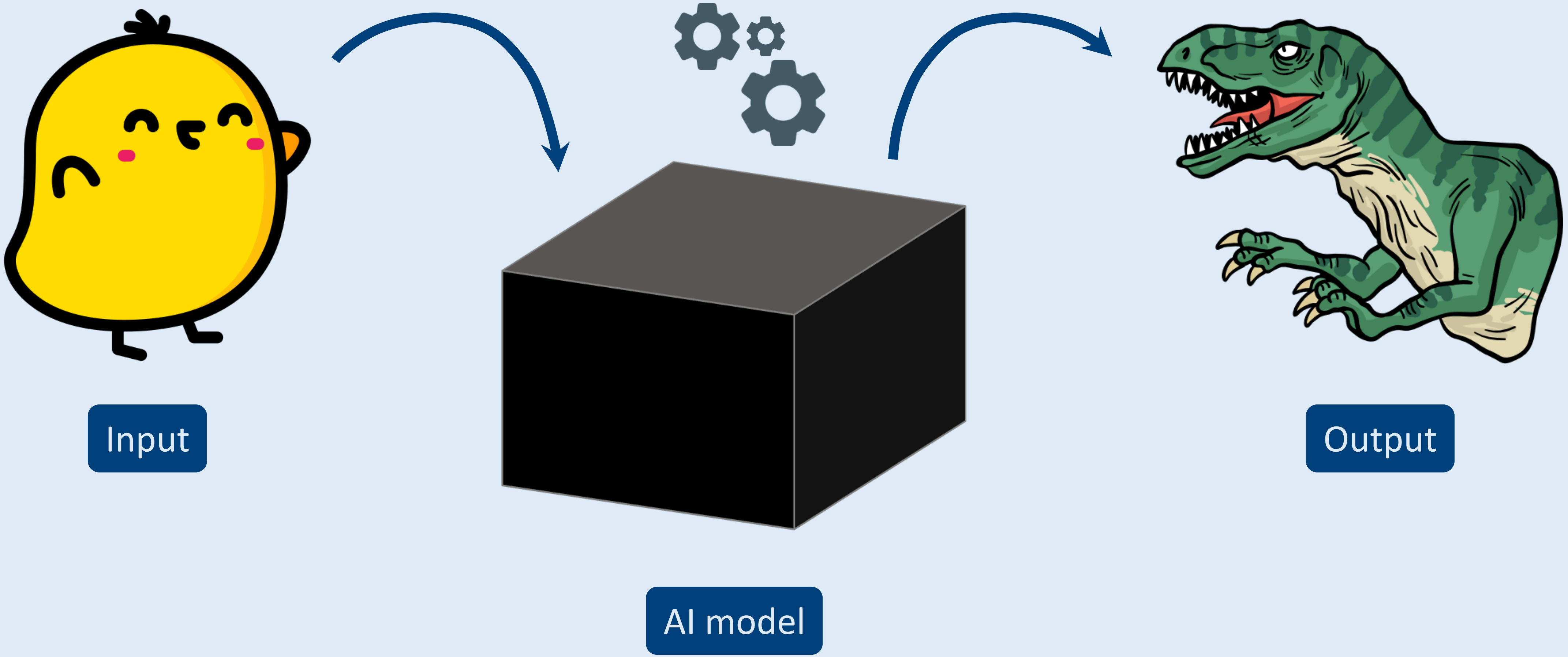
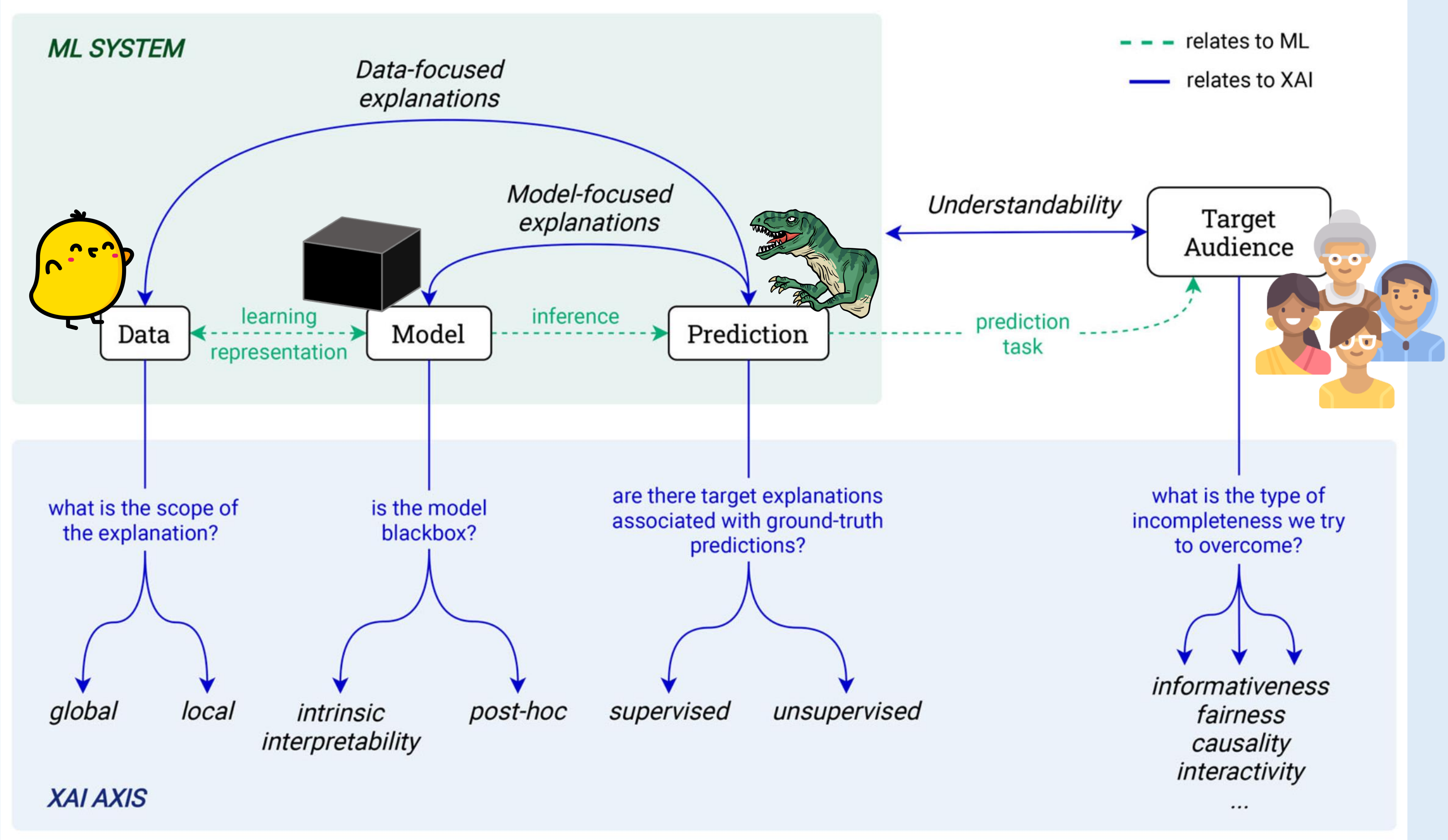
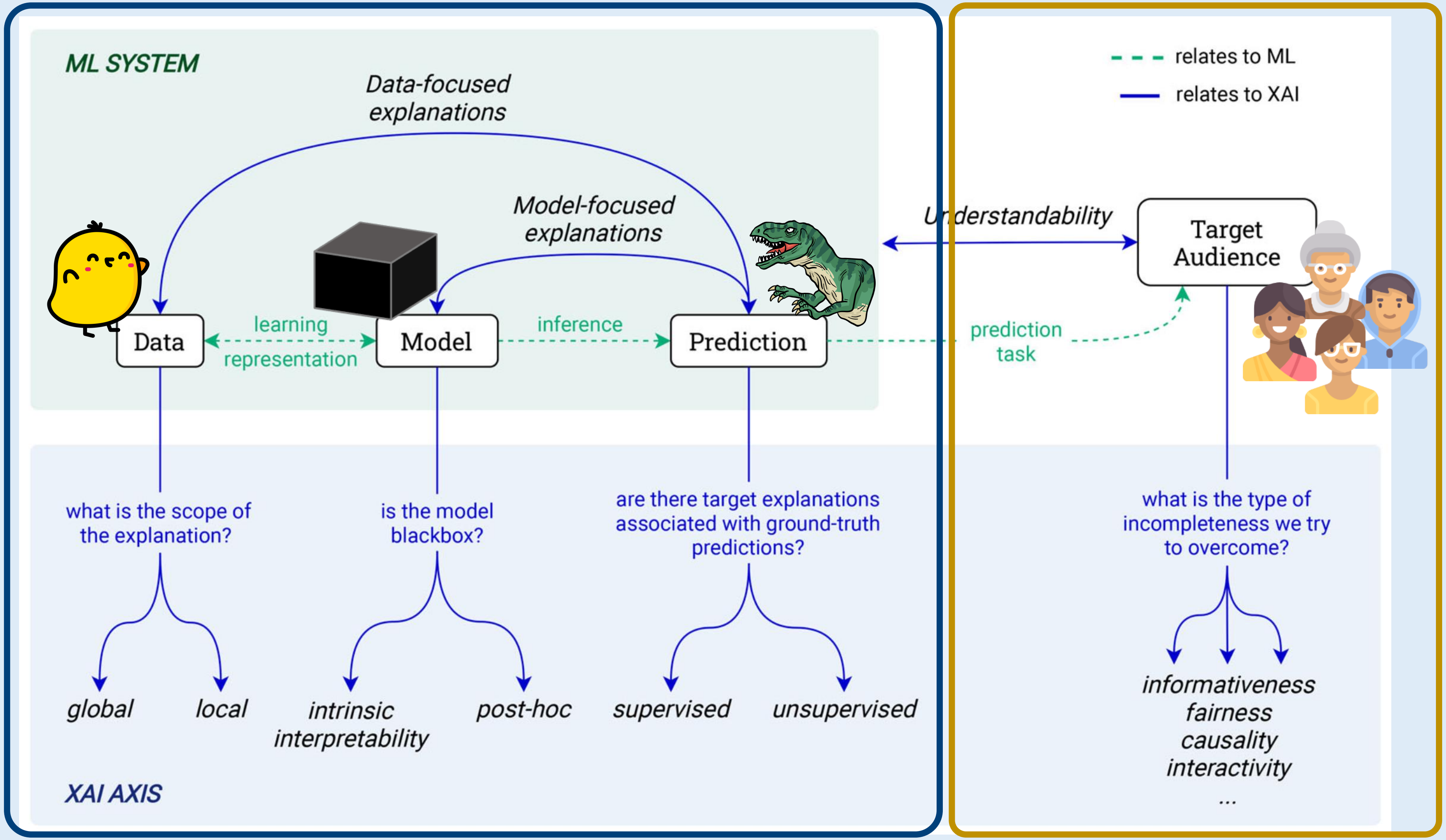
This is an AI model


Input
AI model
Output



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

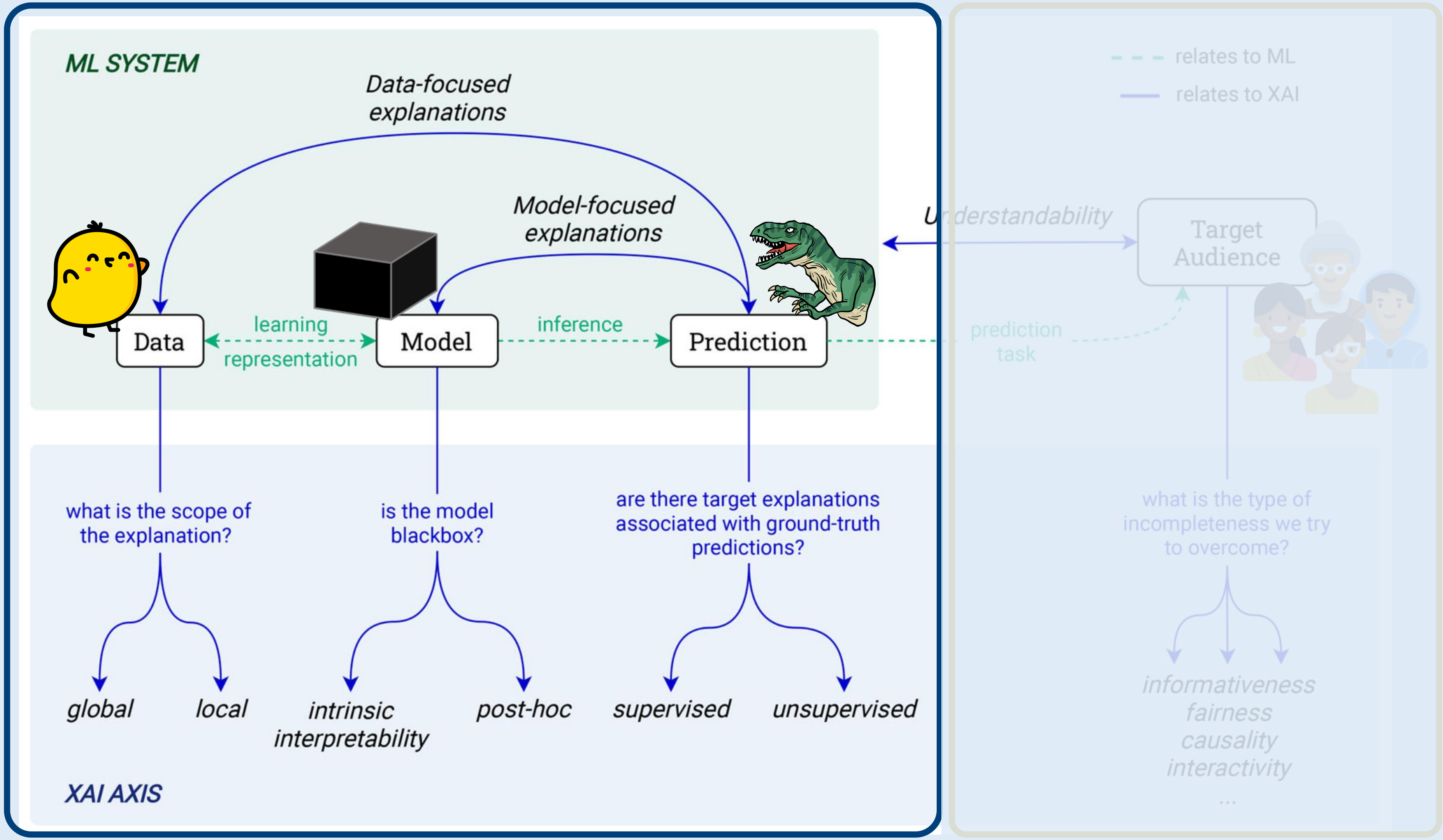
Algorithmic XAI approaches **Human-centred XAI approaches**



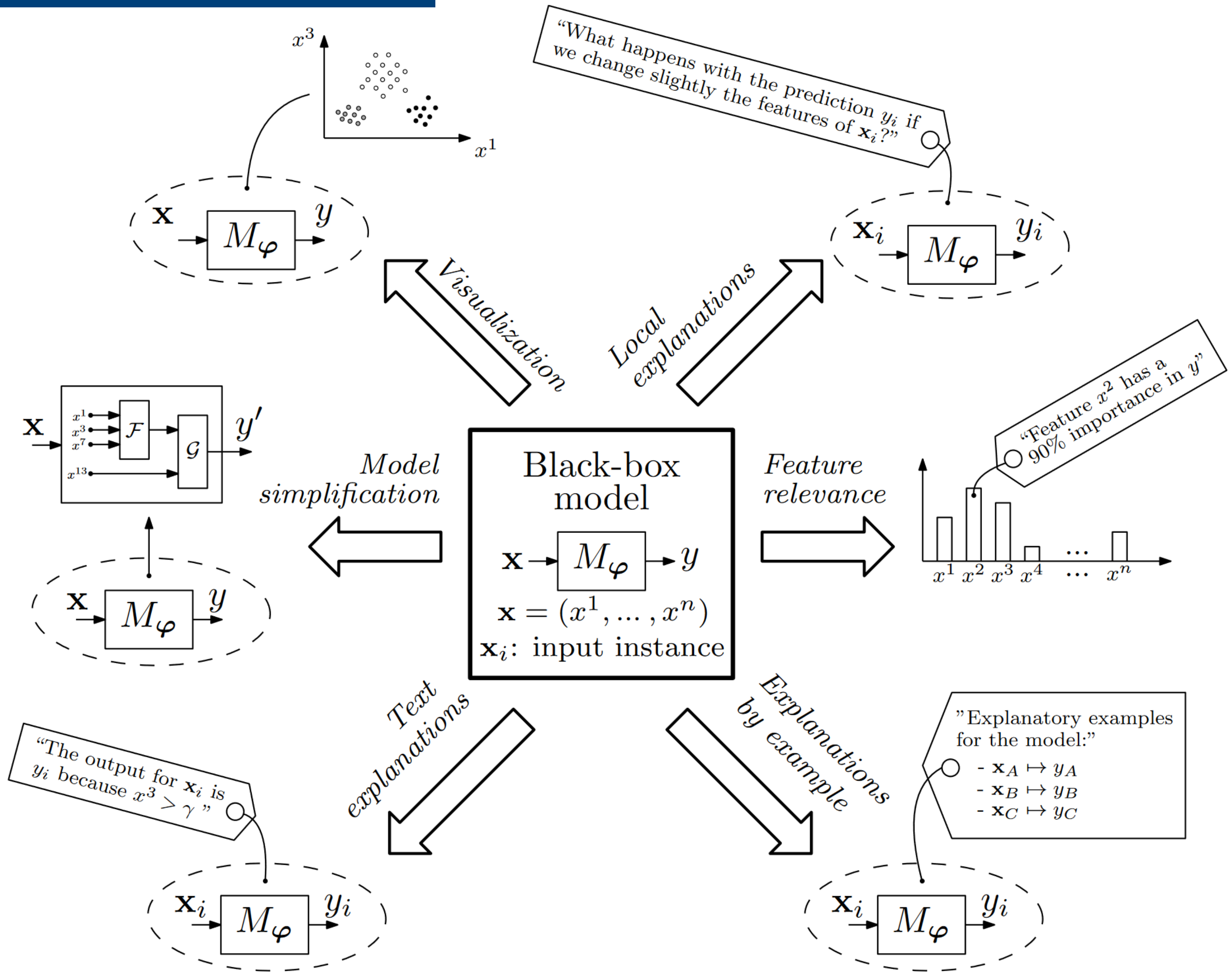
Q. Vera Liao and Kush R. Varshney. 2022. Human-Centered Explainable AI (XAI): From Algorithms to User Experiences.
https://doi.org/10.48550/arXiv.2110.10790

Algorithmic XAI approaches

Human-centred XAI approaches

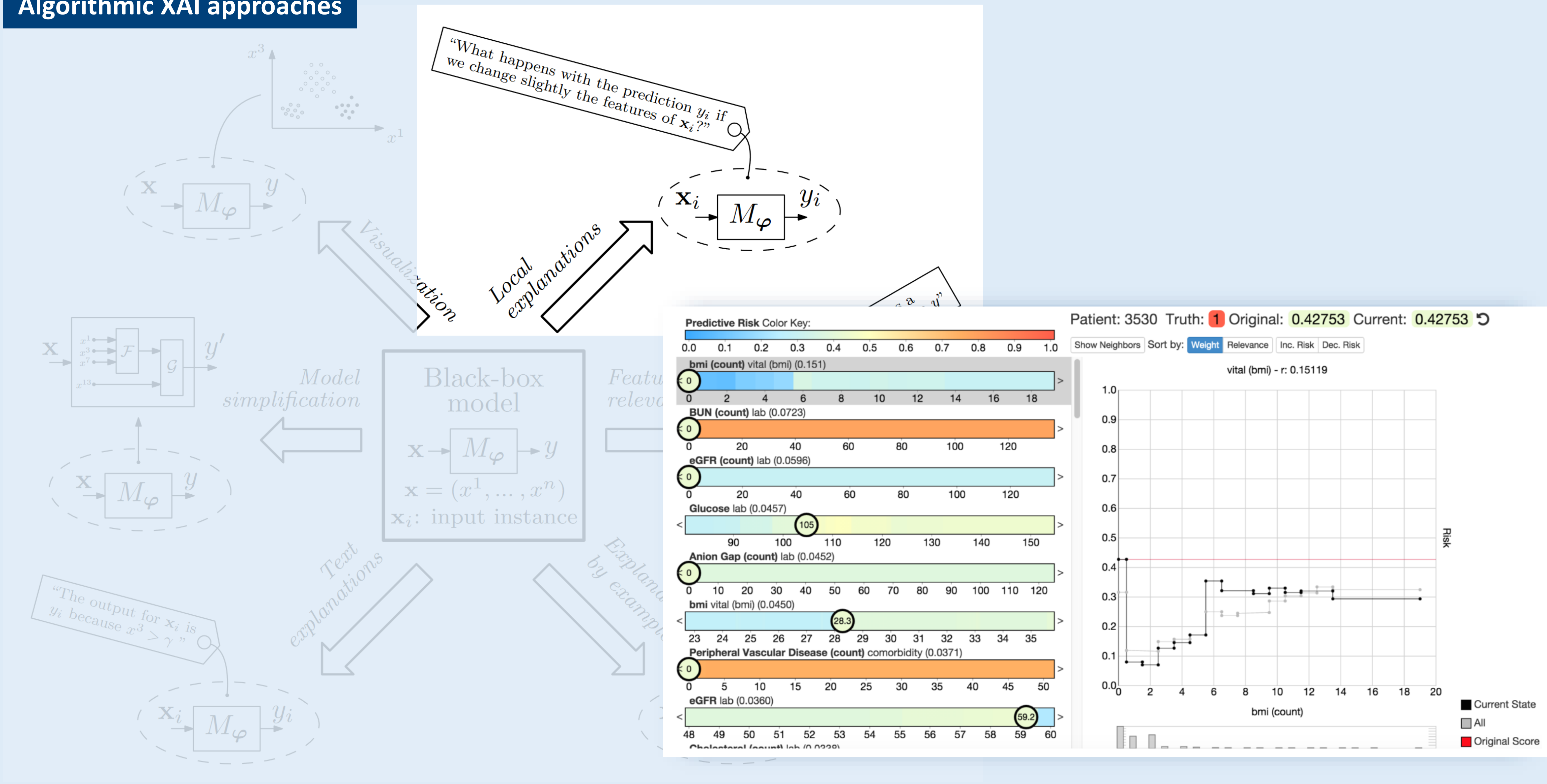


Algorithmic XAI approaches



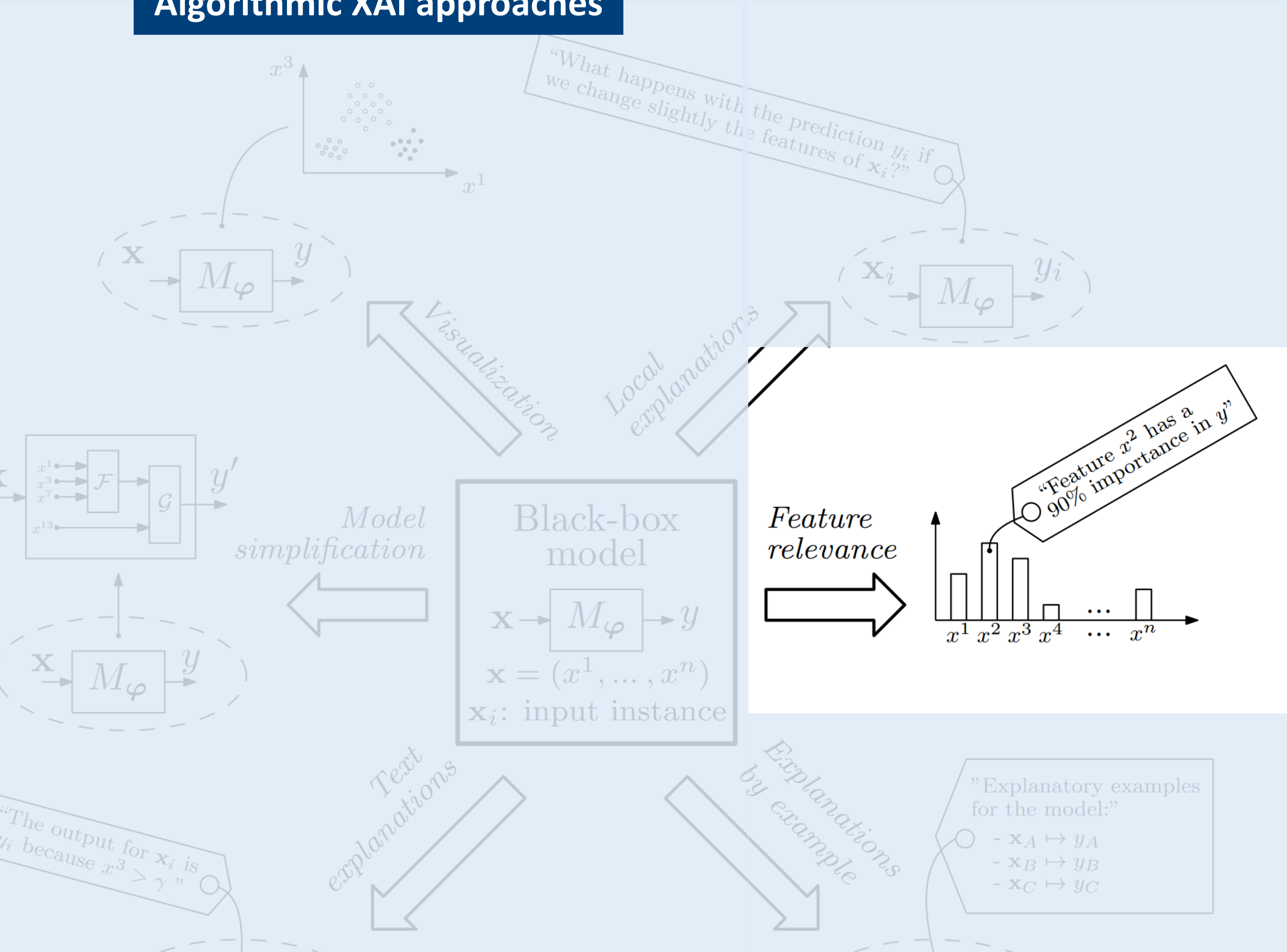
Alejandro Barredo Arrieta et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58: 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

