



HAL
open science

Comment rendre l'intelligence artificielle explicable ?

Corentin Boidot

► **To cite this version:**

Corentin Boidot. Comment rendre l'intelligence artificielle explicable ?. Science en Theizh 2023, Apr 2023, Brest, France. 2023. hal-04207415

HAL Id: hal-04207415

<https://hal.science/hal-04207415>

Submitted on 18 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

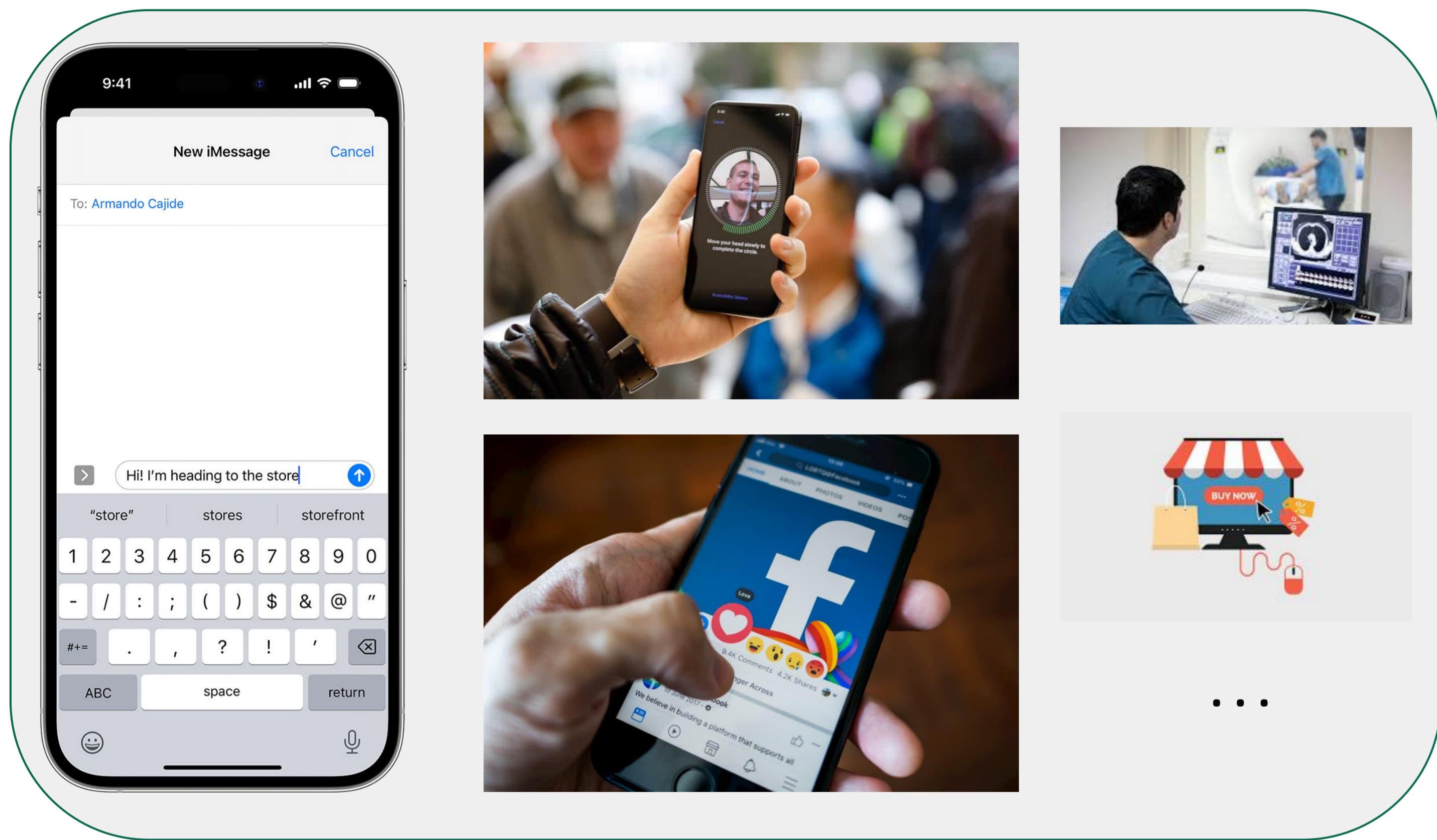
Comment rendre l'intelligence artificielle explicable ?

