Transparency and Control for AI Systems in Education





Jeroen Ooge jeroenooge.be





About Me

Human-centred explainable AI, particularly in education

KU LEUVEN

PhD in computer science 2019-2023

My short career

MSc fundamental mathematics MSc applied informatics 2012-2019

Private tutor, educational author

Assistant professor in computer science 2023-now





Jeroen Ooge jeroenooge.be











Darius Afchar, Alessandro Melchiorre, Markus Schedl, Romain Hennequin, Elena Epure, and Manuel Moussallam. 2022. Explainability in Music Recommender Systems. *Al Magazine* 43, 2: 190–208. <u>https://doi.org/10.1002/aaai.12056</u>



Alejandro Barredo Arrieta et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58: 82–115. <u>https://doi.org/10.1016/j.inffus.2019.12.012</u>

Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. 2019. A Survey of Methods for Explaining Black Box Models. *ACM Computing Surveys* 51, 5: 1–42. <u>https://doi.org/10.1145/3236009</u>





Hassan Khosravi, Simon Buckingham Shum, Guanliang Chen, Cristina Conati, Yi-Shan Tsai, Judy Kay, Simon Knight, Roberto Martinez-Maldonado, Shazia Sadiq, and Dragan Gašević. 2022. Explainable Artificial Intelligence in education. Computers and Education: Artificial Intelligence 3: 100074. <u>https://doi.org/10.1016/j.caeai.2022.100074</u>





Darius Afchar, Alessandro Melchiorre, Markus Schedl, Romain Hennequin, Elena Epure, and Manuel Moussallam. 2022. Explainability in Music Recommender Systems. *Al Magazine* 43, 2: 190–208. <u>https://doi.org/10.1002/aaai.12056</u>

Human-Centred Explainability Algorithmic Explainability - - relates to AI AI SYSTEM Data-focused relates to XAI explanations Model-focused Understandability ш Target explanations Audience inference learning prediction Prediction Model Data < task representation are there target explanations what is the type of what is the scope of is the model associated with ground-truth incompleteness we try blackbox? the explanation? predictions? to overcome? informativeness global supervised local intrinsic post-hoc unsupervised fairness interpretability causality interactivity XAI AXIS ...

Darius Afchar, Alessandro Melchiorre, Markus Schedl, Romain Hennequin, Elena Epure, and Manuel Moussallam. 2022. Explainability in Music Recommender Systems. *Al Magazine* 43, 2: 190–208. <u>https://doi.org/10.1002/aaai.12056</u>

Human-Centred Explainability

Different people and contexts need different explainability solutions

Research goal: Design tailored explanations and evaluate how they affect people's behaviour (e.g., trust, understanding, motivation, learning)





Textual prompt

Exam question









... of kies zelf je volgende oefening



Competent 💡

Novice 💡

Advanced beginner 💡

Control















No explanation

Jeroen Ooge, Shotallo Kato, and Katrien Verbert. 2022. Explaining Recommendations in E-Learning: Effects on Adolescents' Trust. In 27th International Conference on Intelligent User Interfaces (IUI '22), 93–105. <u>https://doi.org/10.1145/3490099.3511140</u>



Visual explanations can increase initial trust but may not be the most important factor for building it

Jeroen Ooge, Shotallo Kato, and Katrien Verbert. 2022. Explaining Recommendations in E-Learning: Effects on Adolescents' Trust. In 27th International Conference on Intelligent User Interfaces (IUI '22), 93–105. <u>https://doi.org/10.1145/3490099.3511140</u>



How is your new level determined?

Wiski estimates your level and the difficulty of exercises. Both change when solving exercises.

Your level remained similar after solving the exercise series. Then, it increased even further because of your feedback.













Reflection on own mastery and recommendations r







How is your new level determined?

Wiski estimates your level and the difficulty of exercises. Both change when solving exercises. Your level remained similar after solving the exercise series.

Then, it increased even further because of your feedback.



Seeing the impact of control can increase initial trust, but control mechanisms by themselves do not necessarily





Jeroen Ooge, Arno Vanneste, Maxwell Szymanski, and Katrien Verbert. Under review.



What-if explanations can encourage teenagers to try harder exercises, but do not necessarily affect initial trust, metacognition, motivation, or performance

0.5

Slider values

0.6

0.7

0.8

0.9

1.0

Proportion WHAT-IF - Proportion CONTROL

0.4

0.2

0.3

0.0

0.1

Jeroen Ooge, Arno Vanneste, Maxwell Szymanski, and Katrien Verbert. Under review.







Jeroen Ooge, Joran De Braekeleer, and Katrien Verbert. 2024. Nudging Adolescents Towards Recommended Maths Exercises with Gameful Rewards. In: Artificial Intelligence in Education. AIED 2024. Lecture Notes in Computer Science, vol 14830. Springer, Cham. <u>https://doi.org/10.1007/978-3-031-64299-9_28</u>





Anissa Faik. Bringing a New Perspective to the Classroom: Detecting and Explaining Student Outliers. Master's thesis, 2023.





Maxwell Szymanski, Jeroen Ooge, Robin De Croon. Vero Vanden Abeele & Katrien Verbert. 2024. Feedback, Control, or Explanations? Supporting Teachers With Steerable Distractor-Generating AI. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*, March 2024, 690–700, <u>https://doi.org/10.1145/3636555.3636933</u>





... of kies zelf je volgende oefening



Proficient 💡

Competent 💡

Novice 💡

Advanced beginner 💡

Control





Check out **jeroenooge.be** and reach out to collaborate*!



Explain With Tail

> one who has been fully solution for surgery known it is effective, there is to unknown about why an anothesis answer (TED-E2, 2015). But figuring e mathematics behind AI is hard. Now what? analy, researchers have developed to and algorithmics techniques that give about what is happening insides black brons; (Adadi and Berrada, 2018; do Arritse et al., 2019). That al., 2019: Guddeti et al., 2019; Monravon

- * *researchers*: involve me in projects *teachers*: participate in studies, exchange ideas

Transparency and Control for AI Systems in Education





Jeroen Ooge jeroenooge.be



