

Mimetic Models: Ethical Implications of AI that Acts Like You

**Reid McIlroy-Young, Jon Kleinberg, Siddhartha Sen, Solon Barocas, Ashton
Anderson**

CS Dept. University of Toronto, Cornell University & Microsoft Research

09/01/2023 - Saša Radosavljevic

Définitions

Modèle mimétique :

Algorithme entraîné sur une base de données d'un certain individu dans un domaine particulier ayant pour objectif de prédire le comportement de cet individu dans ce domaine.

Ethique :

Science qui traite des principes régulateurs de l'action et de la conduite morale.

Résolution d'un conflit de valeurs.

Contexte



Figure 1 : Modèle de Hikaru Nakamura, Chess.com

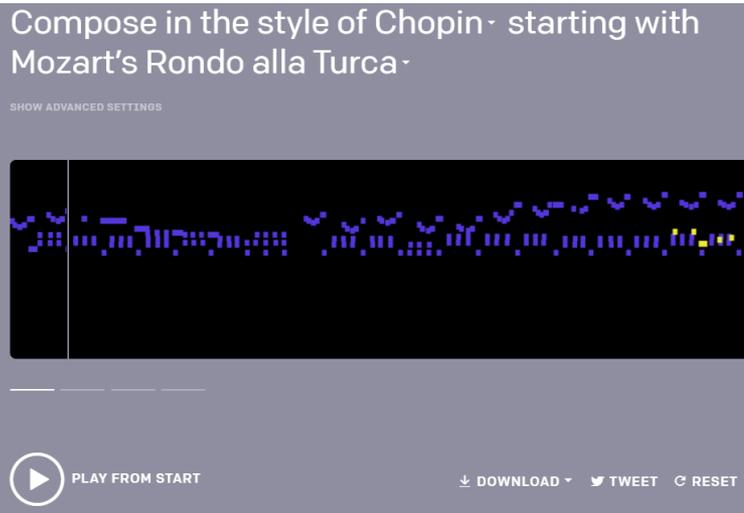


Figure 2 : MuseNet, OpenAI

Paint Monna Lisa in the style of Van Gogh

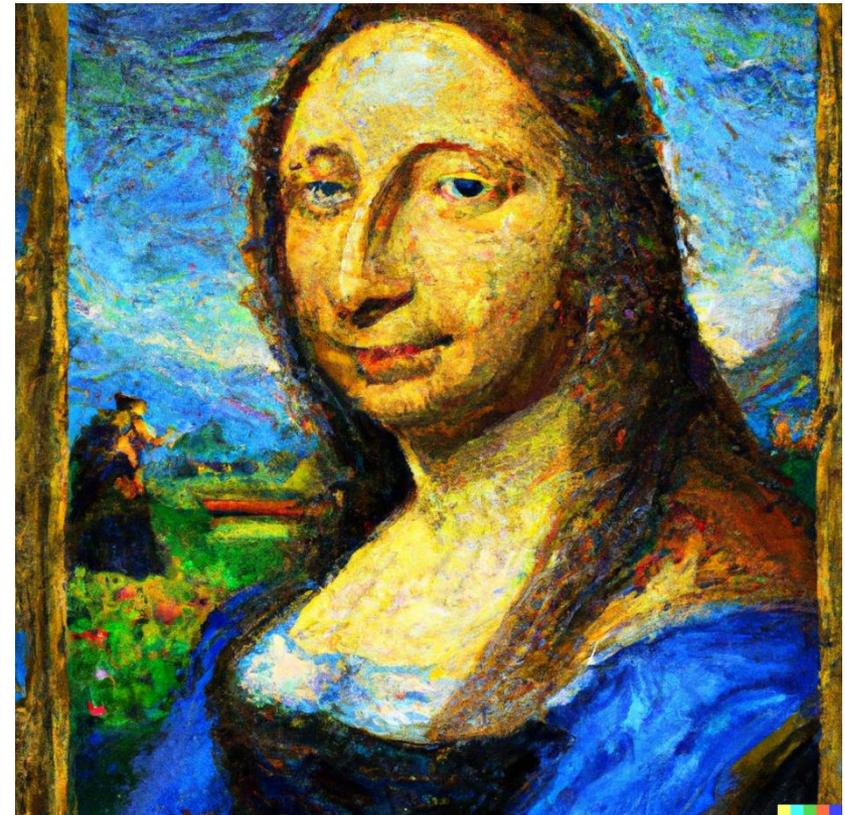


Figure 3 : DALLE-2, OpenAI

Précédents



Figure 3 : Conduite autonome, Tesla



Responsabilités ?

