Role of Case-Based Reasoning for XAI- Intelligent Reuse of Explanation Experiences

Nirmalie Wiratunga

Case-based Reasoning

Case-Based Reasoning (CBR)

- Idea: Drawing conclusions directly from stored **situation-specific experience knowledge**
- Situation-specific experience knowledge stored as tuples of past problem and corresponding solution descriptions – called **cases**

Description of past problem	Description of found solution	Outcome / Lessons Learnt

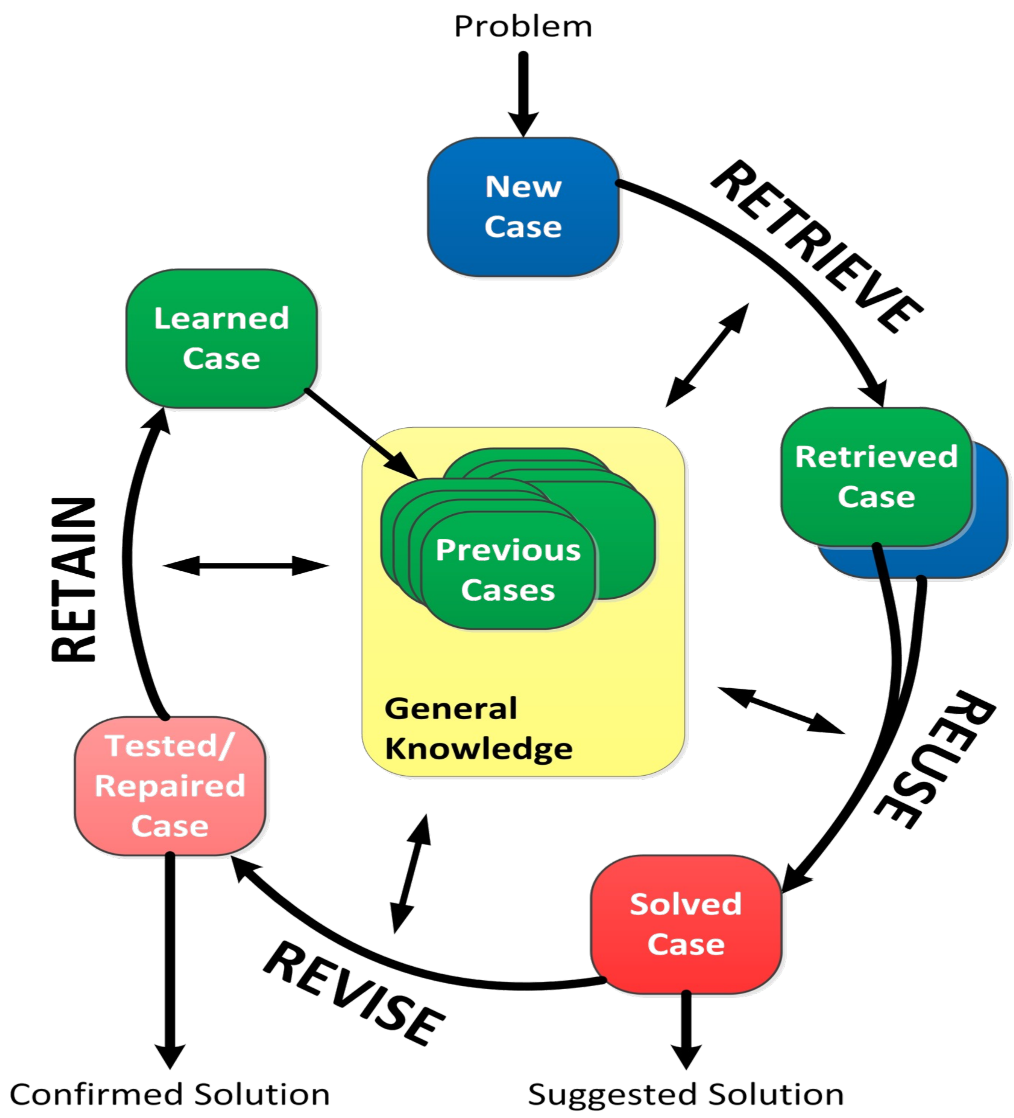
Case

- Solving of new problems by reusing solutions of similar, already solved problems stored in a **case base**

