

Title page

- Explainable AI is Dead, Long Live Explainable AI!
 - Titre aguicheur, aime pas, deuxième mieux
 - Hypothesis-driven Decision Support using Evaluative AI
- Papier pour aide à la décision
- Tim miller, chercheur australien, il publie sur des sujets assez varié en XAI
 - XAI 4 Multi agent, XAI 4 RL
 - Humain-AI interaction
 - AI assisted décision support

Introduction

Quick summary

- Changement paradigme XAI aide décision

Framework evaluative AI :

- Centrée sur l'humain
- Au dela des recommandation
- En évaluant les hypothèse du décideur
- **Mitiger l'Excès de confiance dans ces systems de recommandation**

Over/Under reliance

Définition

- Aide à la décision, retrouve deux phénomènes
- Expliquer les défininions :
 - Excès de confiance : accepter les recommandations, même si faux
 - Manque de confiance : inverse : rejeter les recommandation, même si vrais
- == Automation bias :
 - Qui a généralement des gros impacts
 - **Pourquoi j'ai choisi ce papier**
 - La confiance excessive → Problématique
 - Au quotient
 - correcteur d'orthographe
 - Recrutement ; Prêt bancaire ect
 - Enjeux plus sérieux : Unité de soin intensif, aviation, centrale nucléaire
 - Plus généralement : craches boursier instantané
 - Algo mesure du risque de violence conjugale → Excès de confiance → erreur d'estimation du danger
 - Echo domain sciences cognitives licence, manque un peu

Causes

- Manque d'engagement cognitif, l'esprit humain cherche à minimiser l'effort
- Algorithmic aversion : OSEF
- Ajout XAI qui explique
 - Une sorte de biais de confirmation sur les explications qu'on vas accepter ou rejeter

Solutions

- Solution 1 : forcer l'engagement cognitif
 - par exemple, forcer les gens à donenr leur décision avant celle de la machine
 - pas giga efficace
 - pas trop apprécié
- Solution 2 : Un paradigme shift en XAI 🤔💡
 - C'est ce que propose ce papier
 - => Avant de parler de cela on vas définir des critères plus claire

How we make decisions?

- Ce qu'on fait naturellement c'est identifier, comparer, choisir
- Des gens ont plus réfléchis à la chose
 - **dans notre cadre**, pour les system d'aide à la décision, on peut résumé tout ça par ... DIAPO

10 cardinal decision issue

Bon system d'aide à la décision, besoin de

- Option: identifier, lister, réaliste/fesable
- Opinion & Possibilité : Proba et impact positif / négatif possible pour chaque options
- Compromis: comparer ce qu'on tout ce qu'on a dit au dessus
- Understand: comprendre le systeme d'aide à la décision

Does current decision support align with those criteria?

- On vas comparer plusieurs system d'aide à la décision et voir si ils respectent ces critères

No explanotory information

- Cas classique d'automatisation des décisions : *décrire un peu*
- Comme dit précédement, les gens tendent à soit ignorer le system soit accepter des mauvaise décisions
- Le décideur : Calibration de la confiance uniquement sur :
 - l'accuracy du model
 - Son expertise
- => Novice : se repose sur le systems // expert : utilise leur propre expertise
- Mais ne coche aucune des cases
 - N'aide pas à identifier les autres options probable
 - Opinion uniquement autour de la recommandation
 - N'aide pas à faire des compromis
 - Pas de expliquabilité
- Mais est-ce que c'est quand même utile ?? OUI
 - Si on est d'accord avec le system tout roule
 - Si on est pas d'accord, on peut reconsidérer le choix
 - Et toujours faire de meilleurs décision

With explanatory information

- Cette fois ci avec un outil de XAI
- Coche plus de case
 - Avec les outils XAI, on peut comprendre le modèle
 - et faire des compromis par exemple avec SHAP ou les counterfactual
 - N'aide pas à identifier les autres options probables
 - Jugement et possibilité uniquement autour de la recommandation
- => Mais c'est toujours pour défendre la recommandation

- Est-ce que c'est quand même utile ?? OUI
 - Pour les même raison que précédemment :
 - Si on est pas d'accords on peut regarder et ça donne potentiellement de meilleurs décision
 - Mais en pratique c'est pas souvent le cas
- Un model interprétable coche uniquement la dernière case

