Confiance humain-IA pour la prise de décision : définitions, facteurs et évaluation au travers de prisme académique et industriel

Oleksandra Vereschak,

en collaboration avec Gilles Bailly et Baptiste Caramiaux

IHM & IA 2023

07/06/2023

Contexte : la prise de décision avec l'IA

- IA: apprentissage automatique, réseaux de neurones, traitement automatique des langues…

Contexte : la prise de décision avec l'IA

- IA: apprentissage automatique, réseaux de neurones, traitement automatique des langues…

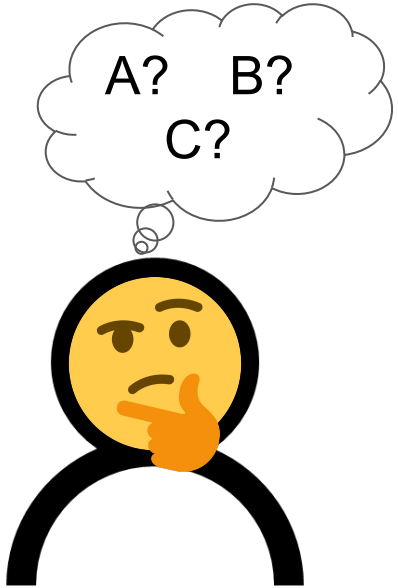
- Tâche : classification ou régression

Contexte : la prise de décision avec l'IA

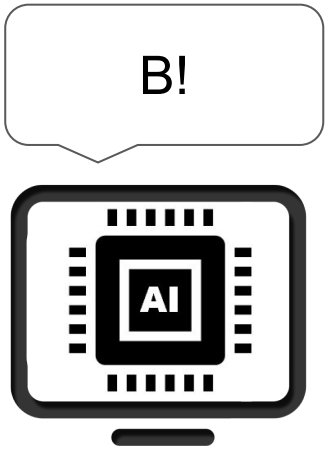
- IA: apprentissage automatique, réseaux de neurones, traitement automatique des langues…
- Tâche : classification ou régression
- Domaine: domaines à haut risque (médecine, finance, justice)

Contexte : la prise de décision avec l'IA

- IA: apprentissage automatique, réseaux de neurones, traitement automatique des langues…
- Tâche : classification ou régression
- Domaine: domaines à haut risque (médecine, finance, justice)



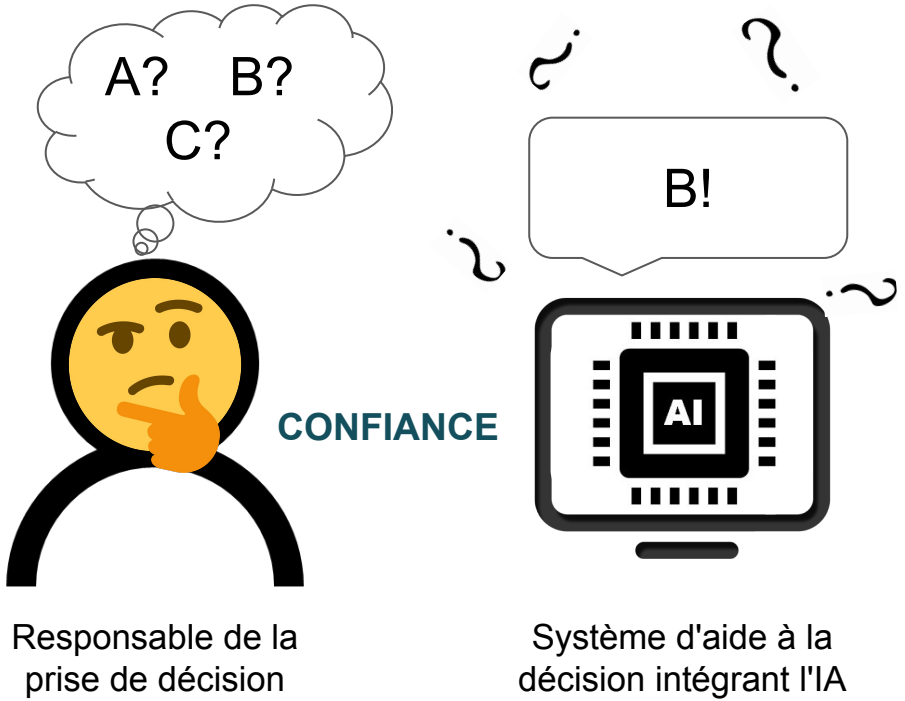
Responsable de la prise de décision



Système d'aide à la décision intégrant l'IA

Contexte : la prise de décision avec l'IA

- IA: apprentissage automatique, réseaux de neurones, traitement automatique des langues…
- Tâche : classification ou régression
- Domaine: domaines à haut risque (médecine, finance, justice)



Les facteurs qui influencent la confiance humain-IA

Point de vue **académique** : une revue systématique de plus de 5000 articles

Work-in-progress

Les facteurs qui influencent la confiance humain-IA

Point de vue **académique** : une revue systématique de plus de 5000 articles

Point de vue de l'**industrie** : entretiens avec des professionnels de l'IA et des personnes impactées par les décisions

Work-in-progress

Les facteurs qui influencent la confiance humain-IA

1. Le contexte socio-technologique
2. Le développement et la conception des systèmes
3. Les préférences et les expériences des personnes