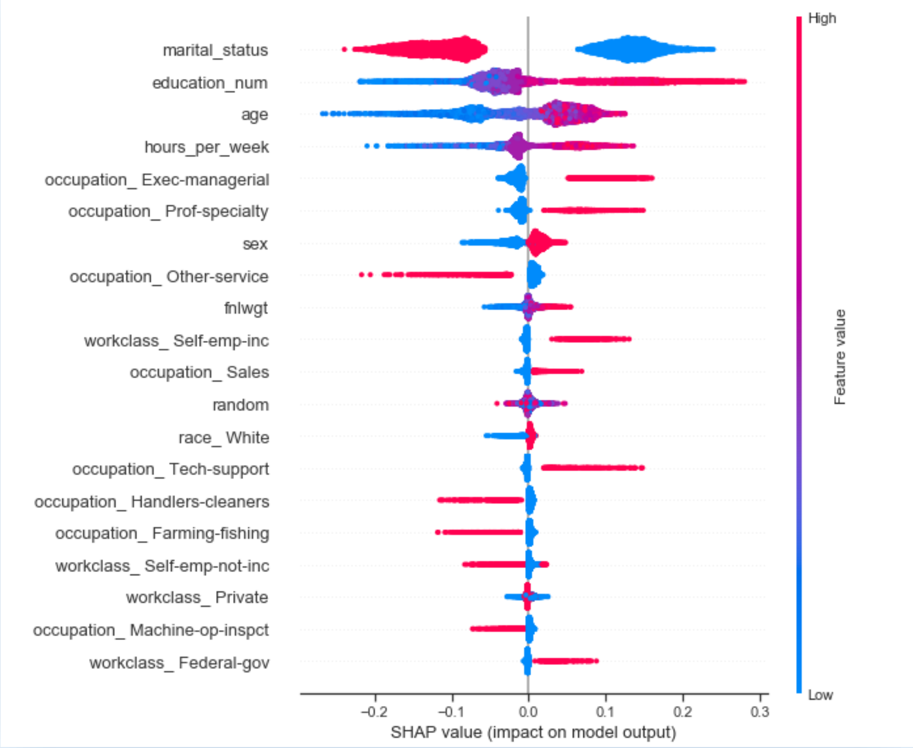
Algorithmic XAI approaches



J. Krause, A. Perer, and K. Ng. 2016. Interacting with predictions: Visual inspection of black-box machine learning models. 5686–5697.
https://doi.org/10.1145/2858036.2858529

Algorithmic XAI approaches

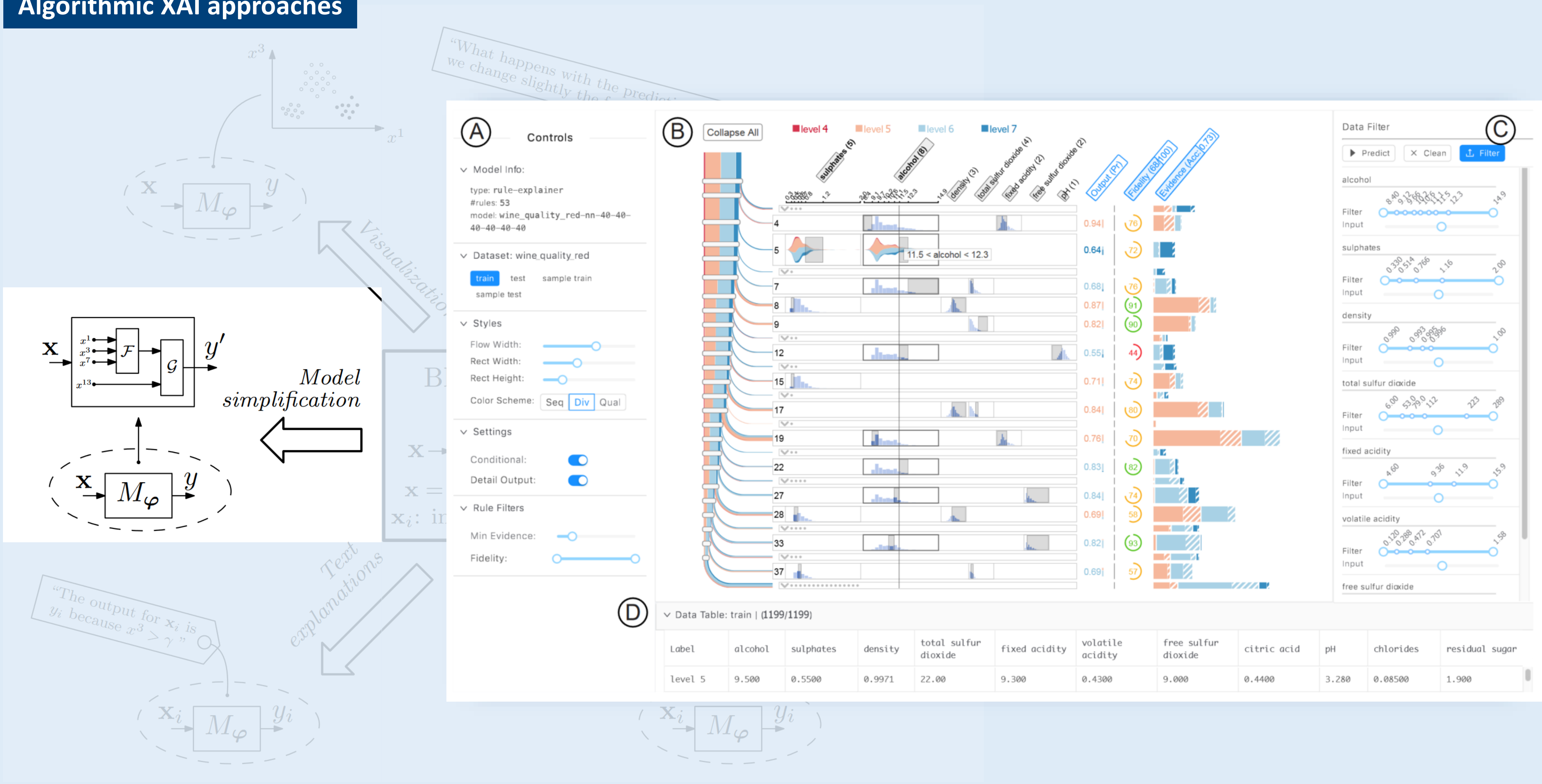




Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*. https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD '16), 1135–1144. https://doi.org/10.1145/2939672.2939778

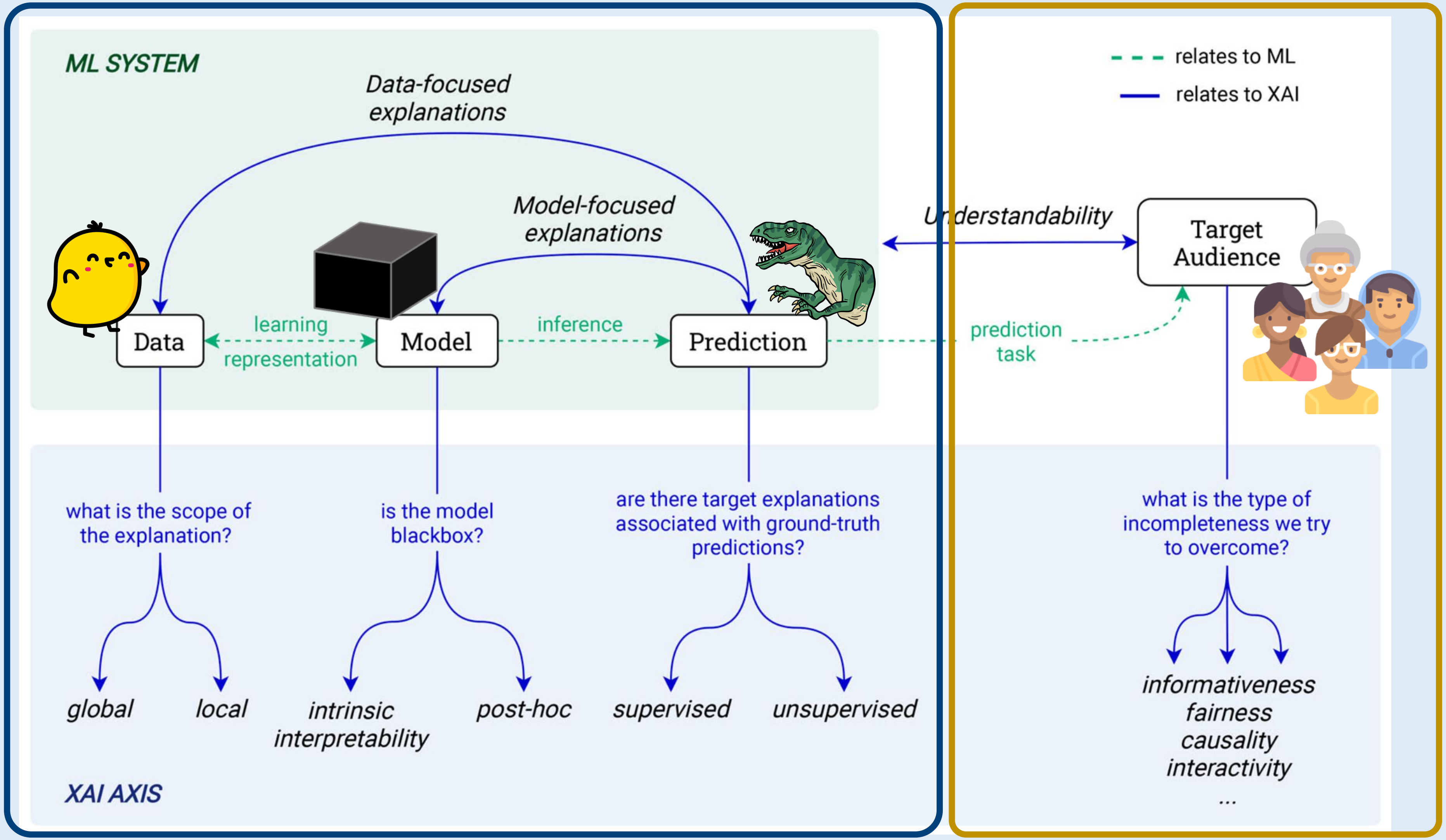
Algorithmic XAI approaches



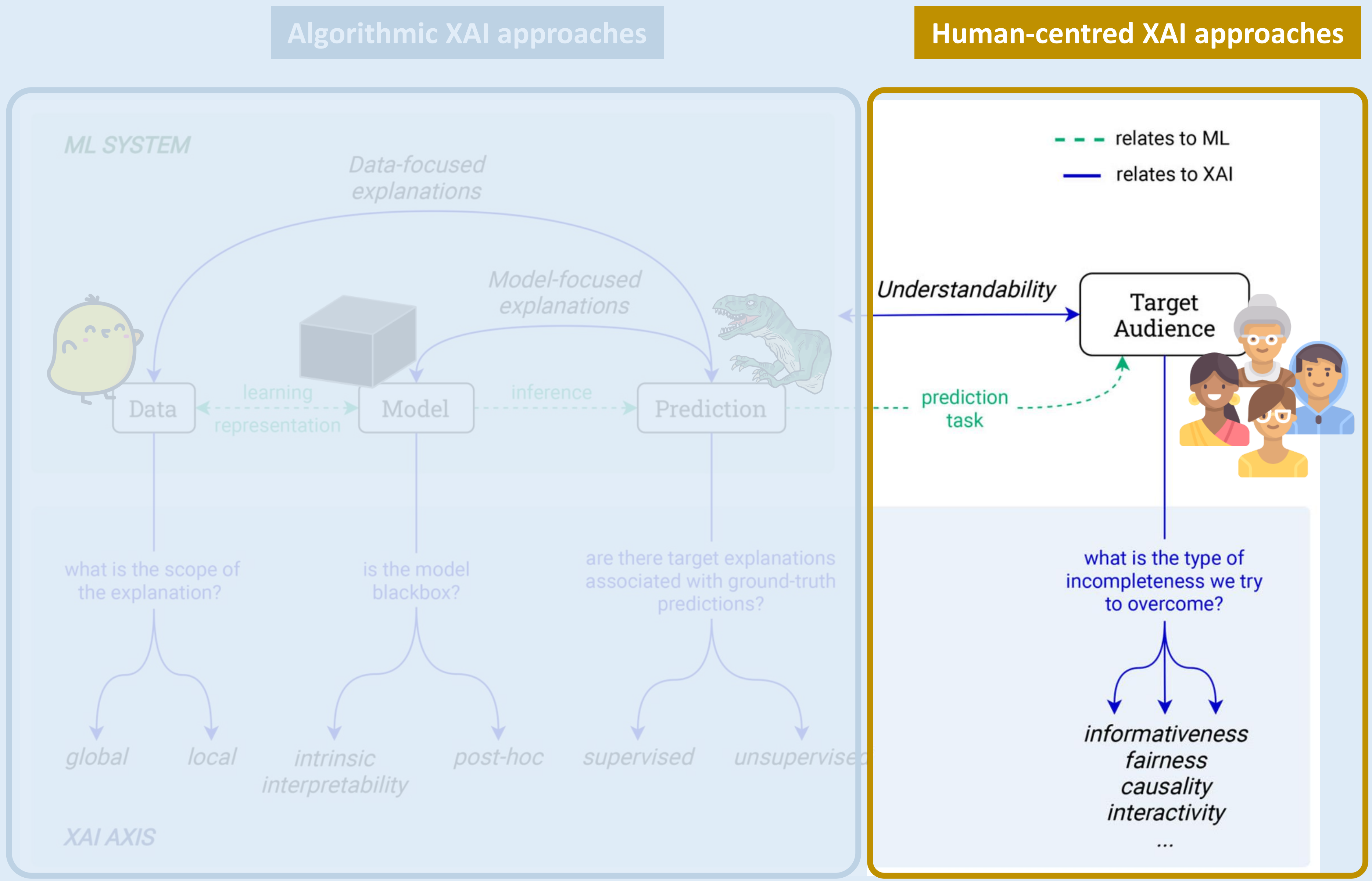
Yao Ming, Huamin Qu, and Enrico Bertini. 2019. RuleMatrix: Visualizing and Understanding Classifiers with Rules. *IEEE Transactions on Visualization and Computer Graphics* 25, 1: 342–352. https://doi.org/10.1109/TVCG.2018.2864812

Algorithmic XAI approaches

Human-centred XAI approaches

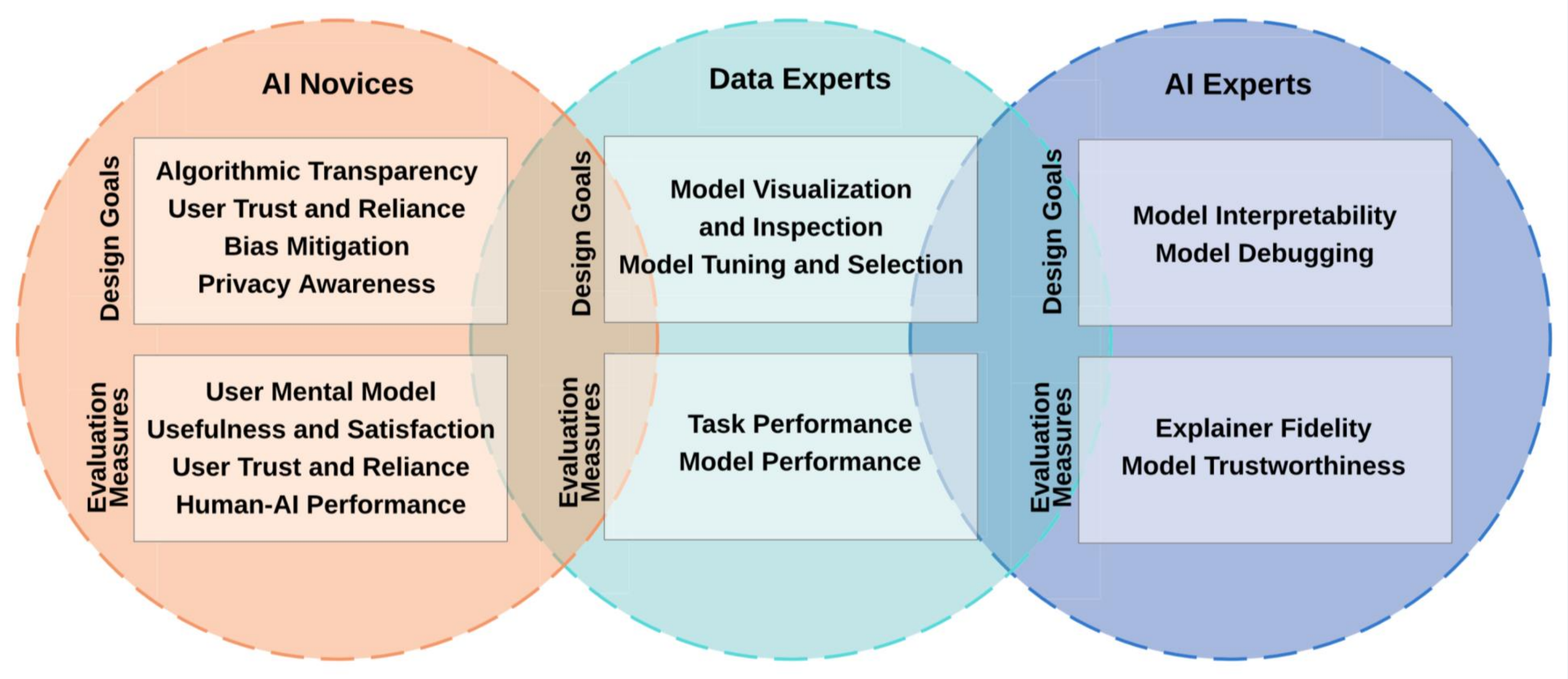


Q. Vera Liao and Kush R. Varshney. 2022. Human-Centered Explainable AI (XAI): From Algorithms to User Experiences.
https://doi.org/10.48550/arXiv.2110.10790



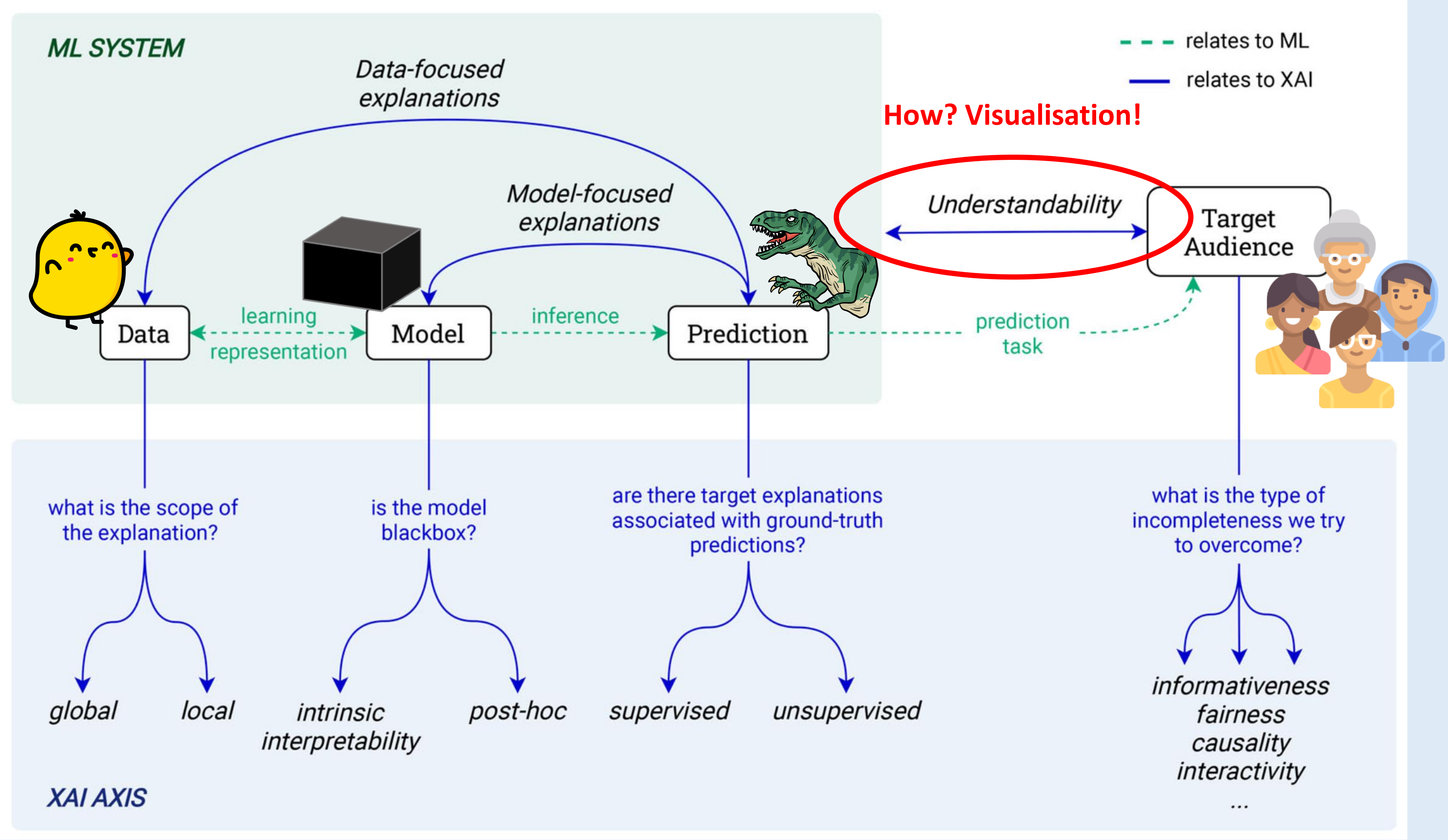
Algorithmic XAI approaches
Human-centred XAI approaches
ML SYSTEM
Data-focused explanations
Model-focused explanations
Data
learning
representation
Model
inference
Prediction
what is the scope of the explanation?
is the model blackbox?
are there target explanations associated with ground-truth predictions?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

Human-centred XAI approaches



"XAI presents as much of a **design challenge** as an algorithmic challenge" (Q. Vera Liao and Kush R. Varshney, 2022)

Sina Mohseni, Niloofar Zarei, and Eric D. Ragan. 2021. A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. *ACM Transactions on Interactive Intelligent Systems* 11, 3–4: 24:1-24:45. https://doi.org/10.1145/3387166