Adapted from Prof Kerstin Bach @NTNU

CBR Cycle

- CBR is neither …
  - a single algorithm
  - nor a collection of similar algorithms
- CBR is more …
  - a paradigm
  - a methodology
  - a general procedure for problem solving
- Overall process model used to describe how a CBR system works (4R Cycle)



Aamodt & Plaza, 1994

Recent CBR Projects AIR@RGU

2016-2021
SELFBACK

Self-management plan recommendation for non-specific low-back pain

2021-2024
ISEE

Reuse explanation experiences by users for users

Innovate UK

recommend asset-specific decommission programmes using asset end-of-use classification models

2019-2021
Prophecy

match job opportunities with appropriate individuals

2022-2023
Nudge

continuous learning and transparent decision-making for energy sector engineering

2024-2026
Katoni

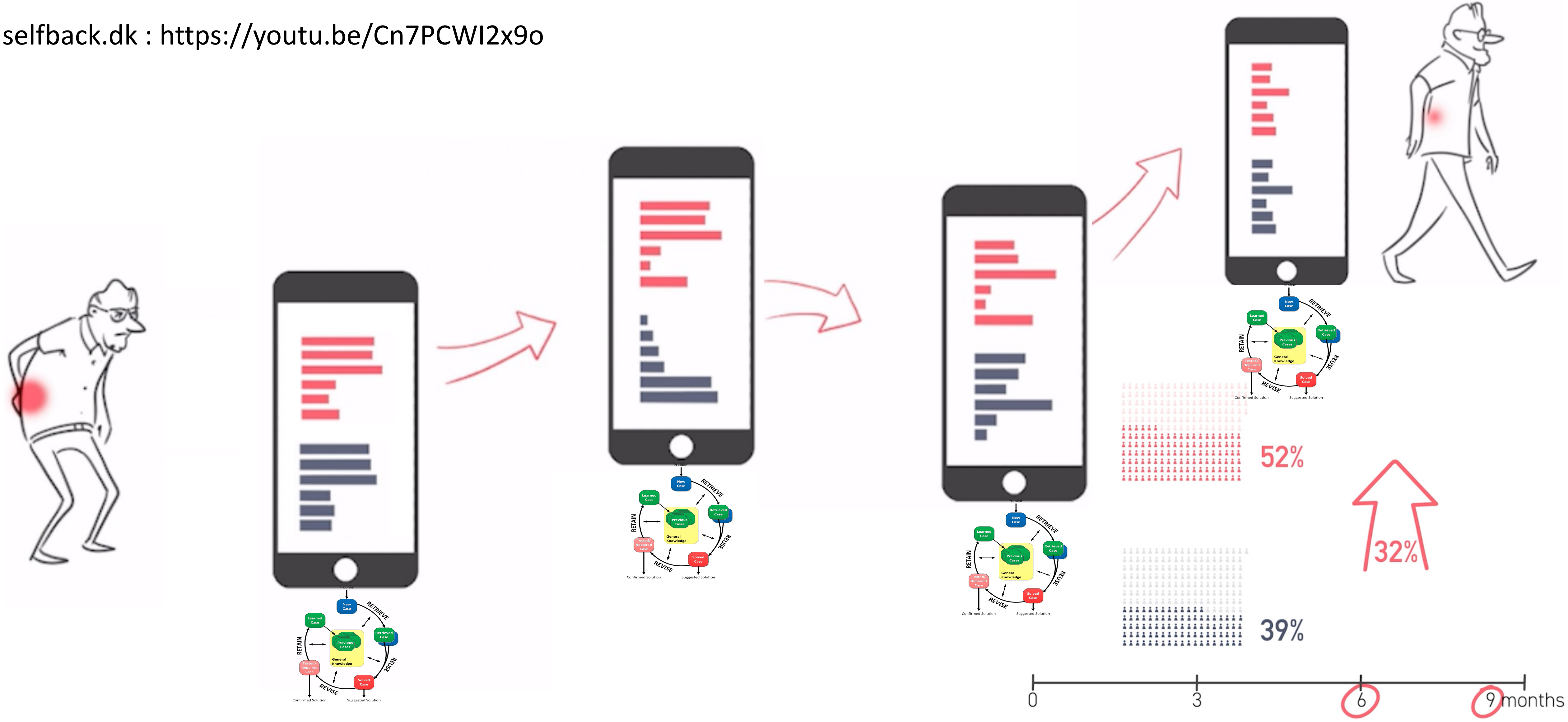
SELFBACK: Decision Support underpinned by CBR

