

Bridging the Gap between Human-Computer Interaction and Machine-Learning on Explainable AI: Initial Observations and Lessons Learned

Comblent la distance entre l'interaction humain-machine et le machine learning sur l'IA explicable : premières observations et leçons apprises

Julien ALBERT*, Adrien BIBAL**, Benoît FRENAY*, Bruno DUMAS*

*NaDI/PReCISE, Faculty of Computer Science, University of Namur, Belgium

**University of Colorado Anschutz Medical Campus, USA

TRAIL
TRUSTED AI LABS
BY DIGITALWALLONIA.AI / SPW-RECHERCHE

 **Wallonie**
service public
SPW

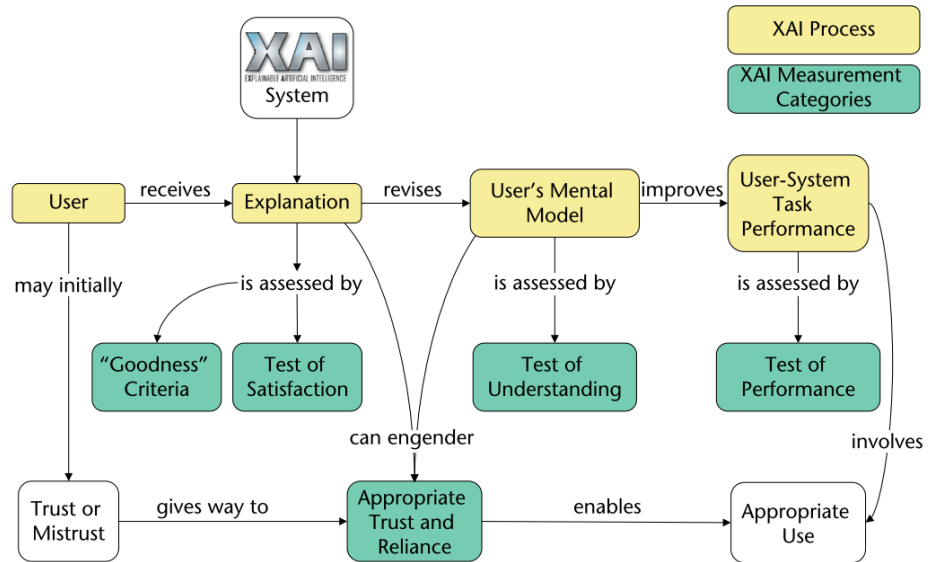
 **PReCISE**
Namur Digital Institute
NADI



UNIVERSITÉ
DE NAMUR

Explainable Artificial Intelligence (XAI)

- XAI regroups “movements, initiatives and efforts made in response to AI transparency and trust concerns” [Adadi2018]
- Its goal is to develop methods and tools “to explain or present in understandable terms to a human” the working of AI systems [Doshi-Velez2017]
- XAI research must involve multiple disciplines, especially Human-Computer Interaction (HCI) [Liao2021]



Explanation process from [Gunning2019]

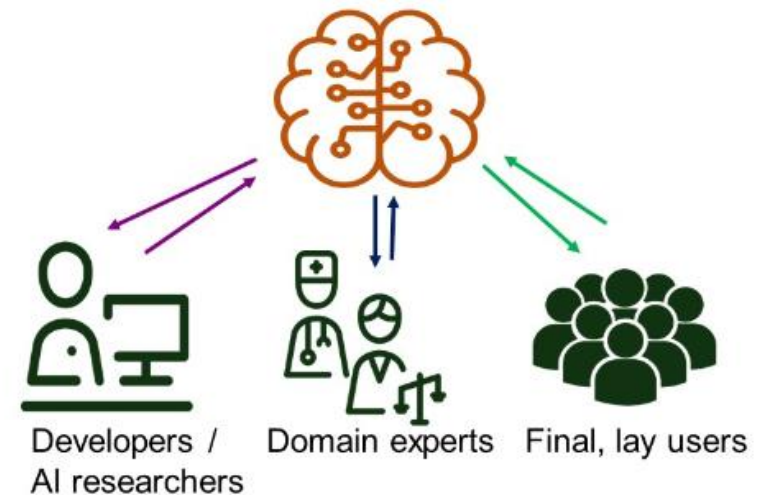
IHM '22 – Workshop HCI and XAI





- What: organization of a workshop during [IHM '22 conference](#)
 - With about 30 participants from various domains of HCI and ML fields
- Why: to encourage exchanges of ideas and to foster collaborations between HCI and ML researchers
- How:
 - Presentations of research activities on XAI
 - World Café as a moment of meeting and discussion
- <https://projects.info.unamur.be/ihm-xai/index-en.html>



