



HAL
open science

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

Ikram Chraibi Kaadoud, Lina Fahed, Philippe Lenca

► To cite this version:

Ikram Chraibi Kaadoud, Lina Fahed, Philippe Lenca. Explainable AI: a narrative review at the crossroad of Knowledge Discovery, Knowledge Representation and Representation Learning. Twelfth International Workshop Modelling and Reasoning in Context (MRC) @IJCAI 2021, Aug 2021, Montréal, Canada. 2021. hal-03345286

HAL Id: hal-03345286

<https://hal.science/hal-03345286v1>

Submitted on 15 Sep 2021

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.

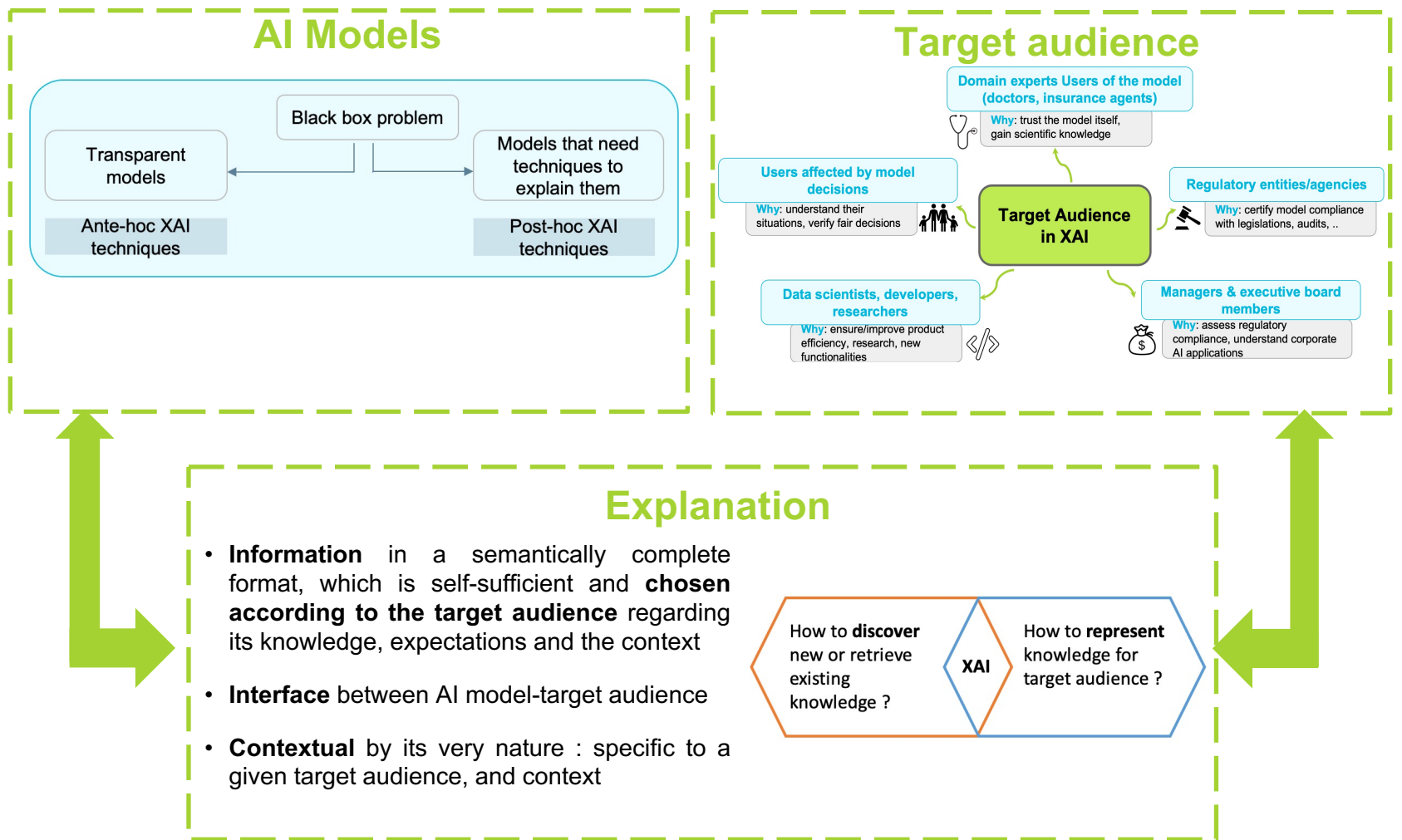


Explainable AI: a narrative review at the crossroad of Knowledge Discovery, Knowledge Representation and Representation Learning

Explainable AI (XAI)

Goals:

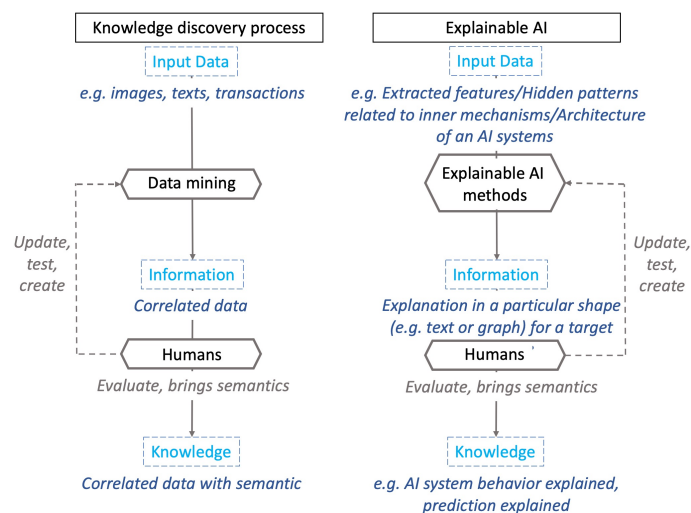
- 1) **Directly design explainable models and results**
- 2) **Make AI models more intelligible and accessible:**
 - For a given **question** about an **AI model inner mechanisms and/or results**, provide an **explanation** to a **target audience**



Knowledge and representation for XAI...questions asked since 1960

Knowledge Discovery process (KDP)

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



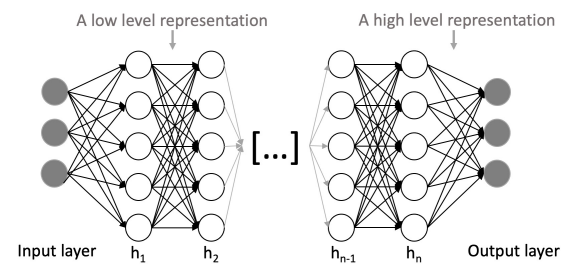
Two schematic ways for data transformation into knowledge: On the left, within a Knowledge Discovery process, and on the right within a XAI process

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 useful 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 x th hidden layer in the network.

Data	What to learn?	RL subfield	Examples of XAI applications
Agent's actions	Agent's actions	State RL	Agent behavior Explanation
Manifold	Manifold	Manifold RL	Similarities and distance in subset of data
Multi-view data	Multi-view data	Multi-view RL	Recommender systems
Networks	Networks	Network RL	Concepts and relations extraction

Examples for XAI applications using RL subfield

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

Authors

Ikram Chraïbi Kaadoud*
Lina Fahed*
Philippe Lenca*

* IMT Atlantique, Brest, France

Partners

