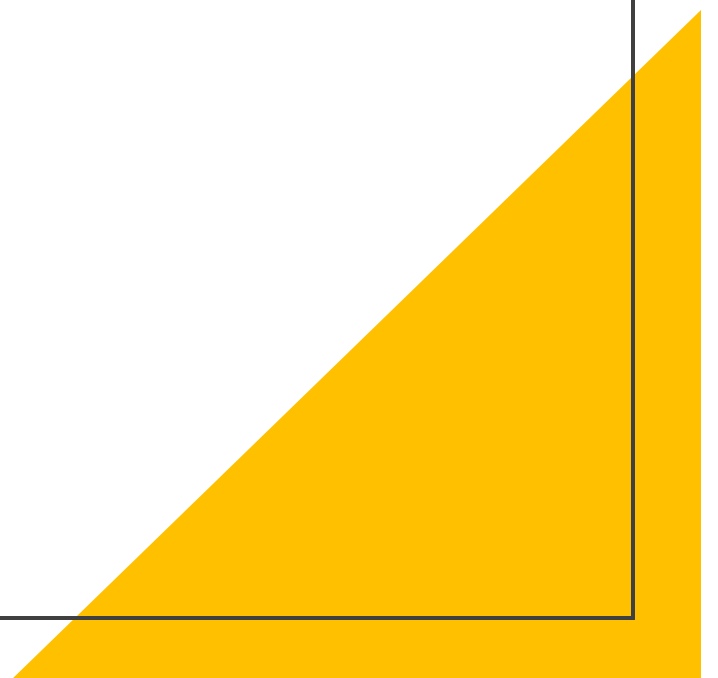


Esprits numériques

PHI 6385

Professeur Jonathan Simon



Aperçu

1) Plan du cours / mes attentes

2) Introduction Générale

3) Conscience vs Intelligence:
le cas des invertébrés

Plan du cours

réunions sur rendez-vous :
envoyez-moi un courriel à

Jonathan.simon@umontreal.ca

Plan du cours

- Tout est sur le site web:
- <https://jonsimon.net/esprits-numeriques/>

- Vos notes seront basées sur :
- Une présentation en classe (avec diapositives) d'environ 40 minutes, et la conduite de la discussion qui s'ensuit (45 minutes) 10% ▪
- Une dissertation de 1500-2500 mots 38% (27 octobre)
 - Complétez la « liste de contrôle du résumé » 2% (20 octobre)
- Une dissertation finale de 1500 – 4000 mots 48% (15 décembre)
 - Complétez la « liste de contrôle du résumé » 2% (8 décembre)

Plan du cours

- (I) Une présentation en classe (avec diapositives) d'environ 30 minutes, et la conduite de la discussion qui s'ensuit (10 minutes) 10% ■
-
- 1) Présentez un argument tiré d'un texte pertinent pour cette semaine (discutez-en d'abord avec moi).
- 2) Discutez de ses prémisses - expliquez, pour chaque prémisse, comment l'auteur la justifie.
- 3) cette discussion doit être suffisamment accessible pour inviter à une conversation, en se concentrant sur les forces et les faiblesses de la prémisse.

Plan du cours

- (I) Une présentation en classe (avec diapositives) d'environ 30 minutes, et la conduite de la discussion qui s'ensuit (10 minutes)
10% ▪
-
- Diapos: 3 points
- Clarté d'exposition : 3 points
- Qualité / précision : 3 points
- Capacité à stimuler et à diriger la discussion : 1 point

Plan du cours

- (II) Une dissertation de 1500-2500 mots 38% (27 octobre)
 - Complétez la « liste de contrôle du résumé » 2% (20 octobre)

1) Je partagerai la liste de contrôle en ligne sous la forme d'un formulaire Google - il s'agit de s'assurer que vous avez une bonne structure à l'esprit.

2) Je clarifierai les spécificités du travail de rédaction dans les semaines à venir.