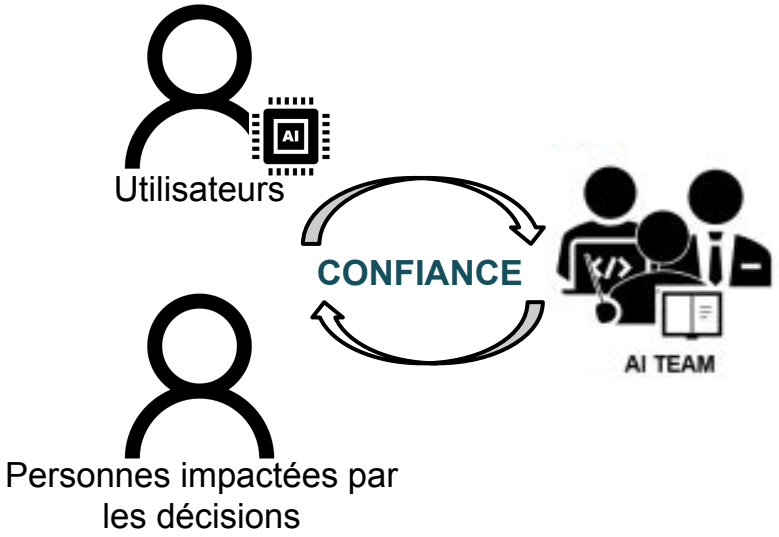
Les facteurs qui influencent la confiance humain-IA

Le contexte socio-technologique

Les facteurs qui influencent la confiance humain-IA : **INDUSTRIE**

Le contexte socio-technologique

Confiance entre humains : focus sur la confiance dans les professionnels de l'IA



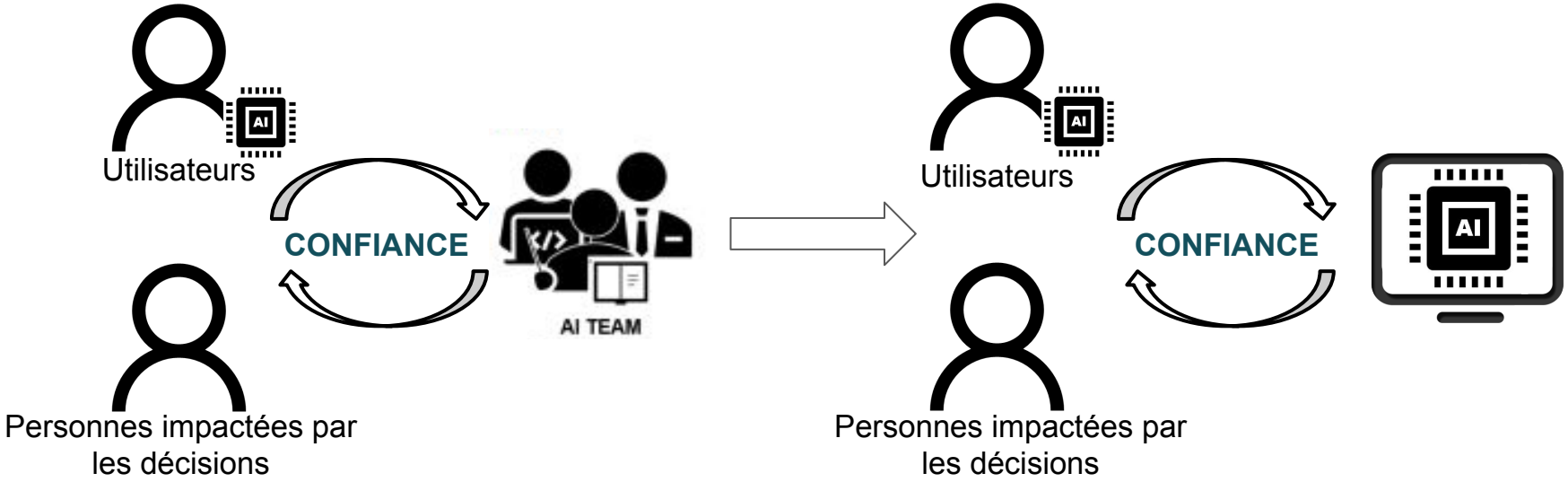
Work-in-progress

Les facteurs qui influencent la confiance humain-IA : **INDUSTRIE**

Le contexte socio-technologique

Confiance entre humains : focus sur la confiance dans les professionnels de l'IA

P4: Si les utilisateurs (c'est-à-dire les clients) font confiance à l'équipe, leur confiance dans l'IA "*[...] est établie avant que le système n'existe*".