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

Explaining AI

with tailored interactive visualisations

Jeroen Ooge

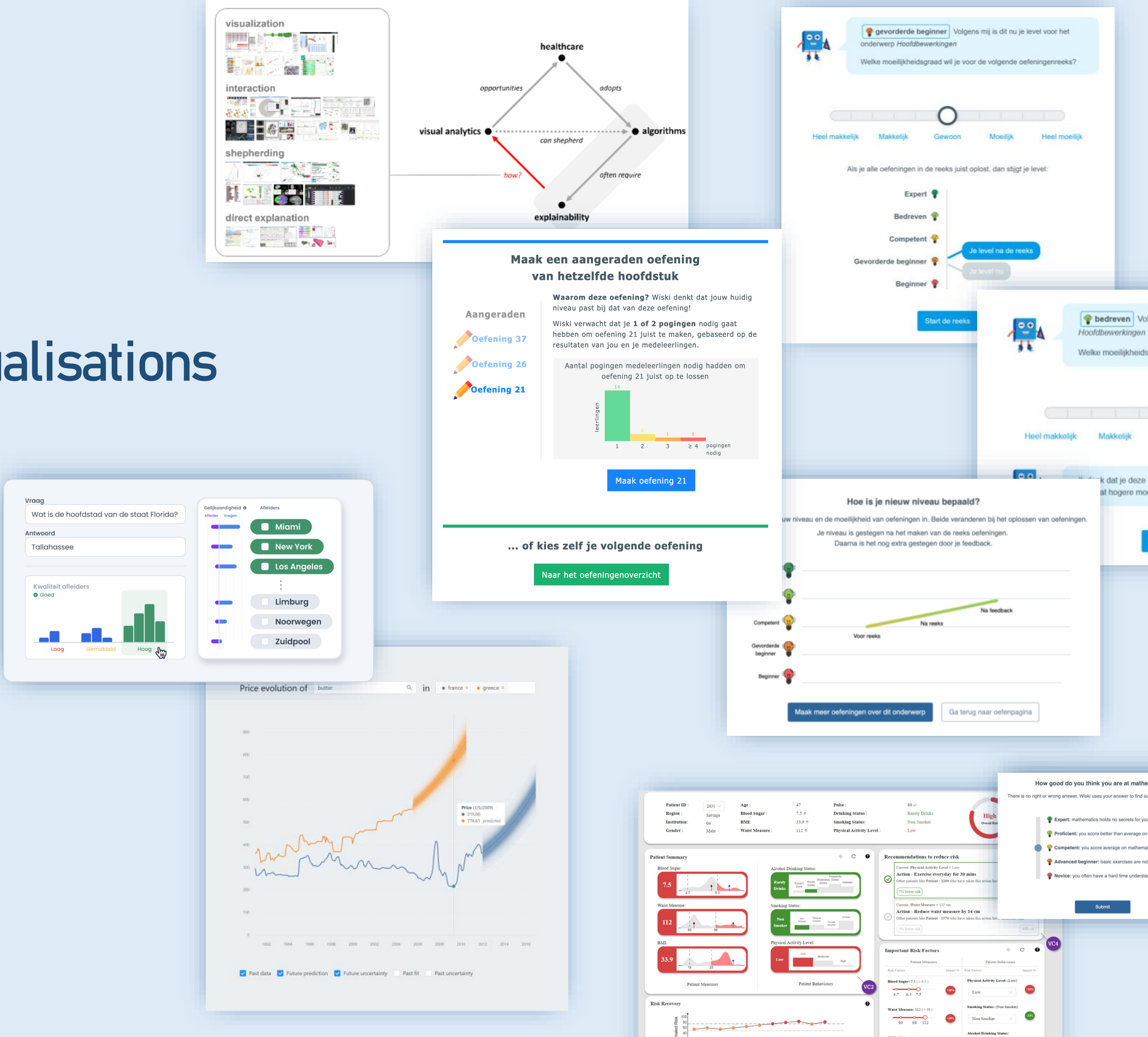
jeroenooge.com

KU LEUVEN

Katrien Verbert

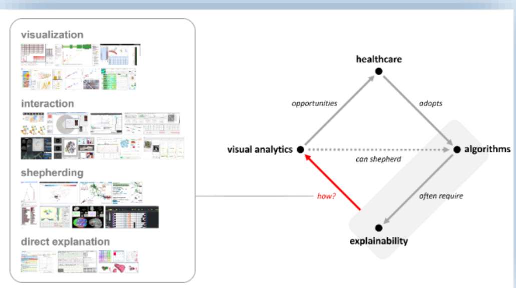
augment.cs.kuleuven.be

KU LEUVEN

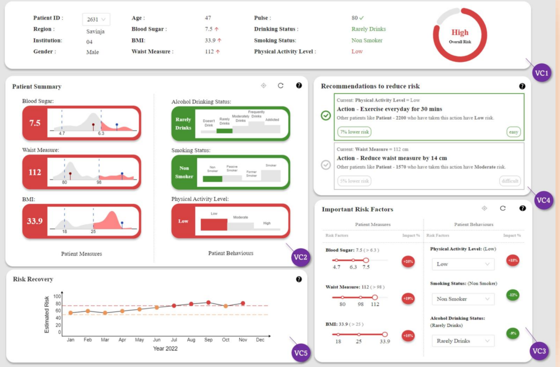


Explainable AI through visualisation

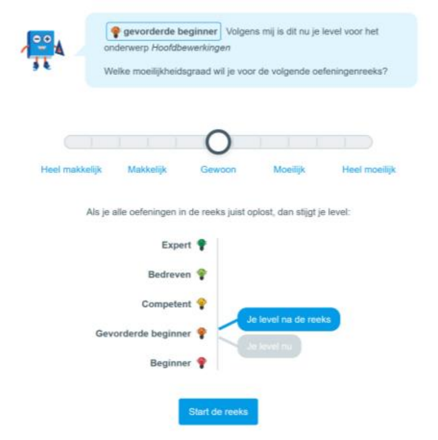
Visual analytics

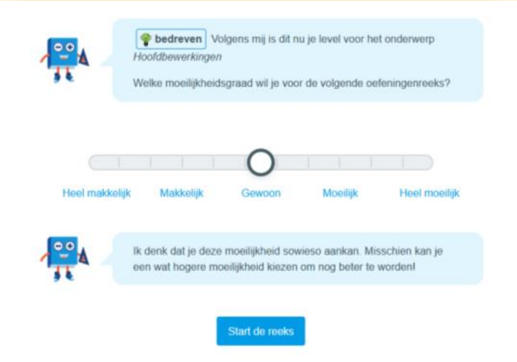


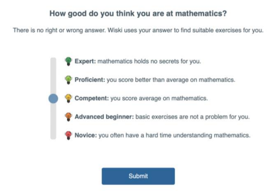
Transparency: justification

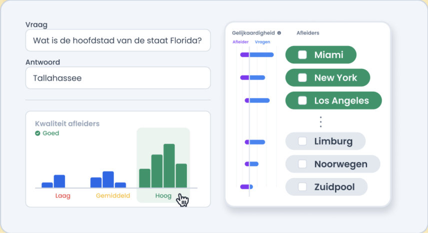


Transparency: control



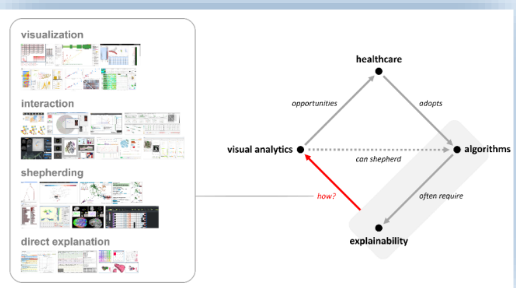




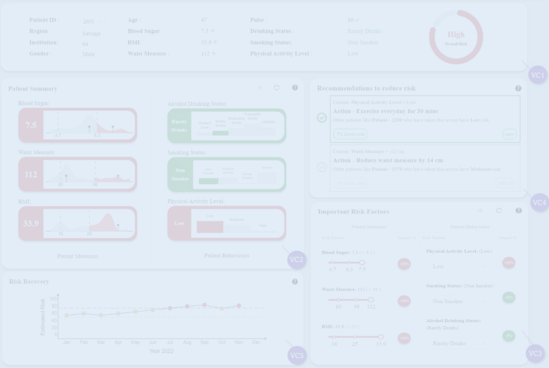


Explainable AI through visualisation

Visual analytics

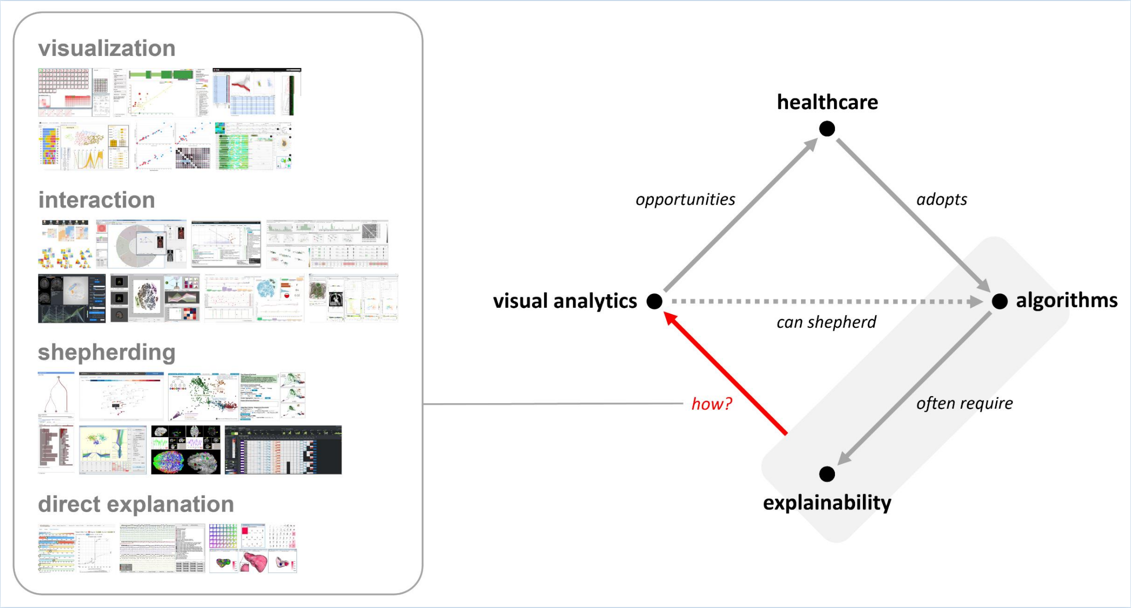


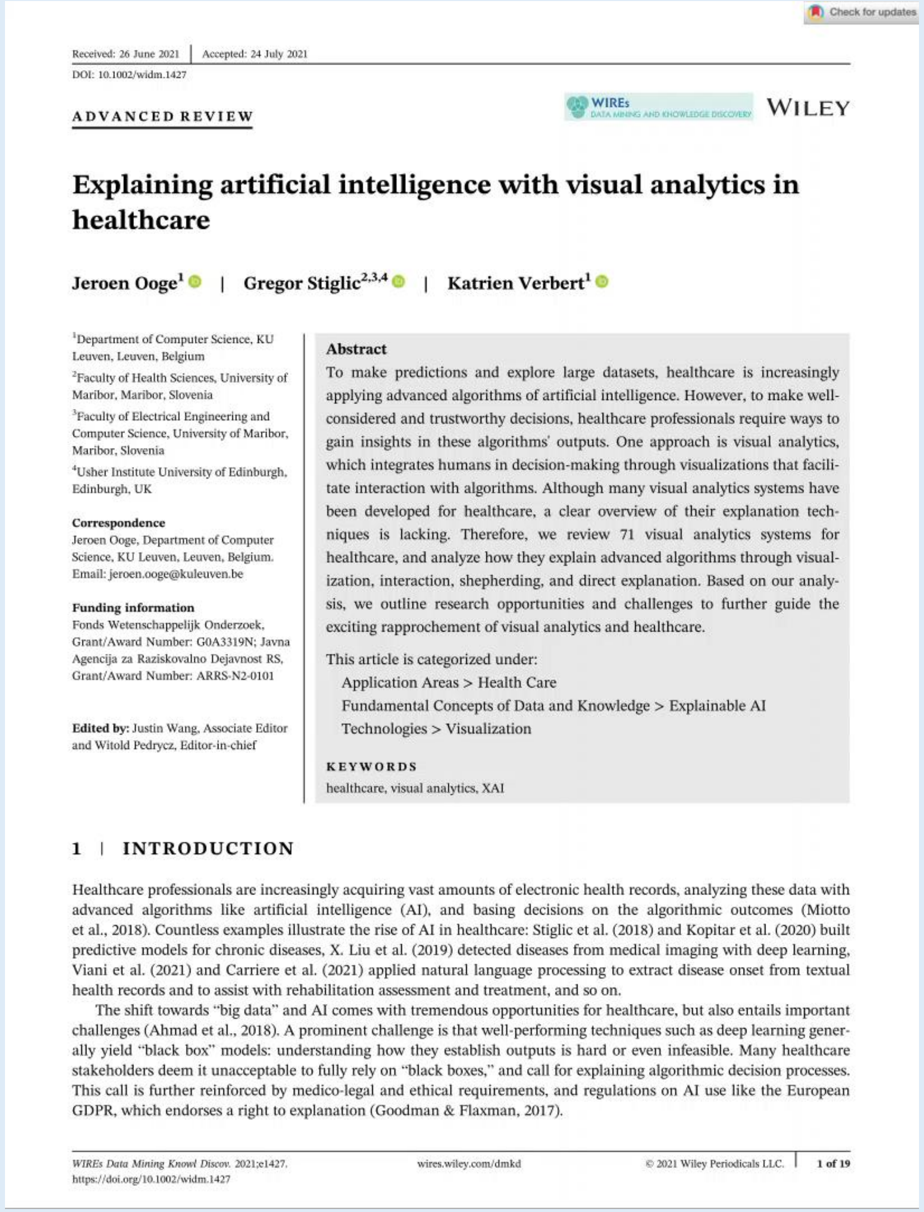
Transparency: justification



Transparency: control

Visual analytics systems targeting laypeople, supporting shepherding, or containing direct explanations are rare. (WIREs 2021)





Check for updates

Received: 26 June 2021 | Accepted: 24 July 2021

DOI: 10.1002/widm.1427

ADVANCED REVIEW

WIREs DATA MINING AND KNOWLEDGE DISCOVERY WILEY

Explaining artificial intelligence with visual analytics in healthcare

Jeroen Ooge[1] | Gregor Stiglic[2,3,4] | Katrien Verbert[1]

[1]Department of Computer Science, KU Leuven, Leuven, Belgium

[2]Faculty of Health Sciences, University of Maribor, Maribor, Slovenia

[3]Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia

[4]Usher Institute University of Edinburgh, Edinburgh, UK

Correspondence
Jeroen Ooge, Department of Computer Science, KU Leuven, Leuven, Belgium.
Email: jeroen.ooge@kuleuven.be

Funding information
Fonds Wetenschappelijk Onderzoek, Grant/Award Number: G0A3319N; Javna Agencija za Raziskovalno Dejavnost RS, Grant/Award Number: ARRS-N2-0101

Edited by: Justin Wang, Associate Editor and Witold Pedrycz, Editor-in-chief

Abstract

To make predictions and explore large datasets, healthcare is increasingly applying advanced algorithms of artificial intelligence. However, to make well-considered and trustworthy decisions, healthcare professionals require ways to gain insights in these algorithms' outputs. One approach is visual analytics, which integrates humans in decision-making through visualizations that facilitate interaction with algorithms. Although many visual analytics systems have been developed for healthcare, a clear overview of their explanation techniques is lacking. Therefore, we review 71 visual analytics systems for healthcare, and analyze how they explain advanced algorithms through visualization, interaction, shepherding, and direct explanation. Based on our analysis, we outline research opportunities and challenges to further guide the exciting rapprochement of visual analytics and healthcare.

This article is categorized under:
Application Areas > Health Care
Fundamental Concepts of Data and Knowledge > Explainable AI
Technologies > Visualization

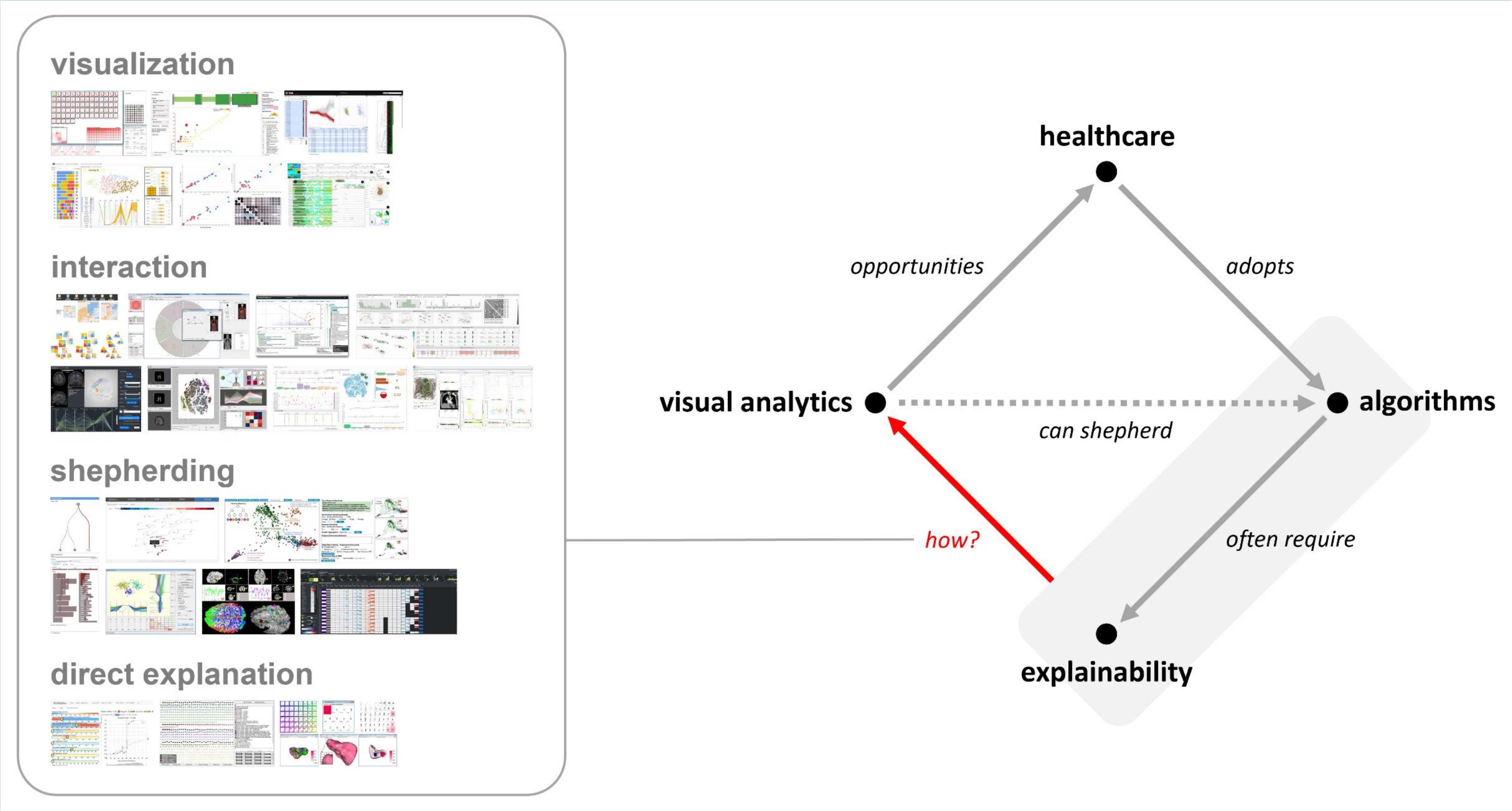
KEYWORDS
healthcare, visual analytics, XAI

1 | INTRODUCTION

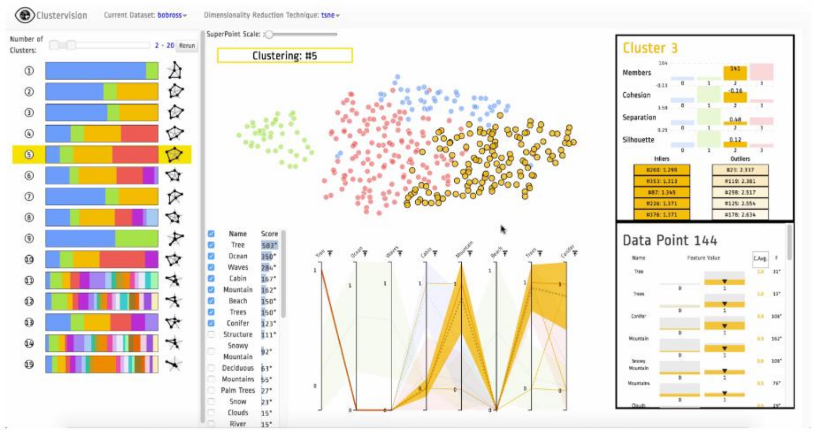
Healthcare professionals are increasingly acquiring vast amounts of electronic health records, analyzing these data with advanced algorithms like artificial intelligence (AI), and basing decisions on the algorithmic outcomes (Miotto et al., 2018). Countless examples illustrate the rise of AI in healthcare: Stiglic et al. (2018) and Kopitar et al. (2020) built predictive models for chronic diseases, X. Liu et al. (2019) detected diseases from medical imaging with deep learning, Viani et al. (2021) and Carriere et al. (2021) applied natural language processing to extract disease onset from textual health records and to assist with rehabilitation assessment and treatment, and so on.

The shift towards "big data" and AI comes with tremendous opportunities for healthcare, but also entails important challenges (Ahmad et al., 2018). A prominent challenge is that well-performing techniques such as deep learning generally yield "black box" models: understanding how they establish outputs is hard or even infeasible. Many healthcare stakeholders deem it unacceptable to fully rely on "black boxes," and call for explaining algorithmic decision processes. This call is further reinforced by medico-legal and ethical requirements, and regulations on AI use like the European GDPR, which endorses a right to explanation (Goodman & Flaxman, 2017).

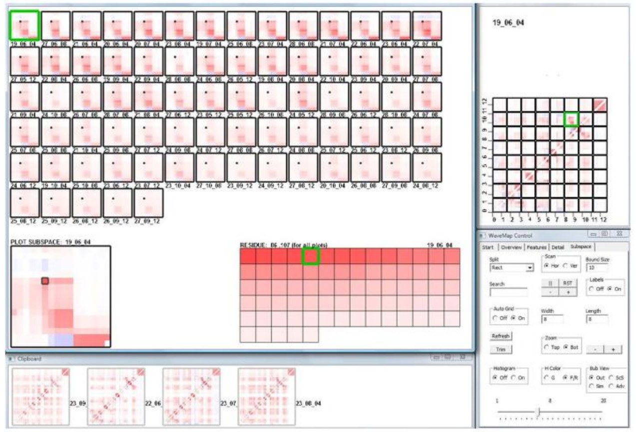
Visual analytics systems targeting laypeople, supporting shepherding, or containing direct explanations are rare. (WIREs 2021)



