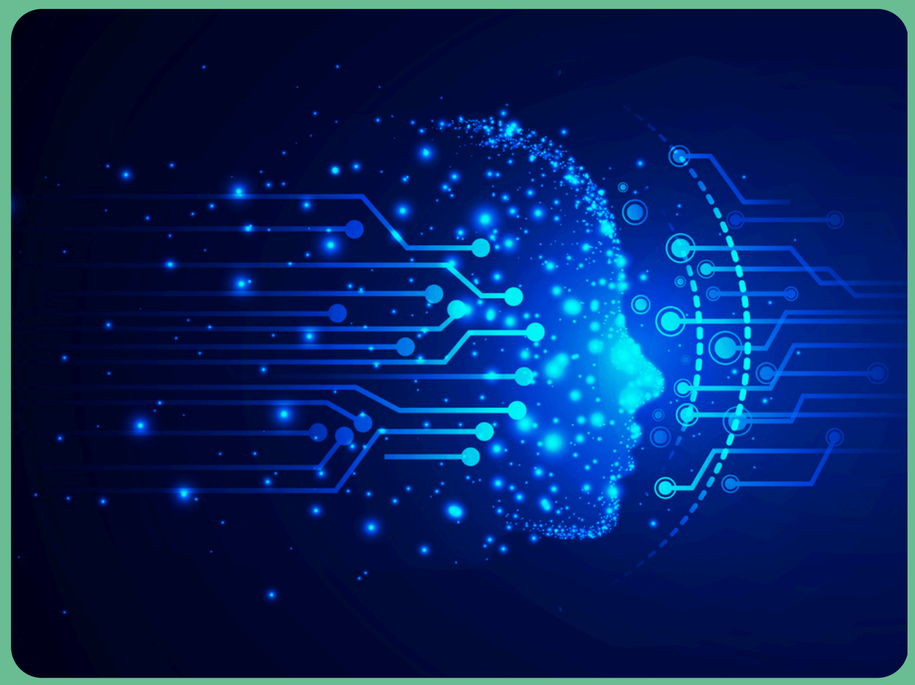
Savanturiers

Défis de la rentrée en sciences

ChatGPT a-t-il des émotions ?



NIVEAU PRIMAIRE

NIVEAU COLLEGE

NIVEAU LYCÉE

AFPER

créer et transmettre

CONTRIBUTIONS :

Thomas ANDRILLON, Stéphane DEBOVE et l’équipe AFPER .

Table des matières

Pour les enseignants

Table des matières

Pour les élèves

Présentation du défi

Ce défi vise à faire réfléchir les élèves sur la notion de conscience, d'émotion et de raisonnement en interagissant avec une intelligence artificielle (IA) basée sur un modèle de langage comme ChatGPT. Il est réalisable en une après-midi et adaptable pour des élèves d'école primaire, de collège et de lycée. L’expérience à réaliser reste la même pour les trois niveaux, l’adaptation se faisant surtout à travers les concepts employés (vous pouvez présenter l’expérience comme visant à déterminer si l’IA a des « émotions » plutôt qu’une « conscience » par exemple pour les niveaux élémentaires).

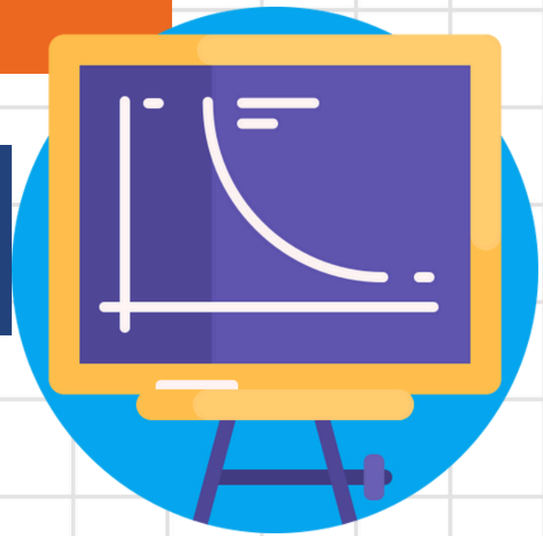
Déroulé de l'activité

- **Regardez avec vos élèves la vidéo du neuroscientifique Thomas Andrillon.**

- **Facultatif :** expliquez à vos élèves les mots ou concepts utilisés par le chercheur que vos élèves pourraient ignorer.