Article L123-1/2 : Le conducteur n'est pas responsable en cas d'infraction avec un système de conduite autonome de niveau 3

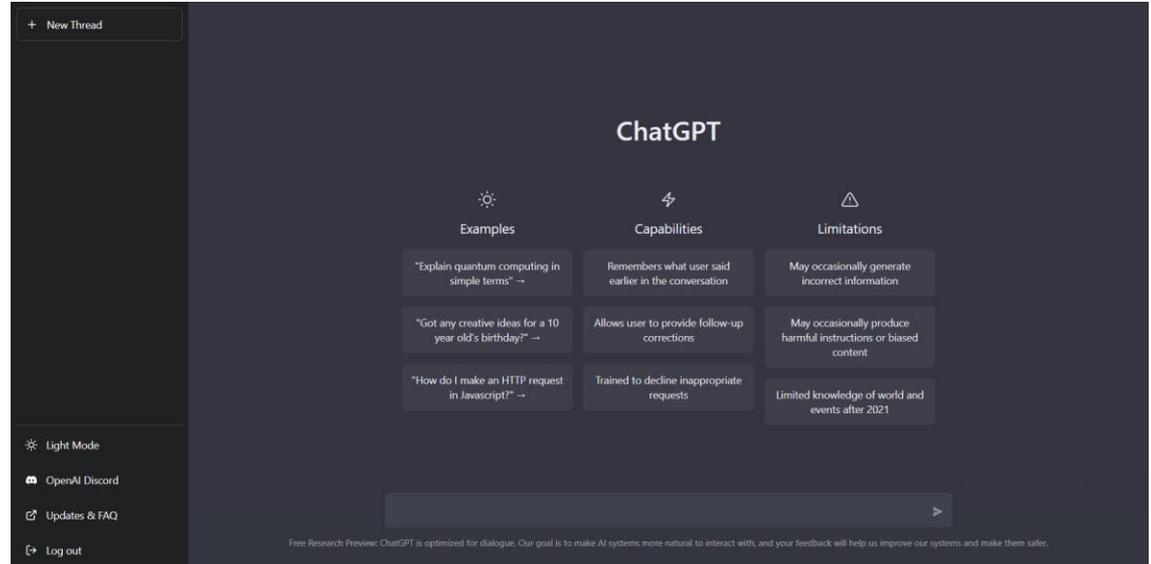


Figure 4 : ChatGPT, OpenAI



Utilisation ?

Interdiction dans le milieu scolaire, New York, USA

Problèmes

- Enjeux sociaux du mimétisme informatique
- Problèmes d'éthique

Exemple

Deepfakes : Répandre de la désinformation (voix ou image)

Répondeur de mail : Risque de décevoir une personne

Utilisations

- Mime en tant que moyen
 - Préparation en vue d'un tournoi
 - Préparation d'un entretien
- Mime en tant que fin
 - Substitution d'une personne
 - Événement contrefactuels
- Mime de soi-même
 - Démultiplier sa productivité
 - Filtrer des interactions sociales

Mime en tant que moyen

Préparation d'un tournoi d'Echecs

Analyse de son adversaire [1]

Problèmes :

- Différences entre analyse des parties de son adversaire et analyse du mime ?
- Avantages réels ?

Réponses :

- Extrapolation exigeante en temps et capacités cognitives
- Capacités d'extrapolation du ML bien supérieures aux humains
- Le mime effectue le travail de l'homme à sa place

Implications :

- Equité entre les joueurs
- Réglementations compliquées

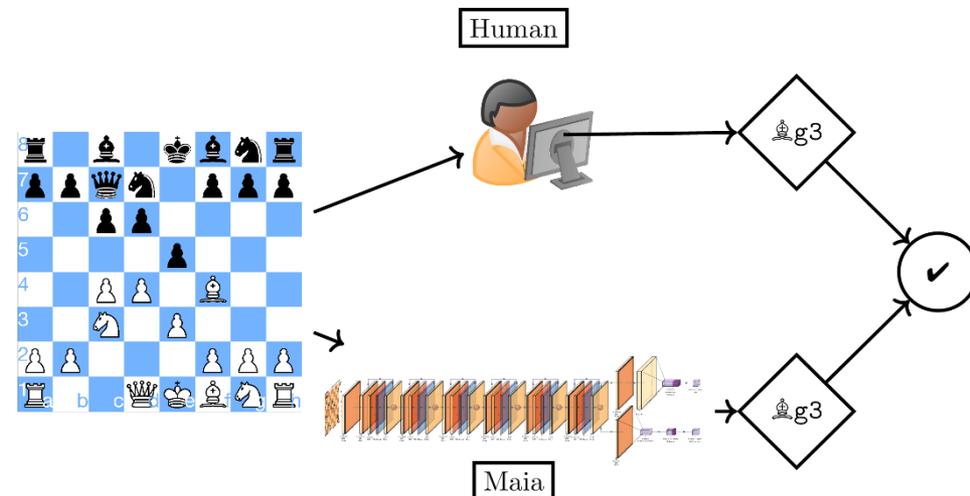


Figure 6 : MaiaChess.com

Mime en tant que moyen

Préparation d'un entretien

Changer son allocution en fonction de la personnalité du recruteur

Problèmes :

- Jeu à somme non nulle
- Manipulation et exploitation

Réponses :

- Utilisation volontaire d'un mime
- Évaluation complètement différente des compétences
- Complément à un interview classique

Implications :

- Plus de relation de symétrie
- Objectif d'un entretien détourné



Figure 7 : Entretien biaisé

Mime en tant que fin

Substitution d'une personne

Remplacer un tuteur en ligne par son mime quand il n'est pas disponible

Problèmes :