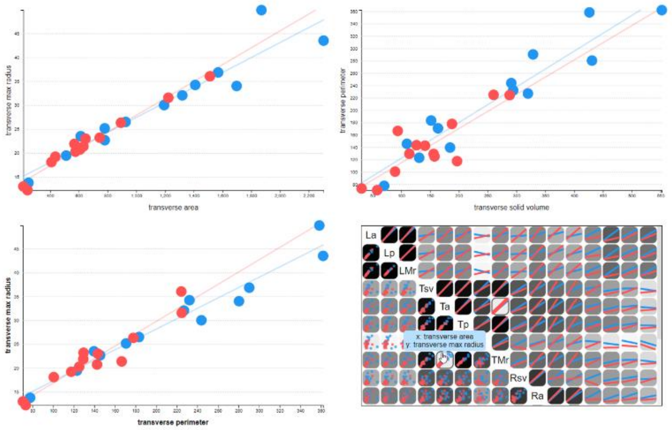
Visualisation



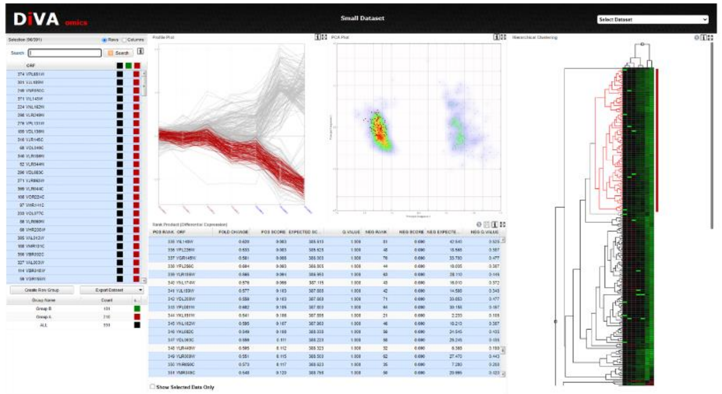
(a) Clustering



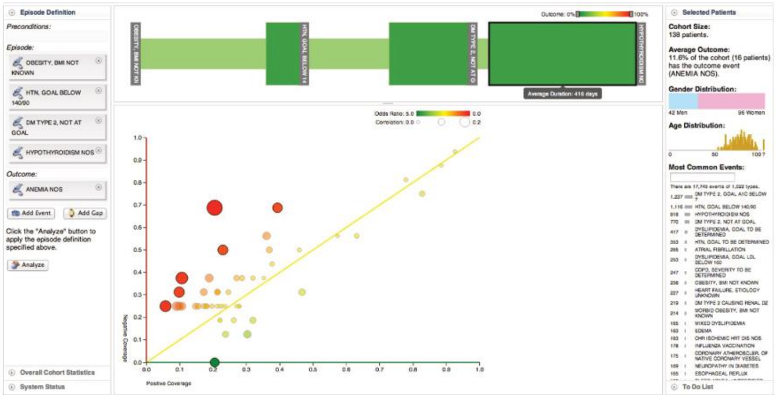
(b) Similarity



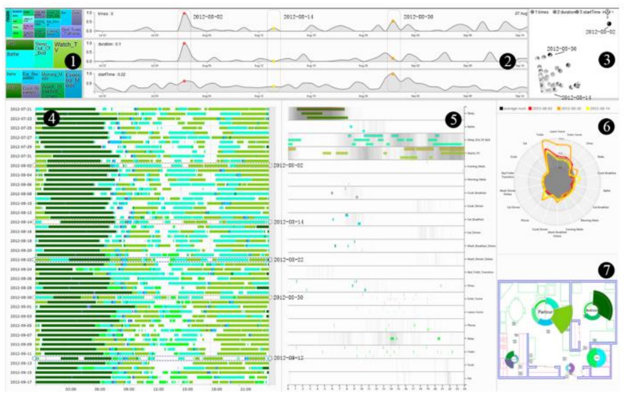
(c) Classical statistics



(d) Dimension reduction

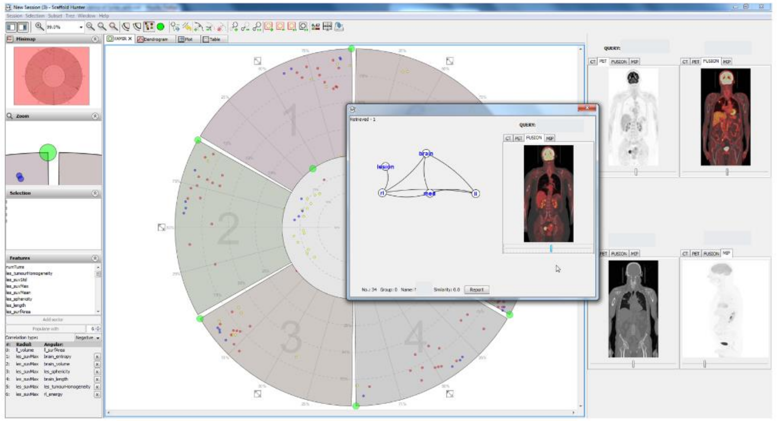


(e) Data mining

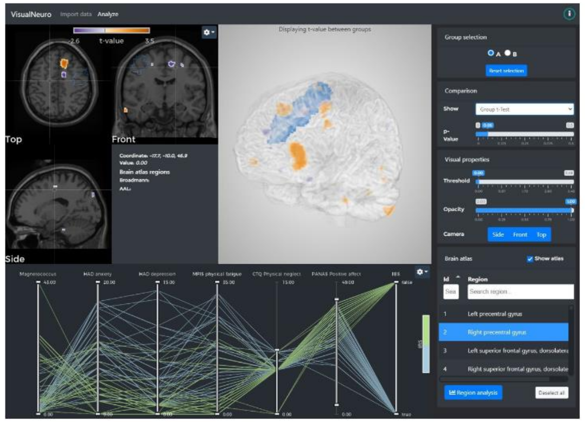


(f) Anomaly detection

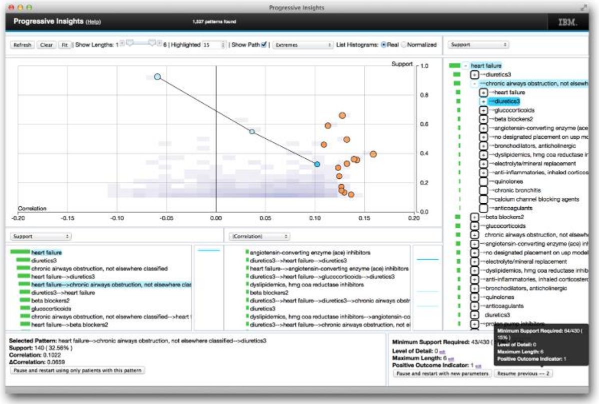
Interaction



(a) Abstract/elaborate



(b) Filter

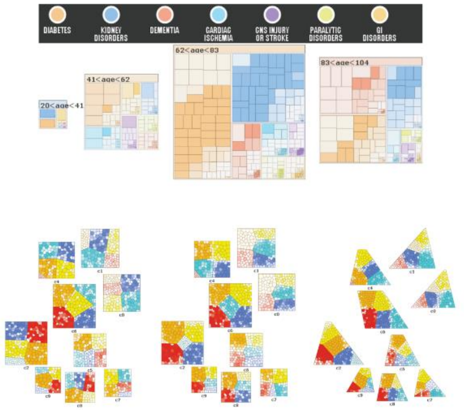


(c) Reconfigure

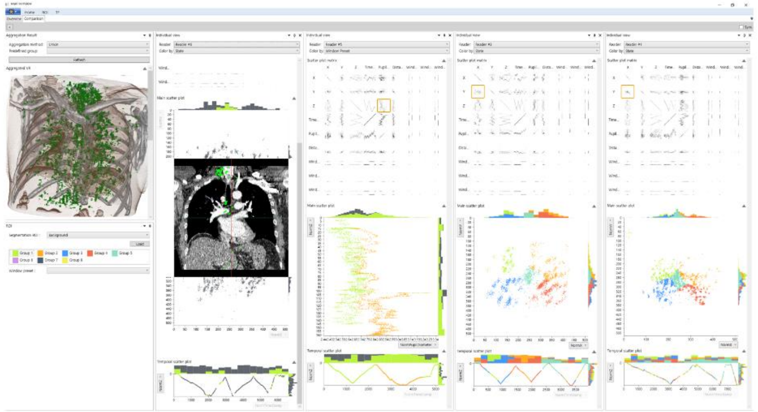
(d) Encode

(e) Select

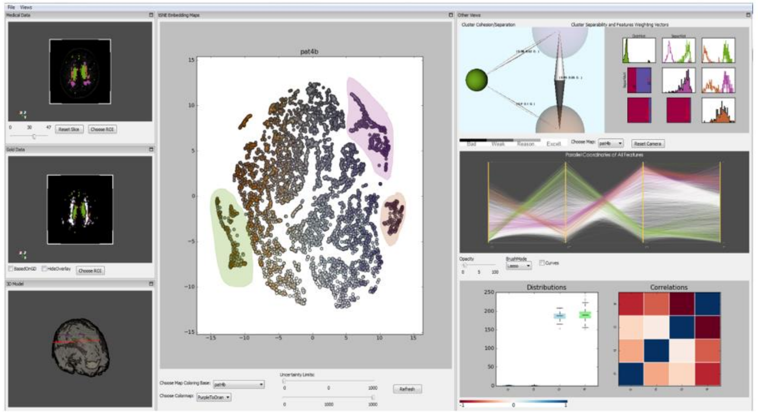
Interaction



(a) Connect



(b) Explore



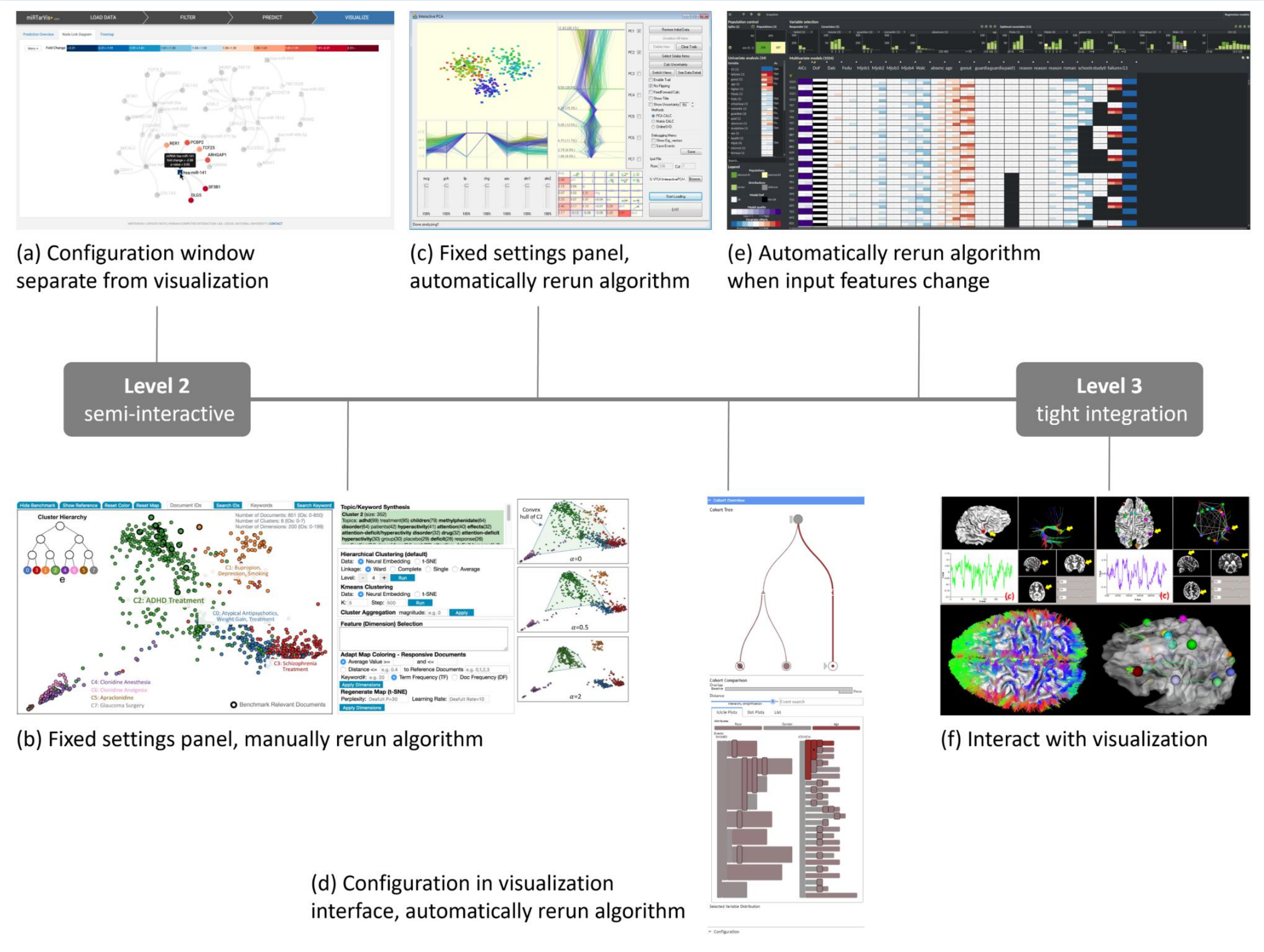
(c) Select + Connect

(a) Abstract/

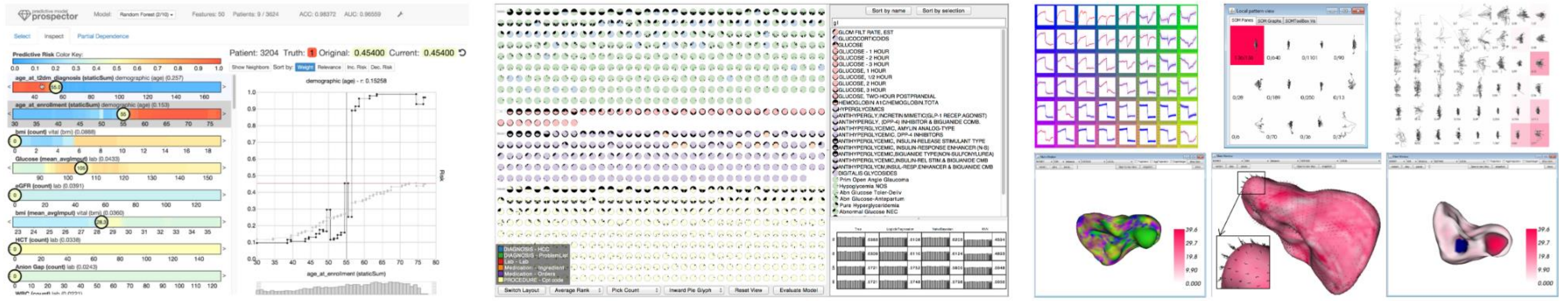
(d) Encode

(e) Select

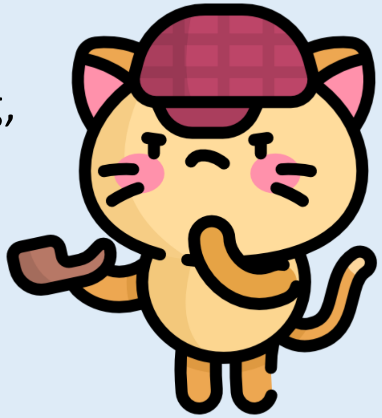
Shepherding



Direct explanation



(a) Partial dependence plot (b) Feature importance (c) Visualization



Visual analytics systems targeting **laypeople**, supporting shepherding, or containing direct explanations are rare! (WIREs 2021)

Experts react differently to an unknown prediction model;
6 evolving themes affect their trust in the model. (TREX 2021)

Trust in Prediction Models: a Mixed-Methods Pilot Study on the Impact of Domain Expertise

Jeroen Ooge* Katrien Verbert†

Department of Computer Science
KU Leuven

Abstract

People's trust in prediction models can be affected by many factors, including domain expertise like knowledge about the application domain and experience with predictive modelling. However, to what extent and why domain expertise impacts people's trust is not entirely clear. In addition, accurately measuring people's trust remains challenging. We share our results and experiences of an exploratory pilot study in which four people experienced with predictive modelling systematically explore a visual analytics system with an unknown prediction model. Through a mixed-methods approach involving Likert-type questions and a semi-structured interview, we investigate how people's trust evolves during their exploration, and we distil six themes that affect their trust in the prediction model. Our results underline the multi-faceted nature of trust, and suggest that domain expertise alone cannot fully predict people's trust perceptions.

1 Introduction

Intelligent systems like visual analytics systems are increasingly incorporating artificial intelligence to support end-users in decision-making [9, 17, 22]. To make well-informed decisions, it is vital that people appropriately trust the underlying models [12, 14]. Therefore, lots of research has been dedicated to trust in human-computer interaction [e.g., 4, 8, 29, 32], specifically in information visualisation [3] and explainable artificial intelligence [e.g., 2, 30, 33].

However, trust is a slippery concept because it is related to many factors [15]. One example is domain expertise, which can refer to artificial intelligence or the application domain in question. Previous studies have shown that both facets can influence people's trust in an automated system [2, 15, 28]. Other factors that might affect trust include the way in which information is visualised [25], age [21], uncertainty [31], cognitive load [35], model accuracy [34], algorithmic transparency [20], and the point in time on which the intelligent system is used [18, 24, 26, 27, 28]. As a consequence of this long list of influential factors, measuring trust is very challenging. Researchers have proposed Likert-scales that capture people's trust in an automated system [e.g., 10, 11, 19, 23], often inspired by the psychological literature on trust relations between humans [16]. However, there is still debate about these scales' validity, and even about whether explainable artificial intelligence should focus on trust in the first place [6].

In this paper, we share our results and experiences of a mixed-methods pilot study with four participants who are familiar with predictive modelling, and active in agrifood domains. Our research contribution is threefold:

1. To measure people's expertise and trust in a prediction model, we propose a mixed-methods approach that goes beyond using single Likert-type questions, yet remains feasible in real-life studies;
2. We illustrate that only knowing people's expertise in predictive modelling does not suffice to predict their trust in a prediction model;
3. By thematically analysing the transcripts of semi-structured interviews, we extract six factors that might influence people's trust in a prediction model.

*e-mail: jeroen.ooge@kuleuven.be
†e-mail: katrien.verbert@kuleuven.be

2 Research Motivation

Research on trust in intelligent systems often subdivides people into those who are familiar with a certain topic ("experts"), and those who are not ("non-experts" or "laypeople") [e.g., 28, 35]. The research goal is then to find differences between, and similarities within those groups. We were interested in the latter, particularly in whether people experienced with predictive modelling have similar trust perceptions when they explore a visual analytics system without knowing the underlying prediction model. Inspired by studies on trust evolution over time [18, 24, 26, 27, 28], we decided to show people increasingly more visual information about a prediction outcome, and to capture their trust evolution. Specifically, our research questions were as follows:

- **RQ1**. Do people experienced with predictive modelling have similar trust levels and evolutions for an unknown prediction model?
- **RQ2**. What influences trust in an unknown prediction model for people experienced with predictive modelling?

To make fair comparisons, we needed participants with similar backgrounds. We chose to target people in agrifood because research on trust and uncertainty visualisation is limited in this domain [13].

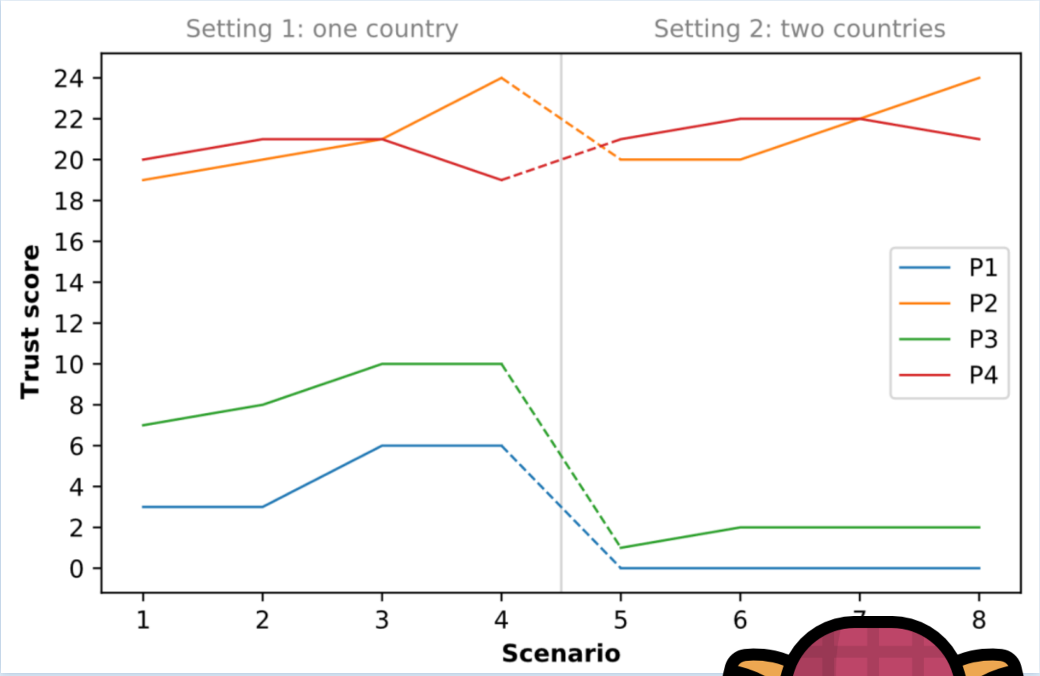
3 Materials and Methods

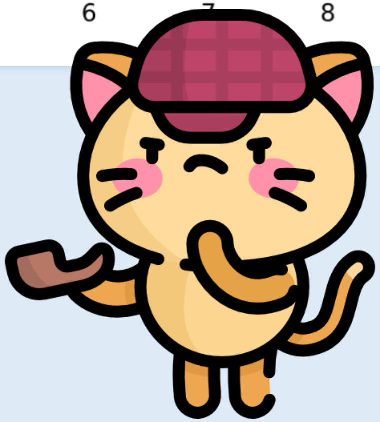
This section presents how we conducted our study. We first describe our visual analytics system and overall study design. Then, we provide more details on how we measured expertise and trust.

3.1 Visual Analytics System

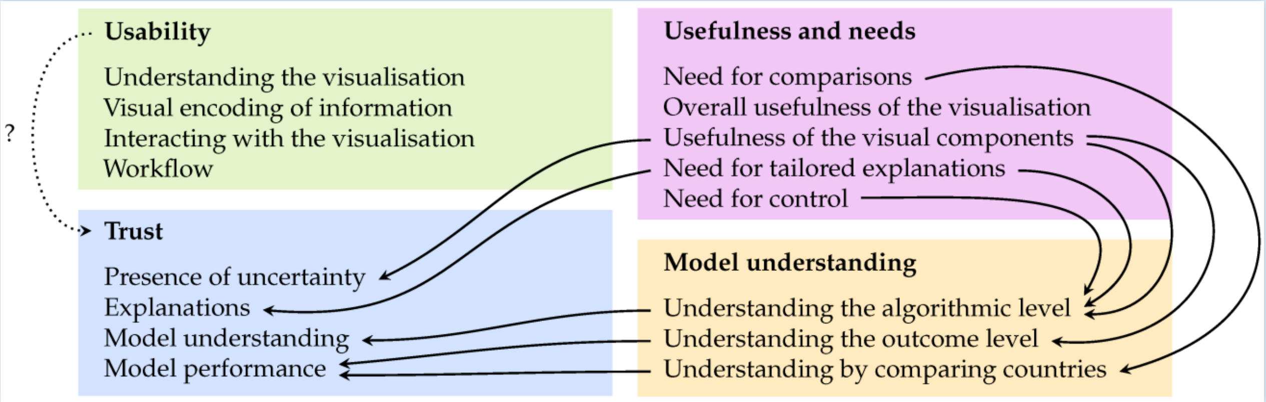
We developed a simple visual analytics system for exploring the price evolution of food products in various European countries. For each country, we fit a third-degree polynomial to the available past data with linear regression and least-squares estimation, used a five-year extrapolation as prediction, and computed prediction intervals at levels 50 to 99 with increments of five. Obviously, more sophisticated techniques for forecasting time series exist; we used linear regression only for illustration purposes. Fig. 1 shows our dashboard with at the bottom five checkboxes that enable visual components related to the prediction outcome and model: *Past data*, *Future prediction*, *Future uncertainty*, *Past fit*, and *Past uncertainty*. The past data were visualised as a full line, future prediction and past fit as dashed lines, and the uncertainty as stacked coloured bands (also called fans). Our system was built with Meteor, React and D3, and is available at `https://github.com/BigDataGrapes-EU/product-prices-public`.

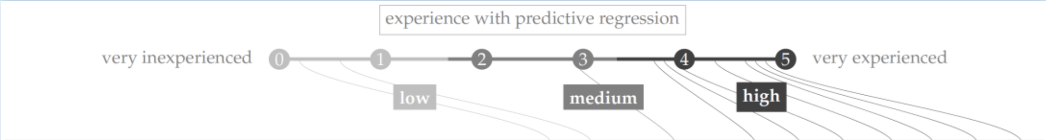
Experts react differently to an unknown prediction model;
6 evolving themes affect their trust in the model. (TREX 2021)



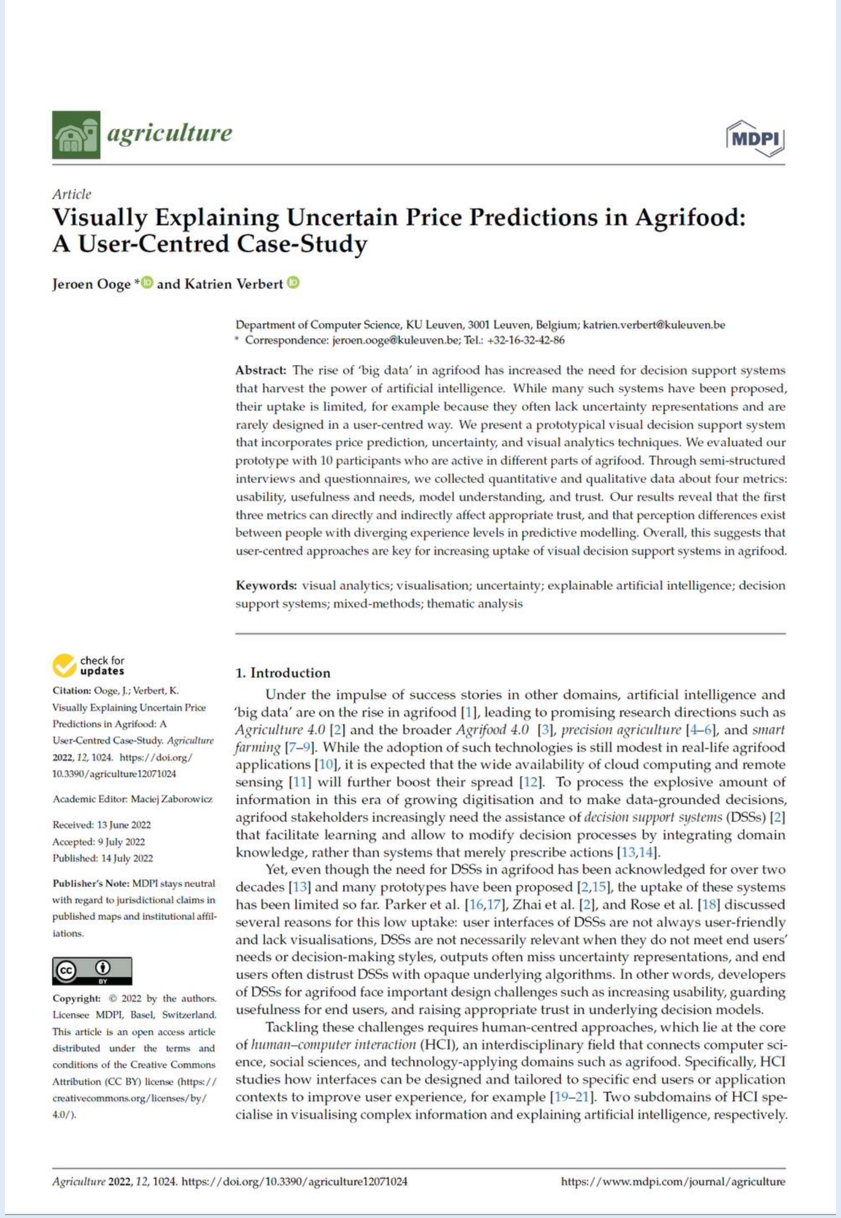


Usability, usefulness and needs, and model understanding affect appropriate trust. User-centred approaches are key for uptake of visual DSSs. (Agriculture 2022)





	P7	P2	P3	P8	P6	P5	P10	P9	P1	P4
Understanding the visualisation										
Past fit and uncertainty are not understood	•	•								
Need for control										
More control over the prediction model						•			•	•
Need for comparisons										
Comparing countries is relevant	•			•				•	•	
Comparing products is relevant			•			•	•	•	•	
Comparing prediction models is relevant				•					•	
Need for tailored explanations										
Explaining the past data									•	•
Explaining the model's development process		•			•	•			•	•
Explaining the prediction model		•	•	•		•	•		•	•
Understanding the algorithmic level										
Visual components gradually improve mental model			•	•	•	•	•	•	•	•



agriculture MDPI

Article

Visually Explaining Uncertain Price Predictions in Agrifood: A User-Centred Case-Study

Jeroen Ooge * and Katrien Verbert

Department of Computer Science, KU Leuven, 3001 Leuven, Belgium; katrien.verbert@kuleuven.be
* Correspondence: jeroen.ooge@kuleuven.be; Tel.: +32-16-32-42-86

Abstract: The rise of 'big data' in agrifood has increased the need for decision support systems that harvest the power of artificial intelligence. While many such systems have been proposed, their uptake is limited, for example because they often lack uncertainty representations and are rarely designed in a user-centred way. We present a prototypical visual decision support system that incorporates price prediction, uncertainty, and visual analytics techniques. We evaluated our prototype with 10 participants who are active in different parts of agrifood. Through semi-structured interviews and questionnaires, we collected quantitative and qualitative data about four metrics: usability, usefulness and needs, model understanding, and trust. Our results reveal that the first three metrics can directly and indirectly affect appropriate trust, and that perception differences exist between people with diverging experience levels in predictive modelling. Overall, this suggests that user-centred approaches are key for increasing uptake of visual decision support systems in agrifood.

