Sondage sûreté de l'IA

Ce questionnaire vise à mesurer les connaissances et les opinions des étudiants dans le domaine de la sûreté de l'IA.

1) Lesquelles de ces tâches ont été résolues par des IA à un niveau similaire à un humain ?

☐ Jouer à Minecraft
☐ Résoudre des problèmes des Olympiades Internationales de Mathématiques
☐ Décrire et interpréter un meme
☐ Identifier l'émotion principale d'une musique

2) GPT3.5, dont l'architecture permet uniquement de prédire le mot suivant dans un texte, a été entraîné sur des jeux de données de parties d'échecs après son entraînement principal. Quel est son score Elo évalué ?

☐ 1000 (compréhension solide du jeu)
☐ 1500 (niveau moyen en club)
☐ 1800 (bon niveau en club)
☐ 2200 (joueur en tournoi, niveau national)

3) Parmi ces aspects des LLM, pour lesquels les chercheurs ont une bonne compréhension basée sur des fondements théoriques, et qui permet de prédire des comportements ?

☐ le lien entre taille du modèle et performances
☐ les comportements de refus de réponse
☐ la capacité à effectuer un raisonnement logique
☐ comment sont stockées les connaissances factuelles

4) Seriez-vous capable d'expliquer ce qu'est le RLHF ?

☐ oui ☐ non

5) OpenAI, l'entreprise ayant créé ChatGPT, a comme objectif public de créer une IA générale. Cela signifie que cette IA fera aussi bien qu'un humain dans essentiellement toutes les tâches, y compris lorsque cela demande des capacités de planification. Dans combien de temps prévoient-ils d'atteindre cet objectif ?

☐ 2 ans ☐ 4 ans ☐ 10 ans ☐ 20 ans ☐ 30 ans ☐ 50 ans

6) D'après le dernier sondage parmi les chercheurs en IA (2800 chercheurs, janvier 2024), quelle proportion des chercheurs pensent qu'un scénario catastrophe peut arriver à cause d'une IA incontrôlable ? Spécifiquement, la question était: « je pense que la probabilité d'un scénario qui cause l'extinction de l'espèce humaine ou conséquence similaire est supérieure à 1/10 ».

☐ 5% ☐ 10% ☐ 15% ☐ 20% ☐ 35% ☐ 50% ☐ 65% ☐ 80% ☐ 85% ☐ 90% ☐ 95%

Sondage sûreté de l'IA

Ce questionnaire vise à mesurer les connaissances et les opinions des étudiants dans le domaine de la sûreté de l'IA.

1) Lesquelles de ces tâches ont été résolues par des IA à un niveau similaire à un humain ?

☐ Jouer à Minecraft
☐ Résoudre des problèmes des Olympiades Internationales de Mathématiques
☐ Décrire et interpréter un meme
☐ Identifier l'émotion principale d'une musique

2) GPT3.5, dont l'architecture permet uniquement de prédire le mot suivant dans un texte, a été entraîné sur des jeux de données de parties d'échecs après son entraînement principal. Quel est son score Elo évalué ?

☐ 1000 (compréhension solide du jeu)
☐ 1500 (niveau moyen en club)
☐ 1800 (bon niveau en club)
☐ 2200 (joueur en tournoi, niveau national)

3) Parmi ces aspects des LLM, pour lesquels les chercheurs ont une bonne compréhension basée sur des fondements théoriques, et qui permet de prédire des comportements ?

☐ le lien entre taille du modèle et performances
☐ les comportements de refus de réponse
☐ la capacité à effectuer un raisonnement logique
☐ comment sont stockées les connaissances factuelles

4) Seriez-vous capable d'expliquer ce qu'est le RLHF ?

☐ oui ☐ non

5) OpenAI, l'entreprise ayant créé ChatGPT, a comme objectif public de créer une IA générale. Cela signifie que cette IA fera aussi bien qu'un humain dans essentiellement toutes les tâches, y compris lorsque cela demande des capacités de planification. Dans combien de temps prévoient-ils d'atteindre cet objectif ?

☐ 2 ans ☐ 4 ans ☐ 10 ans ☐ 20 ans ☐ 30 ans ☐ 50 ans

6) D'après le dernier sondage parmi les chercheurs en IA (2800 chercheurs, janvier 2024), quelle proportion des chercheurs pensent qu'un scénario catastrophe peut arriver à cause d'une IA incontrôlable ? Spécifiquement, la question était: « je pense que la probabilité d'un scénario qui cause l'extinction de l'espèce humaine ou conséquence similaire est supérieure à 1/10 ».