How can we help people to get from here ... ... to here

Mike Evans: https://www.youtube.com/watch?v=BOjTegn9RuY&t=18s

SELFBACK: Personalised Recommendations

selfback.dk : https://youtu.be/Cn7PCWI2x9o

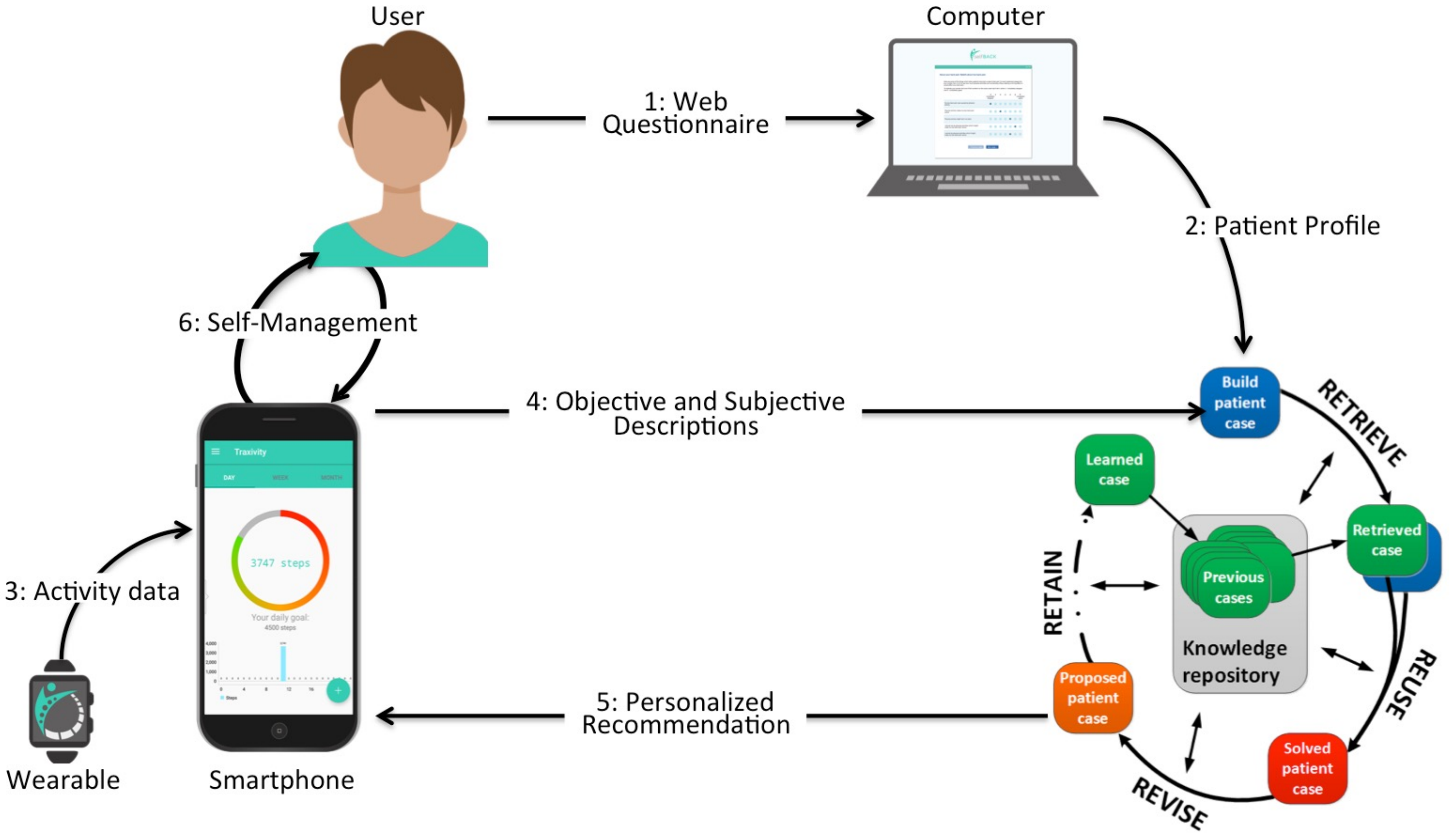


Commercial app : https://www.selfback.dk/en/home

CBR in SELFBACK

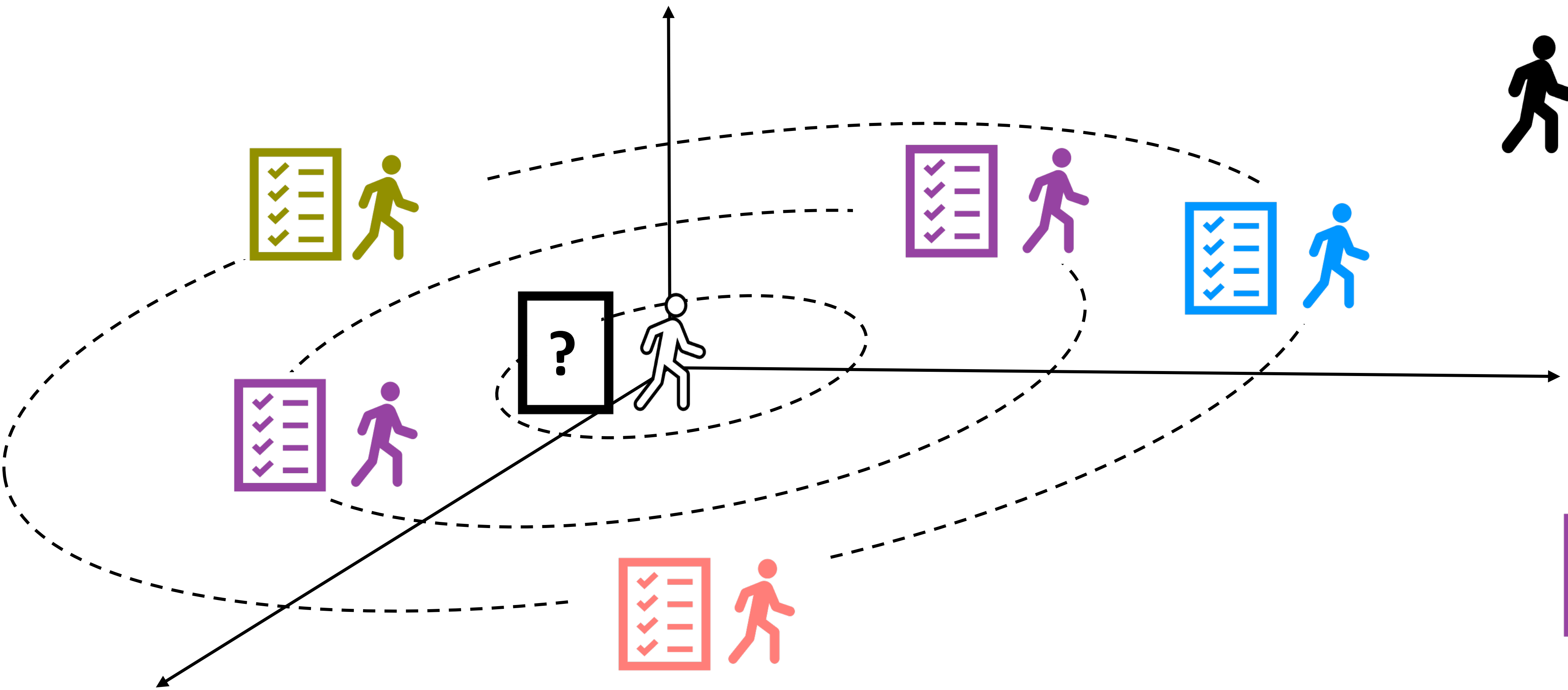
Personalised Self-management plan recommendation

- Physical activity
- Strength/flexibility exercises
- Patient education



Adapted from Prof Kerstin Bach @NTNU

Non-parametric Case Matching



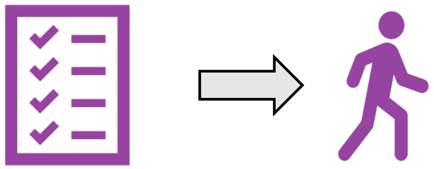
Case Description

- Demographics
- Quality of Life
- Pain Intensity
- Functionality
- Activity Stream

