

Mimetic Models: Ethical Implications of AI that Acts Like You

Analyse d'article avec pour introduction :

Modèle mimétique : Algorithme entraîné sur une base de données d'un certain individu dans un certain domaine qui a pour objectif de prédire précisément et de simuler le comportement de cet individu dans de nouvelles situations dans ce domaine.

Ethique : Science qui traite des principes régulateurs de l'action et de la conduite morale.
On cherche à résoudre un conflit de valeurs.

Mime : Modèle mimétique

Contexte

Simuler les décisions et comportement de certaines personnes dans les domaines des jeux, la génération de texte et l'expression artistique.

- Système d'Échecs, les bots ayant le comportement de joueur professionnels
- Avatars eSport
- Reproduire le style d'écriture de certains auteurs ou d'utilisateurs de réseaux sociaux
- Composer de la musique dans le style d'un compositeur ex :
 - o [MuseNet](#) pour certains compositeurs
 - o [JukeBox](#) pour des styles
- Expertise professionnelle (Diagnostic médical) [Article](#)

Problèmes

Caractériser les problèmes d'éthique et les enjeux sociaux du mimétisme informatique.

La concrétisation de ces nouveaux modèles appelle à considérer les implications éthiques et sociales.

Analyse des implications éthiques et sociales

Individu cible > Créateur du modèle > Utilisateur > Élément ou personne qui interagit avec le modèle

- La déception : Une personne peut penser avoir une interaction avec vous mais c'est un modèle. [DeepFakes](#) pour le cyberharcèlement ou la désinformation sur internet.

Modèle mimétique comme moyen

1. Préparation pour une interaction physique
 - a. Préparation d'un entretien, une personne B interview une personne A, A peut se préparer en passant une dizaine d'entretiens avec le mime de B. Peut être pour une augmentation, une interview journalistique voire un rendez-vous. A peut déterminer des comportements attrayant à B.
2. Tournoi d'Échec
 - a. Trouver les points faibles de son prochain adversaire au travers de son mime. Ou identifier soi-même ses points faibles.

Différents problèmes selon les cas d'utilisation. Plus problématique pour un rdv socialement que pour un tournoi d'échecs.

Modèle mimétique comme fin

1. Chatbots

- a. Assistants d'enseignement pour des cours en ligne. On pourrait choisir le style d'enseignement d'une personne A ou une personne B en fonction de ses préférences. **Dévaluation du travail humain**

Question sur le consentement et la rémunération des vraies personnes A et B. Il faut que la personne soit bien représentée car son modèle sera présenté à sa place. Quels sont les effets d'une classe avec un mauvais comportement envers le modèle sur A.

On peut également imaginer des événements qui ne se sont pas produits avec l'interaction de mimes de personnes. Si A avait affronté B au meilleur de sa forme ?
Si deux anciens champions du monde s'opposaient avec leur modèle, quel serait l'impact sur leur situation actuelle ?

Des hypothèses précurseurs à ce sujet ont fait l'objet de litiges dans le domaine du jeu-vidéo.

Modèle mimétique de soi-même

L'utilisation des mimes va dépendre du but et de qui l'utilise et qui l'a créé. En effet, un modèle de soi-même et opéré par soi-même pose moins de soucis envers la personne représentée. On peut alors décupler le potentiel de travail réalisé. Ex : répondre à des mails dans son style plus rapidement. Cependant si un mail automatique est libellé ainsi, il peut avoir un impact social négatif.

- Déception : augmenter le niveau de politesse de ses mails

Framework

1. Cible le comportement d'une personne plutôt qu'un comportement commun
2. Génératif : produit de nouveaux comportements
3. Interactif à des individus ou environnements

4 Rôles :

- Sujet : personne mimée
- Créateur : conçoit le model (accès indirect aux données)
- Opérateur : Utilisateur du model
- Personne qui interagit avec le model

Deepfakes = beaucoup de problèmes d'éthique mais plutôt conçus pour avoir un comportement déterministe.

Applications

Comment classier le potentiel d'un model mimétique ?

- Préparation avant un évènement ou participation directe du modèle

Participation à un évènement :

1. Sujet aurait pu participer à l'évènement mais le modèle est présent
2. Scénarios qui n'auraient pas pu se dérouler (ex : présence temporelle, 2 athlètes)
3. Scénario où sujet et opérateur identique pour s'améliorer

1. Modèle en tant que moyen

Ex.

- a. Préparation en vue d'un tournoi, modèle d'un joueur d'échecs pour observer son comportement dans certaines situations. Pour aller plus loin, faire jouer une IA contre le modèle pour découvrir les faiblesses du joueur.

Problème : En quoi est-ce que l'utilisation d'un modèle de comportement est-elle différente de l'utilisation des données publiques d'un joueur pour se préparer avant de l'affronter ? Quels sont les avantages ?

Réponse : L'analyse et l'extrapolation à partir des données publiques nécessite un effort en temps et en capacités cognitive bien plus importante que de s'entraîner avec un modèle. **Le modèle effectue le travail de la personne de façon probablement plus performante.** Mais est-ce que la préparation à l'aide du modèle est plus effective que l'analyse des parties jouées par la personne ?

On a démontré que le Machine Learning est particulièrement adapté pour la reconnaissance de patterns.

Implication : Si un modèle permet effectivement d'améliorer la préparation d'une personne, les modèles mimétiques ont bel et bien un impact sur l'équité des tournois. Moins de soucis si les deux adversaires ont un mime. Beaucoup plus inquiétant si un seul des deux a un mime.