Keywords: visual analytics; visualisation; uncertainty; explainable artificial intelligence; decision support systems; mixed-methods; thematic analysis

check for updates

Citation: Ooge, J.; Verbert, K. Visually Explaining Uncertain Price Predictions in Agrifood: A User-Centred Case-Study. *Agriculture* 2022, *12*, 1024. https://doi.org/10.3390/agriculture12071024

Academic Editor: Maciej Zaborowicz

Received: 13 June 2022
Accepted: 9 July 2022
Published: 14 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Under the impulse of success stories in other domains, artificial intelligence and 'big data' are on the rise in agrifood [1], leading to promising research directions such as *Agriculture 4.0* [2] and the broader *Agrifood 4.0* [3], *precision agriculture* [4–6], and *smart farming* [7–9]. While the adoption of such technologies is still modest in real-life agrifood applications [10], it is expected that the wide availability of cloud computing and remote sensing [11] will further boost their spread [12]. To process the explosive amount of information in this era of growing digitisation and to make data-grounded decisions, agrifood stakeholders increasingly need the assistance of *decision support systems* (DSSs) [2] that facilitate learning and allow to modify decision processes by integrating domain knowledge, rather than systems that merely prescribe actions [13,14].

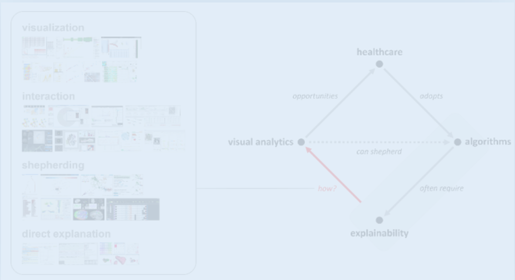
Yet, even though the need for DSSs in agrifood has been acknowledged for over two decades [13] and many prototypes have been proposed [2,15], the uptake of these systems has been limited so far. Parker et al. [16,17], Zhai et al. [2], and Rose et al. [18] discussed several reasons for this low uptake: user interfaces of DSSs are not always user-friendly and lack visualisations, DSSs are not necessarily relevant when they do not meet end users' needs or decision-making styles, outputs often miss uncertainty representations, and end users often distrust DSSs with opaque underlying algorithms. In other words, developers of DSSs for agrifood face important design challenges such as increasing usability, guarding usefulness for end users, and raising appropriate trust in underlying decision models.

Tackling these challenges requires human-centred approaches, which lie at the core of *human–computer interaction* (HCI), an interdisciplinary field that connects computer science, social sciences, and technology-applying domains such as agrifood. Specifically, HCI studies how interfaces can be designed and tailored to specific end users or application contexts to improve user experience, for example [19–21]. Two subdomains of HCI specialise in visualising complex information and explaining artificial intelligence, respectively.

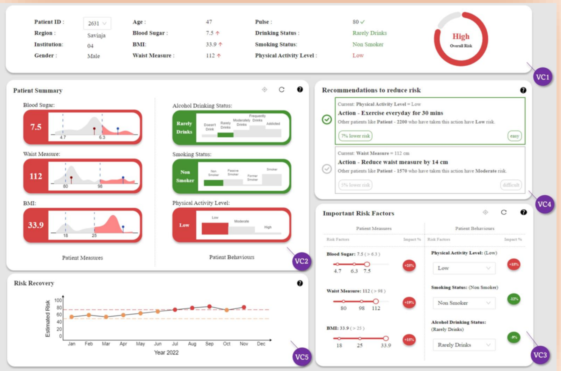
Agriculture 2022, *12*, 1024. https://doi.org/10.3390/agriculture12071024 https://www.mdpi.com/journal/agriculture

Explainable AI through visualisation

Visual analytics

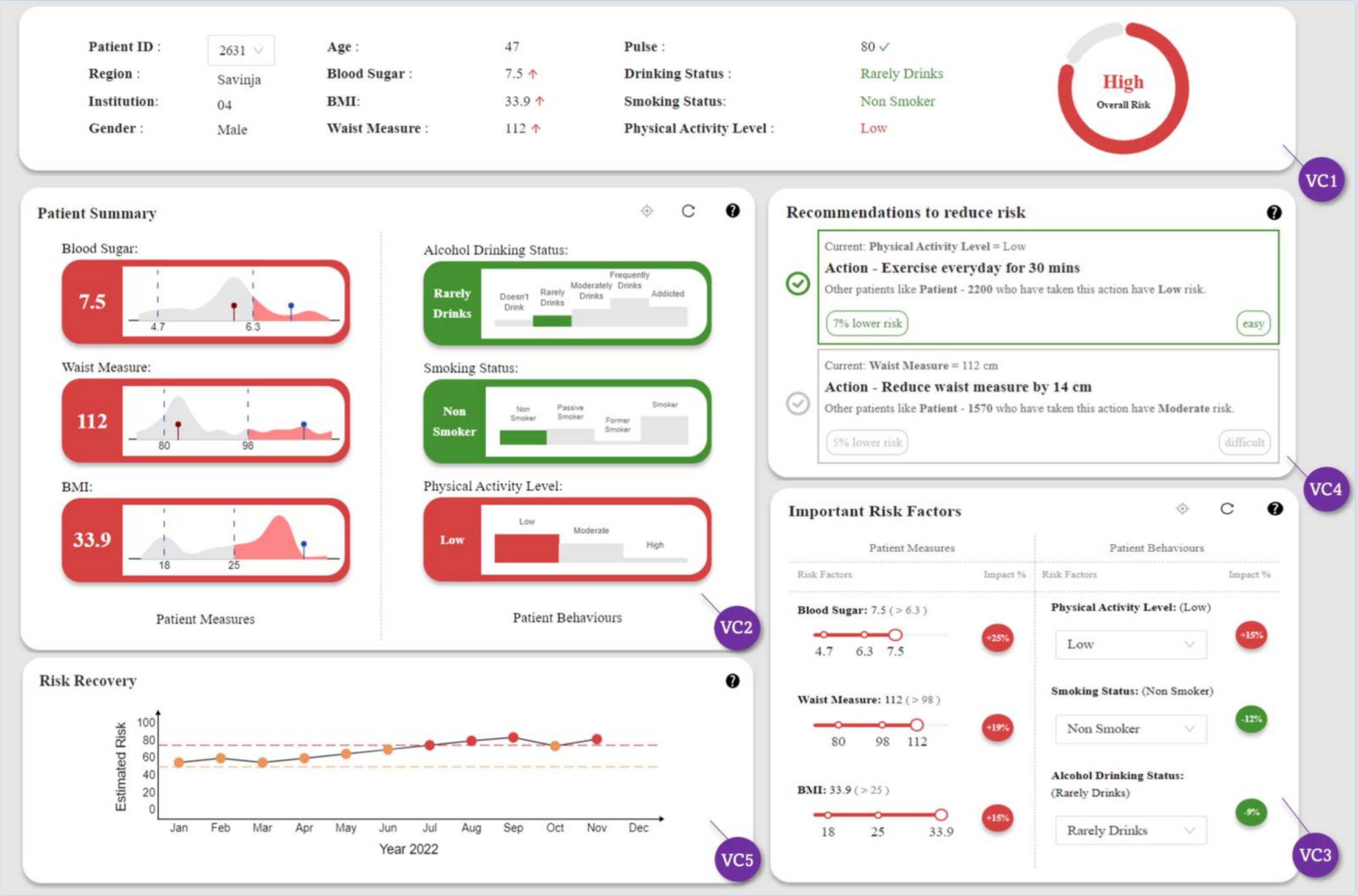


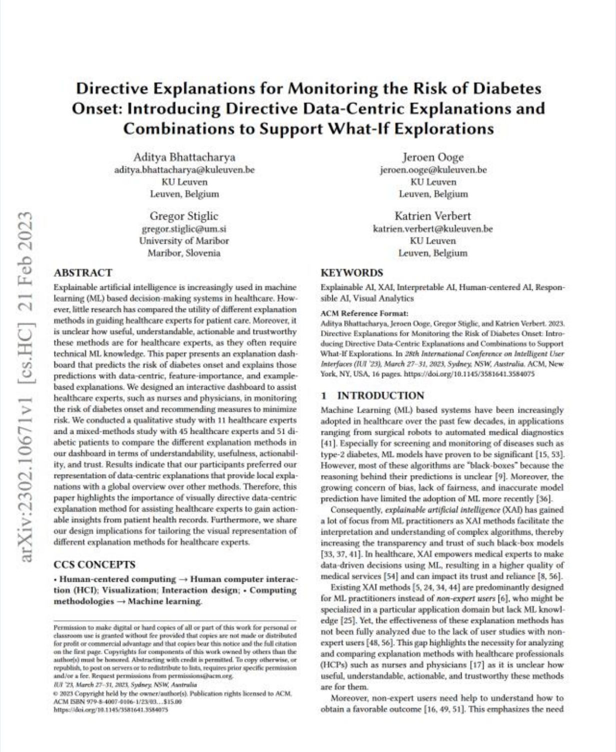
Transparency: justification



Transparency: control

Participants preferred data-centric explanations that provide local explanations with a global overview over other methods. (IUI 2023)





arXiv:2302.10671v1 [cs.HC] 21 Feb 2023

Directive Explanations for Monitoring the Risk of Diabetes Onset: Introducing Directive Data-Centric Explanations and Combinations to Support What-If Explorations

Aditya Bhattacharya
aditya.bhattacharya@kuleuven.be
KU Leuven
Leuven, Belgium

Jeroen Ooge
jeroen.ooge@kuleuven.be
KU Leuven
Leuven, Belgium

Gregor Stiglic
gregor.stiglic@um.si
University of Maribor
Maribor, Slovenia

Katrien Verbert
katrien.verbert@kuleuven.be
KU Leuven
Leuven, Belgium

ABSTRACT

Explainable artificial intelligence is increasingly used in machine learning (ML) based decision-making systems in healthcare. However, little research has compared the utility of different explanation methods in guiding healthcare experts for patient care. Moreover, it is unclear how useful, understandable, actionable and trustworthy these methods are for healthcare experts, as they often require technical ML knowledge. This paper presents an explanation dashboard that predicts the risk of diabetes onset and explains those predictions with data-centric, feature-importance, and example-based explanations. We designed an interactive dashboard to assist healthcare experts, such as nurses and physicians, in monitoring the risk of diabetes onset and recommending measures to minimize risk. We conducted a qualitative study with 11 healthcare experts and a mixed-methods study with 45 healthcare experts and 51 diabetic patients to compare the different explanation methods in our dashboard in terms of understandability, usefulness, actionability, and trust. Results indicate that our participants preferred our representation of data-centric explanations that provide local explanations with a global overview over other methods. Therefore, this paper highlights the importance of visually directive data-centric explanation method for assisting healthcare experts to gain actionable insights from patient health records. Furthermore, we share our design implications for tailoring the visual representation of different explanation methods for healthcare experts.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI); Visualization; Interaction design; • Computing methodologies → Machine learning.**

IUI '23, March 27–31, 2023, Sydney, NSW, Australia

ACM ISBN 979-8-4007-0106-1/23/03...$15.00
https://doi.org/10.1145/3581641.3584075

KEYWORDS

Explainable AI, XAI, Interpretable AI, Human-centered AI, Responsible AI, Visual Analytics

ACM Reference Format:
Aditya Bhattacharya, Jeroen Ooge, Gregor Stiglic, and Katrien Verbert. 2023. Directive Explanations for Monitoring the Risk of Diabetes Onset: Introducing Directive Data-Centric Explanations and Combinations to Support What-If Explorations. In *28th International Conference on Intelligent User Interfaces (IUI '23), March 27–31, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3581641.3584075

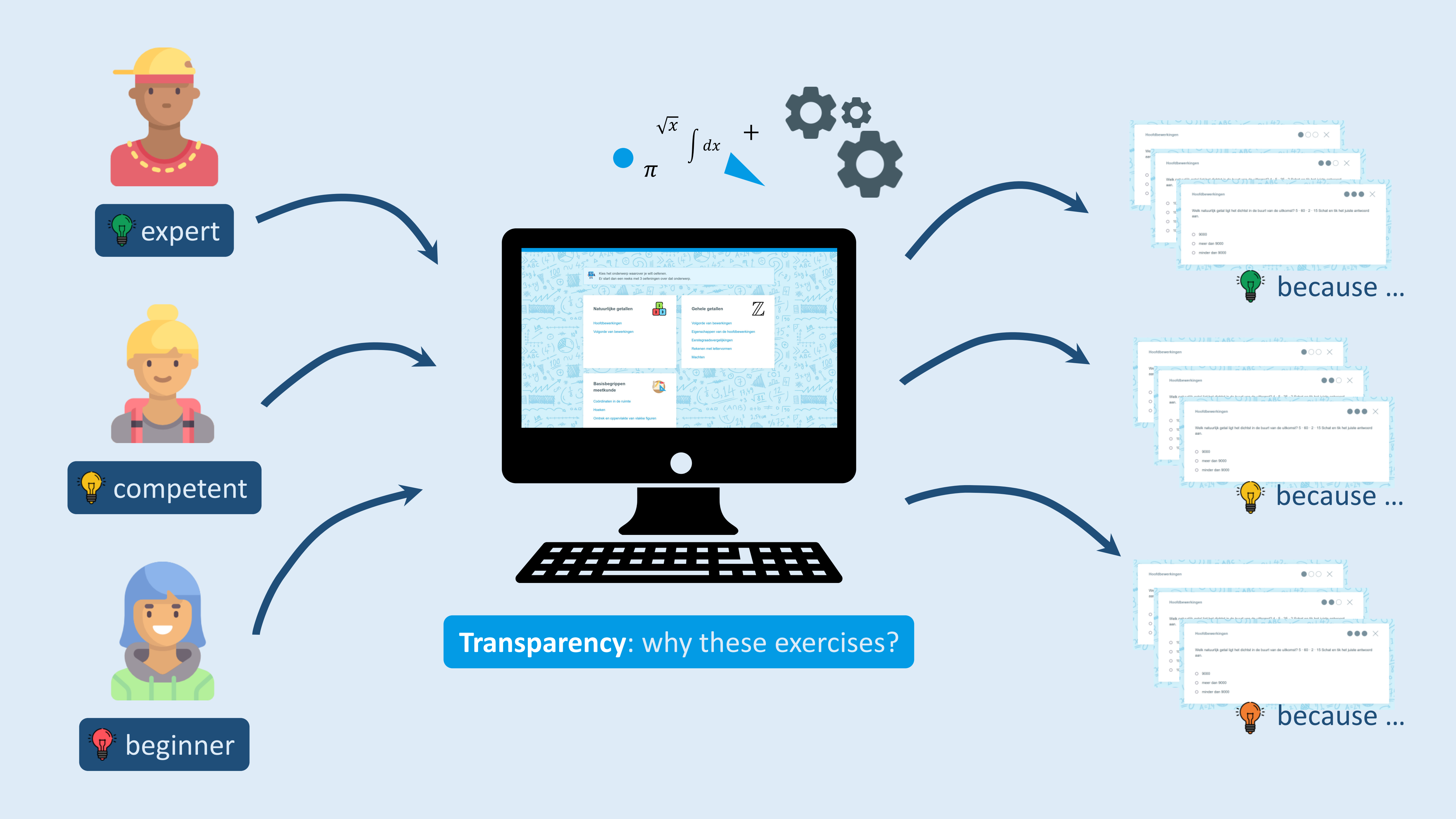
1 INTRODUCTION

Machine Learning (ML) based systems have been increasingly adopted in healthcare over the past few decades, in applications ranging from surgical robots to automated medical diagnostics [41]. Especially for screening and monitoring of diseases such as type-2 diabetes, ML models have proven to be significant [15, 53]. However, most of these algorithms are "black-boxes" because the reasoning behind their predictions is unclear [9]. Moreover, the growing concern of bias, lack of fairness, and inaccurate model prediction have limited the adoption of ML more recently [36].

Consequently, *explainable artificial intelligence* (XAI) has gained a lot of focus from ML practitioners as XAI methods facilitate the interpretation and understanding of complex algorithms, thereby increasing the transparency and trust of such black-box models [33, 37, 41]. In healthcare, XAI empowers medical experts to make data-driven decisions using ML, resulting in a higher quality of medical services [54] and can impact its trust and reliance [8, 56].

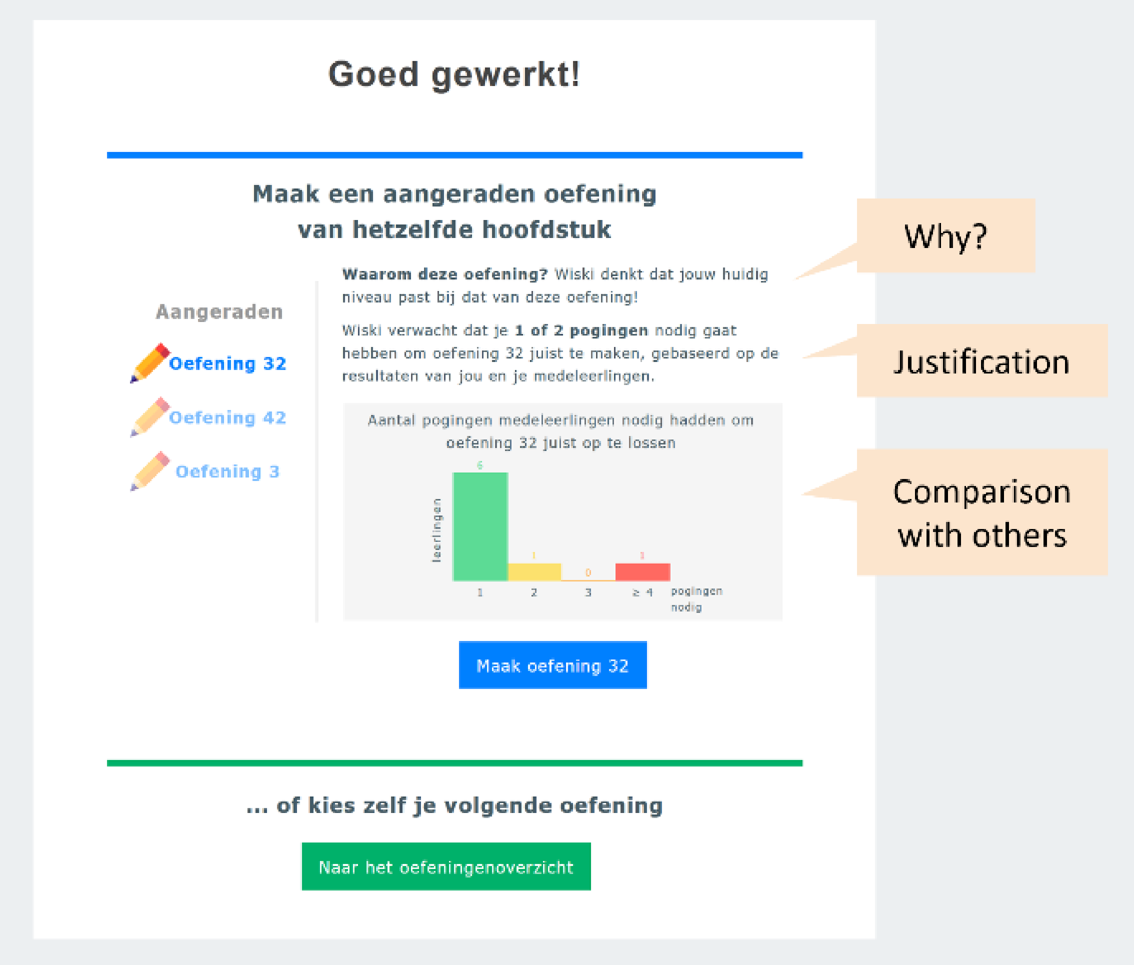
Existing XAI methods [5, 24, 34, 44] are predominantly designed for ML practitioners instead of *non-expert users* [6], who might be specialized in a particular application domain but lack ML knowledge [25]. Yet, the effectiveness of these explanation methods has not been fully analyzed due to the lack of user studies with non-expert users [48, 56]. This gap highlights the necessity for analyzing and comparing explanation methods with healthcare professionals (HCPs) such as nurses and physicians [17] as it is unclear how useful, understandable, actionable, and trustworthy these methods are for them.

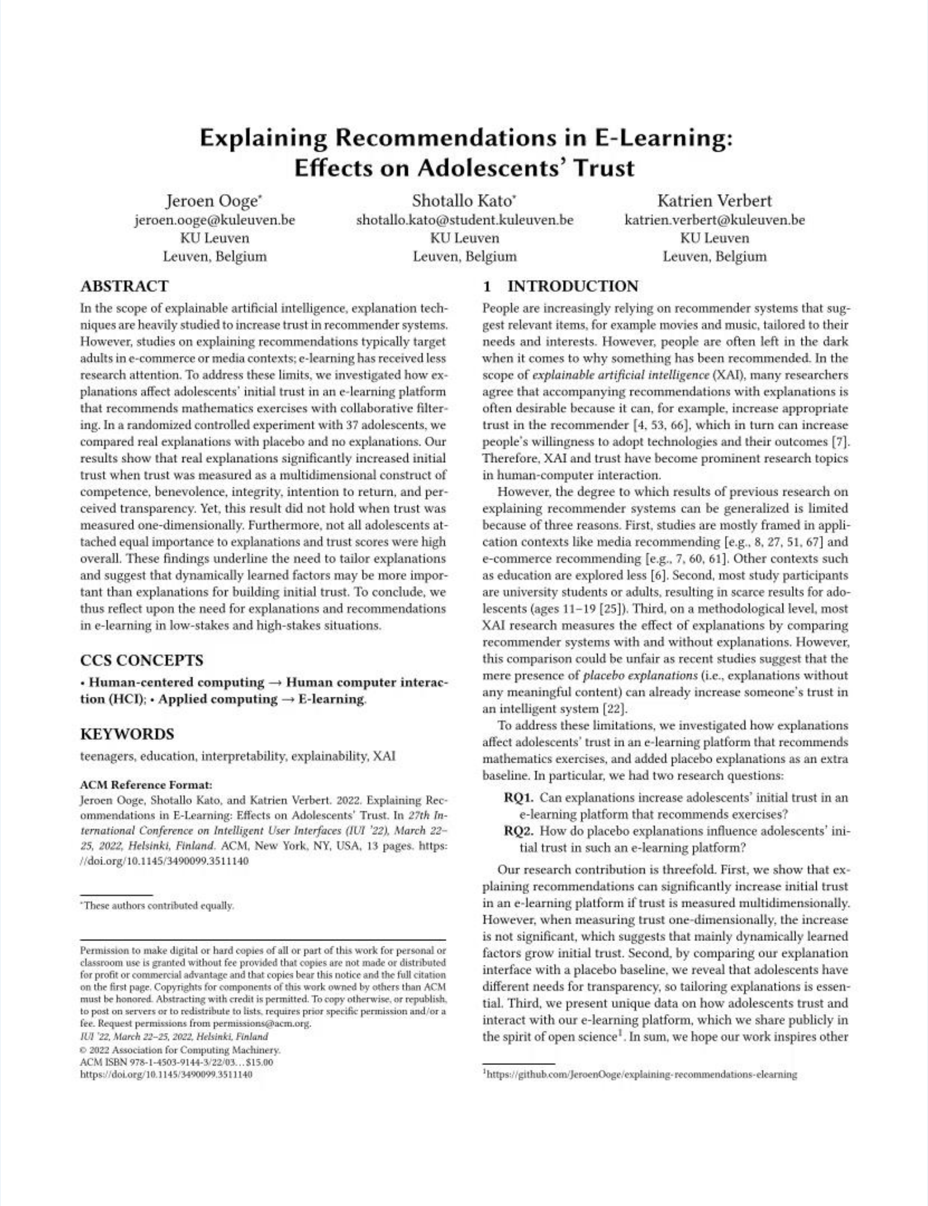
Moreover, non-expert users need help to understand how to obtain a favorable outcome [16, 49, 51]. This emphasizes the need



expert
competent
beginner
$\sqrt{x}$
$\int dx$
π
+
because ...
because ...
because ...
Transparency: why these exercises?

