



explAIIn

Explain AI@Imperial Workshop

25<sup>th</sup> April 2018

Francesca Toni

Department of Computing

Imperial College London

# Why explanation?

The EU General Data Protection Regulation (GDPR) is the most important change in data privacy regulation in 20 years - we're here to make sure you're prepared.

## SECTION 1

### TRANSPARENCY AND MODALITIES

### *Article 12*

*Transparent information, communication and modalities for the exercise of the rights of the data subject*

## Chapter 3: Designing artificial intelligence

Access to, and control of, data

Anonymisation

Strengthening access and control

Box 2: Open Banking

Intelligible AI

Technical transparency

Explainability



“...the development of intelligible AI systems is a fundamental necessity if AI is to become an integral and trusted tool in our society.”

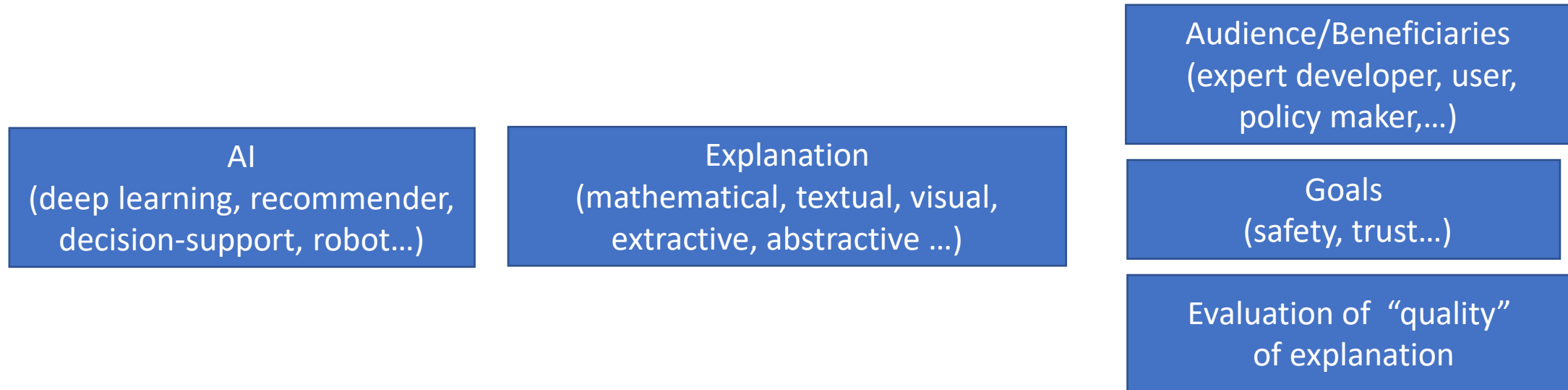


## ICML | 2018

- [5th Workshop on Fairness, Accountability, and Transparency in Machine Learning \(FAT/ML 2018\)](#)
- [3rd Workshop on Human Interpretability in Machine Learning \(WHI 2018\)](#)
- [Workshop on Interpretable & Reasonable Deep Learning and its Applications \(IReDLiA 2018\)](#)
- [Workshop on Explainable Artificial Intelligence \(XAI 2018\)](#)

# What is an explanation?

- 1) Inform and help understand why a particular conclusion was reached
  - 2) provide grounds to contest the conclusion if undesired
  - 3) Inform and help understand what could be changed to get a desired conclusion
- (Wachter et al 2017)



Explanability =

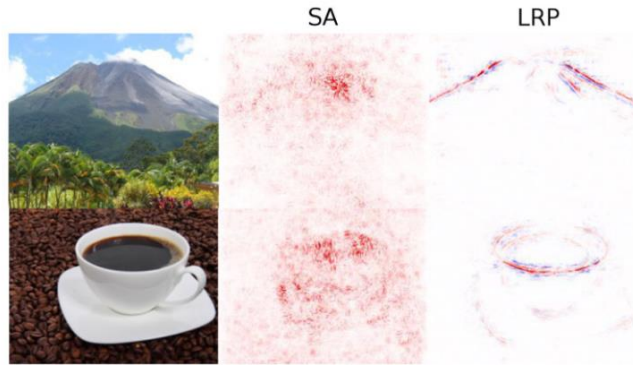
Transparency, Interpretability, Verifiability, Comprehensibility