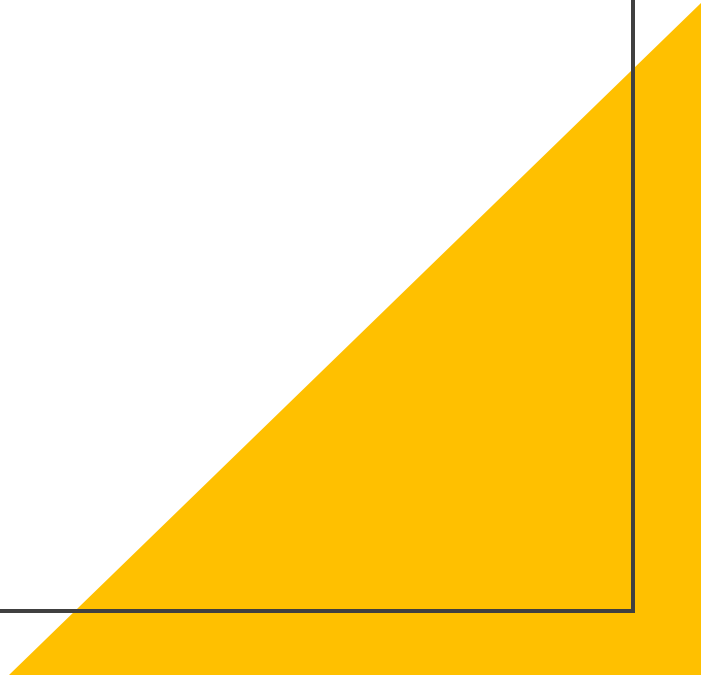
Plan du cours

- (III) Une dissertation finale de 1500 – 4000 mots 48% (15 décembre)
 - Complétez la « liste de contrôle du résumé » 2% (8 décembre)

1) Je partagerai la liste de contrôle en ligne sous la forme d'un formulaire Google - il s'agit de s'assurer que vous avez une bonne structure à l'esprit.

2) Je clarifierai les spécificités du travail de rédaction dans les semaines à venir.

Introduction Générale



Introduction Générale

- Les machines peuvent-elles être conscientes ? Sont-elles déjà conscientes ?



Introduction Générale

- Les machines peuvent-elles être conscientes ? Sont-elles déjà conscientes ?



Introduction Générale

1

Qu'est-ce que la conscience ?
(Conscience vs Intelligence)

2

La conscience est-elle une question de calcul ou de substrat ?

3

Quelle méthodologie pouvons-nous utiliser pour déterminer ce qu'est la conscience, ou quels systèmes sont conscients ?

Introduction Générale

4

Si la conscience est computationnelle, de quel type de calcul s'agit-il ?

5

Les réseaux neuronaux artificiels seront-ils un jour conscients, ou le sont-ils déjà ?

6

Les réseaux neuronaux artificiels peuvent-ils réellement voir / penser / assister / créer, ou effectuent-ils *juste* une simulation de ces choses ?

7

Si les êtres artificiels peuvent être conscients, ont-ils un statut moral ?

Introduction Générale

- 1) Conscience vs Intelligence



L'intelligence: Comportement flexible ou adaptable, orienté vers un objectif



La conscience : La subjectivité, la phénoménologie, l'effet que cela fait, le sentiment - le fait que la lumière intérieure soit allumée.

Introduction Générale

- Conscience vs Intelligence
- La principale ligne de débat dans la philosophie de l'intelligence artificielle : ce qu'il faudra pour que les systèmes artificiels soient vraiment *intelligents*. (on peut également distinguer l'intelligence étroite et l'intelligence générale)
- Mais certains soutiennent qu'il peut y avoir une conscience sans intelligence (invertébrés) et que nous pouvons au moins concevoir une intelligence sans conscience (zombies)