Adolescents have different transparency needs; explanations may not be the most important to build initial trust. (IUI 2022)





Explaining Recommendations in E-Learning: Effects on Adolescents' Trust

Jeroen Ooge*
jeroen.ooge@kuleuven.be
KU Leuven
Leuven, Belgium

Shotallo Kato*
shotallo.kato@student.kuleuven.be
KU Leuven
Leuven, Belgium

Katrien Verbert
katrien.verbert@kuleuven.be
KU Leuven
Leuven, Belgium

ABSTRACT

In the scope of explainable artificial intelligence, explanation techniques are heavily studied to increase trust in recommender systems. However, studies on explaining recommendations typically target adults in e-commerce or media contexts; e-learning has received less research attention. To address these limits, we investigated how explanations affect adolescents' initial trust in an e-learning platform that recommends mathematics exercises with collaborative filtering. In a randomized controlled experiment with 37 adolescents, we compared real explanations with placebo and no explanations. Our results show that real explanations significantly increased initial trust when trust was measured as a multidimensional construct of competence, benevolence, integrity, intention to return, and perceived transparency. Yet, this result did not hold when trust was measured one-dimensionally. Furthermore, not all adolescents attached equal importance to explanations and trust scores were high overall. These findings underline the need to tailor explanations and suggest that dynamically learned factors may be more important than explanations for building initial trust. To conclude, we thus reflect upon the need for explanations and recommendations in e-learning in low-stakes and high-stakes situations.

CCS CONCEPTS

• Human-centered computing → Human computer interaction (HCI); • Applied computing → E-learning.

KEYWORDS

teenagers, education, interpretability, explainability, XAI

ACM Reference Format:
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), March 22–25, 2022, Helsinki, Finland.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3490099.3511140

*These authors contributed equally.

IUI '22, March 22–25, 2022, Helsinki, Finland

ACM ISBN 978-1-4503-9144-3/22/03…$15.00
https://doi.org/10.1145/3490099.3511140

1 INTRODUCTION

People are increasingly relying on recommender systems that suggest relevant items, for example movies and music, tailored to their needs and interests. However, people are often left in the dark when it comes to why something has been recommended. In the scope of *explainable artificial intelligence* (XAI), many researchers agree that accompanying recommendations with explanations is often desirable because it can, for example, increase appropriate trust in the recommender [4, 53, 66], which in turn can increase people's willingness to adopt technologies and their outcomes [7]. Therefore, XAI and trust have become prominent research topics in human-computer interaction.

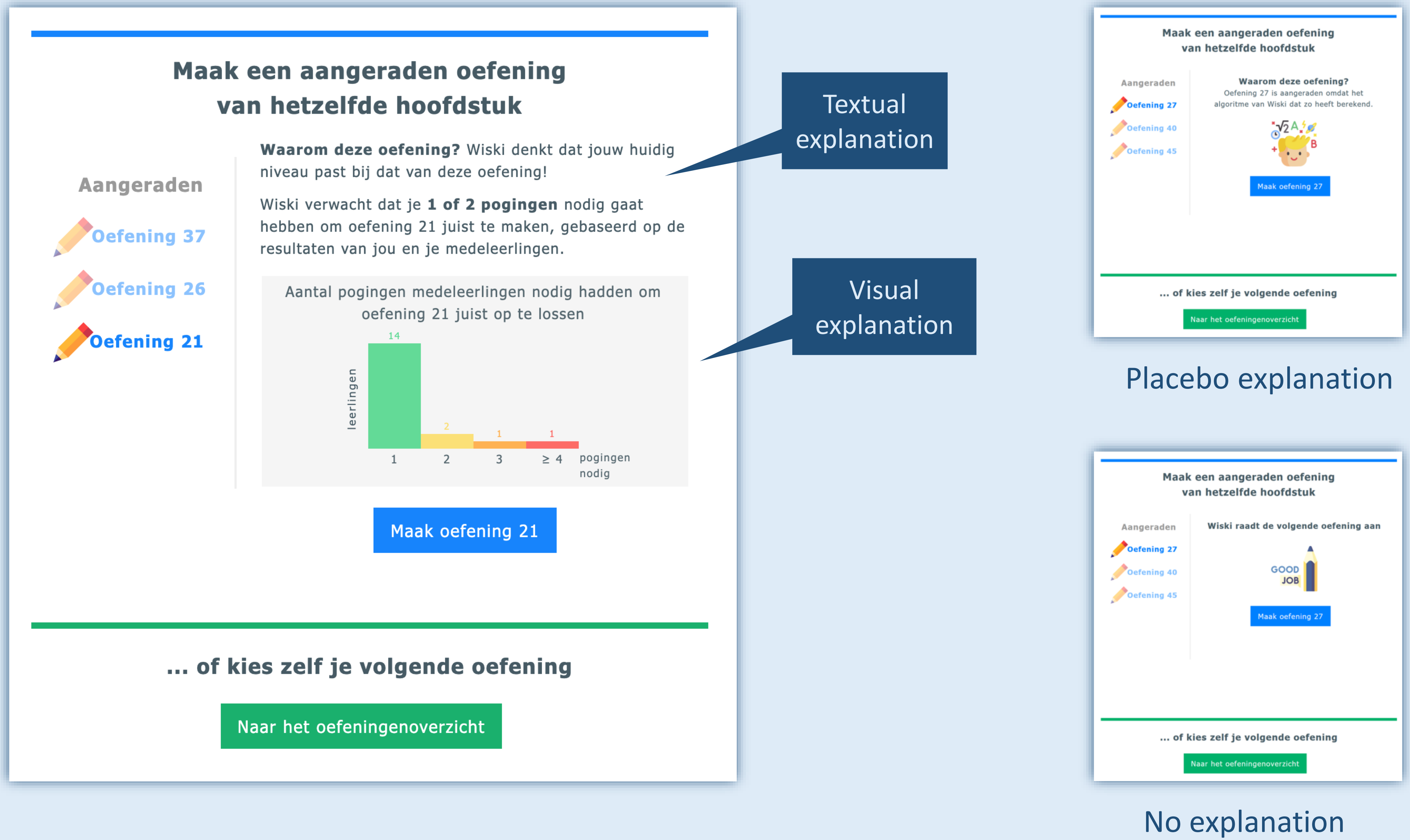
However, the degree to which results of previous research on explaining recommender systems can be generalized is limited because of three reasons. First, studies are mostly framed in application contexts like media recommending [e.g., 8, 27, 51, 67] and e-commerce recommending [e.g., 7, 60, 61]. Other contexts such as education are explored less [6]. Second, most study participants are university students or adults, resulting in scarce results for adolescents (ages 11–19 [25]). Third, on a methodological level, most XAI research measures the effect of explanations by comparing recommender systems with and without explanations. However, this comparison could be unfair as recent studies suggest that the mere presence of *placebo explanations* (i.e., explanations without any meaningful content) can already increase someone's trust in an intelligent system [22].

To address these limitations, we investigated how explanations affect adolescents' trust in an e-learning platform that recommends mathematics exercises, and added placebo explanations as an extra baseline. In particular, we had two research questions:

RQ1. Can explanations increase adolescents' initial trust in an e-learning platform that recommends exercises?
RQ2. How do placebo explanations influence adolescents' initial trust in such an e-learning platform?

Our research contribution is threefold. First, we show that explaining recommendations can significantly increase initial trust in an e-learning platform if trust is measured multidimensionally. However, when measuring trust one-dimensionally, the increase is not significant, which suggests that mainly dynamically learned factors grow initial trust. Second, by comparing our explanation interface with a placebo baseline, we reveal that adolescents have different needs for transparency, so tailoring explanations is essential. Third, we present unique data on how adolescents trust and interact with our e-learning platform, which we share publicly in the spirit of open science[1]. In sum, we hope our work inspires other

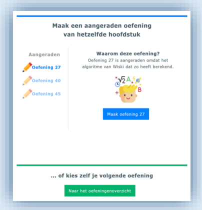
[1]https://github.com/JeroenOoge/explaining-recommendations-elearning



Maak een aangeraden oefening
van hetzelfde hoofdstuk
Aangeraden
Oefening 37
Oefening 26
Oefening 21
Waarom deze oefening? Wiski denkt dat jouw huidig niveau past bij dat van deze oefening!
Wiski verwacht dat je 1 of 2 pogingen nodig gaat hebben om oefening 21 juist te maken, gebaseerd op de resultaten van jou en je medeleerlingen.
Aantal pogingen medeleerlingen nodig hadden om oefening 21 juist op te lossen
14
2
1
1
leerlingen
1
2
3
≥ 4
pogingen nodig
Maak oefening 21
... of kies zelf je volgende oefening
Naar het oefeningenoverzicht
Textual explanation
Visual explanation
Maak een aangeraden oefening
van hetzelfde hoofdstuk
Aangeraden
Oefening 27
Oefening 40
Oefening 45
Waarom deze oefening?
Oefening 27 is aangeraden omdat het algoritme van Wiski dat zo heeft berekend.
Maak oefening 27
... of kies zelf je volgende oefening
Naar het oefeningenoverzicht
Placebo explanation
Maak een aangeraden oefening
van hetzelfde hoofdstuk
Aangeraden
Oefening 27
Oefening 40
Oefening 45
Wiski raadt de volgende oefening aan
GOOD JOB
Maak oefening 27
... of kies zelf je volgende oefening
Naar het oefeningenoverzicht
No explanation

How was **trust** in the recommendations affected?

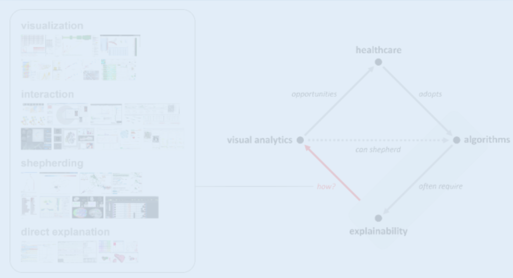
Real vs no explanation



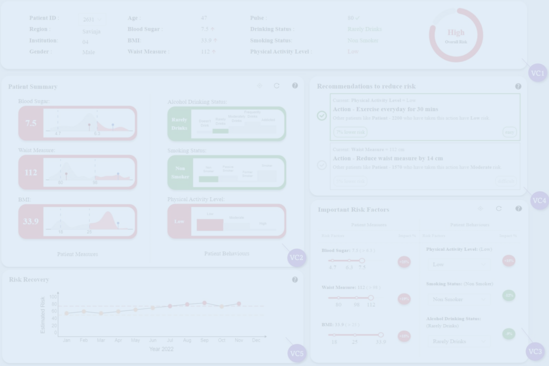
Placebo vs no explanation

Explainable AI through visualisation

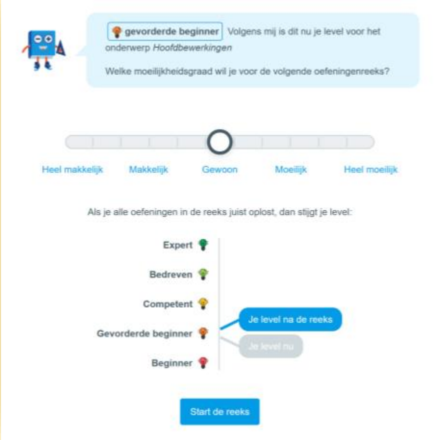
Visual analytics

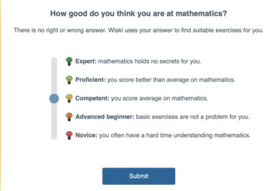


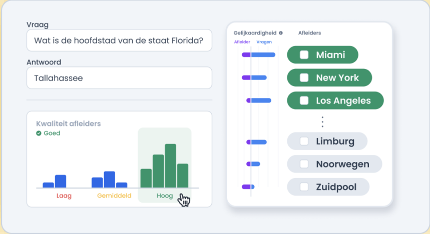
Transparency: justification



Transparency: control

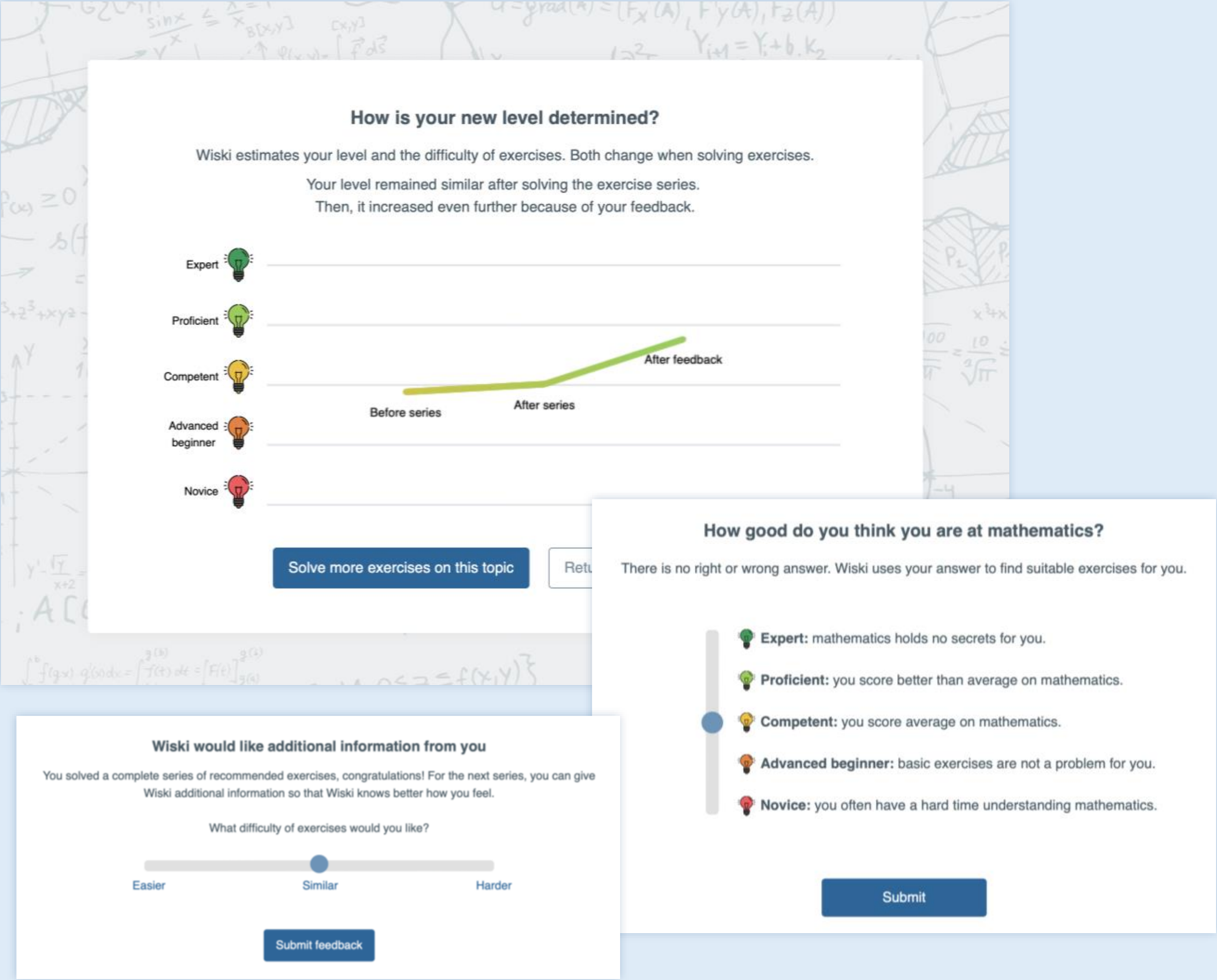







expert
competent
beginner
Control: I want other exercises

Control mechanisms do not necessarily increase trust; showing the impact of control is essential. (IUI 2023)



Steering Recommendations and Visualising Its Impact: Effects on Adolescents' Trust in E-Learning Platforms

Jeroen Ooge
KU Leuven
Department of Computer Science
Leuven, Belgium
jeroen.ooge@kuleuven.be

Leen Dereu
KU Leuven
Department of Computer Science
Leuven, Belgium
leen.dereu@student.kuleuven.be

Katrien Verbert
KU Leuven
Department of Computer Science
Leuven, Belgium
katrien.verbert@kuleuven.be

ABSTRACT

Researchers have widely acknowledged the potential of control mechanisms with which end-users of recommender systems can better tailor recommendations. However, few e-learning environments so far incorporate such mechanisms, for example for steering recommended exercises. In addition, studies with adolescents in this context are rare. To address these limitations, we designed a control mechanism and a visualisation of the control's impact through an iterative design process with adolescents and teachers. Then, we investigated how these functionalities affect adolescents' trust in an e-learning platform that recommends maths exercises. A randomised controlled experiment with 76 middle school and high school adolescents showed that visualising the impact of exercised control significantly increases trust. Furthermore, having control over their mastery level seemed to inspire adolescents to reasonably challenge themselves and reflect upon the underlying recommendation algorithm. Finally, a significant increase in perceived transparency suggested that visualising steering actions can indirectly explain why recommendations are suitable, which opens interesting research tracks for the broader field of explainable AI.

CCS CONCEPTS

• Human-centered computing → Human computer interaction (HCI); • Applied computing → E-learning.

KEYWORDS

education, technology-enhanced learning, teenagers, explainable AI, XAI, controllability, inspectability

ACM Reference Format:
Jeroen Ooge, Leen Dereu, and Katrien Verbert. 2023. Steering Recommendations and Visualising Its Impact: Effects on Adolescents' Trust in E-Learning Platforms. In *28th International Conference on Intelligent User Interfaces (IUI '23), March 27–31, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3581641.3584046

IUI '23, March 27–31, 2023, Sydney, NSW, Australia

ACM ISBN 979-8-4007-0106-1/23/03…$15.00
https://doi.org/10.1145/3581641.3584046

1 INTRODUCTION

Recommender systems have long been actively studied to help reduce information overload in contexts where people are searching for relevant content. To better anticipate people's changing preferences and needs, researchers have increasingly acknowledged the importance of control mechanisms with which people can actively steer recommendations [43]. Studies have shown that being able to control recommendations can increase satisfaction with, perceived understanding of, and trust in a recommender system, which can in turn increase acceptance of recommendations [51]. At the same time, too much control can overwhelm people and incur high cognitive loads [7, 9].

However, most research on controlling recommender systems is limited because of two reasons. First, studied target audiences typically consist of adults, whereas in practice younger audiences such as adolescents (ages 12–19 [25]) are just as much, if not more, exposed to recommendation algorithms. Second, recommender systems are most often studied within contexts such as multimedia, e-commerce, and other services, and it is unclear whether findings therein always transfer to other application domains. In a high-stakes domain such as education, for example, it is crucial to properly understand the effects of control mechanisms, especially now that e-learning platforms are increasingly recommending learning content to personalise learning. Thus, it is important to design control mechanisms fit for an educational context; reflect on how much control students, teachers, and other parties should get; and find suitable ways to communicate the impact of steering.

To address these limitations, we conducted a study on how adolescents trust an e-learning platform when they can steer recommended exercises and see their control's effects. Our research questions were as follows:

RQ1. How does the ability to control recommended exercises affect students' trust in an e-learning platform?

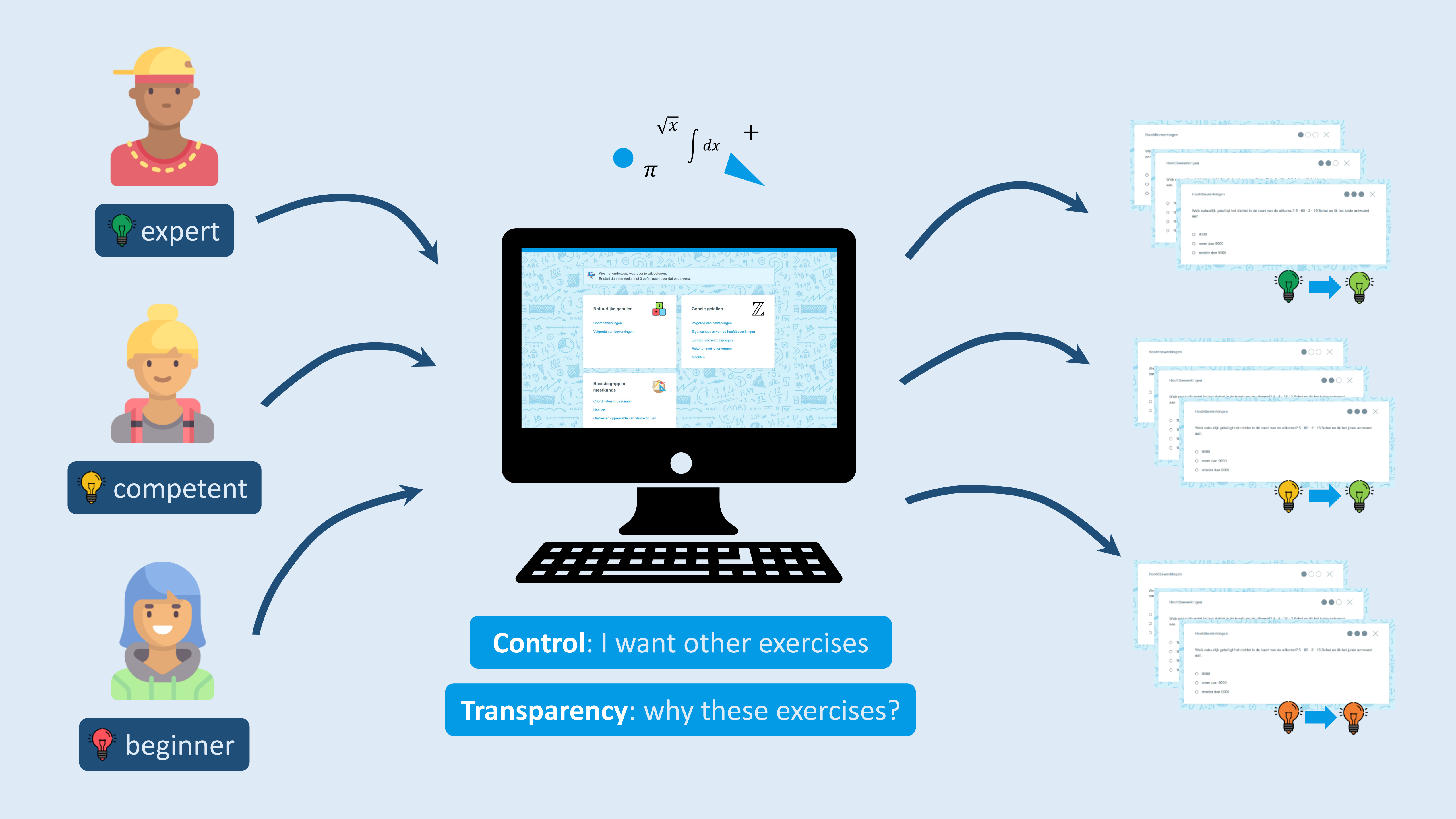
RQ2. How is students' trust in an e-learning platform affected when they see a visual representation of their impact when controlling recommended exercises?

Our research contribution is threefold. First, we present a control mechanism and a visualisation of its impact, which have been found useful and usable by adolescents in a user-centred design process. Second, we discovered that a control mechanism does not necessarily change trust, neither when measured directly, nor when measured as a construct of competence, benevolence, integrity, intention to return, and perceived transparency. We also found, however, that a control mechanism can stimulate adolescents to reflect more upon their mastery level and the underlying recommendation system. Third, we show that visualising the control's impact can

How was **trust** in the recommendations affected?

