PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR
ANR

Quand l'IA explique, l'humain reprend le contrôle

Etude neurophysiologique du sentiment d'agentivité

Solène LE BARS, PhD – Technology & Innovation Leader | Future of HealthCare

20/11/2025

ENS | PSL

Capgemini

Agenda

Introduction
- Interactions Humain-IA
- Sentiment d’agentivité
- Problématique

Méthode
- Cas d’usage : conduite autonome
- Mesures
- Protocole expérimental

Résultats
- Ratings
- EEG / ERP

Discussion & perspectives

01

Introduction

Introduction : Interactions humain - IA

L’intelligence artificielle est omiprésente

- Manufacturing & procédés industriels
- Outils RH & recrutement
- Aide à la decision médicale
- Boom LLM & outils conversationnels
- Armement militaire …

Un usage quotidien de l’IA

Deux tiers de la population mondiale utilisent quotidiennement l’IA. Elle assiste ou supervise des millions d’actions (allant du triage automatique de mails à la recommandation de contenus)

66%

Introduction : Interactions humain - IA

L'intelligence artificielle est omiprésente

- Manufacturing & procédés industriels
- Outils RH & recrutement
- Aide à la decision médicale
- Boom LLM & outils conversationnels
- Armement militaire …

Quid de la responsabilité humaine & des répercussions cognitives ?

- IA = black box ? Intérêt de l'IA explicable ?
- "Out of the Loop Process"
- Diminution de l'engagement de l'agent humain
- Diminution des performances cognitives et de l'agentivité

Vers l'IA explicable (XAI)

La XAI déploie des techniques et des méthodes spécifiques pour garantir que chaque décision prise au cours du processus de ML/DL peut être retracée et expliquée.

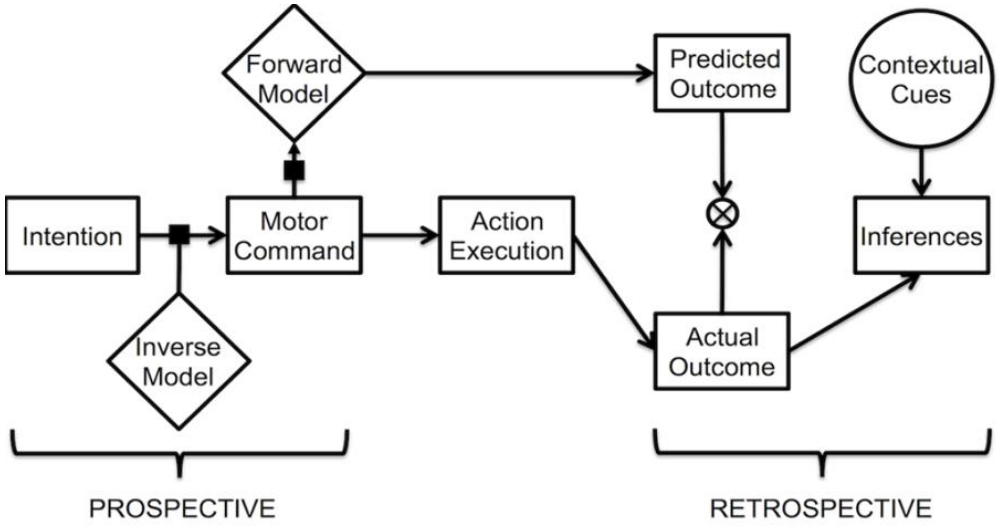
Elle permet notamment aux utilisateurs humains de comprendre et de faire confiance aux résultats issus des algorithmes.

Introduction : Sense of agency

Pourquoi c'est important ?

