d'échec (truqué) Ce n'est pas le sujet...



2025-02-15

l'Informatique Approximative

└ Aperçu historique rapide

└ Préhistoire

Préhistoire

Les automates Les golems Le canard de Vaucanson Le joueur d'échec (truqué) Ce n'est pas le sujet...

De Boole à Minsky

Leibniz, Boole : développement de la logique Turing John McCarthy, Minsky

2025-02-15

l'Informatique Approximative

└ Aperçu historique rapide

└ De Boole à Minsky

Les hivers de l'IA

Alternance de périodes d'enthousiasme et de gels des financements. De 1974 à 1980 De 1987 à 1993 Règne des systèmes experts Développement de multiples techniques d'apprentissage automatique.



2025-02-15

l'Informatique Approximative

└ Aperçu historique rapide

└ Les hivers de l'IA

Puissance de calcul et transformeurs

Avec les années 2000, les nouvelles infrastructures de calcul et de réseaux, les données massives 2005, le projet Blue Brain 2010 DeepMind 2015 OpenAI 2022 Midjourney ou DALL-E 2, et croissance inédite de ChatGPT (un million d'utilisateurs en seulement cinq jours et cent millions d'utilisateurs en deux mois) 2023 Google Gemini, IA générative et multimodale.



2025-02-15

l'Informatique Approximative

└ Aperçu historique rapide

└ Puissance de calcul et transformeurs

Avec les années 2000, les nouvelles infrastructures de calcul et de réseaux, les données massives 2005, le projet Blue Brain 2010 DeepMind 2015 OpenAI 2022 Midjourney ou DALL-E 2, et croissance inédite de ChatGPT (un million d'utilisateurs en seulement cinq jours et cent millions d'utilisateurs en deux mois) 2023 Google Gemini, IA générative et multimodale.

1 Aperçu historique rapide

2 **Un éventail de techniques**

3 Connexions avec les activités humaines et l'environnement

4 L'IA (génération) est un faux sujet

Inférences

2025-02-15

l'Informatique Approximative
└─ Un éventail de techniques
 └─ Inférences

https://fr.wikipedia.org/wiki/Logique_floue

https://fr.wikipedia.org/wiki/Logique_floue



Logiques spécifiques

https://fr.wikipedia.org/wiki/Logique_floue

2025-02-15

l'Informatique Approximative

└─ Un éventail de techniques

└─ Logiques spécifiques

Algorithmes évolutionnistes

https://fr.wikipedia.org/wiki/Algorithme_

2025-02-15

l'Informatique Approximative

└─ Un éventail de techniques

└─ Algorithmes évolutionnistes

L'IA générative

2025-02-15

l'Informatique Approximative

└─ Un éventail de techniques

└─ L'IA générative

2025-02-15

l'Informatique Approximative

↳ Connexions avec les activités humaines et l'environnement

- Aperçu historique rapide
- Un éventail de techniques
- **Connexions avec les activités humaines et l'environnement**
- L'IA (généraliste) est un faux sujet

1 Aperçu historique rapide

2 Un éventail de techniques

3 Connexions avec les activités humaines et l'environnement

4 L'IA (généraliste) est un faux sujet

Cartographie

Extrait seulement.

Voir : <https://cartography-of-generative-ai.net/>

2025-02-15

l'Informatique Approximative

└─ Connexions avec les activités humaines et l'environnement

└─ Cartographie

L'IA générative

Les outils d'IA générative sont utilisés pour automatiser des écrits, des sons, des images. Ces tâches ne sont pas réalisées par une programmation mais reposent sur des exemples.

Si l'on dispose de beaucoup d'exemples d'un objet souhaité, on peut le reproduire en utilisant un modèle statistique qui se configure lui-même en identifiant les schémas récurrents.

Ce modèle statistique nécessite un jeu d'entraînement.

On pourrait dire que l'IA générative désassemble le langage (ou le son, l'image) pour le ré-assembler selon un calcul de probabilités. Si jusqu'à quelques années en arrière ces outils étaient entrainés pour produire des résultats concrets très XXX, ils sont maintenant capables d'aller au-delà de ces XXX pour produire de nombreux types et styles de contenus.