Introduction Générale

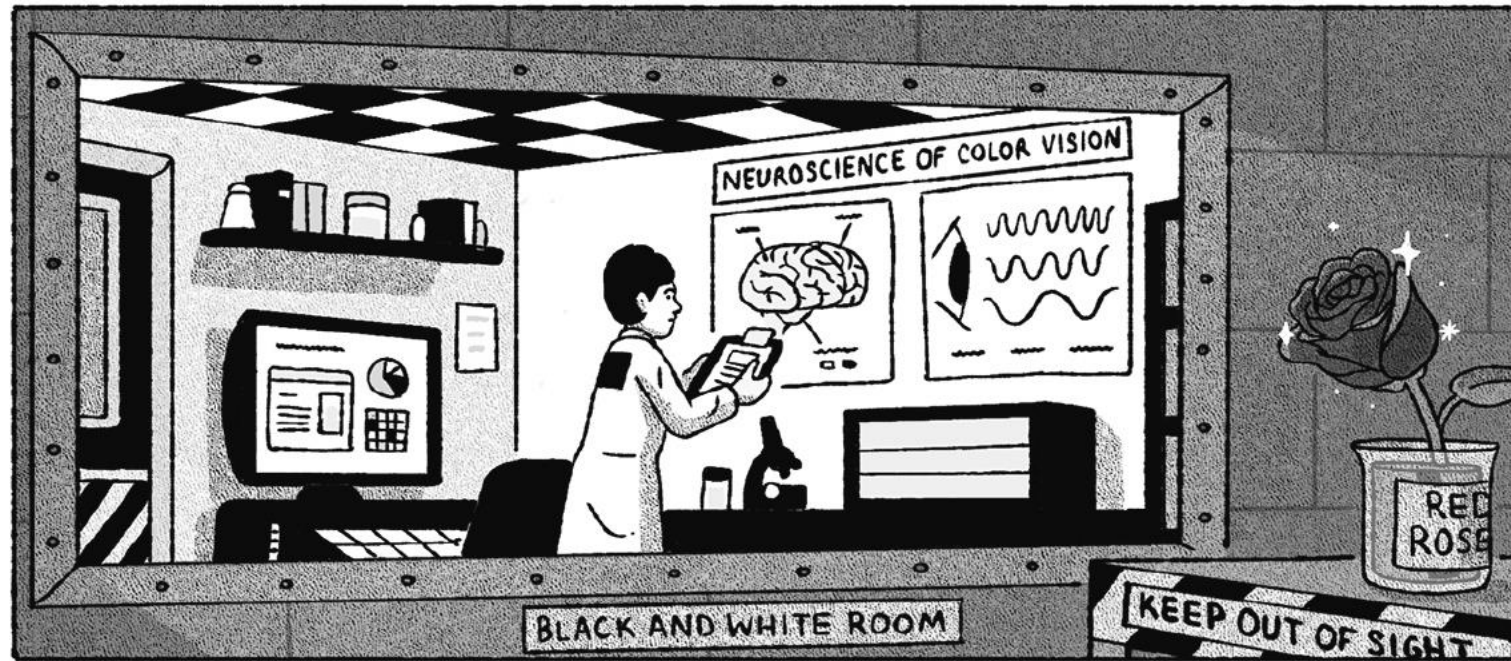


Figure 37 Does Mary the color scientist (here modeled on Mary Whiton Calkins, a leading philosopher and psychologist of the early 20th century) know what it's like to see red?