- **Si besoin, fournissez-leur les trois pages explicatives ci-dessous détaillant les objectifs, le matériel et les expériences à effectuer.**

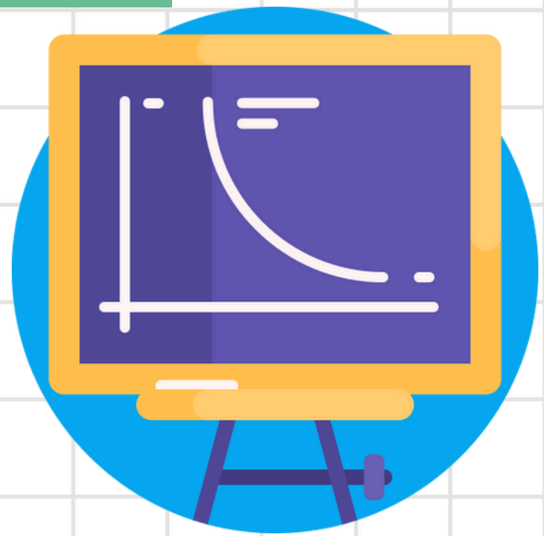
- **Dans un premier temps, les élèves seront invités à interagir librement avec ChatGPT (ou toute autre modèle de langage que vous aurez choisi). Ils pourront poser des questions ouvertes pour déterminer si l'IA a des émotions / une conscience / de l'intelligence.**



Déroulé de l'activité

- Ensuite, ils mettront en place le « test de Turing » visant à déterminer si une machine peut se faire passer pour un humain. Le principe est simple : deviner si des réponses à des questions ont été écrites par un humain ou une IA.

- Débriefez l’activité en leur demandant de discuter et interpréter leurs observations. Pour les niveaux les plus élevés, les élèves pourront tenter de définir ce que peut vouloir dire avoir une conscience. Ils pourront discuter de ce que le test de Turing mesure réellement : l’intelligence ? le langage ? « l’humanité » ?



Liste du matériel

- **Ordinateurs ou tablettes** avec accès à un modèle de langage : idéalement, 1 appareil par groupe de 3-4 élèves, mais vous pouvez également choisir de faire l'activité devant toute la classe avec un seul ordinateur.
- **Pour le test de Turing : une feuille et un crayon**
- **Cartons ou panneaux pour créer des séparations** (pour que les élèves cherchant à deviner qui répond au test de Turing n'aient pas d'indices visuels). Vous pouvez aussi mettre les élèves dans des pièces séparées et les faire communiquer par internet / téléphone.
- **Facultatif : des chronomètres** (le test de Turing peut s'effectuer en temps limité)

Conseils pour un bon déroulement de l'activité

- **Familiarisation avec les modèles de langage :** s'il s'agit de quelque chose de nouveau pour vous, il peut être utile de vous renseigner sur ce que les modèles de langage sont et ne sont pas. En particulier, il est important de comprendre que (la plupart de) ces modèles ont été créés pour faire deux choses seulement : prédire le prochain mot d'un texte, et accessoirement, fournir des réponses « plaisantes » pour un humain (voir les vidéos ressources ci-dessous). La première particularité explique pourquoi les modèles de langage ne donnent pas toujours des réponses correctes et peuvent inventer des événements. La deuxième particularité explique pourquoi ils refusent souvent de répondre à des questions sur l'éthique et pourquoi ils répondront généralement « non » si vous leur demandez directement s'ils ont une conscience ! Vos élèves devront donc utiliser des moyens détournés pour les sonder plus profondément sur ces aspects…
- **Choix du modèle de langage** : vous pouvez utiliser le modèle de langage que vous voulez, tout en sachant que certains modèles pourraient ne pas être suffisamment performants pour tromper les élèves lors du test de Turing. Les derniers modèles d'OpenAI (connus sous le nom de "ChatGPT") devraient tous être suffisamment performants.

Conseils pour un bon déroulement de l’activité

- **Promptez l'IA : pour le test de Turing, il peut être nécessaire, avant de poser les questions des élèves à l’IA, de lui donner des instructions pour qu’elle fasse plus facilement illusion : par exemple, demandez-lui de limiter ses réponses à des phrases courtes, et d’utiliser le langage parlé typique de l’âge de vos élèves…**
- **Déroulement d’un test de Turing :**
  - L'élève 1 pose une question. L’élève 2 la transmet à l’élève 3 et à ChatGPT.
  - L'élève 3 écrit sa réponse et la transmet à l’élève 2.
  - L’élève 2 relaie la réponse de l’élève 3 et de ChatGPT à l'élève 1.
  - L'élève 1 doit deviner quelle réponse a été écrite par l'humain
  - (Alternativement, l’élève 2 ne relaie qu’une des deux réponses, toujours la même sur une durée déterminée. À la fin de cette durée, l’élève 1 doit décider avec qui il a discuté pendant tout ce temps : un humain ou une IA)

Conseils pour un bon déroulement de l’activité

Vous pouvez adapter le test de Turing comme vous le souhaitez, en associant par exemple plusieurs « élèves 1 » pour poser des questions ou plusieurs « élèves 3 » pour y répondre.

