Plan détaillé

Sujet : **Pouvons-nous parler de conscience artificielle pour l'IA ?**

Découpage simple :

- Introduction
- Développement
- Conclusion

Introduction sert à problématiser le sujet, à justifier son existence
|> pose une première pierre de définition
|> donne les deux visions opposées autour de la conscience de l'IA
|> expose le plan, notre démarche et notre méthode

Développement explore le sujet
|> essaye de répondre
|> approfondit le travail de définition de l'introduction
|> ne doit *pas* ouvrir le sujet : se concentre uniquement sur le sujet

Conclusion rappelle notre travail et répond
|> pourquoi parle-t-on de ça ?
|> quelles sont les problèmes ?
|> quelles sont les réponses ?

Introduction

Conscience ?
|> faculté permettant d'avoir connaissance de sa propre existence
|> capacité à avoir des connaissances sur les choses
-> ces définitions font appels à la subjectivité de la chose étudiée
|> compliqué de trancher à cause de la chose considérée comme sujet et non comme objet

Artificielle ?
|> ce qui a été modifiée par l'humain
|> tout l'écosystème a été modifié par l'humain (réchauffement climatique)
-> besoin de définir le degré de modification
|> est aussi une chose non essentielle : l'objet artificielle est un artifice pas nécessaire

IA ?
|> machine intelligente
|> comment définir l'intelligence ?
|> besoin de la conscience pour parler d'intelligence

-> pas suffisant car notre conception de l'IA évolue avec le temps
|> définition technique : algorithme de *machine learning*

Conscience artificielle semble être adaptée pour l'IA
|> impression de conscience, mais c'est une illusion construite par la technique
|> fonctionne bien avec la notion de *machine learning*
|> mais est-ce que la notion de conscience artificielle possède-t-elle un véritable sens ?

Conscience morale
|> conscience signifie en ancien français conscience morale
|> est une forme de prérequis pour établir la conscience (conscience de soi et d'autrui)
-> IA a-t-elle une conscience morale ?
|> non, car même après une phase d'apprentissage, elle continue à être immoral sur des questions morales évidentes (donner des instructions pour construire une bombe)
|> phase d'apprentissage est comparable à notre éducation (à justifier dans développement), mais n'est pas suffisante pour créer (ou faire ressortir) une conscience morale
-> les actions des IA ne permettent pas de parler de conscience morale

Technique derrière une IA

Ici, IA = algorithme de *machine learning*

machine learning, *deep learning*, 7 milliards de paramètre, qu'est-ce que ça veut dire ?
|> théorie mathématique derrière (algèbre linéaire)
|> théorie informatique derrière (réseaux de neurones)
|> mise en pratique (data set, entrainement)
|> défis techniques (alignement, efficacité)

Comment une IA répond-elle à un problème ?
|> que ce problème soit génératif (GPT, Midjourney) ou non (algorithmes de recommandation)
|> récolte des données (entrées par l'utilisateur ou non)
|> approche probabiliste (statistiques, aléatoire)
|> raisonnement intérieur
-> IA ne réfléchit pas d'une manière causale, elle regarde des probabilités

> ✎ IA ou algorithme ?
>
> Bien que les IA génératives et les algorithmes de recommandation sont tous les deux des algorithmes de machine learning, nous n'appelons pas le deuxième IA. Cette différence montre une fausse impression de contrôle (à justifier) sur les IA de recommandation : on pourrait beaucoup plus simplement les maîtriser

Expérience de pensée dite *La chambre chinoise* sur notre notion de conscience
|> fonctionne très bien pour l'IA
-> des statistiques et des probabilités peuvent-elles créées une forme de conscience ?

La notion de conscience artificielle

Conscience de groupes, conscience animale, IA -> exemples de conscience artificielle ?

Caillou du désert qui se déplace tout seul

conscience artificielle = créé par l'humain + artifice de conscience (« fausse conscience »)

Application à l'IA et limites

Besoin de faire attention à l'anthropisation de l'IA

Ce que ça explique bien :

1. IA qui trompe à ses créateurs
2. Problème de l'alignement

Ce que ça limite :

1. Fausse notre relation
2. La régulation de l'IA
3. Qui parlera de conscience pour les algorithmes de recommandation

Conclusion

Nouveau plan ?

1. Technique derrière l'IA (ne change pas), un peu plus court ?
2. L'IA possède-t-elle une conscience ?
    1. recherche d'une définition de la conscience
    2. application de la vision de conscience à l'IA
    3. compliquer de parler de conscience sur les algorithmes de recommendations
    4. compliquer de trancher pour de la vrai conscience -> introduction à la conscience artificielle
3. Création de la notion de conscience artificielle
    1. vision de l'artificielle comme tromperie (mais peut-être vrai)
    2. vision de l'artificielle comme création humaine
    3. marche bien pour l'IA (tous les types)
    4. n'explique pas que l'IA (marche aussi pour la conscience de groupe)

Évoquer les limites dans la conclusion plutôt que dans une full partie
|> évite le côté liste