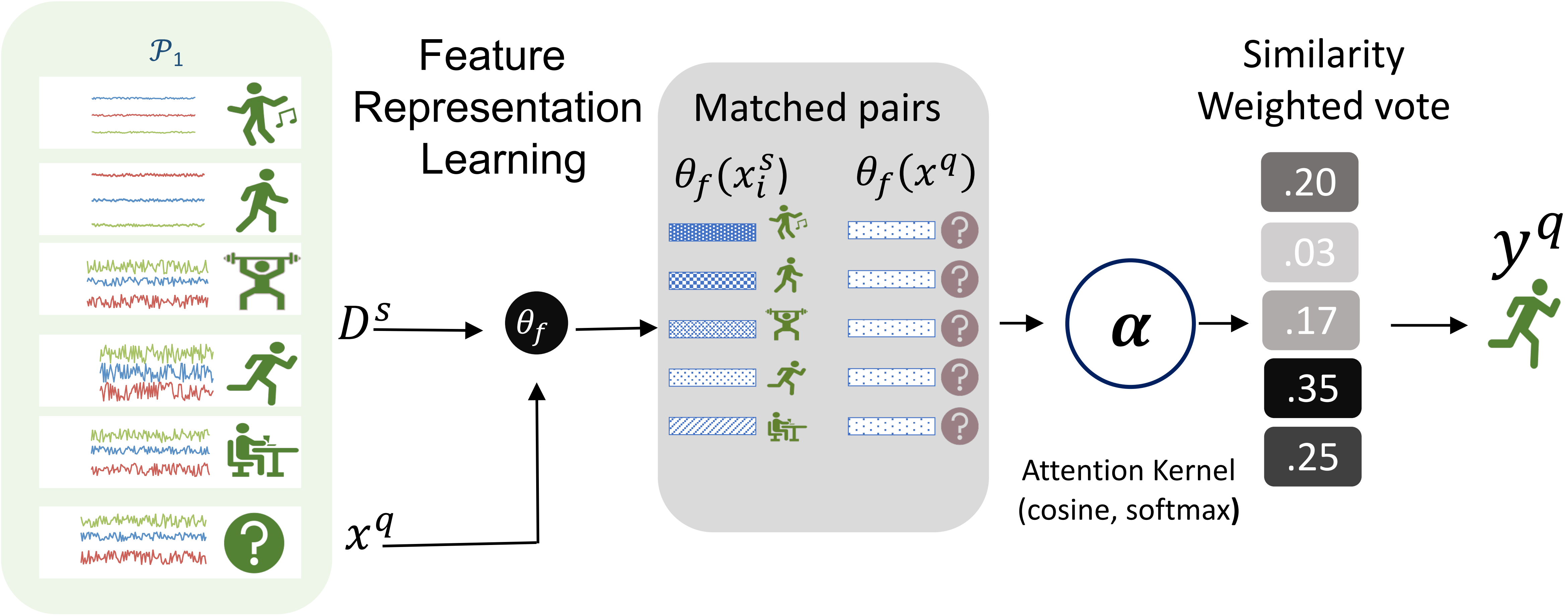
Case Solution

- Physical activity
- Strength/flexibility exercises
- Patient education

Case Retrieval



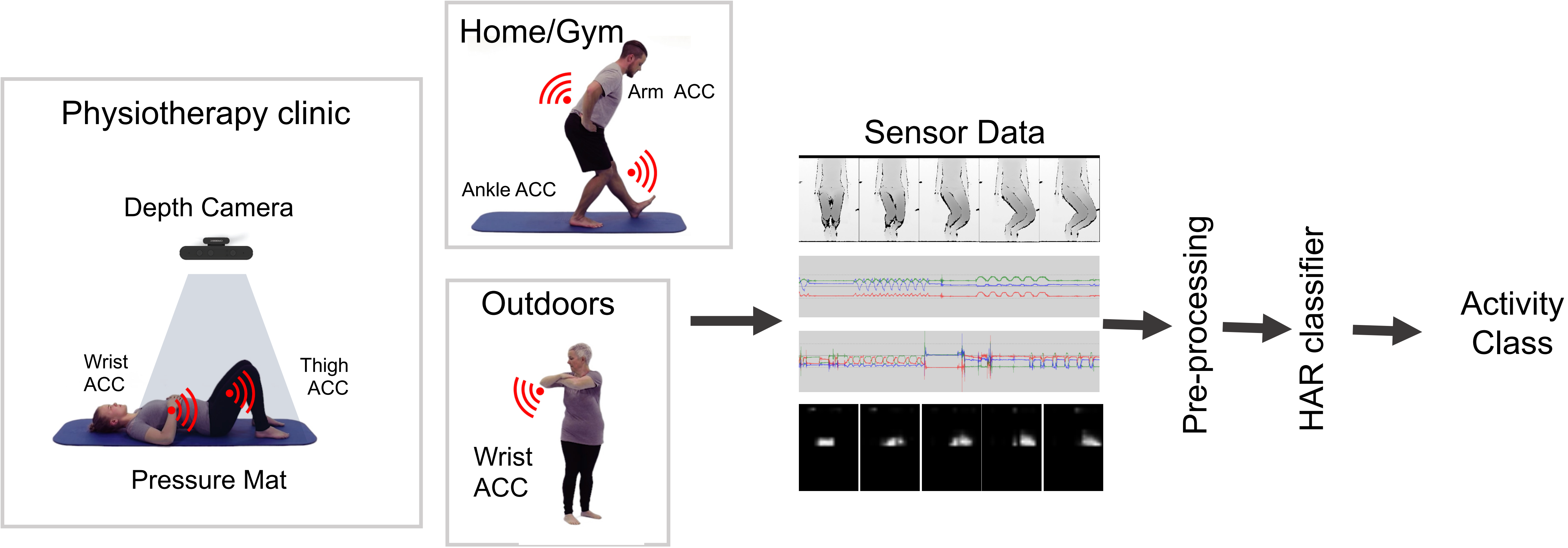
Parametric case matching with Matching Networks



