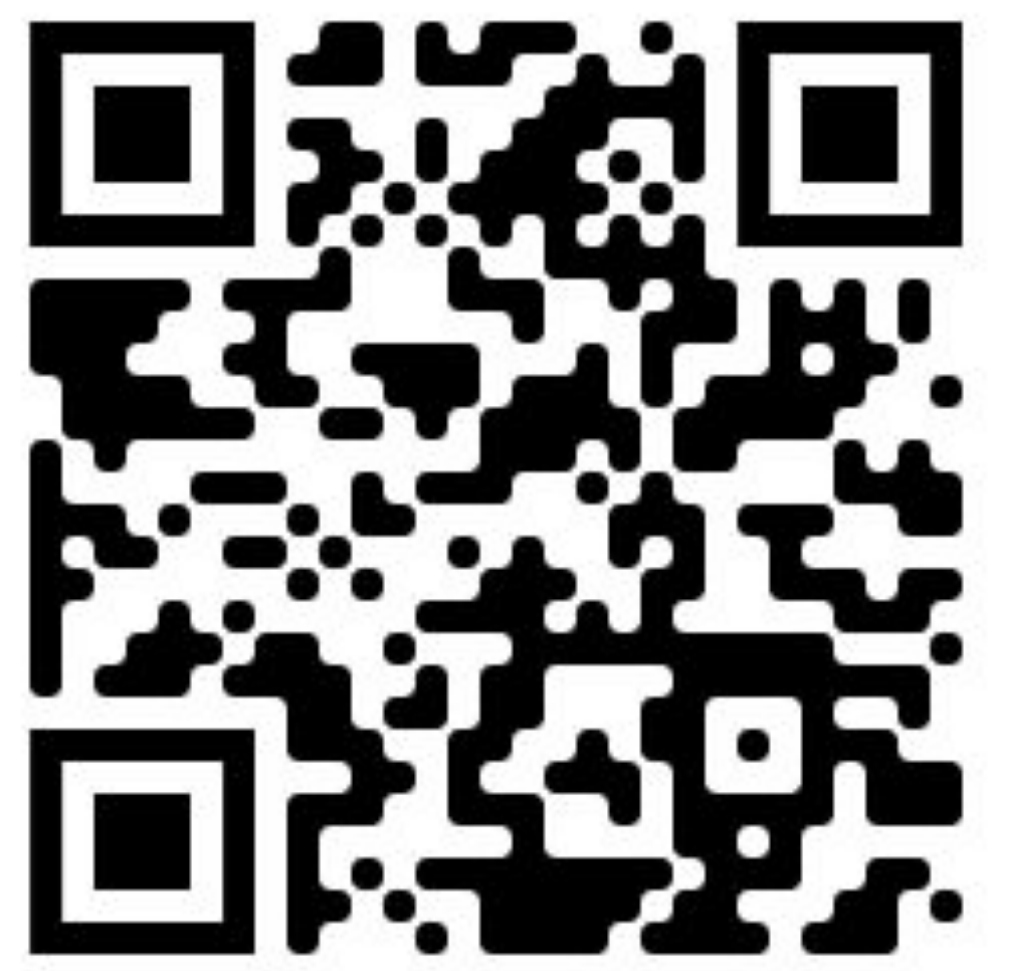


Explainable Artificial Intelligence: Human-centered Counterfactual Explanation



Muhammad Suffian (m.suffian@campus.uniurb.it)
Department of Pure and Applied Sciences, University of Urbino, Italy