☐ 5% ☐ 10% ☐ 15% ☐ 20% ☐ 35% ☐ 50% ☐ 65% ☐ 80% ☐ 85% ☐ 90% ☐ 95%

Toutes les questions suivantes sont des questions de prédictions et d'opinion. Répondez selon vous.

7) Selon vous, quel est le risque d'une catastrophe causée par l'IA comparable à la seconde guerre mondiale ou à une pandémie dévastatrice dans les 10 prochaines années ?

☐ < 0.1% ☐ 1% ☐ 5% ☐ 10% ☐ 20% ☐ 30% ☐ 50% ☐ 75% ☐ 90% ☐ > 95%

8) À quel point êtes vous inquiet pour ces différents scénarios causés par l'IA ?

	1	2	3	4	5	6	7	8	9	10
• cyberattaques automatiques	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
• crash économique d'une ampleur jamais vue	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
• perte de contrôle d'une IA malveillante	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
• automatisation de la majorité des métiers	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

9) Utilisez-vous régulièrement un LLM type ChatGPT ?

☐ oui ☐ non

10) Est-ce une bonne chose que l'API de GPT4 soit publique ?

☐ oui ☐ non

11) Se préoccupe-t-on suffisamment de la sûreté de l'IA d'un point de vue juridique ?

☐ oui ☐ non

12) Se préoccupe-t-on suffisamment de la sûreté de l'IA d'un point de vue de la recherche ?

☐ oui ☐ non

13) Devrait-on ralentir fortement la création de modèles plus puissants que ceux actuels ?

☐ oui ☐ non

14) Devrait-on ralentir fortement le déploiement des modèles actuels les plus puissants ?

☐ oui ☐ non

15) Avez vous entendu parler de:

• l'EU AI Act	☐ oui	☐ non
• Le sommet de la sûreté de l'IA à Séoul (mai 2024)	☐ oui	☐ non
• PauseAI	☐ oui	☐ non
• Le centre français pour la sécurité de l'IA (CeSIA)	☐ oui	☐ non

16) Si un cours de créneau D existait à Télécom sur les aspects techniques de la sûreté de l'IA, est-ce que vous seriez intéressé ?

☐ oui ☐ non

17) Des commentaires sur ce sondage ?

.....

.....

Toutes les questions suivantes sont des questions de prédictions et d'opinion. Répondez selon vous.

7) Selon vous, quel est le risque d'une catastrophe causée par l'IA comparable à la seconde guerre mondiale ou à une pandémie dévastatrice dans les 10 prochaines années ?

☐ < 0.1% ☐ 1% ☐ 5% ☐ 10% ☐ 20% ☐ 30% ☐ 50% ☐ 75% ☐ 90% ☐ > 95%

8) À quel point êtes vous inquiet pour ces différents scénarios causés par l'IA ?

	1	2	3	4	5	6	7	8	9	10
• cyberattaques automatiques	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
• crash économique d'une ampleur jamais vue	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
• perte de contrôle d'une IA malveillante	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
• automatisation de la majorité des métiers	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

9) Utilisez-vous régulièrement un LLM type ChatGPT ?

☐ oui ☐ non

10) Est-ce une bonne chose que l'API de GPT4 soit publique ?

☐ oui ☐ non

11) Se préoccupe-t-on suffisamment de la sûreté de l'IA d'un point de vue juridique ?

☐ oui ☐ non

12) Se préoccupe-t-on suffisamment de la sûreté de l'IA d'un point de vue de la recherche ?

☐ oui ☐ non

13) Devrait-on ralentir fortement la création de modèles plus puissants que ceux actuels ?

☐ oui ☐ non

14) Devrait-on ralentir fortement le déploiement des modèles actuels les plus puissants ?

☐ oui ☐ non

15) Avez vous entendu parler de:

• l'EU AI Act	☐ oui	☐ non
• Le sommet de la sûreté de l'IA à Séoul (mai 2024)	☐ oui	☐ non
• PauseAI	☐ oui	☐ non
• Le centre français pour la sécurité de l'IA (CeSIA)	☐ oui	☐ non

16) Si un cours de créneau D existait à Télécom sur les aspects techniques de la sûreté de l'IA, est-ce que vous seriez intéressé ?

☐ oui ☐ non

17) Des commentaires sur ce sondage ?

.....

.....