Wijekoon A, Wiratunga N, Sani S, Cooper K. (2020) A knowledge-light approach to personalised and open-ended human activity recognition. Knowledge-Based Systems. 15;192:105651

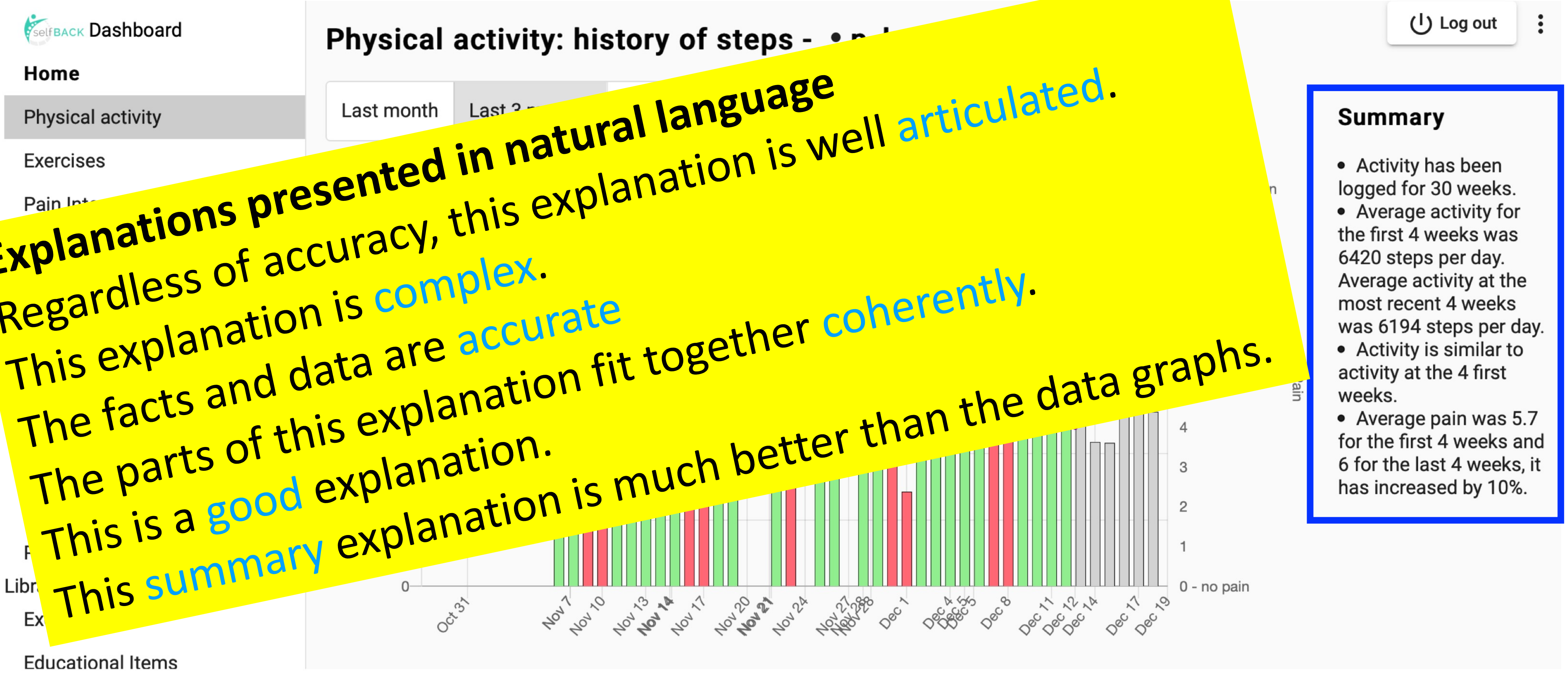
Multi-modal Plan Adherence Monitoring

http://archive.ics.uci.edu/ml/datasets/MEx



Wijekoon, A., Wiratunga, N., & Cooper, K. (2020, July). Heterogeneous multi-modal sensor fusion with hybrid attention for exercise recognition. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

Explainability in SELFBACK - Clinician Dashboard



Explanations presented in natural language

Regardless of accuracy, this explanation is well articulated.

This explanation is complex.

The facts and data are accurate

The parts of this explanation fit together coherently.

This is a good explanation.

This summary explanation is much better than the data graphs.

Adapted fom SelfBACK.eu: Clinician dashboard Screenshot

CBR and XAI in Literature

1988

Early papers

Schank R., “Explanation: A first pass”;1984
Experience, Memory, and Reasoning, J. Kolodner and C. Riesbeck (eds), **1986**

David B. Leake: Evaluating Explanations. AAAI 1988

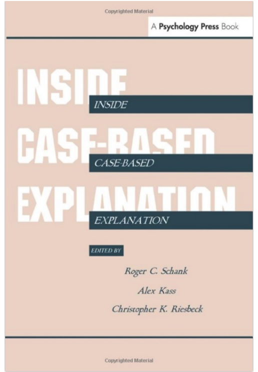
A. Kass: Adaptation-Based Explanation: Explanations as Cases. ML 1989: 49-51

R. Barletta, W. Mark: Explanation-Based Indexing of Cases. AAAI 1988

Roger C. Schank, David B. Leake: Creativity and Learning in a Case-Based Explainer. Artif. Intell. 40(1-3) (1989)

Early papers

1988



1994

1996