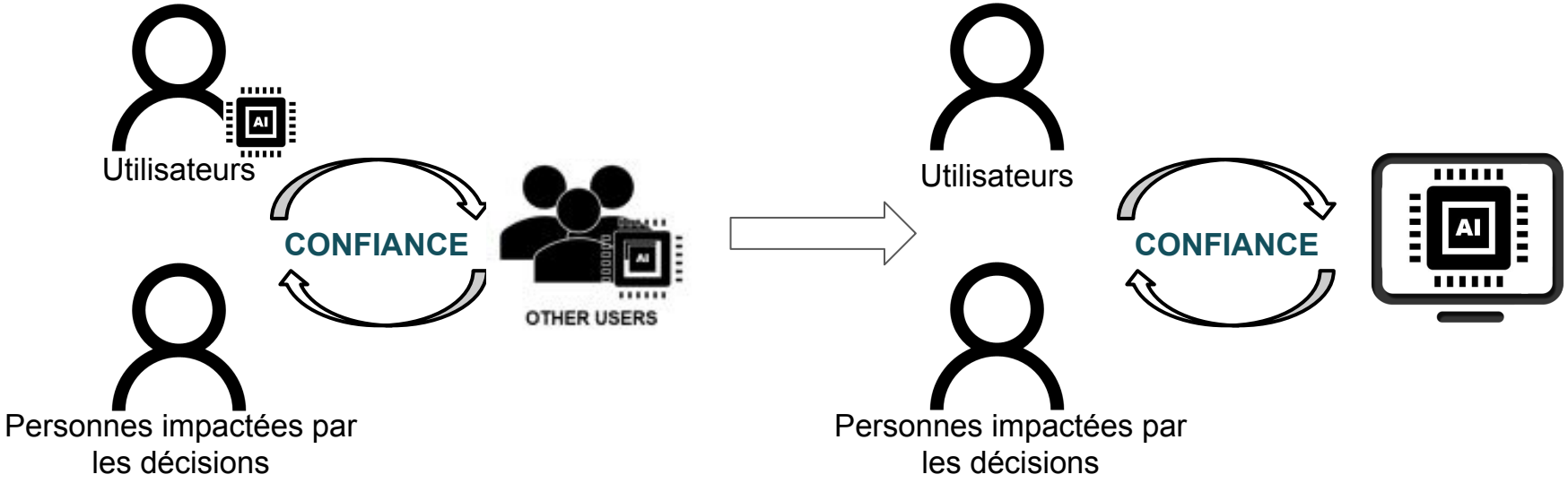


Work-in-progress

Les facteurs qui influencent la confiance humain-IA : ACADÉMIE

Le contexte socio-technologique

Confiance entre humains : peu d'études (5 / 113), axées sur la confiance dans les autres utilisateurs



Work-in-progress

Les facteurs qui influencent la confiance humain-IA : **INDUSTRIE**

Le développement et la conception des systèmes

Transparence : son effet sur la confiance dépend du contexte

Temps limité, DS7 : “*Si les utilisateurs disposaient du temps nécessaire pour lire les explications et les examiner concrètement, ils auraient pris la décision eux-mêmes dès le départ* ”

Les facteurs qui influencent la confiance humain-IA : **INDUSTRIE**

Le développement et la conception des systèmes

Transparence : son effet sur la confiance dépend du contexte et de l'utilisateur

Les développeurs "*sont plus méfiants [à l'égard de leurs modèles d'IA] après avoir utilisé l'explicabilité qu'ils ne l'étaient avant* " (P1)

Les facteurs qui influencent la confiance humain-IA : **INDUSTRIE**

Le développement et la conception des systèmes

Transparence : son effet sur la confiance dépend du contexte et de l'utilisateur

Les développeurs "*sont plus méfiants [à l'égard de leurs modèles d'IA] après avoir utilisé l'explicabilité qu'ils ne l'étaient avant* " (P1)

Les personnes impactées par les décisions recherchent l'interaction, et non davantage d'informations: “*J'aimerais avoir la possibilité de négocier et d'influencer la décision [de l'AI] et de dire "Hé, mais examinez plutôt ceci et cela*" ”(DS4)

Les facteurs qui influencent la confiance humain-IA : **ACADÉMIE**

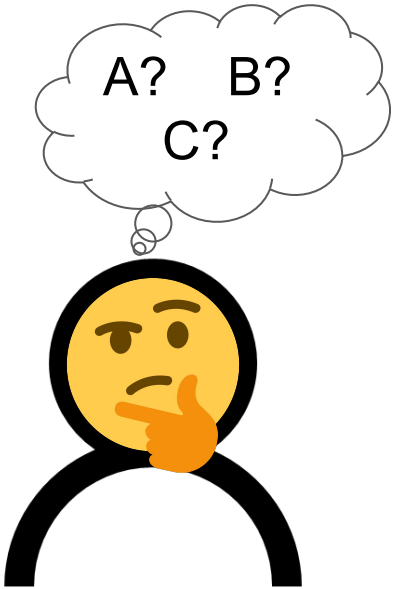
Le développement et la conception des systèmes

Transparence : appel à se concentrer sur la "transparence sociale", plutôt que sur la "transparence technologique"

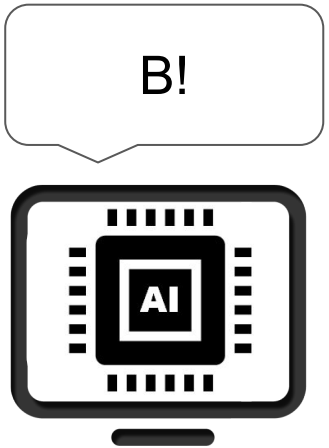
Les facteurs qui influencent la confiance humain-IA : ACADÉMIE

Le développement et la conception des systèmes

Transparence : appel à se concentrer sur la "transparence sociale", plutôt que sur la "transparence technologique"