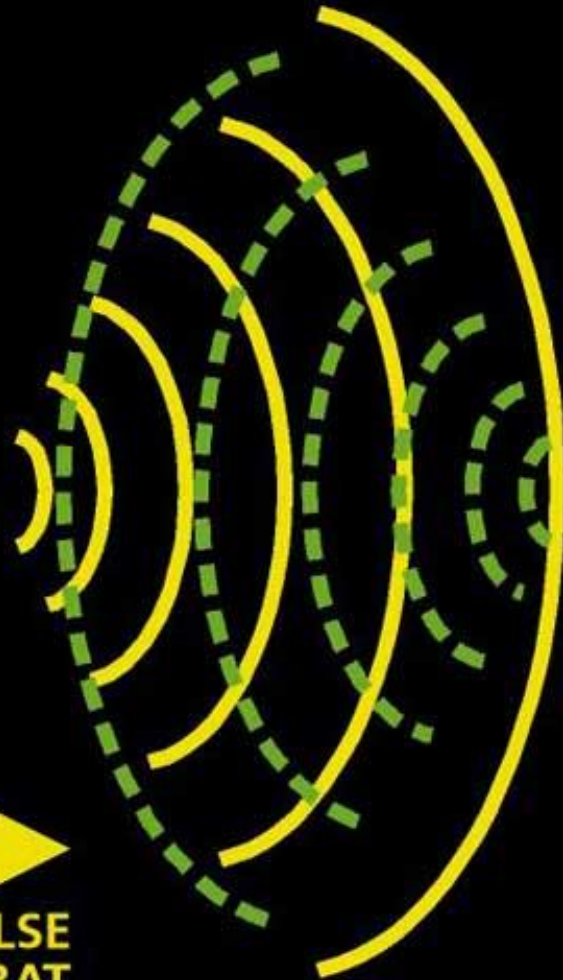


BAT



MOTH

SOUND PULSE
SENT OUT BY BAT



REFLECTED PULSE
RECEIVED BY BAT



2) Calcul ou substrat?

| Main / The Chinese Room Argument

3) Quel
méthodologie?

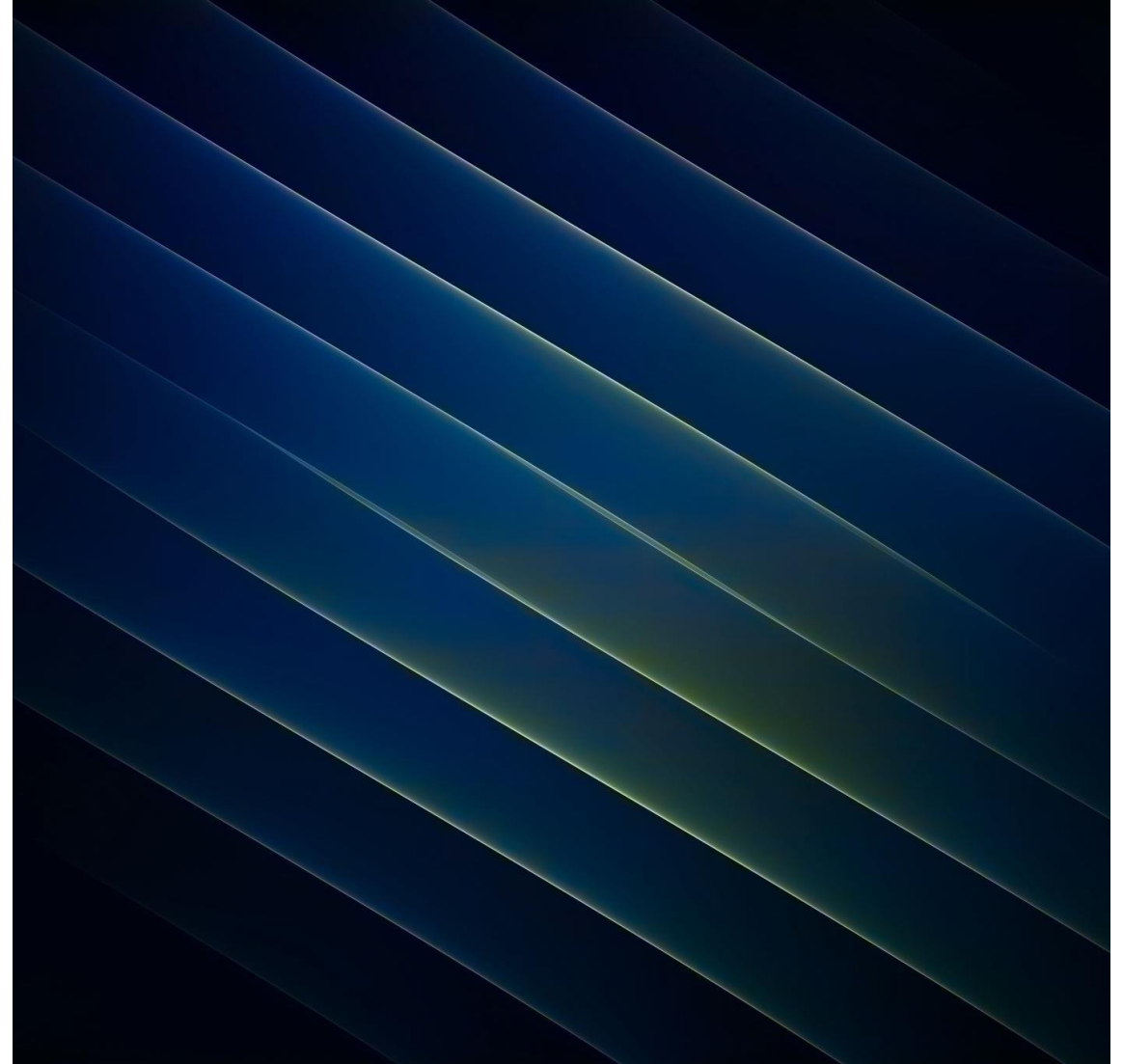
Le comportement seul ?