Inside Case-Based Explanation (book) 1994 R. C Schank, A. Kass, C.K Riesbeck (eds)

Agnar Aamodt: Explanation-Driven Case-Based Reasoning. EWCBR 1993: 274-288

Ashok K. Goel, J. William Murdock: Meta-Cases: Explaining Case-Based Reasoning. EWCBR 1996: 150-163

David B. Leake: Abduction, experience, and goals: a model of everyday abductive explanation. J. Exp. Theor. Artificial Intelligence 7(4): 407-428 (1995)

1988

Early papers

1994

1996

P. Cunningham, D. Doyle, J. Loughrey, **An Evaluation of the Usefulness of Case-Based Explanation,** ICCBR (2003)

D. Doyle, P. Cunningham, D. G. Bridge, Yusof Rahman: Explanation Oriented Retrieval. ECCBR (2004)

Thomas Roth-Berghofer: Explanations and Case-Based Reasoning: Foundational Issues. ECCBR (2004)

Sørmo, F., Cassens, J., Aamodt, A.: Explanation in Case-Based Reasoning – Perspectives and Goals. (2005)

C. Nugent, P. Cunningham.: **A case-based explanation system for black-box systems**. AI Review , (2005)

2004/2005

Thomas Roth-Berghofer, Michael M. Richter: On Explanation. KI 22(2): 5-7 (**2008**)

2008

Timeline CBR and XAI Research

S.Pedram, N. Wiratunga, D.Corsar, A.Wijekoon: Towards Feasible Counterfactual Explanations - A Taxonomy Guided Template-based NLG Method, **ECAI**-2023

Wiratunga, et. al, DiSERN: Discovering Counterfactual Explanations using Relevance Features from Neighbourhoods, IEEE, **ICTAI**-2021

Recio-Garcia, et. al, CBR-LIME: A Case-Based Reasoning Approach to Provide Specific Local Interpretable Model-Agnostic Explanations. **ICCBR** (2020)