- Considérations éthiques et légales
- Biomarqueur de certaines pathologies psychiatriques et neurologiques
- Le SoA vient moduler l'engagement, les performances cognitives et la confiance

Modèle de neurosciences computationnelles



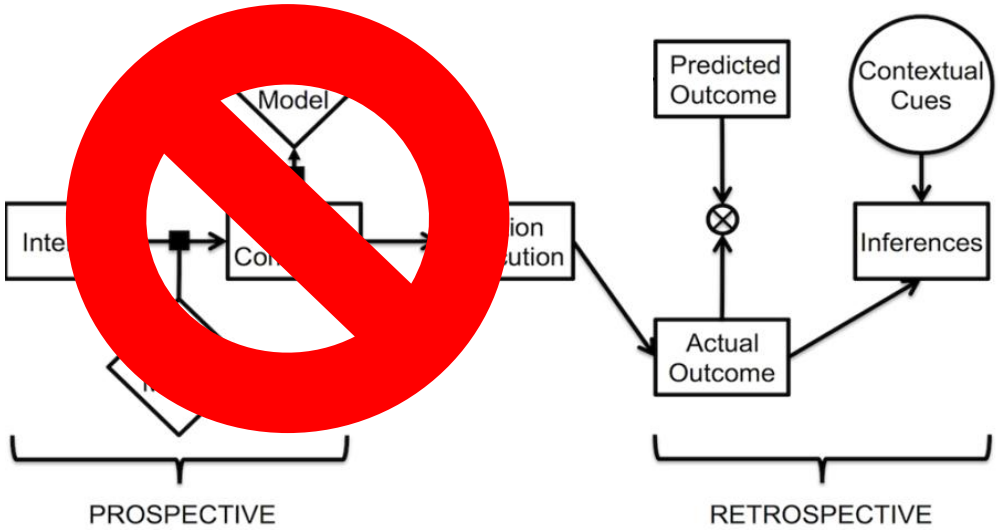
Qu'est-ce que le « sense of agency » ?

« Experience of being in control of one's own actions and, through them, of events in the external world » (Haggard & Tsakiris, 2009)

« *A Bayesian cue integration process may weight the contribution to the sense of agency from voluntary motor commands and from external situational evidence* » (Moore, et al 2009 ; Le Bars et al. 2022)

Introduction : Sense of agency in joint action with HMI / HAI

Modèle de neurosciences computationnelles



Agentivité dans les actions conjointes & IHM / IHAI ?

- Diminution du sentiment d'agentivité global
- Augmentation de la pondération des indices contextuels (car pas / moins d'accès au forward model)

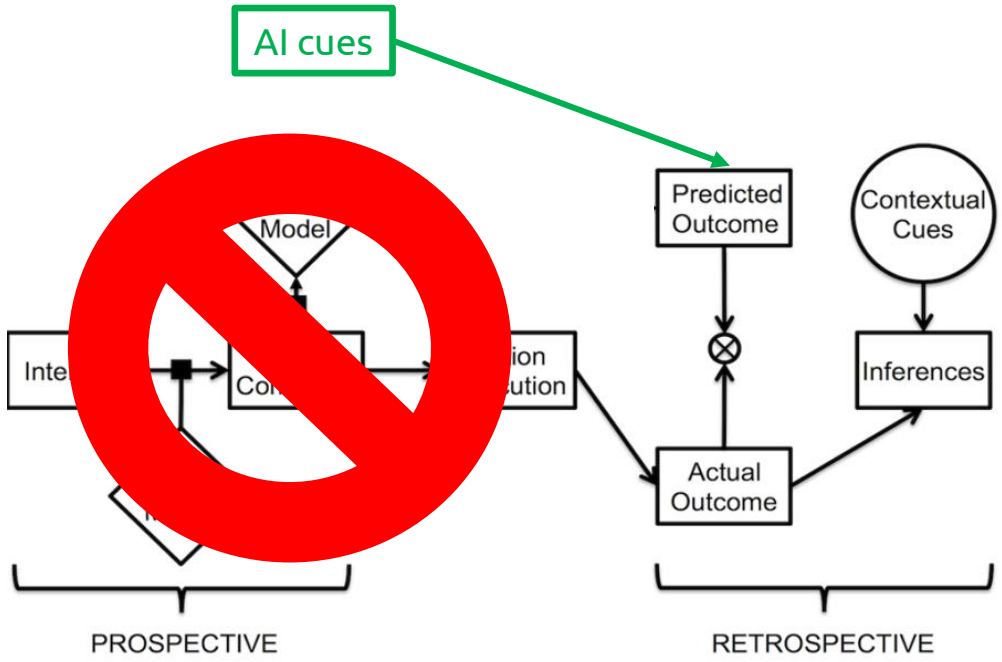
Qu'est-ce que le « sense of agency » ?