1. User Profiles

- Relevant and meaningful XAI requires an understanding of user needs and context [Liao2021]
- Main insights from the discussions
 - Definition of user profiles as a compromise between particular use cases and generalization potential
 - Need for user research and modeling methods
 - Beyond profile, the broader context must also be investigated
 - User perception and cognitive aspects are impactful

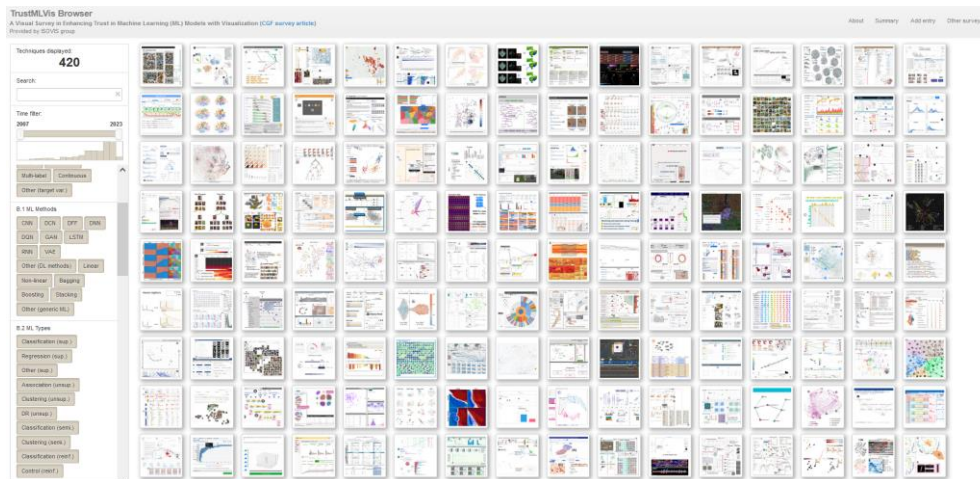


	Developers / AI researchers	Domain experts	Final, lay users
WHY 	Verification Improvement	Learning Adoption	Compliance with legislation
WHAT 	Global model Data representation Why and Why not	Local explanations Why and why not	Why not
HOW 	Intrinsic / Post-hoc	Post-hoc Visualization Natural language	Post-hoc Brief Plain language
EVAL 	Completeness tests Performance	Test of comprehension Performance Survey of trust	Satisfaction questionnaires

Framework proposed by [Ribera2019]

2. Model-Representation-Presentation

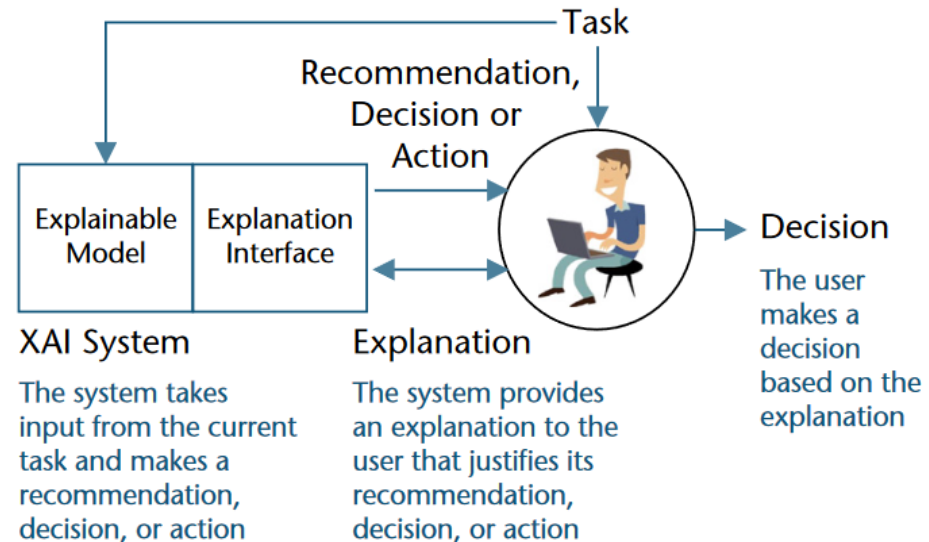
- It is important to distinguish these 3 stages in the interaction between the user and the model [Bibal2016]
- Main insights from the discussions
 - Representation and presentation depend on the addressed profile... But who chooses?
 - Multiple modalities: numerical, rules, textual, visual and mixed
 - Need to deal with the information loss between the stages
 - A unique design choice for the 3 stages is not mandatory