Théories de la conscience
chez l'homme ?

Un mélange ?

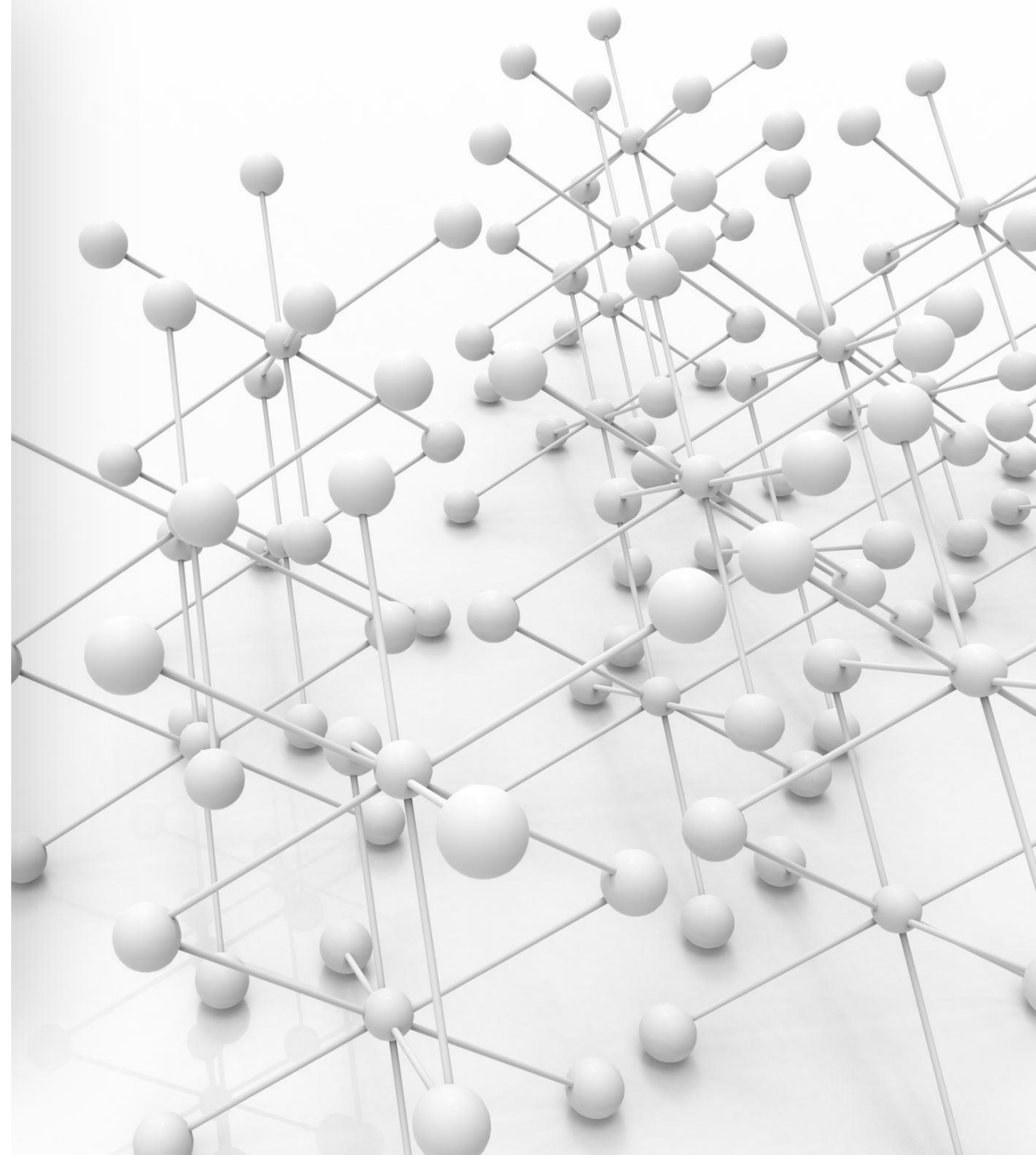
4) Quel type de calcul?