Weller 2017

# Examples

## (A) Image classification

Explaining predictions: "Volcano", "Coffe Cup"



Samek, Wiegand, Müller 2017

## (B) Natural Language Processing

### Review

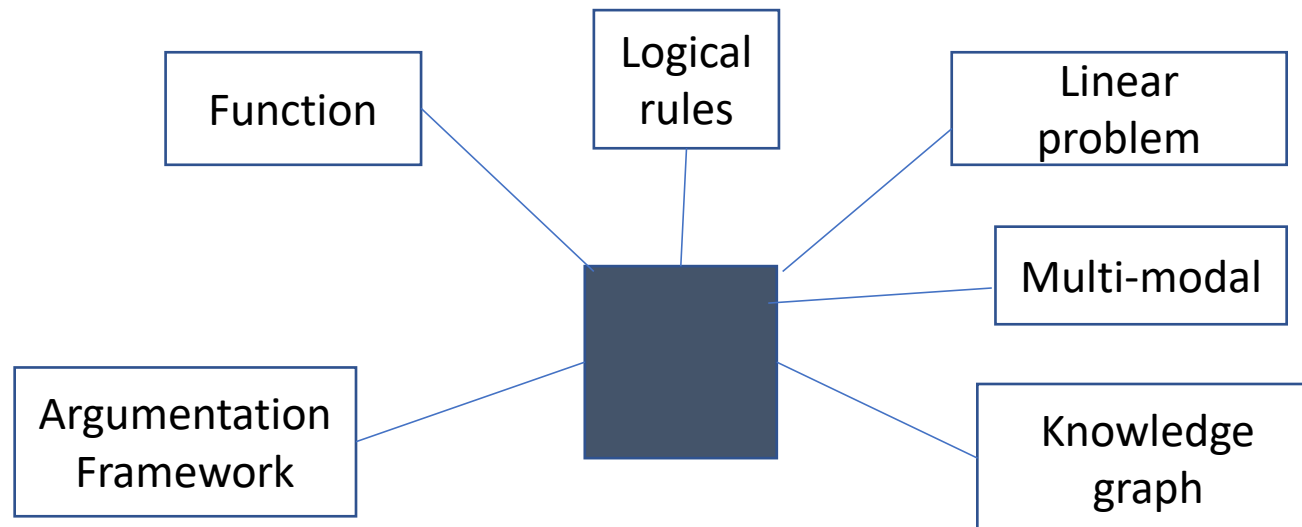
the beer was n't what i expected, and i'm not sure it's "true to style", but i thought it was delicious. **a very pleasant ruby red-amber color** with a relatively brilliant finish, but a limited amount of carbonation, from the look of it. aroma is what i think an amber ale should be - a nice blend of caramel and happiness bound together.

### Ratings

**Look: 5 stars**

Smell: 4 stars

Lei, Barzilay, Jaakkola 2016



**Forms of explanation  
at this workshop**

# Program

## AM

**Richard Evans** - Learning Explanatory Rules from Noisy Data

**Stephen Muggleton** - Ultra-strong machine learning - comprehensibility of programs learned with ILP

**Alessio Lomuscio** - An approach to reachability analysis for feed-forward ReLU neural Networks

**Hajime Morita** - Explainable AI that Can be Used for Judgment with Responsibility

## PM

**Christos Bechlivanidis** - Concreteness and abstraction in everyday explanation

**Seth Flaxman** - Predictor Variable Prioritization in Nonlinear Models: A Genetic Association Case Study

**Erisa Karafili** - Argumentation-based Security for Social Good

**Kristijonas Cyras** - Explaining Predictions from Data Argumentatively

**Oana Cocarascu/Antonio Rago** - Argumentation-Based Recommendations: Fantastic Explanations and How to Find Them

**Yannis Demiris** - Multimodal Explanations in Human Robot Interaction

**Euan Matthews** - The Practicalities of Explanation  ContactEngine