Responsable de la prise de décision



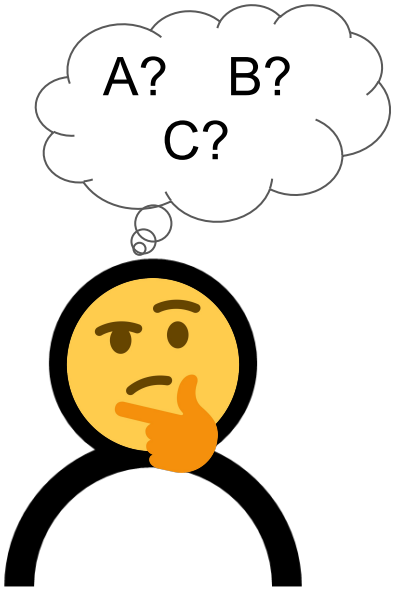
Système d'aide à la décision intégrant l'IA

Work-in-progress

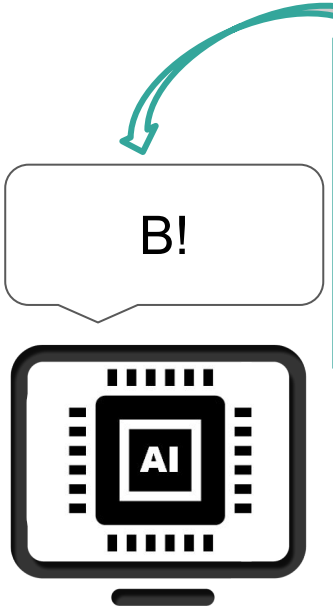
Les facteurs qui influencent la confiance humain-IA : ACADÉMIE

Le développement et la conception des systèmes

Transparence : appel à se concentrer sur la "transparence sociale", plutôt que sur la "transparence technologique"



Responsable de la prise de décision



Système d'aide à la décision intégrant l'IA

Votre responsable a également choisi B dans cette situation, car il s'agit d'un client très fidèle

Work-in-progress

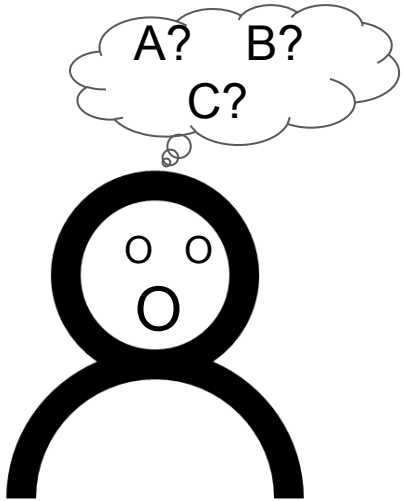
Les facteurs qui influencent la confiance humain-IA

Les préférences et les expériences des personnes

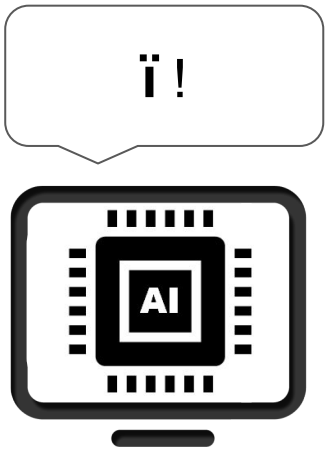
Les facteurs qui influencent la confiance humain-IA : **INDUSTRIE**

Les préférences et les expériences des personnes

Recommandations surprenantes de l'IA : largement discuté (9 / 14)



Responsable de la prise de décision



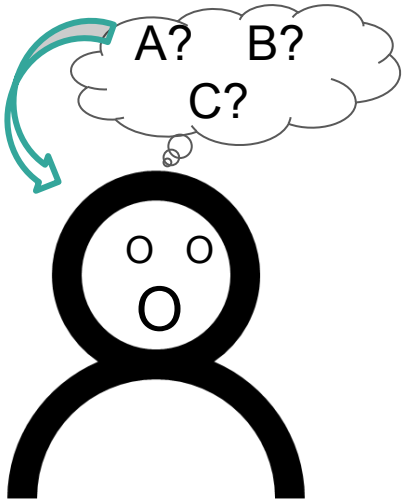
Système d'aide à la décision intégrant l'IA

Work-in-progress

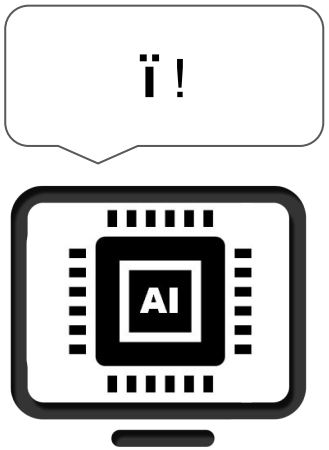
Les facteurs qui influencent la confiance humain-IA : **INDUSTRIE**

Les préférences et les expériences des personnes

Recommandations surprenantes de l'IA : largement discuté (9 / 14)