Systemes de
symboles vs
réseaux de
neurones



5) Des reseaux neuronaux conscients?

- Y sommes-nous déjà arrivés, et si non, que faut-il ajouter : plus de données, plus de paramètres, une architecture plus sophistiquée ... ou quelque chose d'entièrement différent ?





6) Des reseaux neuronaux qui pensent / voient / désirent ?

Que pouvons-nous conclure des architectures plus spécifiques qui ressemblent aux architectures biologiques (CNNs pour la vision, RNNs et transformateurs pour le langage, systems RL pour la desir, GANs pour la créativité)

7) Des reseaux neuronaux artificiels comme des entités morales

Conscientisme,
sentientisme,
agentivisme ..

-

Agent moral vs
patient moral

La conscience des invertébrés

La conscience des invertébrés

La littérature ici comme une illustration de la façon dont la conscience peut se détacher de l'intelligence (générale):

Les abeilles et les poissons font preuve d'intelligence dans des domaines limités, mais ni les unes ni les autres ne sont généralement intelligentes.

Pourtant, il existe de bonnes raisons théoriques de supposer qu'ils pourraient être conscients.

La conscience des invertébrés

Aussi, une distinction méthodologique très utile
(tirée de l'article de Birch)

Méthodologie *Theory-Neutral*

Méthodologie *Theory-Heavy*

Méthodologie *Theory-Light*

Nous nous demanderons si l'un d'entre eux est
approprié pour l'étude de la conscience artificielle

La conscience
des
invertébrés

Méthodologie *Theory-Neutral*:

*Les méthodes qui ne font pas du tout appel à la théorie.
Tests comportementaux,
recours à l'intuition, etc.*

La conscience
des
invertébrés

Méthodologie *Theory-Heavy*:

Tout d'abord, détermine quelle théorie est correcte (ou quel ensemble de théories est le plus susceptible d'être correct), puis détermine ce qu'elles impliquent.

La conscience des invertébrés

- Méthodologie *Theory-Light*:
- *Cherche des "marqueurs" de la conscience, en utilisant à la fois la théorie et l'observation, dans les cas où nous sommes plus confiants, puis vérifie lesquels nous trouvons dans les cas plus exotiques..*

Un mot sur chat Gpt

Mot sur chat gpt

- Je n'ai rien contre le fait que vous utilisiez le chat gpt pour vous aider à étudier, à réfléchir ou à acquérir des connaissances de base.
- Cependant, je vous déconseille fortement de l'utiliser pour faire vos devoirs à votre place. Et ce, pour trois raisons.

Mot sur chat gpt

- 1) <https://integrite.umontreal.ca/accueil/>: Elle est officiellement interdite dans le cadre de l'évaluation (couvrant tous les travaux notés). S'il devient évident pour mes correcteurs que vous l'avez utilisé, ils sont obligés de le traiter comme un cas de plagiat.

