

Journées Madics 8 juillet 2021 - Rennes



Federate ExplaiNability DEfinitions and Research

Nicolas Labroche, Julien Aligon, Michael Baker









Pourquoi l'atelier Fender?

- Fender se propose de rapprocher des spécialistes de l'informatique et des sciences cognitives pour traiter les questions de l'explicabilité
 - o Call for paper de l'an passé (mars 2020, AIJ)

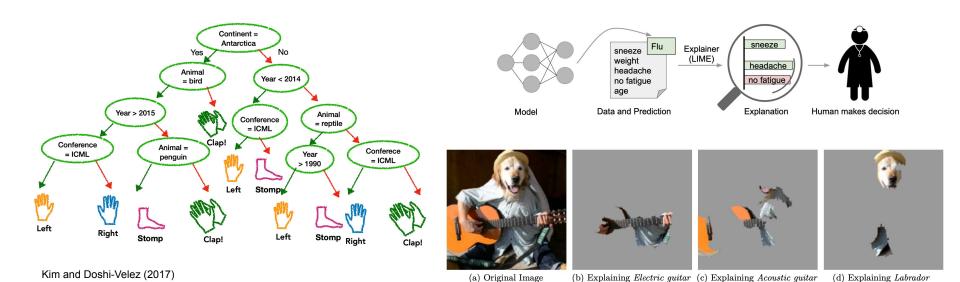
Explanation in Artificial Intelligence: Insights from the Social Sciences

Tim Miller

School of Computing and Information Systems University of Melbourne, Melbourne, Australia tmiller@unimelb.edu.au "There is a rich understanding in philosophy and cognitive & social psychology of how humans explain concepts to themselves and to others"

Pourquoi l'atelier Fender ?

Aborder le problème de l'explicabilité sans se limiter à un type de problème ou un type de données



Ribeiro et al. (2016)

Questions scientifiques abordées

1. Définition d'une explication

O Comment établir une définition croisée INS2I/INSHS de ce qu'est une explication ?

2. Liens entre données et explications

A quel point les explications dépendent de la qualité, diversité des données ?

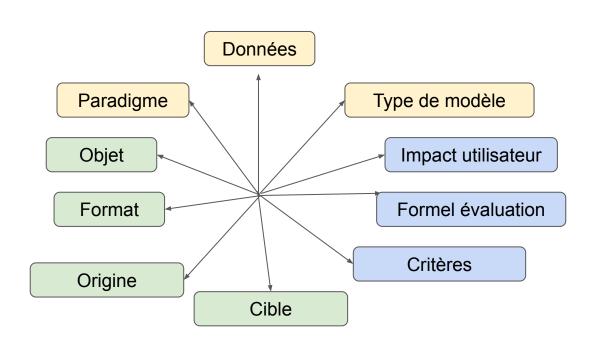
3. Liens avec l'utilisateur

Comment prendre en compte l'humain dans le processus d'explication ?

4. Evaluation d'une explication

A quel point prendre en compte l'humain dans le processus d'évaluation ?

Explicabilité : de nombreuses dimensions d'étude



Comment représenter le problème d'explicabilité ?

Inspiré de Friedrich, Zanker (2011)

Données : informations exploitées

- taille / dimensionnalité / type
- meta données
- modèle de préférence utilisateur