Cette capacité de généralisation se base sur une puissance de traitement bien plus grande et sur des jeux de données beaucoup plus larges et hétérogènes pour pouvoir répondre à n'importe quelle question.

Le résultat est que le développement de l'IA générative est si important



2025-02-15

l'Informatique Approximative

└ Connexions avec les activités humaines et l'environnement

└ L'IA générative

L'IA générative
Les outils d'IA générative sont utilisés pour automatiser des écrits, des sons, des images. Ces tâches ne sont pas réalisées par une programmation mais reposent sur des exemples.
Si l'on dispose de beaucoup d'exemples d'un objet souhaité, on peut le reproduire en utilisant un modèle statistique qui se configure lui-même en identifiant les schémas récurrents.
Ce modèle statistique nécessite un jeu d'entraînement.
On pourrait dire que l'IA générative désassemble le langage (ou le son, l'image) pour le ré-assembler selon un calcul de probabilités. Si jusqu'à quelques années en arrière ces outils étaient entrainés pour produire des résultats concrets très XXX, ils sont maintenant capables d'aller au-delà de ces XXX pour produire de nombreux types et styles de contenus.
Cette capacité de généralisation se base sur une puissance de traitement bien plus grande et sur des jeux de données beaucoup plus larges et hétérogènes pour pouvoir répondre à n'importe quelle question.
Le résultat est que le développement de l'IA générative est si important

Jeux de données

The industrial compilation of training datasets is achieved by extracting content from the largest digital archive in existence : the Internet. The data is obtained through automated 'scraping' of online content published and shared by millions of internet users. The original motivation for this extraction did not foresee today's commercial exploitation by start-ups and platforms, but was driven by the desire for scholarly and non-commercial research. Now that these huge digital archives have been used to generate texts and images on demand, we are faced with a series of paradoxes and controversies within the cultural industries. If, on the one hand, the ideology of big data understands Internet content as a vast repository that can be extracted, processed and automated, on the other hand, this extractivist drive is seen by other cultural actors as a process of massive privatisation of the creativity of millions of Internet users.



2025-02-15

l'Informatique Approximative

└ Connexions avec les activités humaines et l'environnement

└ Jeux de données

The industrial compilation of training datasets is achieved by extracting content from the largest digital archive in existence: the Internet. The data is obtained through automated 'scraping' of online content published and shared by millions of internet users. The original motivation for this extraction did not foresee today's commercial exploitation by start-ups and platforms, but was driven by the desire for scholarly and non-commercial research. Now that these huge digital archives have been used to generate texts and images on demand, we are faced with a series of paradoxes and controversies within the cultural industries. If, on the one hand, the ideology of big data understands Internet content as a vast repository that can be extracted, processed and automated, on the other hand, this extractivist drive is seen by other cultural actors as a process of massive privatisation of the creativity of millions of Internet users.

Ajustement

While datasets are raw materials, they are not enough to refine the current level of personalised interaction offered by online AI services. To make it more user-friendly, it needs pockets of micro-workers dedicated to refining the model : scoring the answers it generates, tagging images or text, annotating and other evaluation processes that involve cognitive work (often consisting of clicking on one of the options displayed on the screen). Large corporations outsource these employment services to third party companies, which in turn export them to countries in the global south with high poverty rates, where the cost/hour of each worker becomes a residual cost. Some of the outsourced companies have been documented operating in refugee camps, training displaced people in Lebanon, Uganda, Kenya and India to perform micro-tasks with data, exploiting the economic hardship of displaced people (Jones, 2022).



2025-02-15

l'Informatique Approximative

- ↳ Connexions avec les activités humaines et l'environnement

- ↳ Ajustement

Ajustement

While datasets are raw materials, they are not enough to refine the current level of personalised interaction offered by online AI services. To make it more user-friendly, it needs pockets of micro-workers dedicated to refining the model : scoring the answers it generates, tagging images or text, annotating and other evaluation processes that involve cognitive work (often consisting of clicking on one of the options displayed on the screen). Large corporations outsource these employment services to third party companies, which in turn export them to countries in the global south with high poverty rates, where the cost/hour of each worker becomes a residual cost. Some of the outsourced companies have been documented operating in refugee camps, training displaced people in Lebanon, Uganda, Kenya and India to perform micro-tasks with data, exploiting the economic hardship of displaced people (Jones, 2022).