« Experience of being in control of one's own actions and, through them, of events in the external world » (Haggard & Tsakiris, 2009)

« A Bayesian cue integration process may weight the contribution to the sense of agency from voluntary motor commands and from external situational evidence » (Moore, et al 2009 ; Le Bars et al. 2022)

Introduction : Problématique & Hypothèses

Comment restaurer l'agentivité dans une IHAI ?



- Suppléance du modèle interne moteur par des indices et explications fournies par l'IA
- Plus l'IA fournira d'explications sur ses "intentions" / "décisions" & plus l'agent se sentira en contrôle au décours de l'interaction

Qu'est-ce que le « sense of agency » ?

« Experience of being in control of one's own actions and, through them, of events in the external world » (Haggard & Tsakiris, 2009)

« A Bayesian cue integration process may weight the contribution to the sense of agency from voluntary motor commands and from external situational evidence » (Moore, et al 2009 ; Le Bars et al. 2022)

02

Méthode

Méthode : Cas d’usage de conduite autonome

Méthode : Comment mesurer le sentiment d’agentivité ?

Méthodes subjectives

Questionnaires :

“A quel point vous sentez-vous en contrôle sur une échelle de 0 à 100”

“A quel point vous sentez-vous responsable de la conséquence / action”

Manque d’objectivité, sujets aux biais, difficilement implémentables dans une intéraction continue…

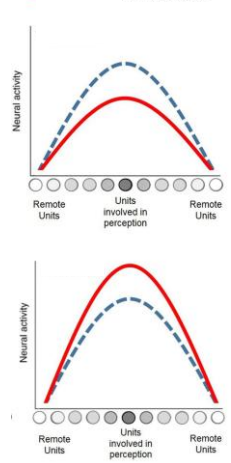
Méthodes objectives

Electroencéphalographie (EEG) : atténuation sensorielle

Diminution de l’activité cérébrale associée au traitement de la conséquence sensorielle d’une action

Electroencéphalographie (EEG) : prediction error

Augmentation drastique de l’activité cérébrale en cas d’erreur / conséquence inattendue résultant de l’action



Méthode : Participants & Matériel

Participants

N = 60

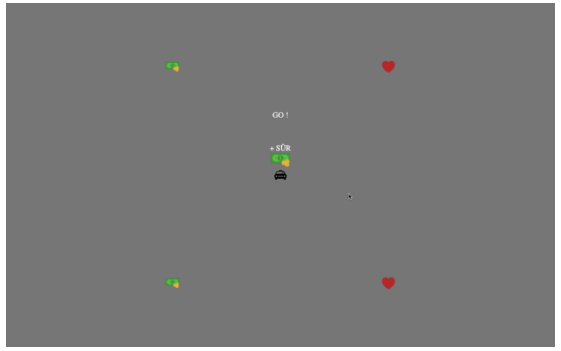
30 participants - expérience “contrôle manuel” vs. “contrôle IA”

30 participants - expérience “IA sans explication” vs. “IA avec explication”

Tâche & Matériel

Objectif : atteindre des cibles pour gagner le plus de points possibles en évitant les évetuels obstacles

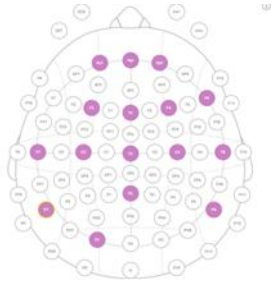
Expérience informatisée synchronisée avec matériel EEG (16 Channels)



Preprocessing EEG

Bandpass [0,1 Hz ; 80 Hz] / ICA

ROI 7 channels ["F3", "Fz", "F4", "C3", "Cz", "C4", "Pz"]



Méthode : Protocole expérimental

2 expériences, 6 conditions distinctes

Condition motrice = contrôle manuel (baseline)

- Effet sonore / conséquence attendue (match) -> atténuation sensorielle ?
- Effet sonore / conséquence inattendue (mismatch) -> prediction error ?