Cognitive forcing

- Cognitive forcing : décideur donne décision avant machine
- C'est le cas qui coche le plus de case dans le paradigme actuel mais y'a toujours des problèmes
 - décideur voie plus d'option : forcé de les chercher
 - Toujours Avec les outils XAI, on peut comprendre le modèle
 - et faire des compromis par exemple avec SHAP ou les counterfactual
- Mais le system est toujours centré sur sa recommandation
- dès que centré autour de la recommandation, on ne fait que partiellement cocher les cases
- => Sortir de ce paradigme de recommandation unique => evaluative AI

The evaluative AI framework

- Décrire : boucle, décideur -> HP -> feedback
- Le paradigme est inversé :
 - c'est la machine qui donne son avis sur la décision du *decision-maker*
 - Et non le décision maker qui donne son avis sur la décision de la machine

Properties

- Exemple d'interface
 - Potentiel mélanome ?
 - Vu sur toutes les options possible
 - interaction avec l'utilisateur
 - hypothèse pour, hypothèse contre

Zoom on properties

- Naturellement leur modèle coche toutes les cases
- Option
- Trade-off
 - Je trouve que c'est ici que ça réussi le mieux
 - Pour ou contre assez clair pour permettre au décideur un bonne overview
 - Et en faite les "bon décideur" sont les personnes qui regarde toujours les arguments qui vont contre leurs conclusions initiales
- => extrapolation sur de l'IRL
 - Quand décision complexe, type choix de stage, orientation
 - Si on reste regarde tous les pous et contre, j'trouve ça devient vite borbier
 - Alors que maybe se fier a l'instincts et identifier les contre serait plus efficace
 - les discussions IRL ?
 - la clé ça reste d'être à l'écoute et tourner autour de l'opinion de l'autre sans forcément directement relate sur des pov personel

Limits

- Pourquoi les gens s'engagerai avec ce system et pas les autres méthode
 - Plus de controle
 - Proche de la manière dont on fait des décisions (identifier, comparer, choisir)
 - X : pas de preuve de ça dans le papier (en psychologie ça serait pas passé, jsp pour Humain-AI interaction)

- Méthode qui charge mentalement le décideur
 - X: toujours la moins aimé surement
 - Mais auteur se défend en disant qu'il y a quand même moins d'info à considérer

Mes critiques

- Les critères sont dur à différencier
 - Y'en a 10 de base, il en garde 6, 1 n'est jamais remplis, et 2 fusionne en 1 car proche (opinion et possibilité)
 - Des fois c'est dur de s'y retrouver, le tableau résumé est pas forcément accords avec que qui est dit dans le texte,
- Quand j'ai été voir la page wikipedia de l'automation bias, elle est assez remplis et l'auteur en parle pas du tout. Y'as pas mal d'autre facteur décrit et j'arrive pas à voir pourquoi y'a pas un mot dessus dans le papier
 - A la place l'intro parle du résonnement abductif pour appuyer son modèle comme un modèle proche de la manière naturel de la décision
 - Alors qu'il aurait eu la place car beaucoup de répétition dans son papier

Mes points forts du papier

- S'attaque à un vrais problème
- Avec une proposition forte, position pas facile à tenir
- Pas d'expérience pour appuyer l'évaluative AI
 - mais donne une liste exhaustive de piste de recherche dans la direction de l'évaluative AI

CONCLUSION

- auteur propose de changer de voix pour le XAI appliqué l'aide à la décision
- Qu'il faut arreter d'expliquer les recommandation et se focus sur l'utilisateur et ces hypothèses
- En se rapprochant de la manière dont on prends naturellement des décisions

--- Je garde pour les questions au cas où ---

Differences with cognitive forcing

- Apparament ça ressemblerai pas mal au technique de cognitive forcing
- Les auteurs essaye plusieurs fois de se différencier à travers le papier
- Ici la clé c'est que le décideur est en position de contrôle face à la machine, "machine in the loop"
- également que ça suit un chemin de décision plus naturelle

Long live XAI

- Le titre est pas vraiment clair au première abord mais il se défend
- L'auteur ne veut pas se séparer de l'XAI ou faire une refonte
- Il veut améliorer une petite branche de l'XAI
 - Evaluative AI \subset XAI
- XAI + approche basé sur la recommandation sont bien et adapté dans certains cas
 - Making decision at scale
- Il faudra toujours un model recommendation based pour n'importe quelle XAI technique
- Outil de XAI existant -> déjà adapté à l'evaluative AI
 - Counterfactuals
 - Feature importance (SHAP)
 - Wiegths of Evidence, case-based reasoning techniques