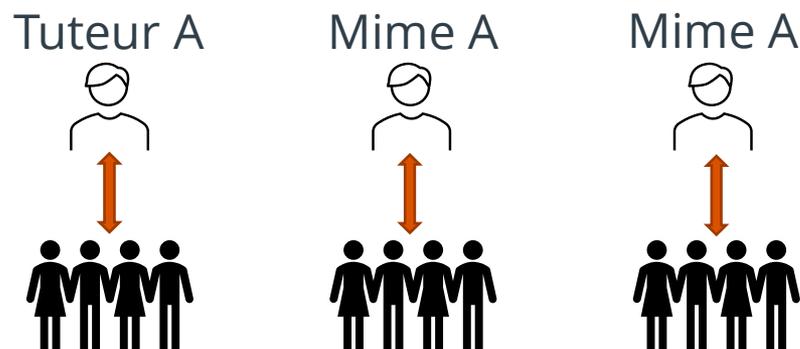
- Le tuteur est-il indispensable ?
- Un seul tuteur suffirait-il à tous les clients ?
- Menace de remplacer une personne par son mime
- Perte de valeur économique et sociale de la présence

Réponses :

- Garder un quota minimum
- Nécessité de transparence lorsqu'un mime est utilisé

Implications :

- Interaction sociale compromise
- Rémunération compliquée
- Menace d'automatisation de tâches que l'on pensait jusqu'à maintenant impossible à automatiser



Mime de soi-même

Effectuer le travail à sa place

Démultiplier sa force de travail

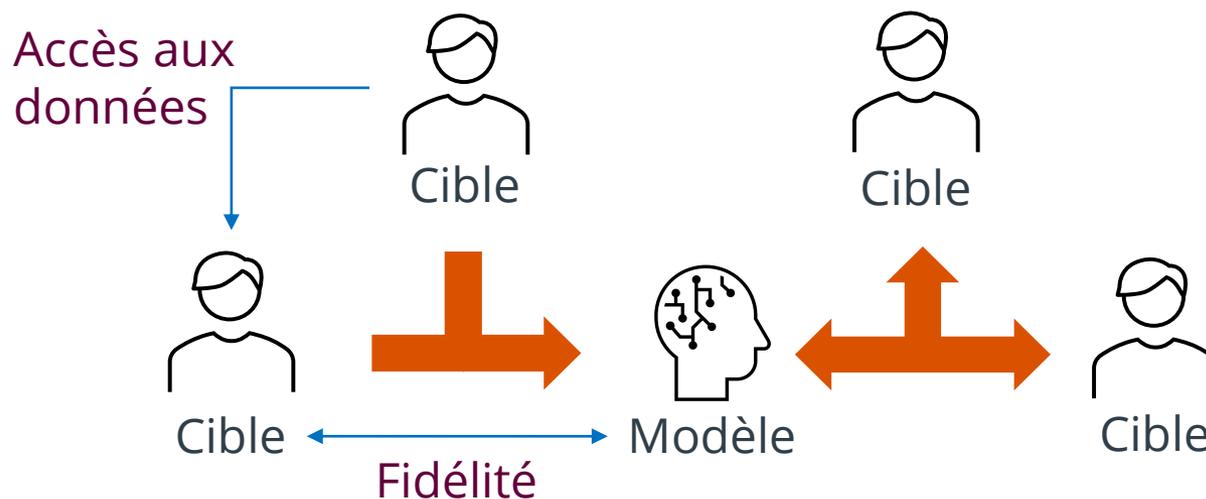


Figure 8 : Structure d'un mime de soi-même

- A la fois Moyen et Fin
- Plus de question sur les données
- Plus de question des droits d'utilisation et consentement

Mime de soi-même

Effectuer le travail à sa place

Démultiplier sa force de travail



Problèmes :

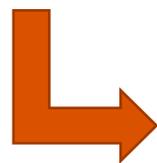
- Signer les mails générés ?
- Valeur monétaire pour un artiste utilisant un mime ?
- Ressenti de réponses meilleurs ou moins bonne que la vraie personne ?

Réponses :

- Mesure du niveau de fidélité
- Distinction entre sortie réelle, sortie du mime et fidélité

Implications :

- Politesse dans un mail avec une fidélité élevée
- Économiquement avantageux grâce à l'extensibilité du mime
- Possibilité d'améliorer le mime (ex: politesse)
- Répandre les biais d'une personne plus facilement en ligne



Création de déception
potentielle

09/01/2023 - Saša Radosavljevic

Conclusion

- Altération des relations
 - Estime de soi
 - Augmentation de la valeur des interactions physiques
 - Déception
 - Attaque par hameçonnage
 - Manipulation par la pratique
 - Vie privée exposée
- **Fidélité**
Précision des réponses du modèle
 - **Modalité**
Domaine d'opération
 - **Généralité**
Ampleur des domaines mimés

Références

- [1] McIlroy-Young, R., Wang, R., Sen, S., Kleinberg, J., and Anderson, A. Learning models of individual human behavior in chess. In Proceedings of the 28th ACM SIGKDD international conference on Knowledge discovery and data mining (2022).