Responsable de la prise de décision



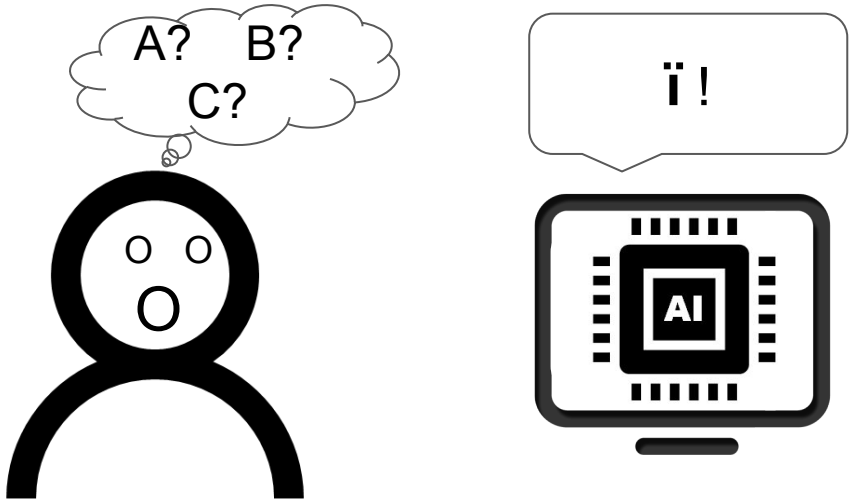
Système d'aide à la décision intégrant l'IA

Work-in-progress

Les facteurs qui influencent la confiance humain-IA : **INDUSTRIE**

Les préférences et les expériences des personnes

Recommandations surprenantes de l'IA : largement discuté (9 / 14), la surprise peut être bonne ou mauvaise.



Responsable de la prise de décision

Système d'aide à la décision intégrant l'IA

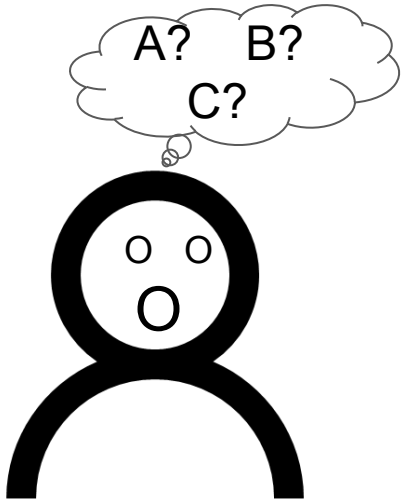
Work-in-progress

Les facteurs qui influencent la confiance humain-IA : **INDUSTRIE**

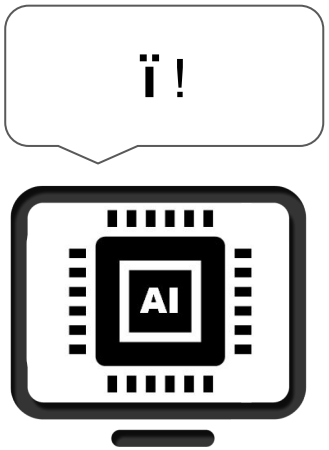
Les préférences et les expériences des personnes

Recommandations surprenantes de l'IA : largement discuté (9 / 14), la surprise peut être bonne ou mauvaise.

P1 : "*bonne surprise, c'est l'IA qui nous apprend [aux humains] des choses que nous ne savions pas*"



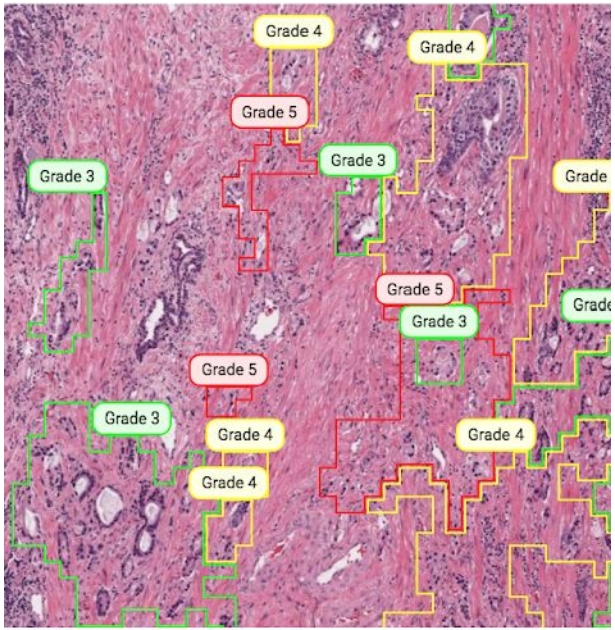
Responsable de la prise de décision



Système d'aide à la décision intégrant l'IA

Contexte : Évaluation de la confiance dans les systèmes d'aide à la décision intégrant l'IA

Contexte : Évaluation de la confiance dans les systèmes d'aide à la décision intégrant l'IA