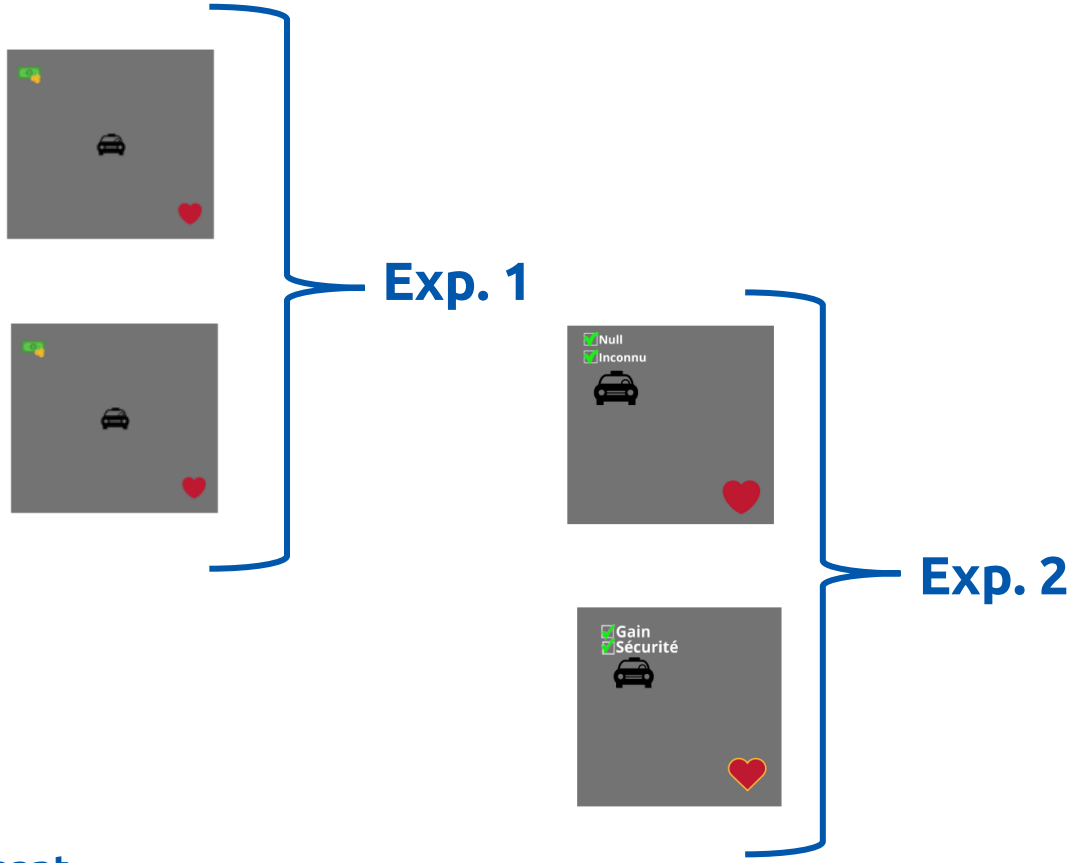
Condition IA <u>sans explication</u> = conduite autonome

- Effet sonore / conséquence attendue (match) -> atténuation sensorielle ?
- Effet sonore / conséquence inattendue (mismatch) -> prediction error ?

Condition IA <u>avec explication</u> = conduite autonome

l'IA qui fournit des explications sur ses intentions distales et/ou proximales

- Effet / conséquence attendue (match) -> atténuation sensorielle ?
- Effet sonore / conséquence inattendue (mismatch) -> prediction error ?