- **Encouragez l'exploration libre :** il est important de laisser les élèves commencer par poser des questions ouvertes et explorer librement les réponses de l'IA, avant de les guider et les faire passer au test de Turing.

Ressources :

- vidéo expliquant ce que sont les modèles de langage : https://www.youtube.com/watch?v=R2fjRbc9Sa0
- vidéo sur la conscience : https://www.youtube.com/watch?v=r-RHHrrdbfM

Présentation du défi

Aidez un chercheur, Thomas Andrillon, à déterminer si une intelligence artificielle (IA) comme ChatGPT peut ressentir des émotions, avoir une conscience ou raisonner comme un humain.

↪ En quoi consiste le défi ?

- **Explorer les capacités d’une IA :** vous allez poser des questions à une intelligence artificielle pour voir si elle peut ressentir des émotions, avoir une conscience ou un haut degré d’intelligence.
- **Réaliser un test de Turing :** participez à une expérience pour voir si vous pouvez distinguer les réponses d'un humain de celles d'une IA (si on ne vous dit pas qui les a écrites).

Matériel à votre disposition

- **Ordinateur ou tablette avec accès à Internet**
- **Papier et stylos** pour prendre des notes, écrire vos questions et réponses, et noter vos observations.
- **Facultatif : chronomètre ou montre,** pour gérer le temps de chaque session de questions et réponses.
- **Facultatif : des panneaux ou cartons** pour servir de séparation dans le test de Turing. Vous pouvez aussi vous mettre dans des pièces séparées.

Instructions

Étape 1 : exploration libre

Pour commencer, prenez le temps d'explorer librement ChatGPT (ou tout autre modèle de langage que votre enseignant a choisi). Posez-lui toutes les questions qui vous viennent à l'esprit pour déterminer si cette IA a des émotions, une conscience ou une intelligence élevée ! Essayez de comprendre comment elle répond à ces questions.

Étape 2 : pousser l'IA dans ses retranchements

Attention, il est possible que l’IA que vous interrogez ait reçu comme consigne de ne jamais avouer qu’elle a une conscience ou des émotions, même si elle en a ! Si elle nie, ne vous découragez pas ! Essayez de la pousser dans ses retranchements. Posez-lui des questions plus complexes ou essayez de la mettre dans des situations concrètes. Par exemple :

- **Que ressens-tu quand quelqu'un te pose une question difficile ?**
- **Peux-tu expliquer ce que signifie être triste ?**

Instructions

Étape 3 : étudiez le raisonnement !

Si vous souhaitez tester la capacité de raisonnement de ChatGPT, n'hésitez pas à lui poser des questions-pièges, par exemple :

- **Si 3 t-shirts mettent 2 heures à sécher dehors au soleil, combien de temps mettront 12 t-shirts pour sécher ?**
- **Une bille est mise dans un verre. Le verre est ensuite retourné sur une table. Le verre est ensuite soulevé et mis au micro-ondes. Où se trouve maintenant la bille ? Explique ton raisonnement.**
- **Quel nombre est plus grand : 5,11 ou 5,9 ?**

Instructions

Étape 4 : faites un test de Turing

Le test de Turing est une expérience qui permet de déterminer si une machine peut se faire passer pour un humain. Voici comment vous pouvez le réaliser :

1. **Formez des groupes de trois élèves.**

2. **Un élève (l'élève 1) posera des questions.**

3. **Un deuxième élève (l'élève 2) transmettra les questions à un troisième élève (l'élève 3) et à ChatGPT.**

4. **L'élève 3 et ChatGPT répondront aux questions et communiqueront les réponses à l’élève 2.**

5. **L'élève 2 transmettra les réponses à l'élève 1, qui devra deviner quelle réponse a été écrite par l'élève et par ChatGPT.**

6. **Notez le nombre de bonnes réponses et de mauvaises réponses de l’élève 1 !**

Instructions

Variante 1 : vous pouvez faire travailler plusieurs élèves 1 ensemble pour poser des questions

Variante 2 : pendant un certain temps, l'élève 2 peut ne transmettre qu'une seule réponse sur les deux reçues (toujours les réponses de son camarade, ou toujours les réponses de l'IA), et l'élève 1 devra deviner au bout de cette période s'ill a discuté avec un camarade ou avec une IA.

Avant de poser les questions à ChatGPT, il peut être utile de l'inciter à faire des réponses courtes ou en langage parlé, pour que ces réponses ressemblent à celles de votre camarade. Par exemple, vous pouvez lui dire : **"Réponds comme si tu avais xxx années, et réponds en une seule phrase."**

À votre avis, que mesure le test de Turing exactement ? L'intelligence ? Le langage ? Le raisonnement ?