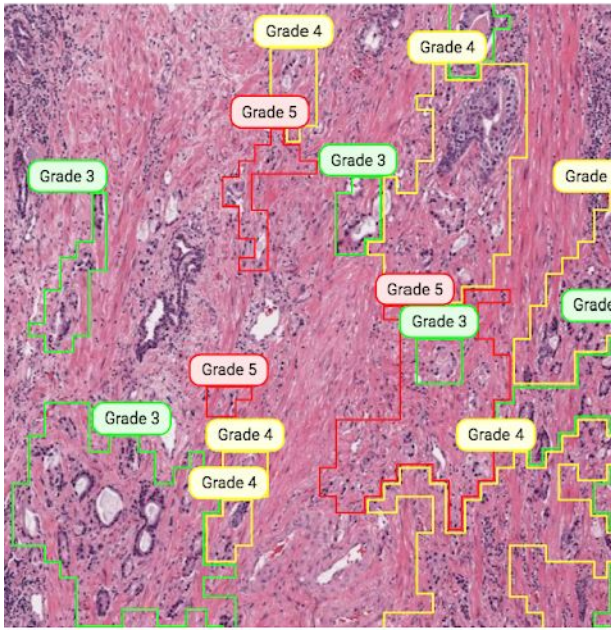
Prédiction du grade global de cancer: 4

Accepter

Rejeter

Cai et al., “Hello AI”, CSCW’19

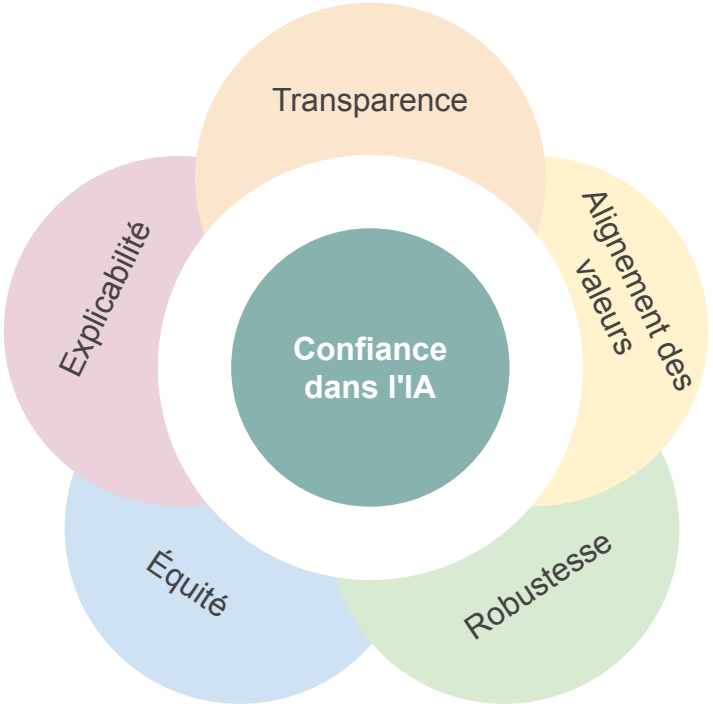
Contexte : Évaluation de la confiance dans les systèmes d'aide à la décision intégrant l'IA



Prédiction du grade global de cancer: 4

Accepter

Rejeter



Cai et al., “Hello AI”, CSCW’19

Contexte : Évaluation de la confiance dans les systèmes d'aide à la décision intégrants l'IA

Comment évaluer la confiance des utilisateurs ?

Overall cancer grade prediction: 4

Agree

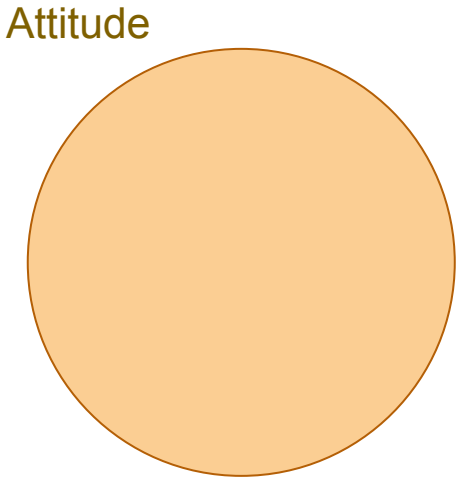
Disagree

Principaux résultats : éléments clés de la confiance

Une attitude affirmant qu'un agent atteindra les objectifs d'un individu dans une situation caractérisée par l'incertitude et la vulnérabilité (n = 10, 12%; Lee and See, 2004)

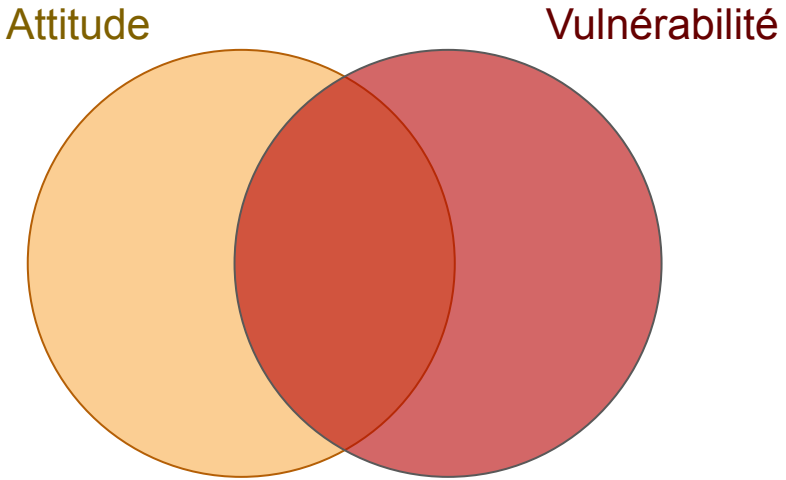
Principaux résultats : éléments clés de la confiance

*Une **attitude** affirmant qu'un agent atteindra les objectifs d'un individu dans une situation caractérisée par l'incertitude et la vulnérabilité*
(n = 10, 12%; Lee and See, 2004)



