Explainable Techniques to Contribute to a more Fair and Transparent Tax AI

Dieter Brughmans
The complexity of AI

AlexNet

ImageNet Classification with Deep Convolutional Neural Networks

61 000 000

AlphaFold

21 000 000

ChatGPT

GPT-3

175 000 000 000
XAI Characteristics

The Counterfactual Explanation

- Why was a declaration predicted as fraudulent?
- The minimal change in input that changes the output.
- “If the country of origin would change from Colombia to USA and the type of goods from bananas to avocados, the predicted class would change from fraudulent to compliant.”
- No limitations on complexity of the model
What is the influence of each feature on the prediction score?
- Based on Game Theory
- Explains score, not predicted class
- Can be biased
Counterfactual Explanations for Customs Fraud Detection

Data Scientist
- Model improvement
- Identify Trends

Targeting Officer
- High-level explanations
- Improve investigation time

Domain Expert
- Aggregated explanations
- New insights from data

Decision Subject
- No explanations
Open Issues and Challenges XAI

▪ Which explanation method to use?
▪ How to choose among explanations: moral hazard
▪ What explanations do users want?
▪ Who should get access to explanations