Control vs no control

Control with impact vs no control


expert
competent
beginner
Natuurlijke getallen
Gehele getallen
Basisbegrippen meetkunde
Control: I want other exercises
Transparency: why these exercises?

AIDA

Aanbevolen oefeningen

Aanbevolen sequentie		
Oefening 15 Deel 1 Het onderwerp en de pv	**Niveau oefening** Gemakkelijk	**Reeds gemaakt?** Nee
Oefening 23 Deel 1 Soorten zinnen	**Niveau oefening** Gemakkelijk	**Reeds gemaakt?** Nee
Oefening 12 Deel 1 Het onderwerp en de pv	**Niveau oefening** Gemakkelijk	**Reeds gemaakt?** Nee
Oefening 35 Deel 1 Over klinkers, medeklinkers	**Niveau oefening** Gemakkelijk	**Reeds gemaakt?** Nee
Oefening 10 Deel 1 Soorten zinnen	**Niveau oefening** Gemakkelijk	**Reeds gemaakt?** Nee

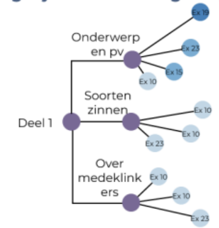
Algemeen | Het onderwerp en de pv | Soorten zinnen | Over klinkers, medeklinkers

Waarom deze sequentie van oefeningen?

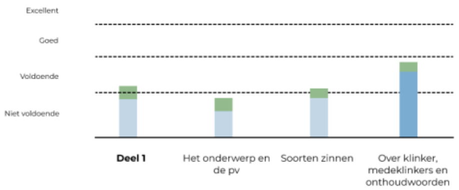
Het systeem zoekt oefening aangepast aan jou niveau waarbij je het meeste voortgang kan boeken.
Je niveau wordt geschat aan de hand van voorgaand gemaakte oefeningen.

1. Jouw niveau & waarom:

2. Meest vergelijkbare oefening aange



Wat als je deze sequentie oplost?



Door komende sequentie te maken stijg je zowel voor 'het onderwerp en de pv', 'soorten zinnen' en 'over klinker, medeklinkers en onthoudwoorden'. Doe zo voort!

Maak sequentie

Aanbevolen sequentie		
Oefening 15 Deel 1 Het onderwerp en de pv	**Niveau oefening** Gemakkelijk	**Reeds gemaakt?** Nee
Oefening 23 Deel 1 Soorten zinnen	**Niveau oefening** Gemakkelijk	**Reeds gemaakt?** Nee
Oefening 12 Deel 1 Het onderwerp en de pv	**Niveau oefening** Gemakkelijk	**Reeds gemaakt?** Nee
Oefening 35 Deel 1 Over klinkers, medeklinkers	**Niveau oefening** Gemakkelijk	**Reeds gemaakt?** Nee
Oefening 10 Deel 1 Soorten zinnen	**Niveau oefening** Gemakkelijk	**Reeds gemaakt?** Nee

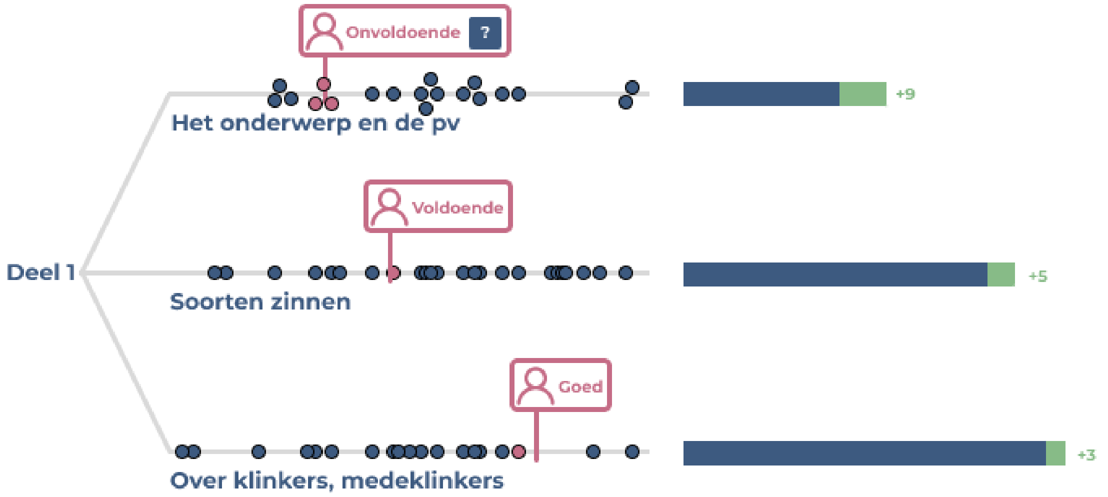
Uitleg

Waarom deze reeks van oefeningen?

Het systeem zoekt oefening aangepast aan jou niveau waarbij je het meeste voortgang kan boeken. Je niveau wordt geschat aan de hand van voorgaand gemaakte oefeningen.

Wat als je de reeks correct oplost?

Zo schat het systeem dat je kennis op de gekozen topics zal evolueren wanneer de aanbevolen reeks corret oplost.

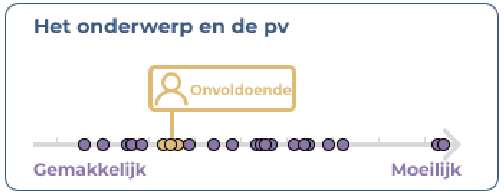


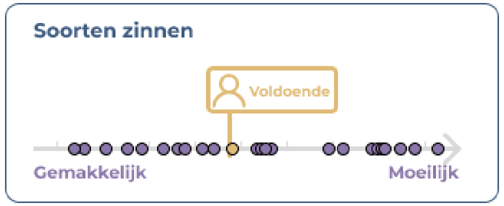
Deze reeks wordt aanbevolen

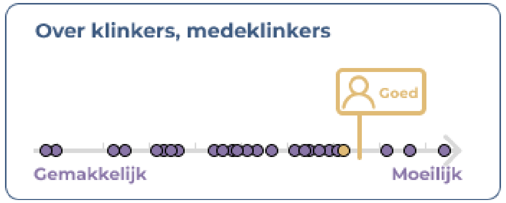
Oefening	Niveau oefening
Oefening 15 Het onderwerp en de pv	**Gemakkelijk**
Oefening 23 Soorten zinnen	**Gemiddeld**
Oefening 12 Het onderwerp en de pv	**Gemakkelijk**
Oefening 35 Over klinkers, medeklinkers	**Moeilijk**
Oefening 10 Het onderwerp en de pv	**Gemakkelijk**

Deze reeks past bij jou, want …

… de aanbevolen oefeningen liggen dichter bij jouw niveau dan de niet aanbevolen oefeningen

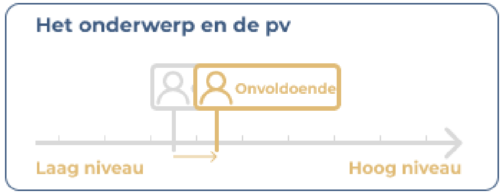


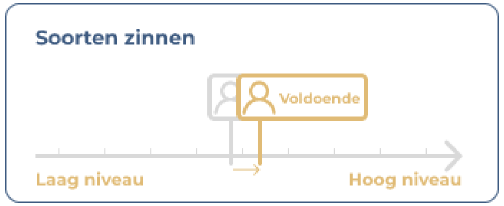


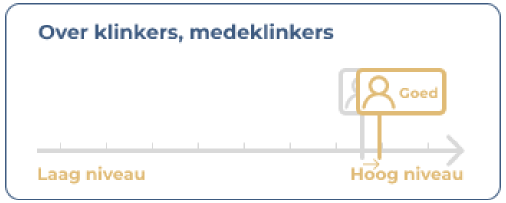


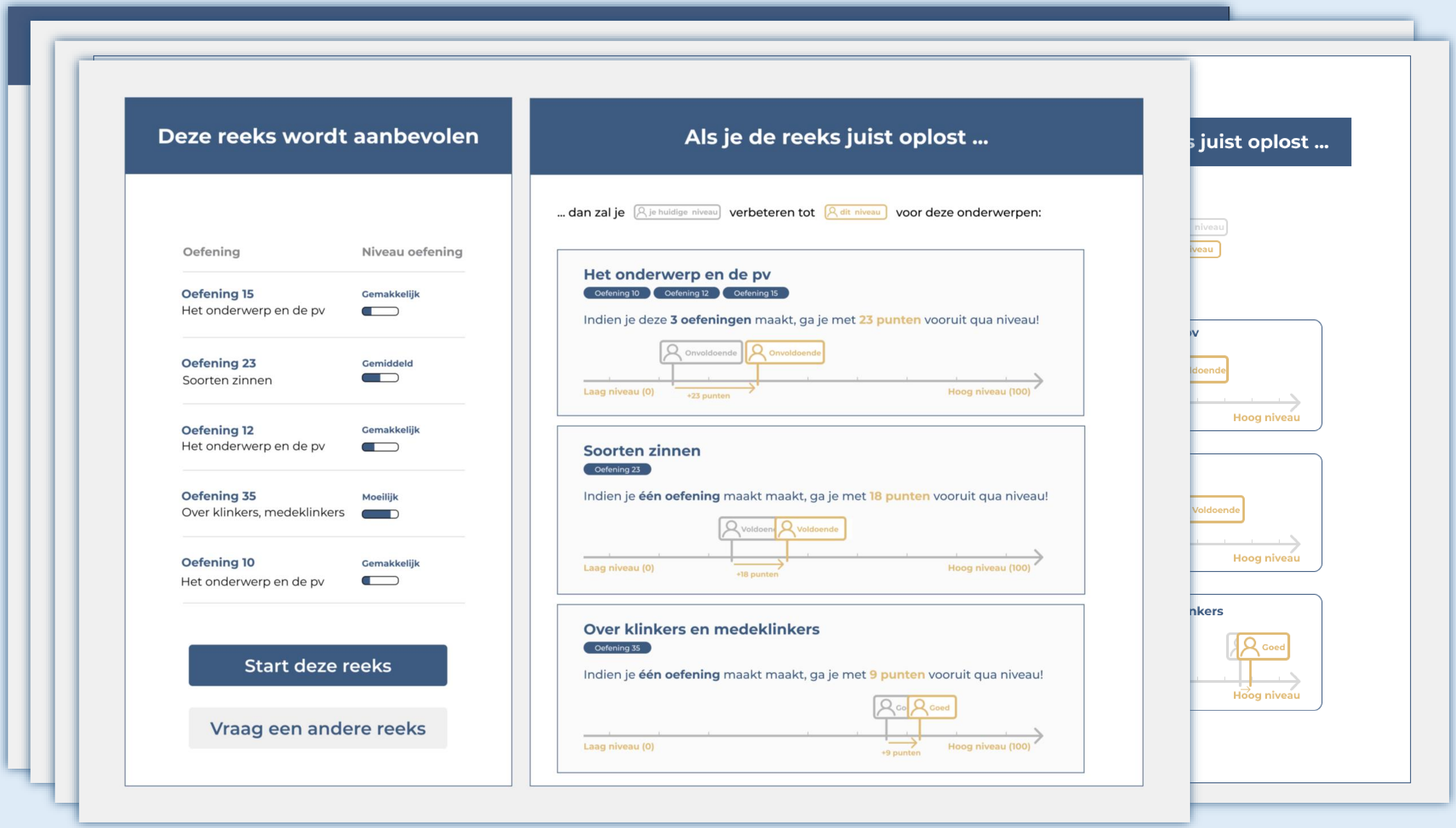
Als je deze reeks juist oplost …

… dan zal je je huidige niveau verbeteren tot dit niveau

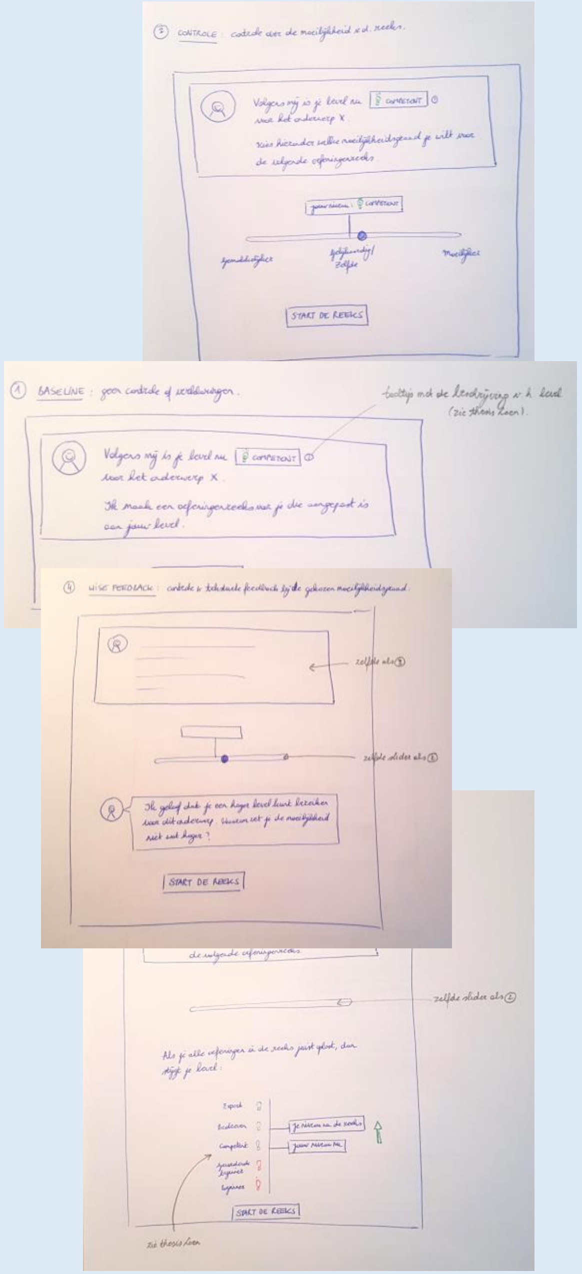


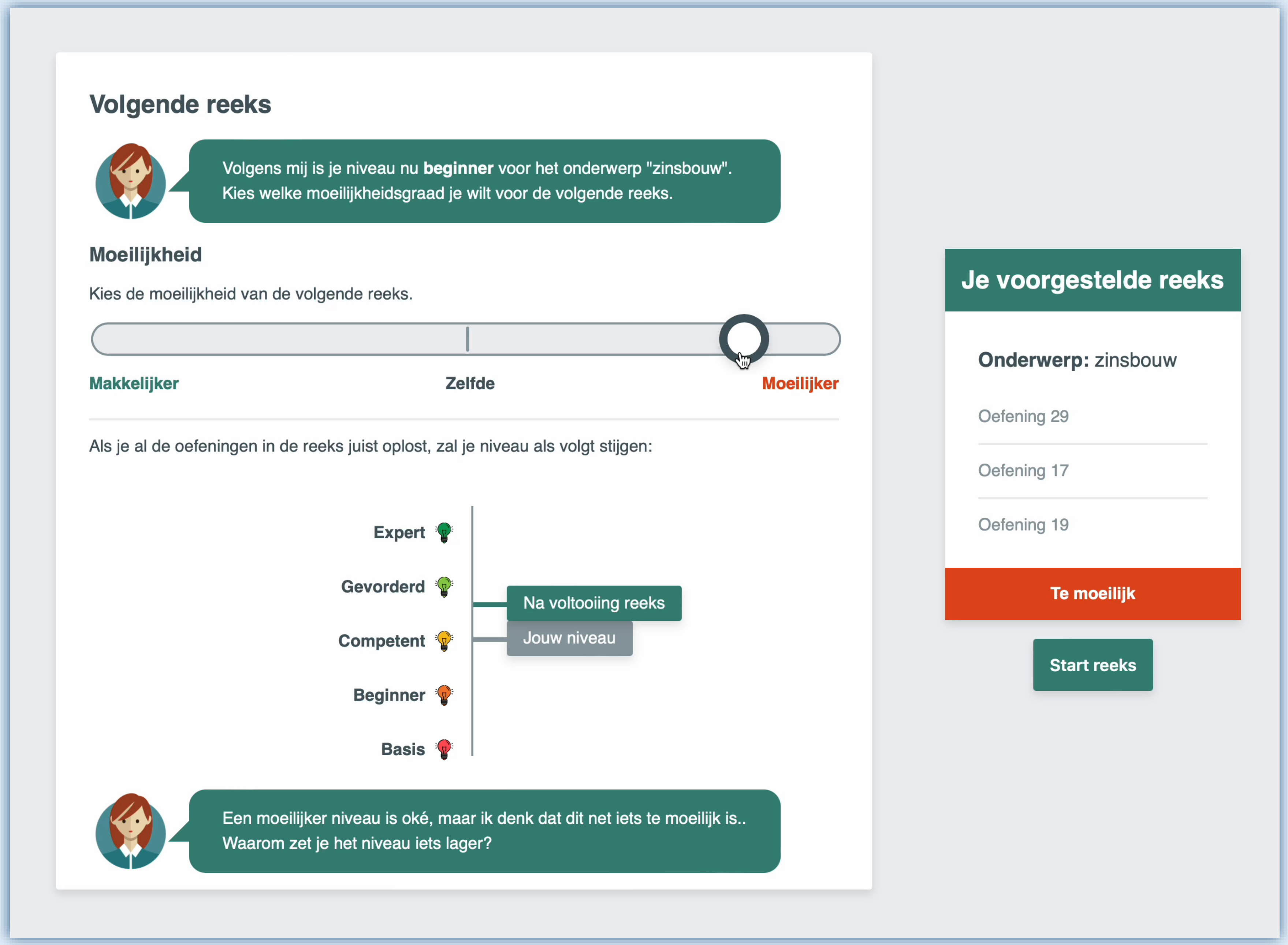


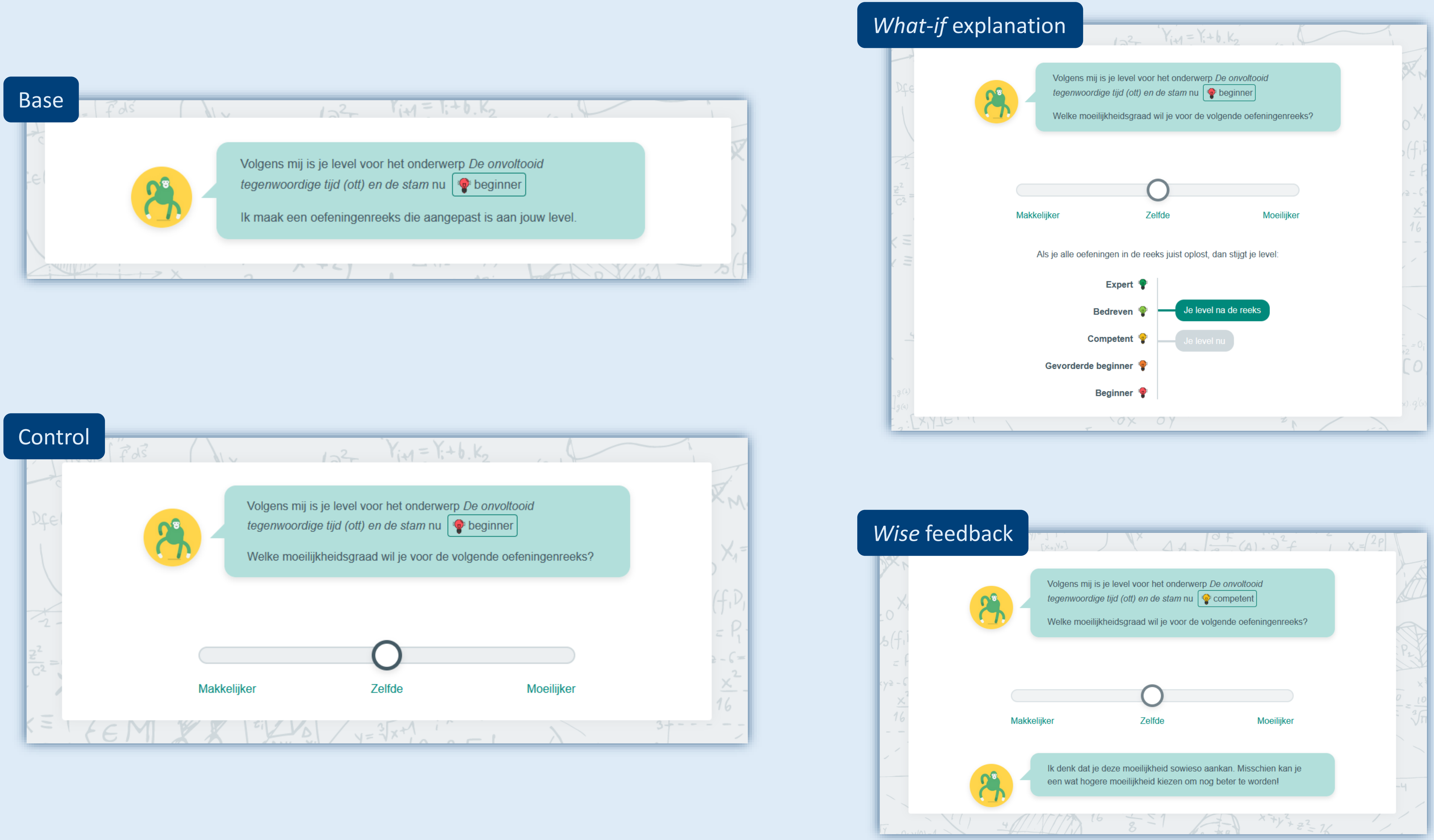




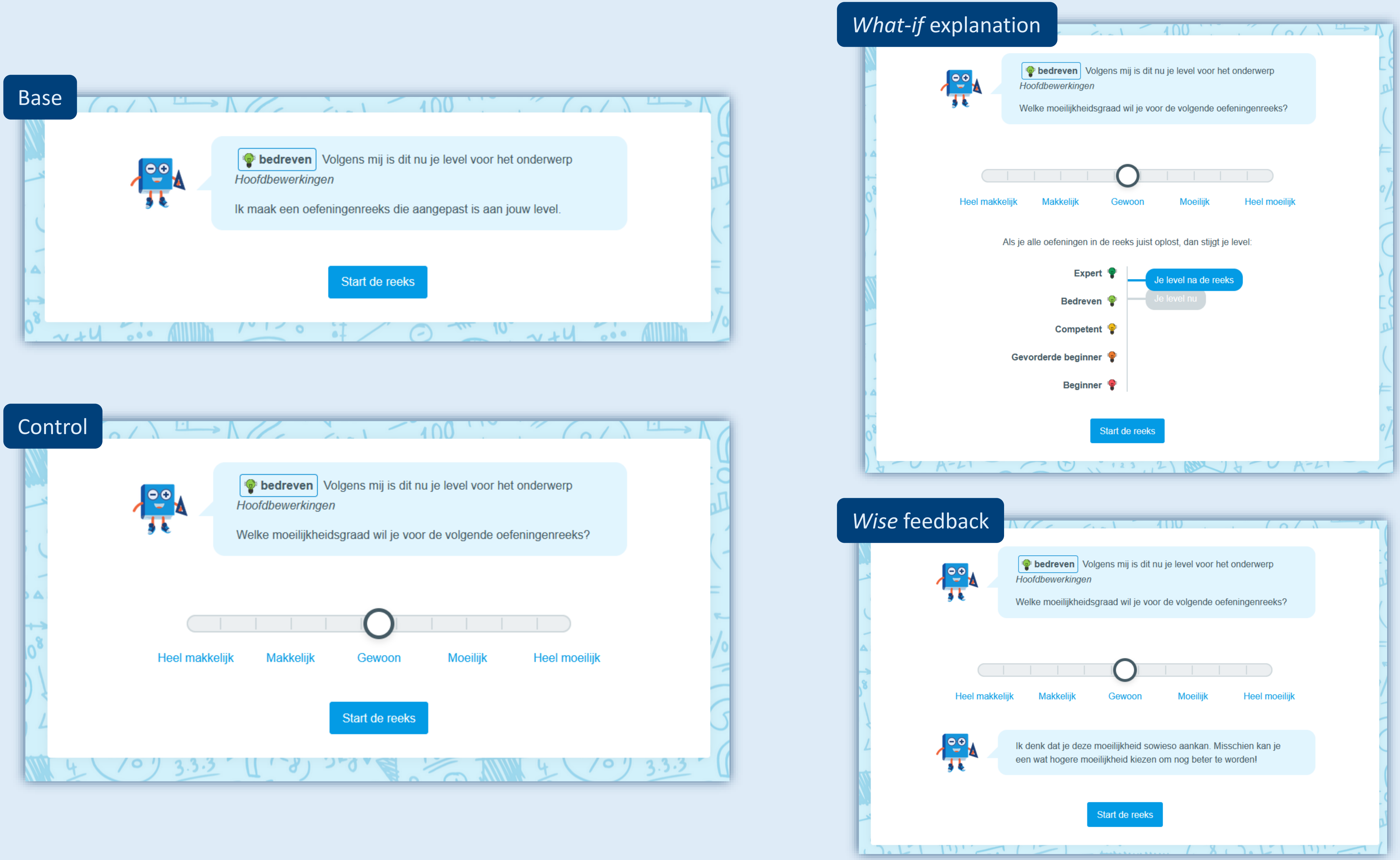
Deze reeks wordt aanbevolen
Oefening
Niveau oefening
Oefening 15
Het onderwerp en de pv
Gemakkelijk
Oefening 23
Soorten zinnen
Gemiddeld
Oefening 12
Het onderwerp en de pv
Gemakkelijk
Oefening 35
Over klinkers, medeklinkers
Moeilijk
Oefening 10
Het onderwerp en de pv
Gemakkelijk
Start deze reeks
Vraag een andere reeks
Als je de reeks juist oplost ...
... dan zal je
je huidige niveau
verbeteren tot
dit niveau
voor deze onderwerpen:
Het onderwerp en de pv
Oefening 10
Oefening 12
Oefening 15
Indien je deze 3 oefeningen maakt, ga je met 23 punten vooruit qua niveau!
Onvoldoende
Onvoldoende
Laag niveau (0)
+23 punten
Hoog niveau (100)
Soorten zinnen
Oefening 23
Indien je één oefening maakt maakt, ga je met 18 punten vooruit qua niveau!
Voldoende
Laag niveau (0)
+18 punten
Hoog niveau (100)
Over klinkers en medeklinkers
Oefening 35
Indien je één oefening maakt maakt, ga je met 9 punten vooruit qua niveau!
Goed
Laag niveau (0)
+9 punten
Hoog niveau (100)




Volgende reeks
Volgens mij is je niveau nu beginner voor het onderwerp "zinsbouw".
Kies welke moeilijkheidsgraad je wilt voor de volgende reeks.
Moeilijkheid
Kies de moeilijkheid van de volgende reeks.
Makkelijker
Zelfde
Moeilijker
Als je al de oefeningen in de reeks juist oplost, zal je niveau als volgt stijgen:
Expert
Gevorderd
Competent
Beginner
Basis
Na voltooiing reeks
Jouw niveau
Een moeilijker niveau is oké, maar ik denk dat dit net iets te moeilijk is..
Waarom zet je het niveau iets lager?
Je voorgestelde reeks
Onderwerp: zinsbouw
Oefening 29
Oefening 17
Oefening 19
Te moeilijk
Start reeks



Base
Volgens mij is je level voor het onderwerp De onvoltooid tegenwoordige tijd (ott) en de stam nu beginner
Ik maak een oefeningenreeks die aangepast is aan jouw level.
Control
Volgens mij is je level voor het onderwerp De onvoltooid tegenwoordige tijd (ott) en de stam nu beginner
Welke moeilijkheidsgraad wil je voor de volgende oefeningenreeks?
Makkelijker
Zelfde
Moeilijker
What-if explanation
Volgens mij is je level voor het onderwerp De onvoltooid tegenwoordige tijd (ott) en de stam nu beginner
Welke moeilijkheidsgraad wil je voor de volgende oefeningenreeks?
Makkelijker
Zelfde
Moeilijker
Als je alle oefeningen in de reeks juist oplost, dan stijgt je level:
Expert
Bedreven
Je level na de reeks
Competent
Je level nu
Gevorderde beginner
Beginner
Wise feedback
Volgens mij is je level voor het onderwerp De onvoltooid tegenwoordige tijd (ott) en de stam nu competent
Welke moeilijkheidsgraad wil je voor de volgende oefeningenreeks?
Makkelijker
Zelfde
Moeilijker
Ik denk dat je deze moeilijkheid sowieso aankan. Misschien kan je een wat hogere moeilijkheid kiezen om nog beter te worden!