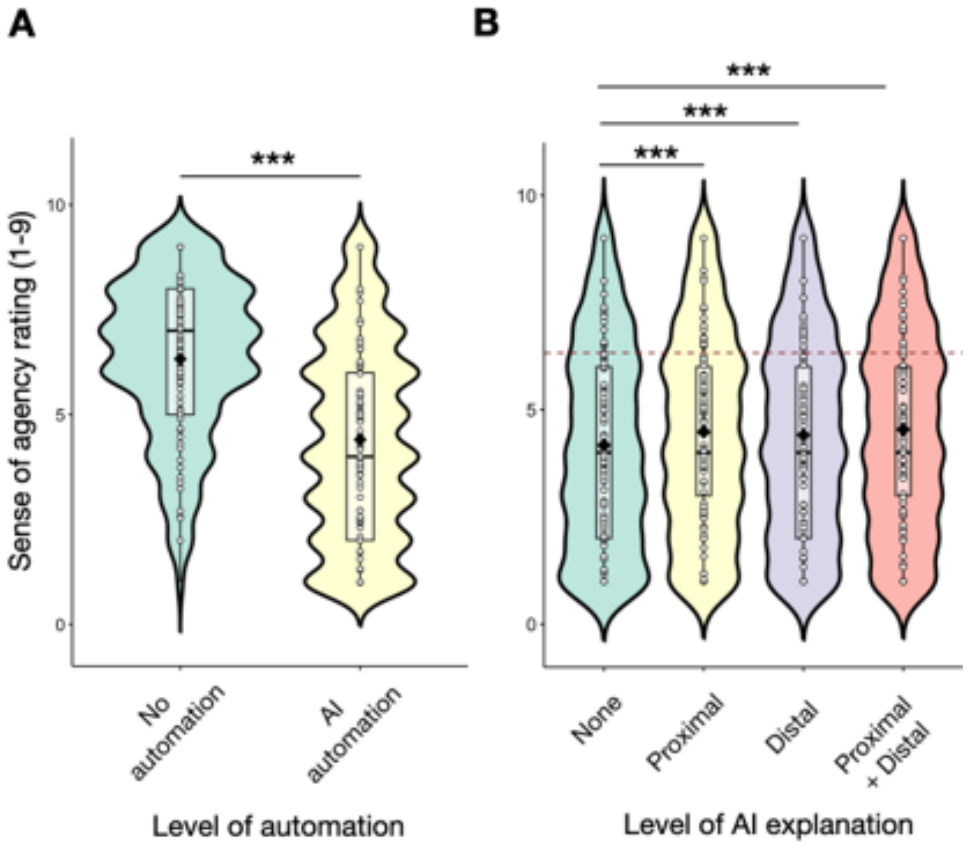


Chaque essai (cible atteinte) était suivi d'un effet sonore attendu ou Surprenant

03

Résultats

Résultats : Ratings



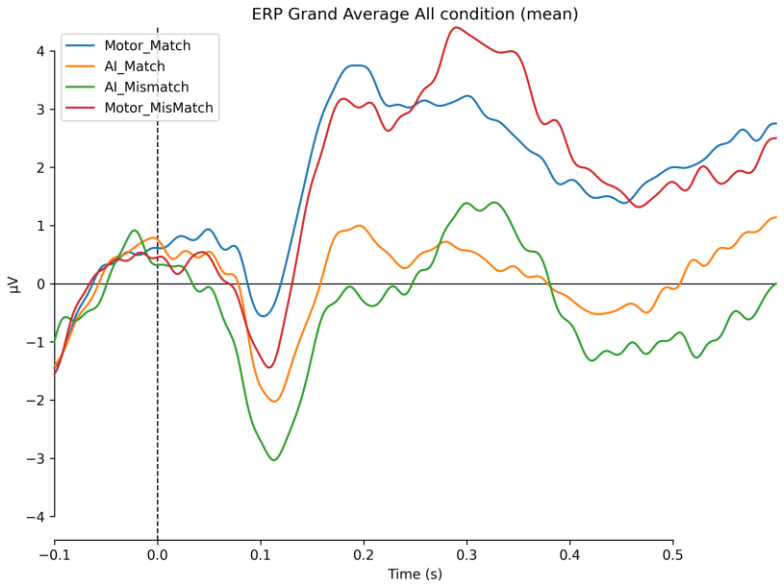
Lecture des données subjectives

Sentiment de contrôle explicite plus fort dans la condition manuelle qu'automatique