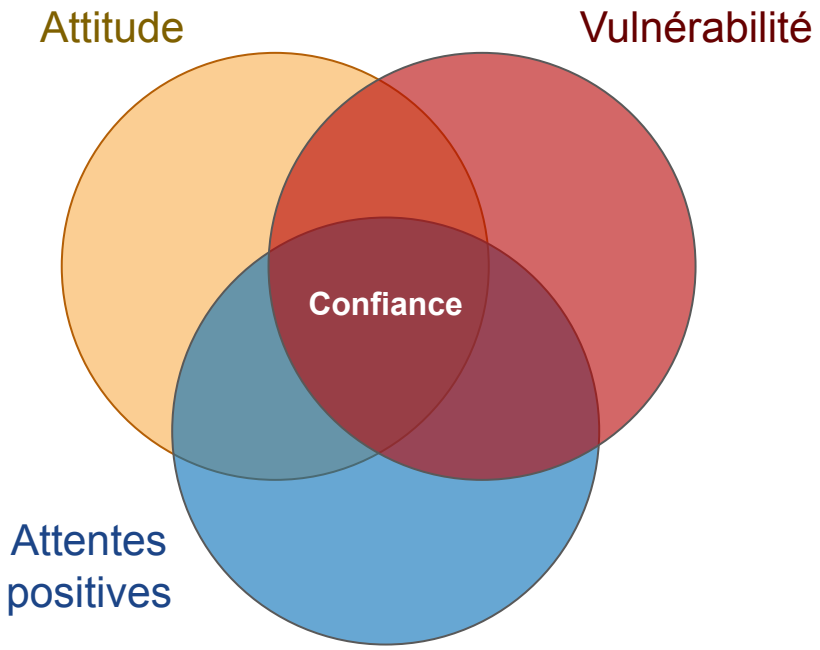
Principaux résultats : éléments clés de la confiance

*Une **attitude** affirmant qu'un agent atteindra les objectifs d'un individu dans une situation caractérisée par **l'incertitude et la vulnérabilité***
(n = 10, 12%; Lee and See, 2004)



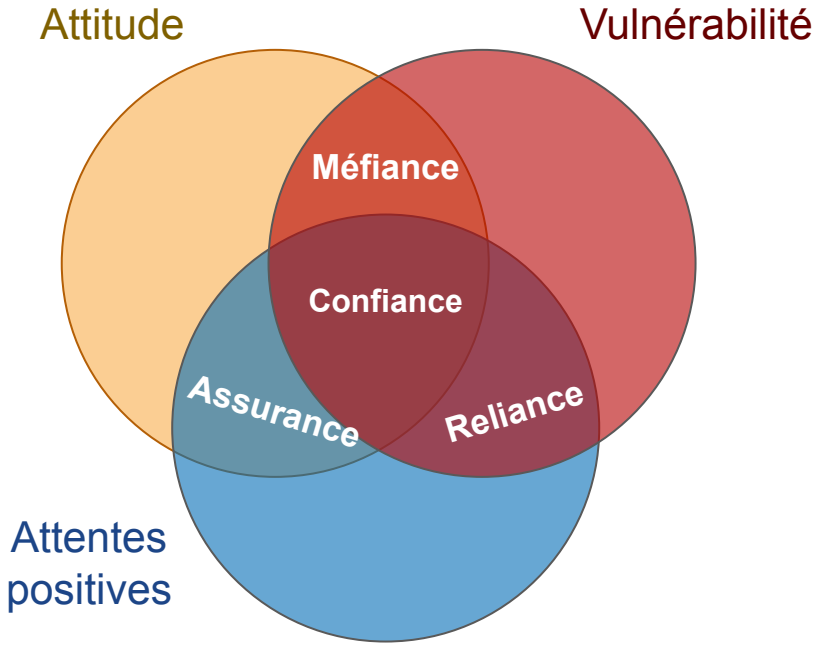
Principaux résultats : éléments clés de la confiance

*Une **attitude** affirmant qu'un agent **atteindra les objectifs** d'un individu dans une situation caractérisée par **l'incertitude et la vulnérabilité***
(n = 10, 12%; Lee and See, 2004)



Principaux résultats : éléments clés de la confiance

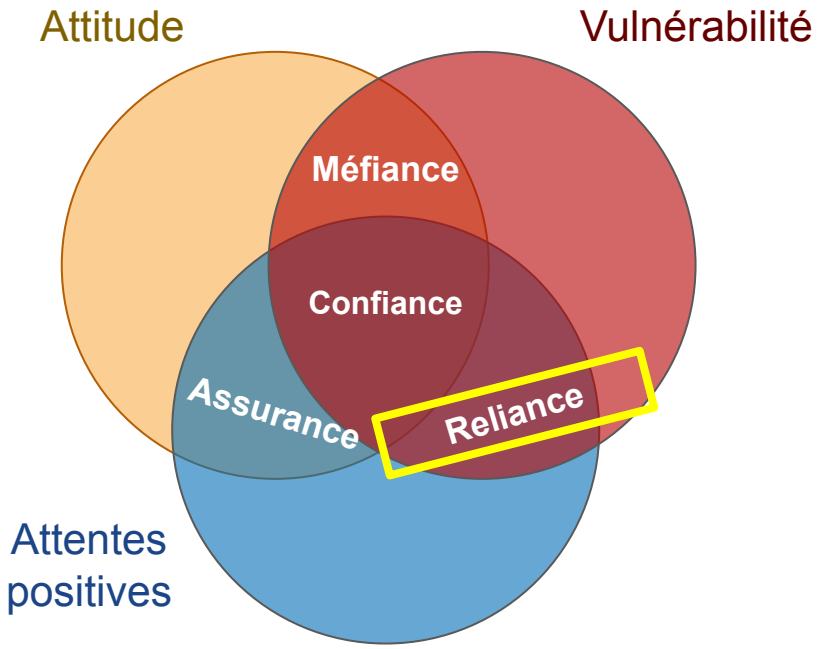
*Une **attitude** affirmant qu'un agent **atteindra les objectifs** d'un individu dans une situation caractérisée par **l'incertitude et la vulnérabilité***
(n = 10, 12%; Lee and See, 2004)