Introduction

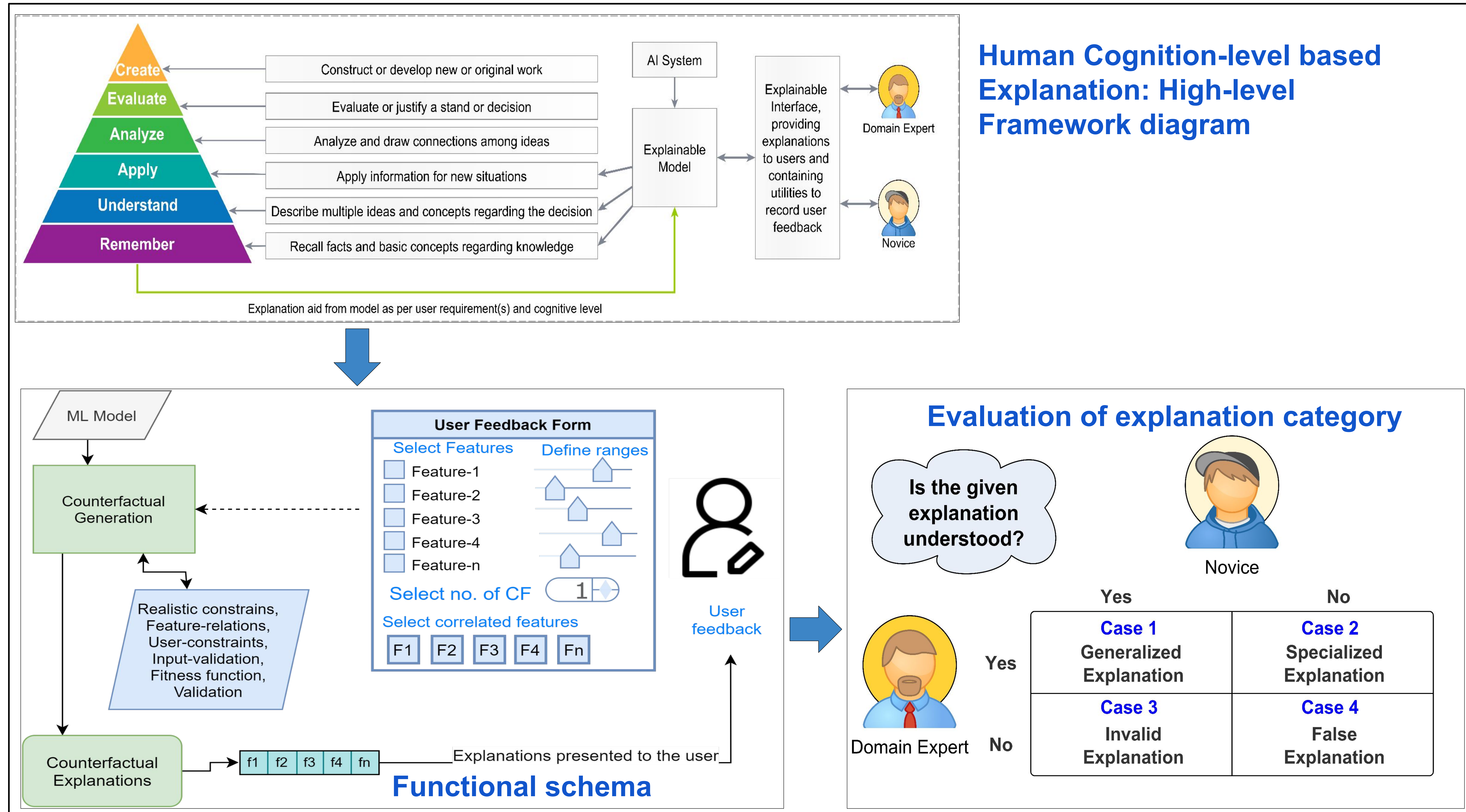
In spite of the considerable progress in Explainable Artificial Intelligence (XAI), inadequate emphasis has been put on the needs and preferences of human users. We bring together cognitive science and artificial intelligence to investigate the human cognition level-based counterfactual explanations to achieve the goals of 'good explanation' under the umbrella of human-centered XAI.

Current XAI

Currently XAI research recognizes that different user prefer different explanations based on their levels of expertise.
Source: Mohseni, Zarei and Ragan, 2018

Our Vision

Our vision and aim is to develop Human-centered XAI framework to generate user preference based explanations with user cognitive levels of understanding. For this task, we use Bloom's Taxonomy to capture human-understanding levels.
source: Revised Bloom's taxonomy 2001.



Conclusion

We investigate the user preferences of explanations for different users (experts, novice) to generate the explanations accordingly for their better understanding.

Acknowledgment

This is an ongoing PhD project at University of Urbino, Italy under the supervision of Alessandro Bogliolo (Digit Srl, Italy) and co-supervision of Jose Maria Alonso-Moral (CiTIUS, Spain).