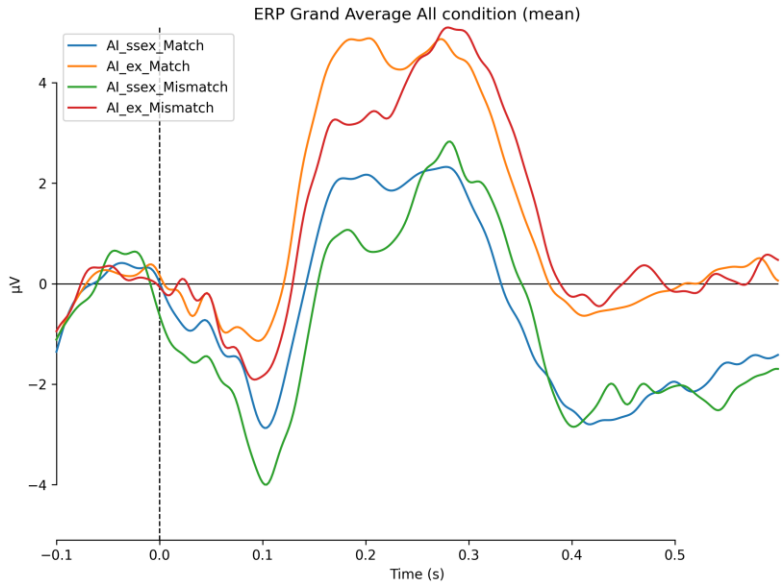
Toutefois, plus l'IA fournit d'explications (proximales + distales) plus ce sentiment explicite de contrôle augmente

Résultats : Analyses ERP / EEG

Moteur vs. AI sans ex.



AI sans ex. AI avec ex



Lecture des données ERP

Signatures neurales distinctes entre conditions motrices, conditions IA avec et sans explication

On observe une atténuation sensorielle dans la condition motrice vs. IA sans explication

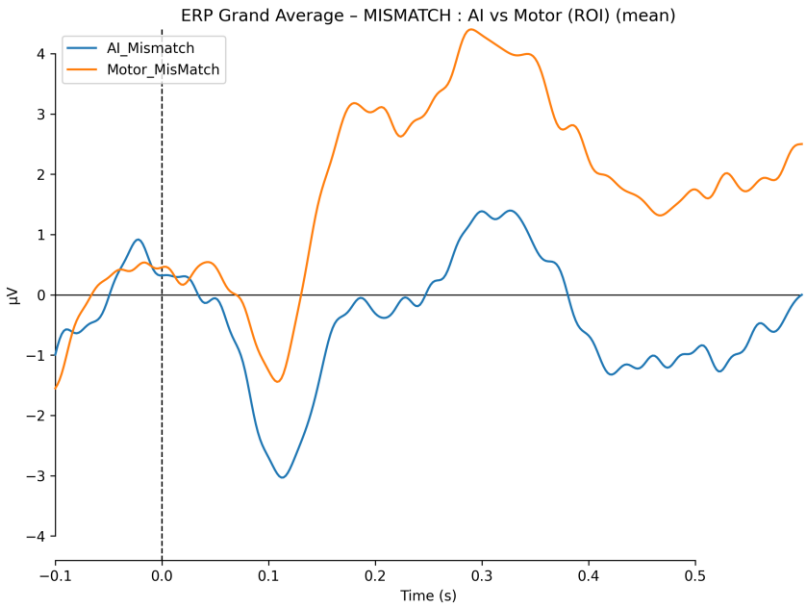
On observe une atténuation sensorielle dans la condition IA avec explication vs. IA sans explication

Prediction errors plus importantes dans la condition motrice vs IA sans explication