Principaux résultats : éléments clés de la confiance

*Une **attitude** affirmant qu'un agent **atteindra les objectifs** d'un individu dans une situation caractérisée par **l'incertitude et la vulnérabilité***
(n = 10, 12%; Lee and See, 2004)

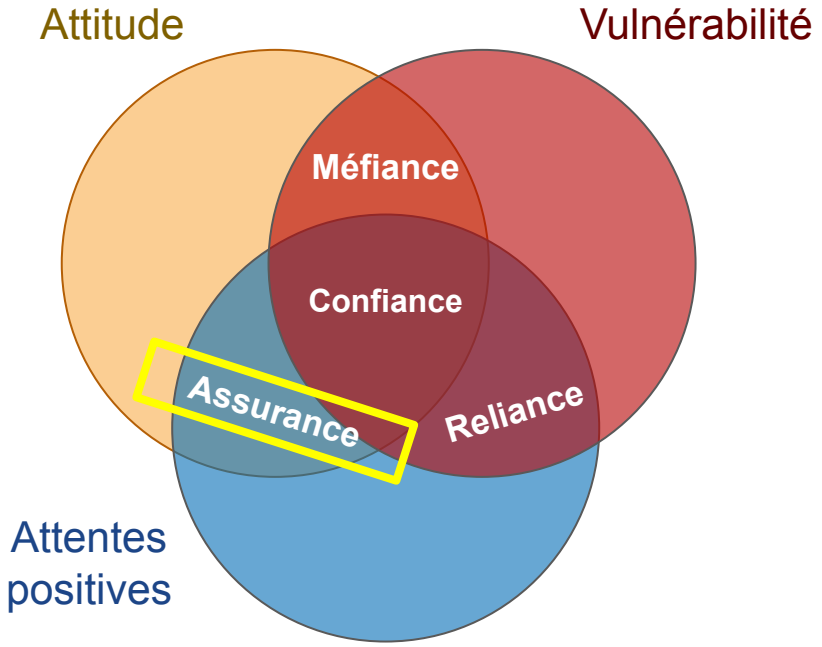
La confiance ne peut pas être observée tout le temps -> questionnaires (**G12**)



Principaux résultats : éléments clés de la confiance

*Une **attitude** affirmant qu'un agent **atteindra les objectifs** d'un individu dans une situation caractérisée par **l'incertitude et la vulnérabilité***
(n = 10, 12%; Lee and See, 2004)

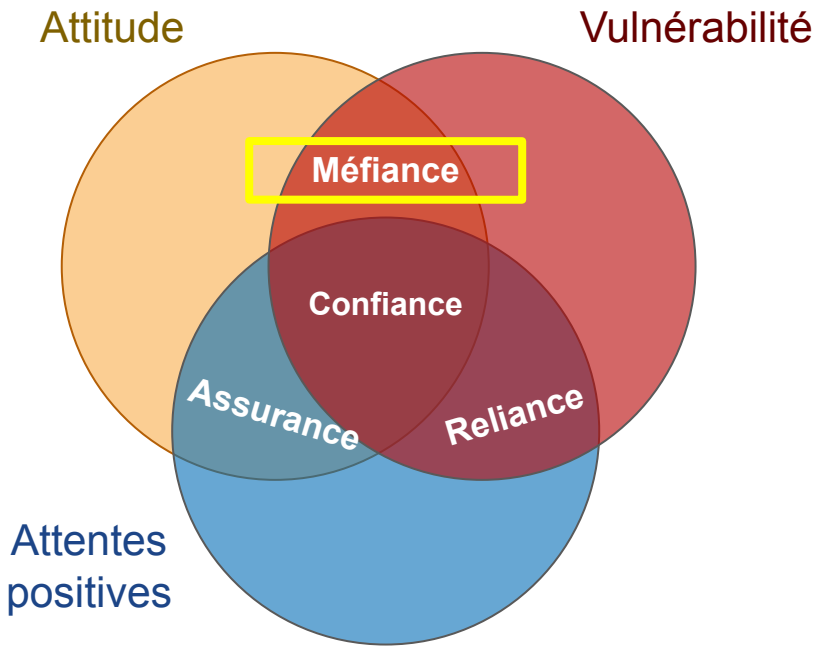
Les décisions prises dans le cadre de l'expérience doivent avoir des conséquences crédibles et réalistes (**G7**)



Principaux résultats : éléments clés de la confiance

*Une **attitude** affirmant qu'un agent **atteindra les objectifs** d'un individu dans une situation caractérisée par **l'incertitude et la vulnérabilité***
(n = 10, 12%; Lee and See, 2004)

L'introduction et la première interaction avec le système doivent être contrôlées afin que les participants puissent avoir des attentes positives (**G9, G10**)



Confiance humain-IA pour la prise de décision : définitions, facteurs et évaluation au travers de prisme académique et industriel

Oleksandra Vereschak,

en collaboration avec Gilles Bailly et Baptiste Caramiaux

IHM & IA 2023

07/06/2023