Corentin Boidot

Crédit Mutuel
ARKEA



Les technologies d'IA



Une approche utile, mais insatisfaisante !

C'est quoi une bonne explication ?

Une question de public !

- Le data scientist \Rightarrow compréhension (modèle), garanties
- L'expert applicatif \Rightarrow compréhension (tâche), confiance
- Les superviseurs \Rightarrow confiance, garanties
- Le grand public \Rightarrow confiance, curiosité, actionnabilité

A priori une explication doit être *claire, et concise*.

On peut ainsi **évaluer** des explications, selon leur format, à condition de *formaliser* ces propriétés désirées.

En pratique, seul une *expérience* permet de comparer objectivement des explications. Il faut alors faire interagir l'humain avec un système d'IA pour évaluer **a posteriori** son utilité.

\rightarrow *C'est le but de mon travail !*

Mesures

- subjectives
 - questionnaires
 - entretiens
- objectives
 - taux d'utilisation de l'explication
 - performance à la tâche
 - vitesse

Expliquer un résultat d'IA

Différents types d'explications (automatiques) pour les IA :

Explain using:

- Simplification**: Comme blueTotalGold est inférieur à 17551 et redTotalExperience est supérieur à 17394, l'équipe rouge devrait certainement gagner.
- Exemple**: **Pronostic IA : équipe rouge**
- Variable**: **Pronostic IA : équipe rouge**
- Information supplémentaire**: **Pronostic IA : équipe rouge**

	blue	red
firstBlood	1	0
Dragons	0	1
Heralds	0	0
TowersDestroyed	0	0
WardsPlaced	16	16
TowersDestroyed	1	3
Kills	5	7
Assists	4	7
TotalGold	16324	16654
TotalExperience	18159	18104
TotalMinionsKilled	231	213
TotalJungleMinionsKilled	60	60

Le match le plus similaire dont je me rappelle est celui-ci. J'estime que l'équipe rouge gagnerait, avec une confiance de 67%. L'équipe rouge a finalement gagné ce match. Si je cherche un match gagné par les rouges, on doit aller chercher au moins le 7e plus proche.

Pronostic IA : équipe rouge

Probabilité estimée : 52%

Pronostic IA : équipe rouge

Victoire prédite : équipe rouge

Probabilité estimée : 67%

Et la recherche suit son cours...

Match n°5

	blue	red	mean
firstBlood	1	0	0.5
Dragons	0.37	0.41	0.39
Heralds	0.23	0.25	0.27
TowersDestroyed	0.05	0.04	0.04
WardsPlaced	21.08	21.22	21.15
TowersDestroyed	2.82	2.75	2.78
Kills	4.32	4.12	4.24
Assists	4.54	4.63	4.58
TotalGold	16496.05	16477.78	16487.02
TotalExperience	17026.73	17044.63	17045.68
TotalMinionsKilled	236.73	217.52	227.12
TotalJungleMinionsKilled	56.08	51.26	56.67

Match n°1

	blue	red	mean
firstBlood	1	0	0.5
Dragons	0.0	0.0	0.0
Heralds	0.0	0.0	0.0
TowersDestroyed	0.0	0.0	0.0
WardsPlaced	18.0	18.0	18.0
TowersDestroyed	2.0	2.0	2.0
Kills	0.0	0.0	0.0
Assists	0.0	0.0	0.0
TotalGold	16038.0	17443.0	16740.5
TotalExperience	17294.0	18275.0	17784.5
TotalMinionsKilled	200.0	203.0	201.5
TotalJungleMinionsKilled	40.0	40.0	40.0

Victoire prédite : équipe rouge

Probabilité estimée : 67%

\Rightarrow On utilise le pronostic sur des matchs de League of Legends comme **tâche** sur laquelle s'évaluent ensemble

- une performance humaine
- une performance IA
- une performance homme-IA (liée à l'explication)

\Rightarrow *Expérience en cours au CERV !*

Envoyez un mail à boidot@enib.fr pour participer d'ici 10 jours, ou flashez le QR code. Pré-requis : une connaissance minimale du MOBA League of Legends.