Screenshot of TrustMLVis Browser , <https://trustmlvis.lnu.se/>

3. Interaction & Actionability

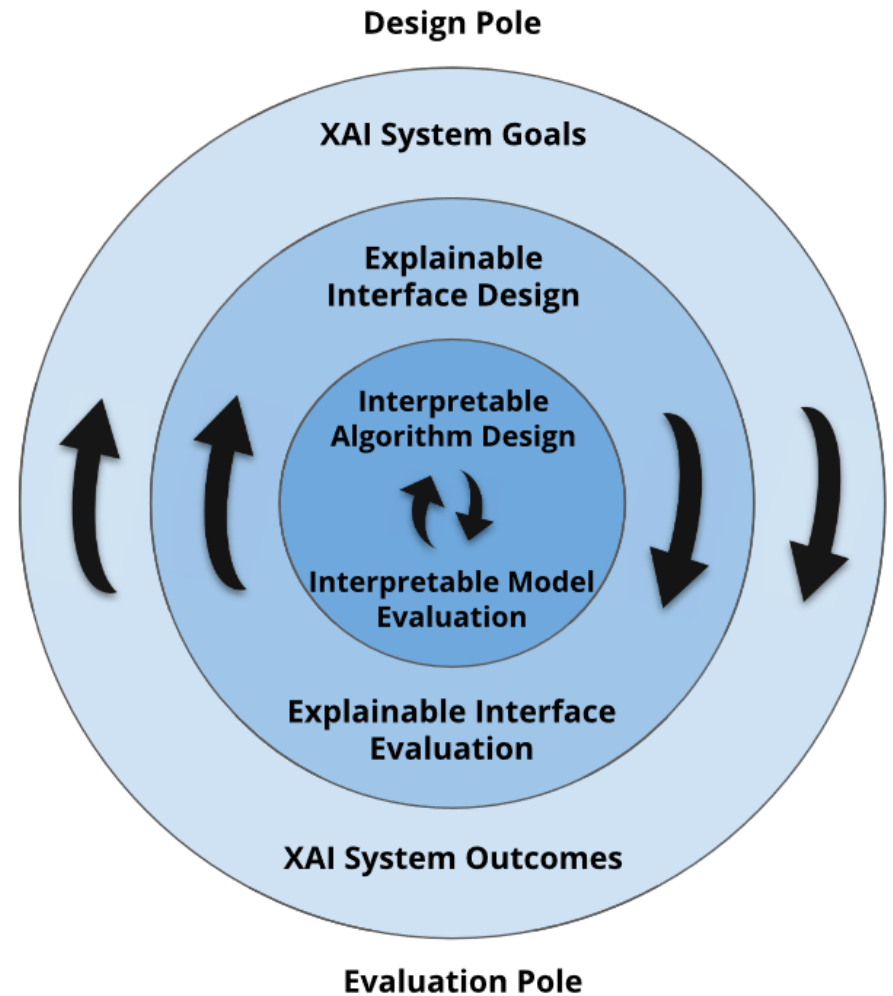
- It's important to put back the explainability in an interaction perspective [Gunning2019]
- Main insights from the discussions
 - Explanations must be actionable w.r.t. the user goal/task
 - Human-in-the-loop ML scenarios are particularly concerned by this
 - Understanding of user profile and context is still essential
 - Interactivity is a desirable feature for XAI systems



Explanation flow from [Gunning2019]