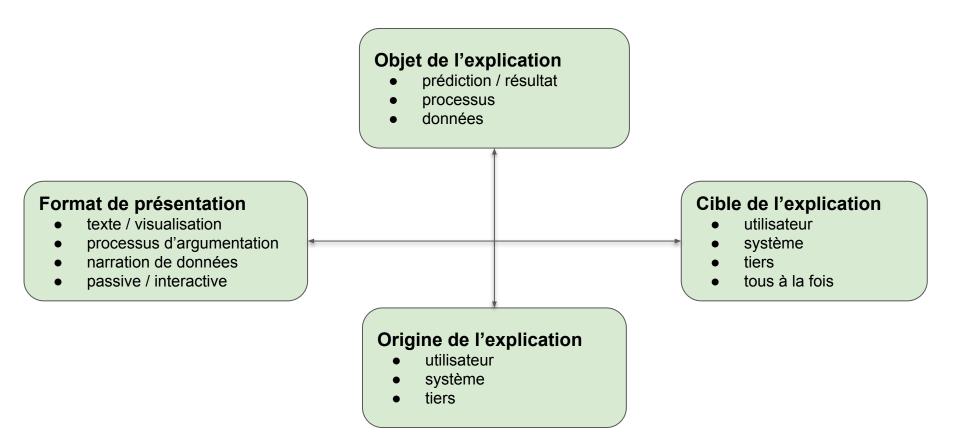
Modèle explicatif

- how / why explanation (MYCIN, 1984)
- white box
- black box

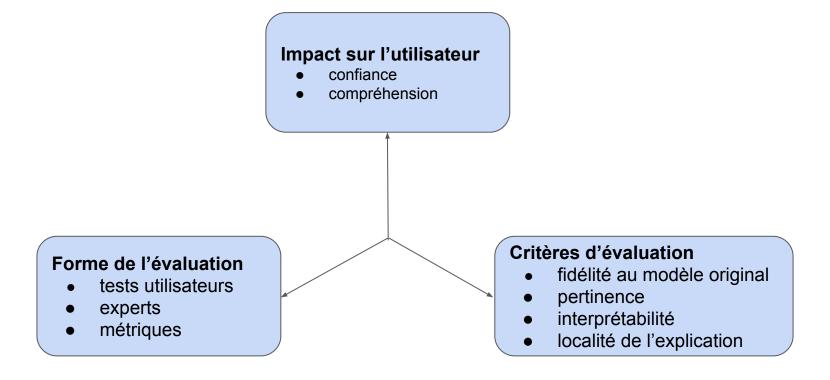
Paradigme de l'explication

- classification
- clustering
- recommendation
- .

Quelle relation avec l'utilisateur?



Quelle évaluation de l'explicabilité ?



Fender: un groupe interdisciplinaire

Informatique

SHS

Données



Jean-yves Antoine Nicolas Canovas Alexandre Chanson Cyril De Runz **Nicolas Labroche** Patrick Marcel Nicolas Ragot Arnaud Soulet Willème Verdeaux



Bernd Amann Isabelle Bloch Christophe Denis Marie-Jeanne Lesot Nicolas Maudet





Julien Velcin



Nedra Mellouli



Vassilis Christophides







Vincent Labbé



Thibault Laugel Xavier Renard



Nafise Gouard



Robin Cugny



Elodie Escriva

Planning de la matinée

- 9h00 Bienvenue et mot d'introduction, Nicolas Labroche, Julien Aligon, Michael Baker
- 9h15 Explainable Anomaly Detection (xAD), Vassilis Christophides (ENSEA ETIS)
- 9h45 Epistémologie et philosophie des sciences de l'apprentissage machine, Christophe Denis (LIP6)
- 10h15 Pause
- 10h30 La robustesse : un nouveau critère d'évaluation des explications en apprentissage automatique, Robin Cugny (IRIT)
- 11h00 Explanation for Humans, for Machines, for Human-Machine Interactions? Alain Mille (LIRIS)
- 11h45 Table ronde

Et aussi

Nous suivre toute l'année :

https://sites.google.com/view/fender-madics/accueil

Évènements à venir

- Institut d'automne en IA explicabilité : http://ia2.gdria.fr/inscriptions21/
- EGC 2022, Blois: https://egc2022.univ-tours.fr

Mais aussi

- INRIA Hybrid Approaches for Interpretable AI: https://project.inria.fr/hyaiai/
- GT explicabilité du GdR IA : https://www.gdria.fr/gt-explicabilite/



EGC 2022 BLOIS 24-28 janvier egc2022.univ-tours.fr



Sihem Amer-Yahia Présidente du comité de programme

Dates de soumission

Résumé: 8 octobre

Article: 15 octobre





















Remerciements

- Ewa Kijak et Vincent Claveau pour l'organisation locale
- Marie Chagnoux pour l'aide à l'organisation et la relecture attentive
- Les membres du groupe Fender
- Les membres du groupe Blobfish-BDTLN!