Base
bedreven Volgens mij is dit nu je level voor het onderwerp
Hoofdbewerkingen
Ik maak een oefeningenreeks die aangepast is aan jouw level.
Start de reeks
Control
bedreven Volgens mij is dit nu je level voor het onderwerp
Hoofdbewerkingen
Welke moeilijkheidsgraad wil je voor de volgende oefeningenreeks?
Heel makkelijk
Makkelijk
Gewoon
Moeilijk
Heel moeilijk
Start de reeks
What-if explanation
bedreven Volgens mij is dit nu je level voor het onderwerp
Hoofdbewerkingen
Welke moeilijkheidsgraad wil je voor de volgende oefeningenreeks?
Heel makkelijk
Makkelijk
Gewoon
Moeilijk
Heel moeilijk
Als je alle oefeningen in de reeks juist oplost, dan stijgt je level:
Expert
Je level na de reeks
Bedreven
Je level nu
Competent
Gevorderde beginner
Beginner
Start de reeks
Wise feedback
bedreven Volgens mij is dit nu je level voor het onderwerp
Hoofdbewerkingen
Welke moeilijkheidsgraad wil je voor de volgende oefeningenreeks?
Heel makkelijk
Makkelijk
Gewoon
Moeilijk
Heel moeilijk
Ik denk dat je deze moeilijkheid sowieso aankan. Misschien kan je een wat hogere moeilijkheid kiezen om nog beter te worden!
Start de reeks

WISKI

Oefenen

Kies het onderwerp waarover je wilt oefenen.
Er start dan een reeks met 3 oefeningen over dat onderwerp.

Natuurlijke getallen

Hoofdbewerkingen

Volgorde van bewerkingen

Gehele getallen

Volgorde van bewerkingen

Eigenschappen van de hoofdbewerkingen

Eerstegraadsvergelijkingen

Rekenen met lettervormen

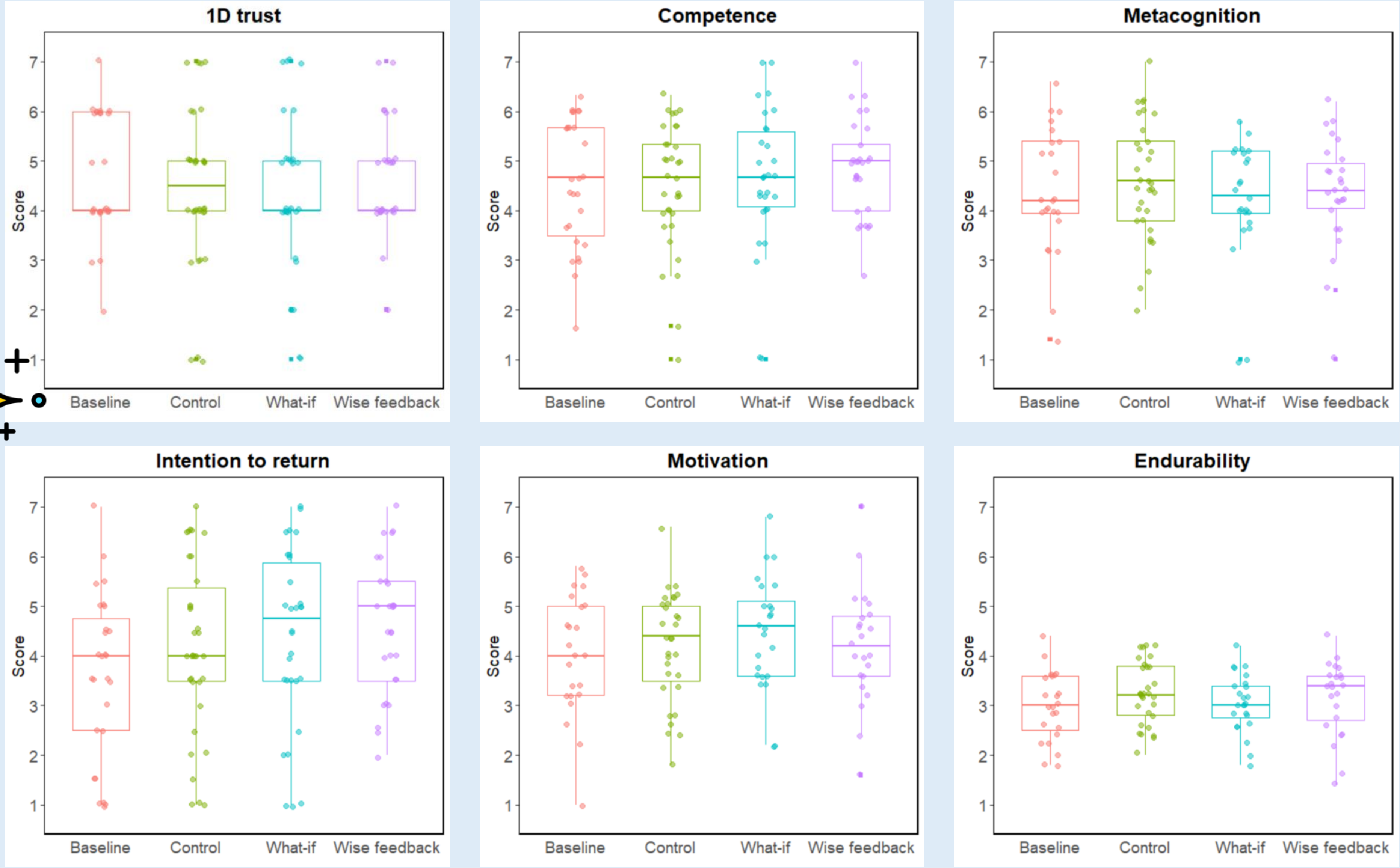
Machten

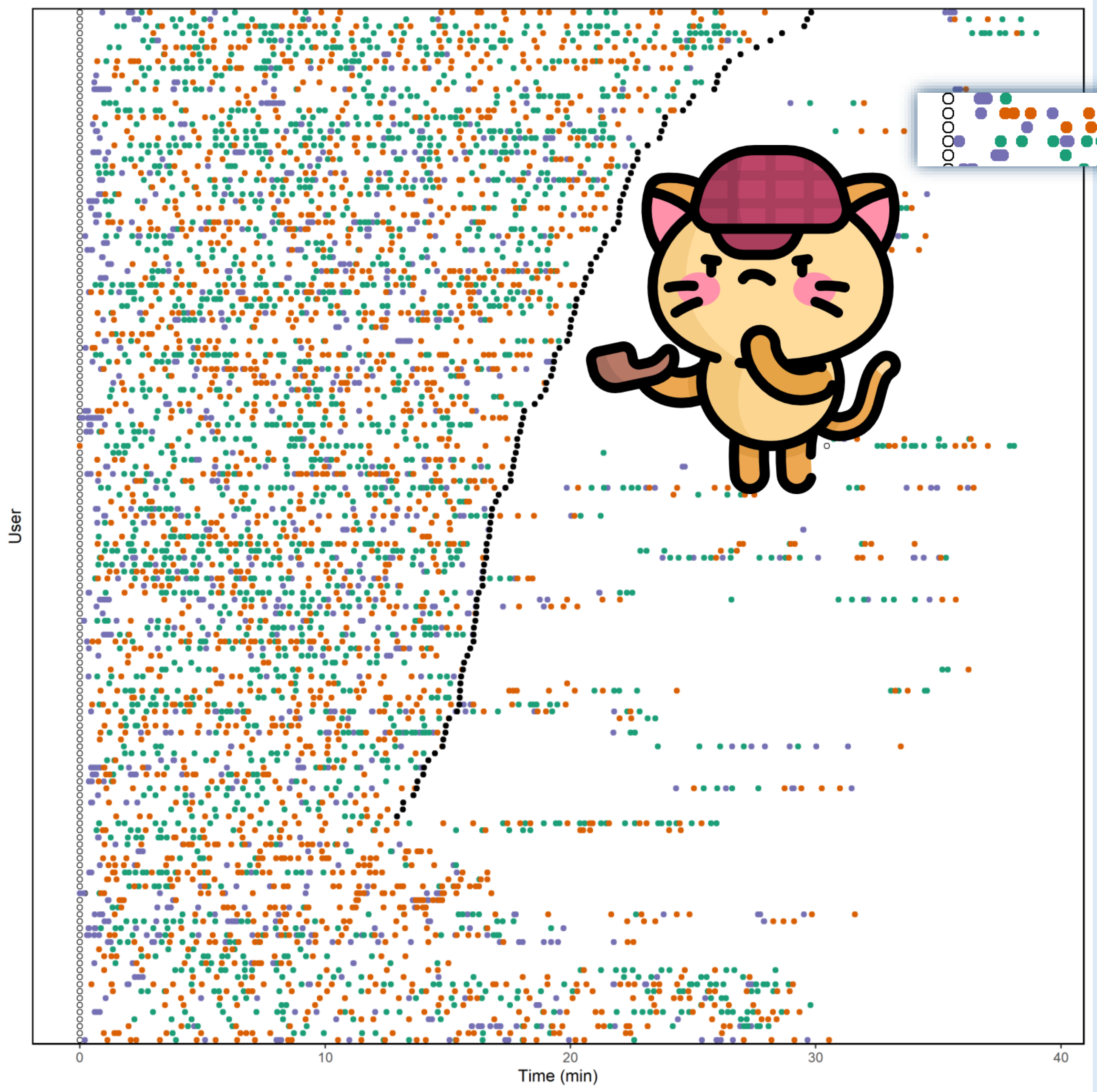
Basisbegrippen meetkunde

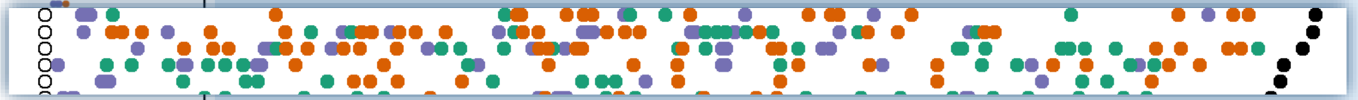
Coördinaten in de ruimte

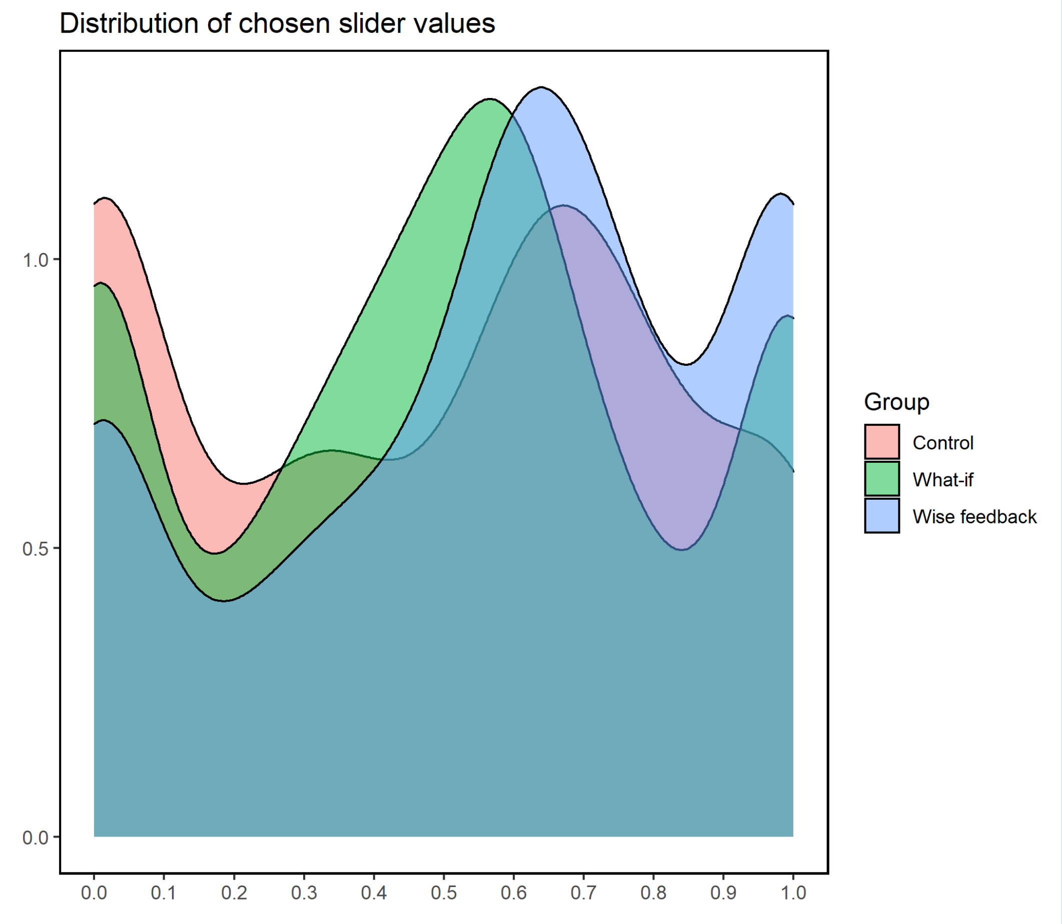
Hoeken

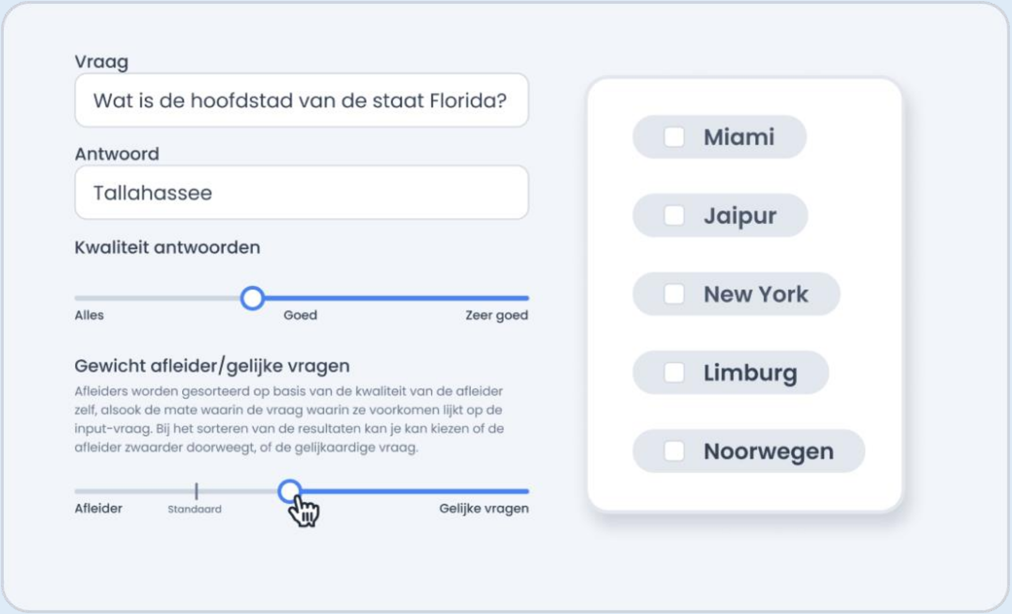
Omtrek en oppervlakte van vlakke figuren

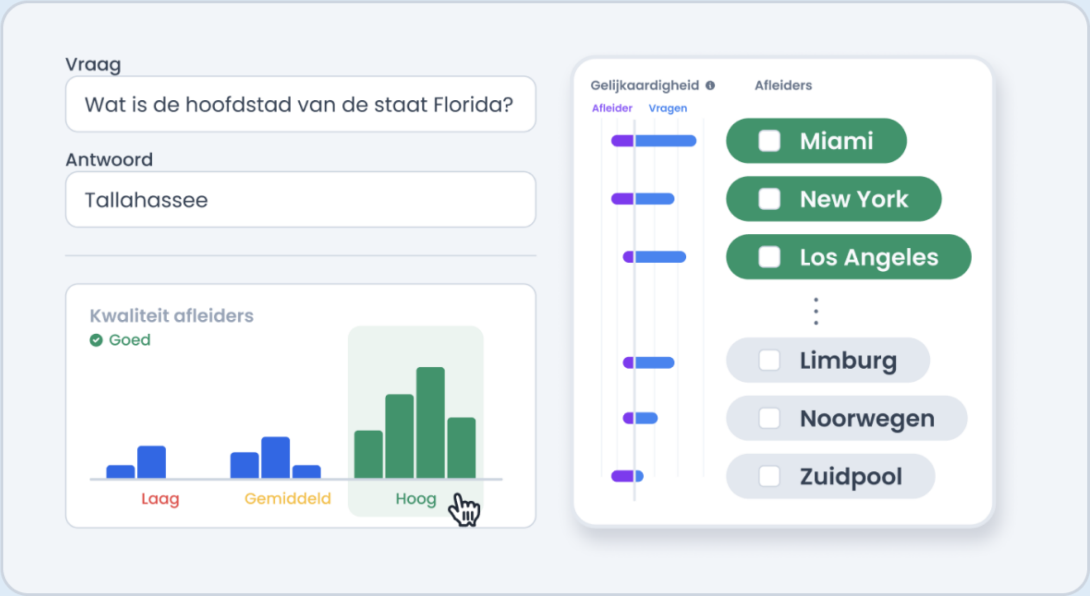

1D trust
Competence
Metacognition
Intention to return
Motivation
Endurability
Score
1
2
3
4
5
6
7
Baseline
Control
What-if
Wise feedback


User
0
10
20
30
40
Time (min)



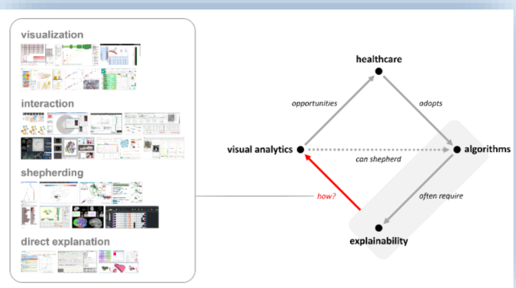

Distribution of chosen slider values
1.0
0.5
0.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Group
Control
What-if
Wise feedback


Vraag
Wat is de hoofdstad van de staat Florida?
Antwoord
Tallahassee
Kwaliteit antwoorden
Alles
Goed
Zeer goed
Gewicht afleider/gelijke vragen
Afleiders worden gesorteerd op basis van de kwaliteit van de afleider zelf, alsook de mate waarin de vraag waarin ze voorkomen lijkt op de input-vraag. Bij het sorteren van de resultaten kan je kan kiezen of de afleider zwaarder doorweegt, of de gelijkaardige vraag.
Afleider
Standaard
Gelijke vragen
Miami
Jaipur
New York
Limburg
Noorwegen

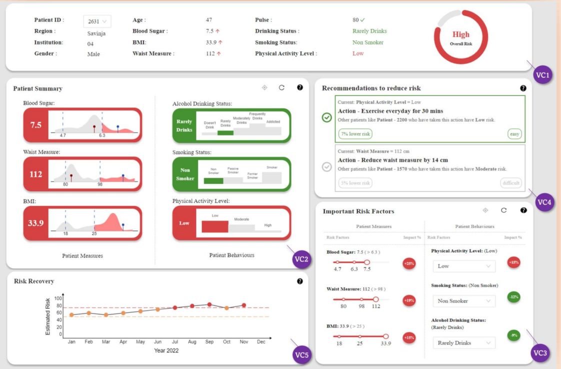

Vraag
Wat is de hoofdstad van de staat Florida?
Antwoord
Tallahassee
Kwaliteit afleiders
Goed
Laag
Gemiddeld
Hoog
Gelijkaardigheid
Afleiders
Afleider
Vragen
Miami
New York
Los Angeles
Limburg
Noorwegen
Zuidpool

Explainable AI through visualisation

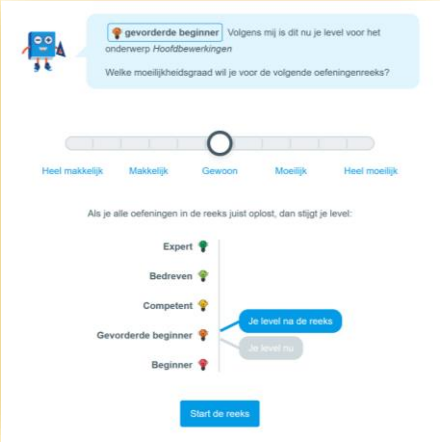
Visual analytics

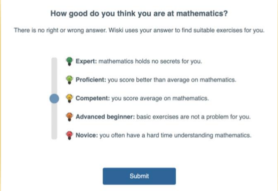


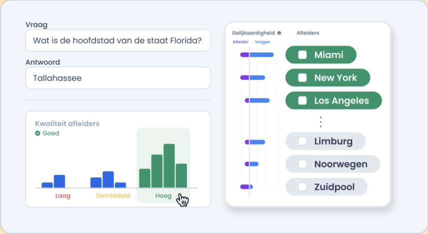
Transparency: justification



Transparency: control

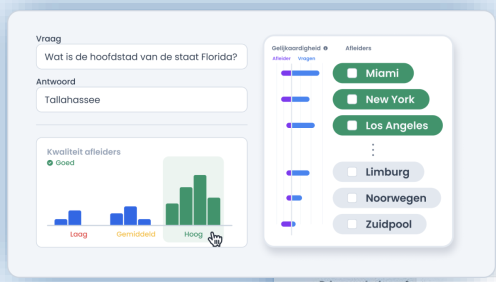






Explaining AI

with tailored interactive visualisations



Jeroen Ooge
jeroenooge.com
KU LEUVEN

Katrien Verbert
augment.cs.kuleuven.be
KU LEUVEN