Filtrage

At a later stage, it is necessary to filter the content generated by the large AI models. The most common tasks at this stage are those aimed at moderating so-called toxic content : hate, political controversy, extreme or explicit sex and violence, originally contained in the training datasets. The moderation work is carried out in the labour markets of countries such as Kenya or Uganda (or even in the large metropolises of the North, among migrant communities) and is dedicated to identifying and classifying texts and images with violent, murderous, rape or child abuse content. Behind the autonomous appearance of AI tools, therefore, we find different levels of human resources displaced to different geographies, precarised and invisible by the technological innovation industry.



2025-02-15

l'Informatique Approximative

- ↳ Connexions avec les activités humaines et l'environnement
 - ↳ Filtrage

Filtrage

At a later stage, it is necessary to filter the content generated by the large AI models. The most common tasks at this stage are those aimed at moderating so-called toxic content : hate, political controversy, extreme or explicit sex and violence, originally contained in the training datasets. The moderation work is carried out in the labour markets of countries such as Kenya or Uganda (or even in the large metropolises of the North, among migrant communities) and is dedicated to identifying and classifying texts and images with violent, murderous, rape or child abuse content. Behind the autonomous appearance of AI tools, therefore, we find different levels of human resources displaced to different geographies, precarised and invisible by the technological innovation industry.

Discours public

The AI panic discourse, channelled by institutes and philanthropic foundations and supported by the "visionary" leaders of start-ups, is covered by the media and has the desired alarming effect on public opinion. At a time when these technologies are being subjected to the first regulatory processes (the first AI law was passed in the European Parliament in early 2024), the agitation of existential threat messages is intended to support the AI industry's demands for self-regulation by public administrations. Meanwhile, in the context of a crisis of standardised formats of credibility, defined for some years now by the term "post-truth", social networks are already filling up with synthetic messages, images or texts generated by these tools. The automation of public discourse and its implications in a context of growing misinformation and political polarisation will be at the core of the media agenda in the coming years.



2025-02-15

l'Informatique Approximative

- └ Connexions avec les activités humaines et l'environnement

- └ Discours public

Discours public

The AI panic discourse, channelled by institutes and philanthropic foundations and supported by the "visionary" leaders of start-ups, is covered by the media and has the desired alarming effect on public opinion. At a time when these technologies are being subjected to the first regulatory processes (the first AI law was passed in the European Parliament in early 2024), the agitation of existential threat messages is intended to support the AI industry's demands for self-regulation by public administrations. Meanwhile, in the context of a crisis of standardised formats of credibility, defined for some years now by the term "post-truth", social networks are already filling up with synthetic messages, images or texts generated by these tools. The automation of public discourse and its implications in a context of growing misinformation and political polarisation will be at the core of the media agenda in the coming years.

Absence d'intelligence

> Les modèles génératifs ne possèdent aucune notion des information manipulées ou produites. > Leurs résultats n'ont de sens que lorsqu'ils sont interprétés et vérifiés par des experts.



2025-02-15

l'Informatique Approximative

└ L'IA (généralive) est un faux sujet

└ Absence d'intelligence

> Les modèles génératifs ne possèdent aucune notion des information manipulées ou produites. > Leurs résultats n'ont de sens que lorsqu'ils sont interprétés et vérifiés par des experts.

Non reproductibilité, absence de preuve

2025-02-15

l'Informatique Approximative

└ L'IA (généralive) est un faux sujet

└ Non reproductibilité, absence de preuve

Consommation massive de données

2025-02-15

l'Informatique Approximative

└ L'IA (généralive) est un faux sujet

└ Consommation massive de données

Problème de droits Problème de jeux de données sans biais

Problème de droits Problème de jeux de données sans biais

Consommation énergétique

2025-02-15

l'Informatique Approximative

└ L'IA (généralive) est un faux sujet

└ Consommation énergétique

C'est fini !

Merci de votre attention.
Des questions ?