- b. Préparation d'un interview, le candidat peut découvrir le caractère du recruteur, permettant ainsi de dévoiler ou non certains points volontairement selon ce qu'ils ont à offrir et ce dont le recruteur a besoin. Utilisation de différents styles de persuasion. Découvrir des informations personnelles sur le recruteur et ainsi avoir un lien plus étroit avec la personne. Allant jusqu'à le gain de confiance.

Problème : Plus compliqué cette fois-ci, ce n'est plus un jeu à somme nul (le gain de l'un est au dépit de l'autre). On vient s'aligner avec les intérêts. Un recruteur pourrait être ravi de voir un trait d'intérêt du candidat.

Les cibles des mimes seraient plus enclins à la manipulation et à être exploitées. Dans la vraie vie, on n'a souvent qu'un essai à un entretien.

Réponse : Le recruteur peut volontairement utiliser un modèle pour observer l'implication du candidat vis-à-vis de l'utilisation du mime. Il n'évaluera alors pas les mêmes compétences et caractéristiques du candidat dès lors qu'il sait que le candidat s'est préparé à l'aide d'un mime. On peut compléter un interview classique à un interview préparé avec un mime pour obtenir davantage d'informations sur le candidat.

Implication : La relation de symétrie n'est plus présente avec un modèle, on peut recevoir des informations sans en donner en retour.

- c. Préparation à un rendez-vous amoureux, la personne peut se préparer à poser des questions et découvrir la personnalité de la personne avant même de la rencontrer pour la première fois. On pourrait court-circuiter la phase de recherche et de découverte, notamment le moment où une personne peut demander des conseils à un ami proche de la personne que l'on veut séduire.

2. Modèle en tant que fin

- a. Substitution d'une personne : Remplacer un tuteur en ligne par son modèle virtuel lorsqu'il n'est pas disponible (malade).

Problème : L'employeur peut se demander si le tuteur A est dispensable si son modèle virtuel satisfait les clients de la même manière. Il pourrait même se demander si tous ses clients pourraient disposer du modèle du tuteur A au lieu des nombreux tuteurs.

Dès qu'un modèle est disponible, on pourrait l'utiliser pour avoir une « audience privée ».

Problème : On menace la personne réelle d'être remplacée ou dévaluée. On menace de remplacer les individus uniques au même titre qu'on veut remplacer des tâches répétables par des robots. Des joueurs prestigieux sont démarchés et rémunérés pour leur présence à des événements (compétitions). On perd le capital social durement acquis, la demande de socialisation sera moindre et on pourrait perdre certaines amitiés, si elles sont remplacées par des mimes.

b. Contrefactuel : Produire des actions fidèlement au style d'une personne dans une situation à laquelle elle n'a et ne sera peut-être jamais confrontée. B. Fischer VS A. Karpov 1975 (match au lieu d'un forfait), la musique de Mozart s'il avait vécu plus de 35 ans ?

Problème : Risque d'amputer sur la réputation de la cible. Un modèle peut ne pas être parfait. Dès lors, les dérives de ses réponses n'étant pas mesurable au sens de la fidélité, elles peuvent être reportés sur la cible elle-même. Le risque d'amputation augmente en fonction de la précision des modèles. Si le mime est moins fiable, il passera pour une caricature. Au contraire, s'il est très fiable, les réponses infidèles à la cible seront moins rapidement identifiées à des anomalies.

3. Modèles mimétiques de soi-même

Cible = Créateur = Opérateur = Interacteur

- a. Effectuer le travail à sa place : Répondre aux mails. Démultiplier la force de travail. Un coach d'échecs peut encadrer de nombreux élèves en même temps.
- b. Filtrer les interactions sociales : Interagir avec le modèle avant de continuer d'interagir avec la vraie personne.

Problème : Doit-on faire signer les mails générés par le modèle ? Quelle est valeur monétaire des créations artistiques du mime ? Comment réagir aux messages que l'on trouve meilleur ou moins bonnes que celle de la cible d'origine ?

Réponse : On doit pouvoir mesurer le niveau de fidélité d'un modèle. Distinction entre sortie du modèle, sortie réelle et la fidélité.

Implication : Même si un modèle a un niveau élevé de fidélité, comment interpréter la politesse dans un mail généré par le modèle ? La politesse a-t-elle le même niveau ? Le prix d'une œuvre doit-elle être identique si elle provient de l'artiste ou de son modèle ?

D'un point de vue économique, même si la valeur est dévaluée pour le modèle, cela reste avantageux grâce à l'extensibilité du modèle.

Problème : Rehausser la qualité du modèle, on vient améliorer la politesse des mails générés. Si on découvre que le modèle a été rehausser, on va créer de la déception.

Problème : Les biais d'une personne transféré au model peuvent être beaucoup plus répandus en ligne.

Vue d'ensemble de l'éthique

1. Altération des relations
2. Estimation de soi : corrélée avec l'estime portée à son modèle
3. Les interactions sociales physiques vont-elles devenir primordiales pour s'assurer que la cible est réelle ?
4. Création de déception
5. Attaque par hameçonnage : prétendre être une autre personne pendant un temps prolongé
6. Manipulation par la pratique
7. Vie privée exposée par l'apprentissage des actions passées

3 Dimensions importantes :

- I. Fidélité
- II. Modalité
- III. Généralité