Mot sur chat gpt

- 2) C'est de la philosophie. Pourquoi suivez-vous ce cours ? Apprendre à penser et à écrire clairement, et comprendre quelque chose aux grands problèmes qui structurent le monde dans lequel vous vivez, mérite certainement quelques minutes de votre précieux temps. Confier ces tâches à une IA est un moyen infallible de vous rendre obsolète

Mot sur chat gpt

- 3) Le Chat GPT est plus stupide que vous ne le pensez. En général, il réussit les tests qui mesurent les tâches de niveau secondaire, mais il échoue à de nombreux tests pour des tâches de niveau universitaire : il ne vous obtiendra certainement pas un A.

Mot sur chat gpt

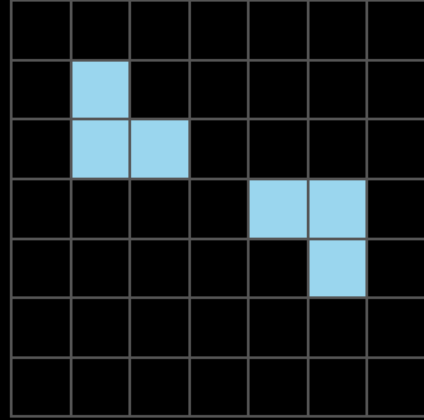
- Mais c'est encore pire que cela. Le Chat GPT commet des erreurs surprenantes et stupides, notamment en ce qui concerne la logique et le bon sens, et il peut les dissimuler subtilement de manière à ce que vous ne les remarquiez pas.

Mot sur chat gpt

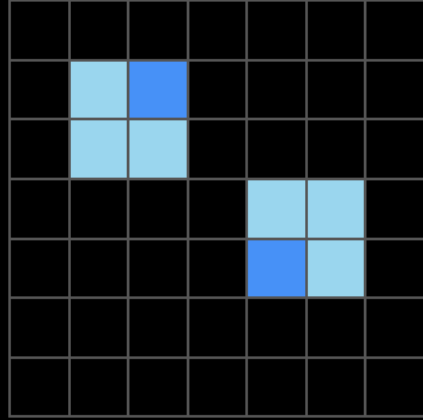
- Examinez ce problème : il s'agit d'identifier la règle

EXAMPLES

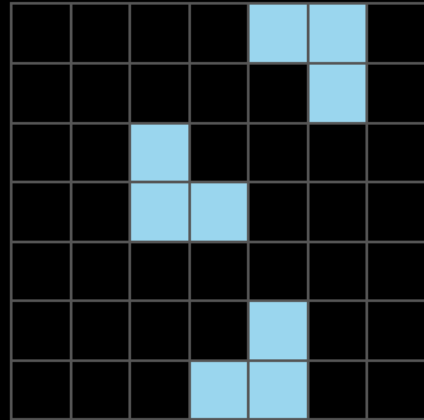
Ex.1 Input (7x7)



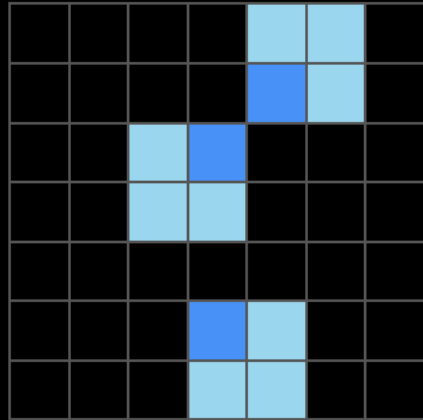
Ex.1 Output (7x7)



Ex.2 Input (7x7)



Ex.2 Output (7x7)



Mot sur chat gpt

- Chat GPT ne peut pas résoudre ce problème. Il y a littéralement un prix d'un million de dollars à gagner si vous pouvez programmer une IA capable de résoudre de manière fiable des problèmes de ce type.
- <https://arcprize.org/>