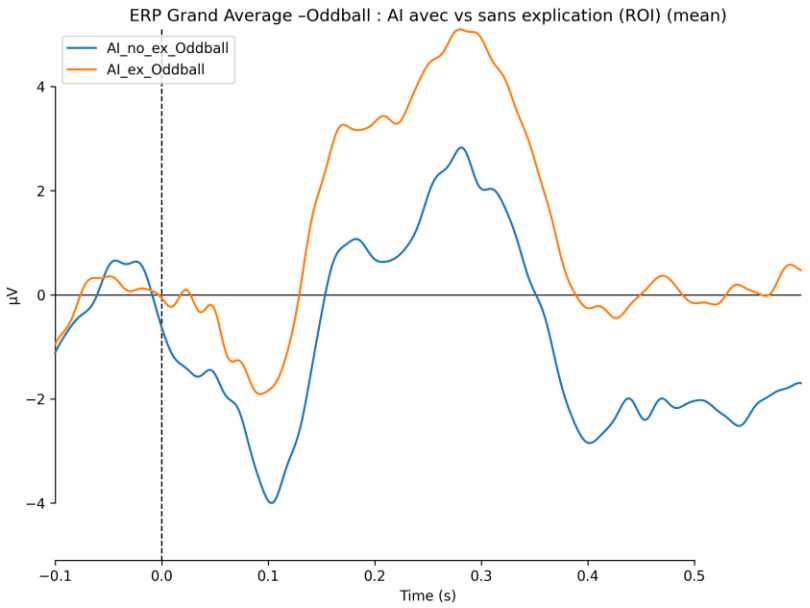
Prediction errors plus importantes dans la condition IA avec explication que IA sans explication

Résultats : Analyses ERP / EEG

Moteur vs. AI sans ex.



AI sans ex. AI avec ex



Lecture des données ERP

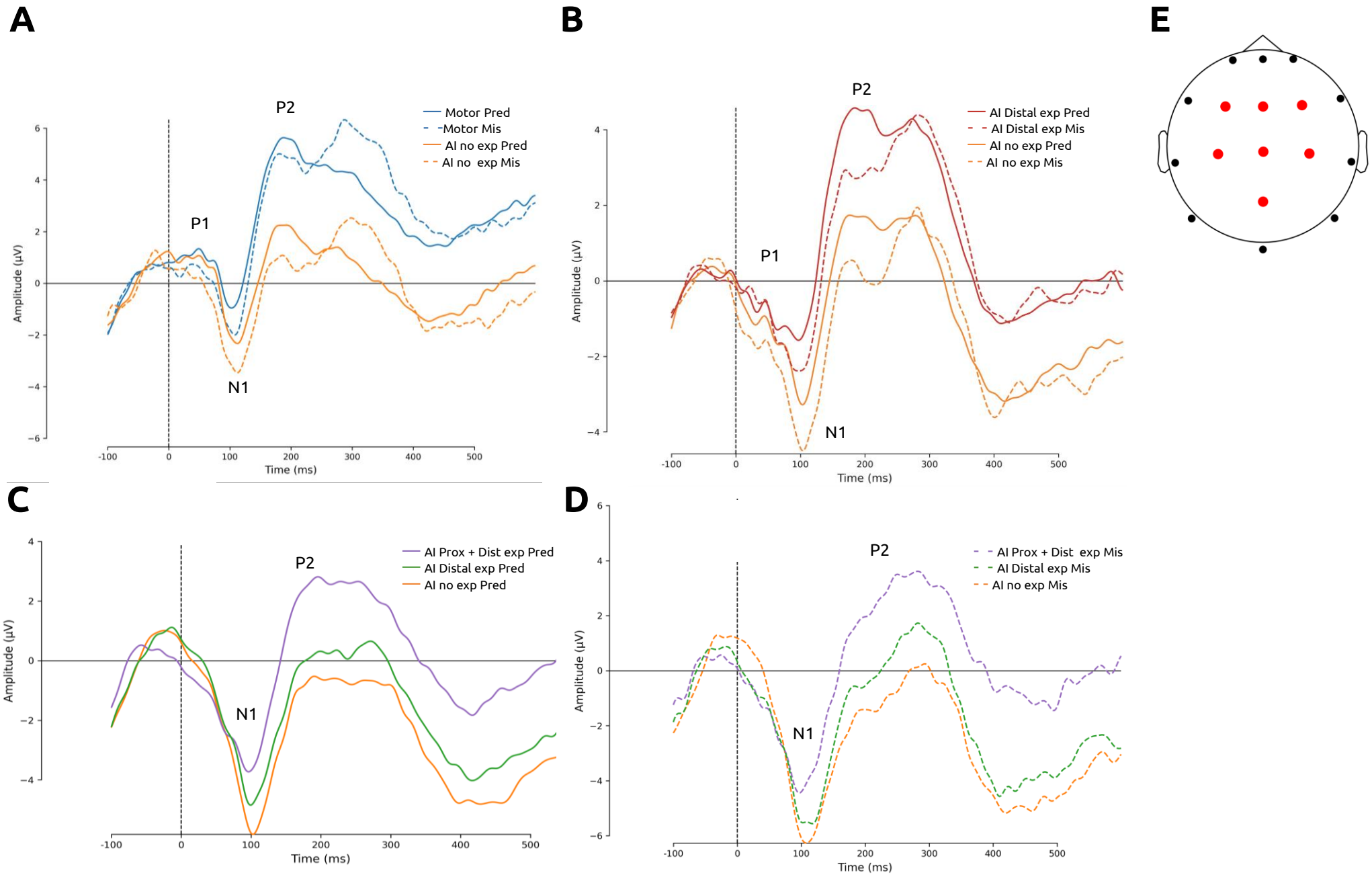
Signatures neurales distinctes entre conditions motrices, conditions IA avec et sans explication

