Explainable AI: a narrative review at the crossroad of Knowledge Discovery, Knowledge Representation and Representation Learning

Ikram Chraibi Kaadoud, Lina Fahed, Philippe Lenca

Explainable AI: a narrative review at the crossroad of Knowledge Discovery, Knowledge Representation and Representation Learning

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Authors

Ikram Chraibi Kaadoud*
Lina Fahed*
Philippe Lenca*

* IMT Atlantique, Brest, France

Partners

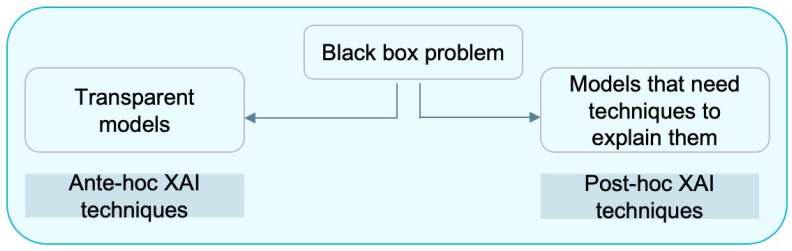
DECIDE
Lab-STICC
cnrs
Région Bretagne
UNION EUROPÉENNE
Fonds européen de développement régional

Explainable AI (XAI)

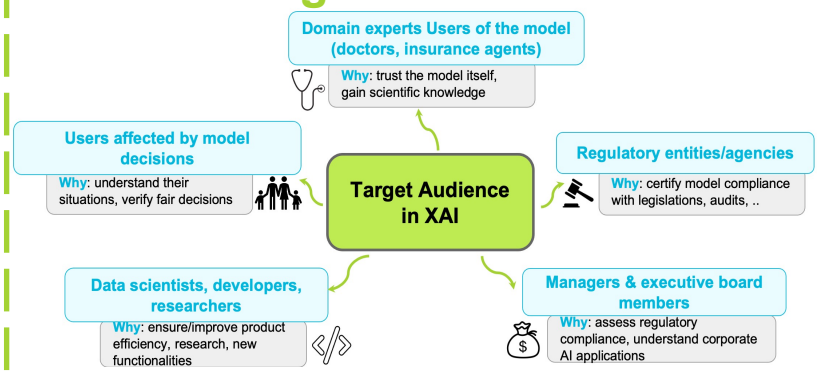
Goals:

1) **Directly design explainable models and results**
2) **Make AI models more intelligible and accessible:**
   - For a given **question** about an **AI model inner mecanisms and/or results**, provide an **explanation** to a **target audience**

AI Models

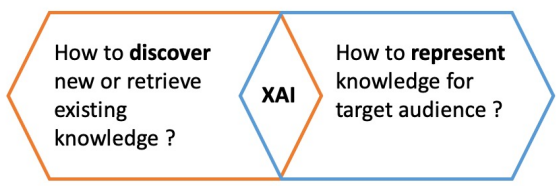


Target audience



Explanation

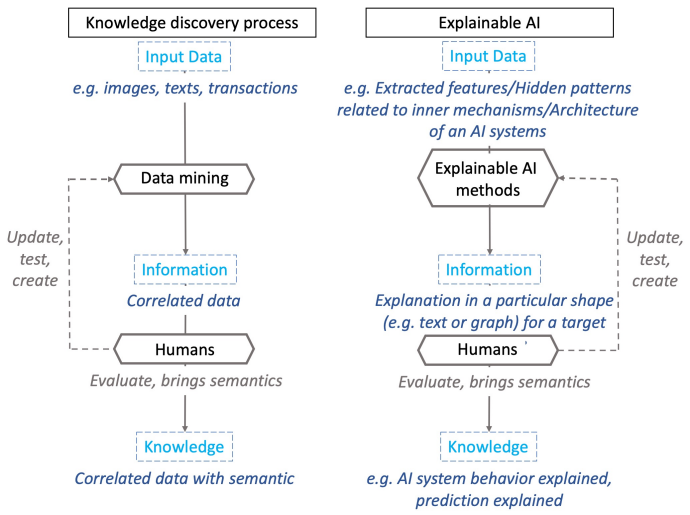
- **Information** in a semantically complete format, which is self-sufficient and **chosen according to the target audience** regarding its knowledge, expectations and the context
- **Interface** between AI model-target audience
- **Contextual** by its very nature : specific to a given target audience, and context



Knowledge and representation for XAI…questions asked since 1960

Knowledge Discovery process (KDP)

- "How to efficiently discover new or retrieve existing knowledge?"



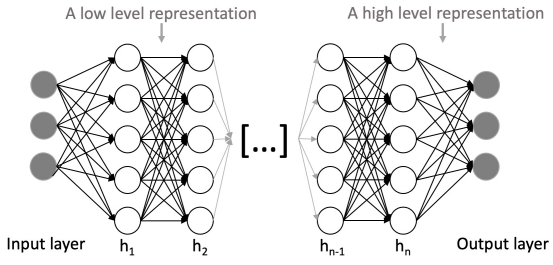
Two schematic ways for data transformation into knowledge:
***On the left,** within a Knowledge Discovery process, and **on the right** within a XAI process*

Knowledge representation (KR)

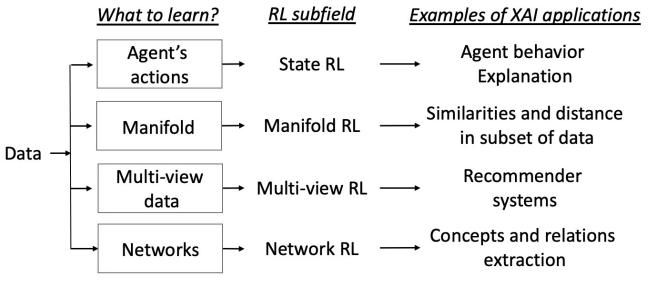
- "How to represent the knowledge efficiently to be able to reason on it?"

Representation Learning (RL)

- "How to efficiently learn representations of the data that make it easier to extract usefull information when building AI models such as classifiers or predictor ?"



Illustrative and schematic representation of the position of a low level representation and a high level representation in a deep neural network.
h_x refers to the xth hidden layer in the network.



Examples for XAI applications using RL subfeild

August 2021 MRC-HCCS @IJCAI 2021 Montréal, Canada

XAI should take advantage of recent works in KDP, KR and RL domains, **as well as older works**

Contact: ikram.chraibi-kaadoud@imt-atlantique.fr