

De la machine de Turing à l'intelligence artificielle

En 1936, bien avant le premier ordinateur, **Turing** décrit une machine imaginaire : un ruban, une tête qui lit et écrit, une table de règles. Elle ne sert à rien pratiquement. Elle sert à définir ce que « calculer » veut dire — et à montrer qu'il existe des problèmes qu'aucune machine ne résoudra jamais. Tout ce chapitre part de là.

1 Calculer : ce que la machine de Turing définit

DÉFINITION

La machine de **Turing** est un modèle abstrait : un ruban de cases, une tête qui lit un symbole, l'efface ou le remplace, et se déplace selon une table de règles finie. Son intérêt n'est pas l'efficacité mais la **définition** : est calculable ce qu'une telle machine peut produire. Elle donne ainsi un sens précis au mot **algorithme** — une suite finie d'instructions non ambiguës.

Le résultat le plus important attaché à ce modèle est **négatif** : certains problèmes n'admettent aucun algorithme, quelle que soit la puissance de la machine. Il y a donc une limite qui ne tient ni au matériel ni au temps disponible. C'est un point utile à garder contre l'idée qu'un ordinateur assez gros finirait par tout résoudre.

2 De la règle écrite à l'apprentissage

PROPRIÉTÉ

Deux façons de faire agir une machine doivent être distinguées. Dans la première, un humain écrit les **règles** : le programme applique ce qui a été prévu. Dans la seconde — l'**apprentissage** —, on fournit des **données** et le système ajuste ses paramètres pour reproduire au mieux les régularités qu'elles contiennent. Personne n'écrit alors la règle appliquée, et personne ne peut toujours l'énoncer après coup.

RÈGLE

Un système qui apprend ne connaît du monde que ce que ses **données** en montrent : il reproduit leurs **biais** sans savoir qu'il en a. Ce qui manque à l'entraînement manque au système, et rien dans son fonctionnement ne le signale — il répond avec la même assurance sur ce qu'il maîtrise et sur ce qu'il n'a jamais vu.

données d'entraînement → régularités apprises → biais reproduits

LE PIÈGE QUI COÛTE LE PLUS DE POINTS

Conclure d'une bonne performance globale qu'un système est fiable pour tous les cas. — Une **moyenne** cache une répartition. Les erreurs peuvent se concentrer sur une catégorie sous-représentée dans les **données** d'entraînement, et le système y échouer systématiquement tout en affichant un score élevé. Un taux global ne dit rien à la personne qui tombe dans la mauvaise catégorie.

Le réflexe : Demander la performance **par catégorie**, jamais seulement le score moyen.

3 Ce qu'un système ne fait pas

PROPRIÉTÉ

Deux limites à retenir. La première : un système d'**apprentissage** trouve des **corrélations** ; il n'établit pas de causes, et prédire n'est pas expliquer. La seconde : ce qu'il **optimise** a été choisi par quelqu'un — maximiser le nombre de bonnes réponses ou minimiser les erreurs les plus graves ne donne pas le même système. Ce choix se fait dans un fichier de configuration ; il n'en est pas moins un arbitrage.

Le mot **intelligence artificielle** recouvre donc des choses très différentes, du programme à règles au système entraîné sur d'immenses corpus. Le seul critère utile en classe n'est pas de savoir si la machine « pense », question qui n'avance à rien : c'est de savoir **d'où vient son comportement** — d'une règle qu'on peut lire, ou de données qu'il faut examiner.

EXAMINER UNE SORTIE D'ALGORITHME

- 1 Déterminer si le système suit des **règles écrites** ou **apprend** sur des données.
- 2 Demander sur quelles **données** il a été entraîné, et ce qui y est sous-représenté.
- 3 Regarder la performance **par catégorie**, et non le score global.
Un score de 95 % peut recouvrir un échec complet sur une catégorie minoritaire.
- 4 Vérifier si le lien exploité est une **corrélation** ou une cause établie.
- 5 Nommer ce que le système **optimise**, et qui l'a choisi.

À VÉRIFIER

- Ai-je dit à quoi sert la machine de **Turing** — définir le calculable ?
- Ai-je mentionné qu'il existe des problèmes sans **algorithme** ?
- Ai-je distingué règle écrite et **apprentissage** ?
- Ai-je expliqué qu'un **biais** vient des **données** ?
- Ai-je séparé **corrélation** et cause ?
- Ai-je dit que ce qui est optimisé résulte d'un choix ?

À RETENIR

- | | |
|---|--|
| <ul style="list-style-type: none">• Machine de Turing : ruban, tête, table de règles. Elle définit le calculable.• Un algorithme : suite finie d'instructions non ambiguës.• Certains problèmes n'admettent aucun algorithme : limite logique.• Règle écrite $\neq$ apprentissage sur des données. | <ul style="list-style-type: none">• Un biais vient des données, et le système ne le signale pas.• Un score global peut masquer un échec sur une catégorie.• Corrélation n'est pas cause : prédire n'est pas expliquer.• Ce qui est optimisé résulte d'un choix humain. |
|---|--|