On observe une atténuation sensorielle dans la condition motrice vs. IA sans explication

On observe une atténuation sensorielle dans la condition IA avec explication vs. IA sans explication

Prediction errors plus importantes dans la condition motrice vs IA sans explication

Prediction errors plus importantes dans la condition IA avec explication que IA sans explication


A
P2
P1
N1
Motor Pred
Motor Mis
AI no exp Pred
AI no exp Mis
Amplitude (µV)
Time (ms)
B
P2
P1
N1
AI Distal exp Pred
AI Distal exp Mis
AI no exp Pred
AI no exp Mis
Amplitude (µV)
Time (ms)
E
C
P2
N1
AI Prox + Dist exp Pred
AI Distal exp Pred
AI no exp Pred
Amplitude (µV)
Time (ms)
D
P2
N1
AI Prox + Dist exp Mis
AI Distal exp Mis
AI no exp Mis
Amplitude (µV)
Time (ms)

04

Discussion & Perspectives

Discussion & perspectives

Synthèse des résultats

- Sentiment d'agentivité explicite (questionnaires) toujours plus fort dans la condition de contrôle manuel qu'automatique
  - Toutefois, plus l'IA fournit d'indices / explications – plus les utilisateurs se disent être en contrôle
- Marqueurs neurophysiologiques d'agentivité (attenuation sensorielle + erreur de prediction) démontrent l'importance des explications fournies par l'IA pour restaurer contrôle et confiance
  - Explications de l'IA sont associées aux mêmes biomarqueurs d'agentivité qu'un contrôle moteur : permettraient-elles de restaurer l'agentivité implicite ?

Prochaines étapes & développements en cours

- Usage du deep learning et classification du niveau de confiance / agentivité neurophysiologique des utilisateurs au décours d'une IHM / IHAI : **accuracy actuelle > 75%**
- Nouvelles métriques "objectives" de confiance / agentivité pour des applications industrielles
- Expérience sur simulateur puis avec voiture autonome Capgemini

Vers des IHM & IHAI centrées utilisateurs ?

L'application de tels résultats lors de la conception d'IHM / XAI permettrait d'améliorer la qualité des interactions et de remettre l'humain au coeur des procédés impliquant l'IA

Valérian CHAMBON,
DR-CNRS ENS-PSL

Eléonore HOUDOYER,
PhD student ENS-PSL

Emilien BONHOMME,
PhD & Data scientist,
Capgemini

Thank you.

Make it real.

ENS | PSL | Capgemini

World Map

DON’T ADD BORDERS OR OUTLINES TO HIGHLIGHT REGIONS OR COUNTRIES.

LOCATIONS SHOULD BE INDICATED WITH A PIN OR DOT AND LABEL