4. Evaluation

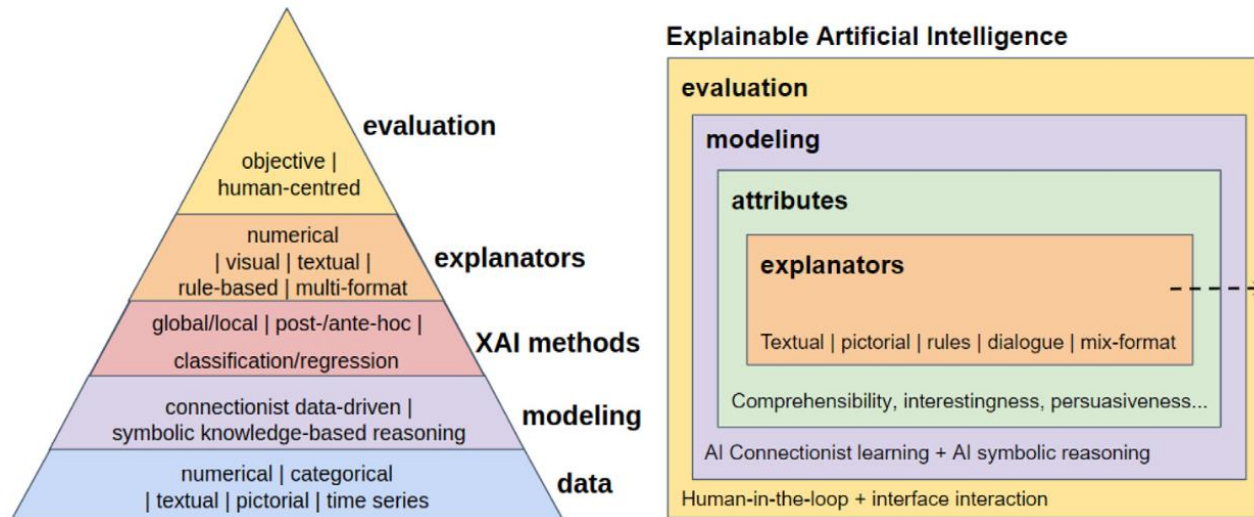
- Although mandatory, evaluating explanations involves conceptual and practical difficulties [Vilone2021]
- Main insights from the discussions
 - Importance of defining a hypothesis:
 - Contextual aspects -> Ensuing needs -> Relevant properties
 - Heuristic-based vs User-based
 - Objectivity vs subjectivity
 - Practical aspects
 - Richness and relevancy of the results
 - Need for mixed methods and robust guidelines



XAI design and evaluation framework from [Mohseni2021]

Conclusion

- Opportunities for collaboration between HCI and ML researchers are numerous!
 - To design methods and frameworks for user research and modeling
 - To better understand the model-representation-presentation path and its implications
 - To improve the user experience when interacting with XAI systems
 - To design robust evaluation methods and guidelines



Current structure of XAI research and ideal structure for [Vilone2021]

THANKS FOR YOUR ATTENTION!

If you are interested in one of those topics,
especially for recommender systems,
don't hesitate to contact me!

julien.albert@unamur.be

References

- [Adadi2018] Amina Adadi and Mohammed Berrada. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6 (2018), 52138–52160.
- [Bibal2016] Adrien Bibal and Benoît Frénay. Interpretability of Machine Learning Models and Representations: An Introduction. In *The European Symposium on Artificial Neural Networks*. 77–82.
- [Doshi-Velez2017] Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608* (2017).
- [Gunning2019] David Gunning and David W. Aha. DARPA’s Explainable Artificial Intelligence Program. *AI Magazine* 40, 2 (2019), 44–58.
- [Liao2021] Q Vera Liao and Kush R Varshney. Human-centered explainable AI (XAI): From algorithms to user experiences. *arXiv preprint arXiv:2110.10790* (2021).
- [Mohseni2021] Sina Mohseni, Niloofar Zarei, Eric D. Ragan. A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. In *ACM Transactions on Interactive Intelligent Systems* 11, iss. 3-4 (2018).
- [Ribera2019] Mireia Ribera and Àgata Lapedriza García. Can We Do Better Explanations? A Proposal of User-Centered Explainable AI. In *IUI Workshops* (2019).
- [Vilone2021] Giulia Vilone and Luca Longo. Notions of Explainability and Evaluation Approaches for Explainable Artificial Intelligence. *Information Fusion* 76 (2021), 89–106.