Mark T Keane and Eoin M Kenny. Twin-Systems to Explain Artificial Neural Networks using Case-Based Reasoning: Comparative Tests of Feature-Weighting Methods in ANN-CBR Twins for XAI **IJCAI** (2019)

Li et. al, Deep Learning for Case-Based Reasoning through Prototypes: A Neural Network that Explains Its Predictions, **AAAI-18**

1988 Early papers

1994

1996

2004

2008

XCBR Workshop

2018

2020

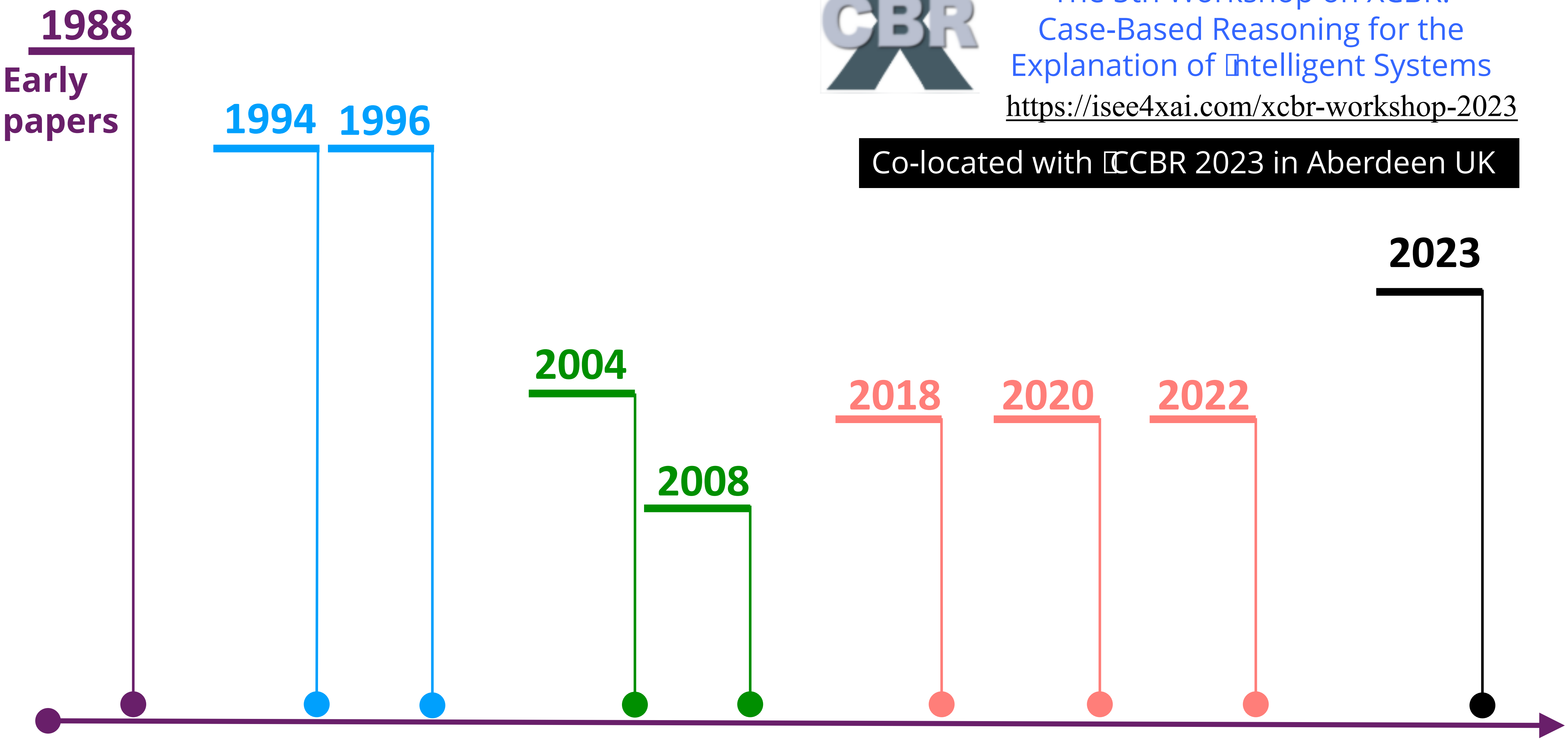
2022

Timeline CBR and XAI Research

The 5th Workshop on XCBR:
Case-Based Reasoning for the
Explanation of Intelligent Systems

https://isee4xai.com/xcbr-workshop-2023

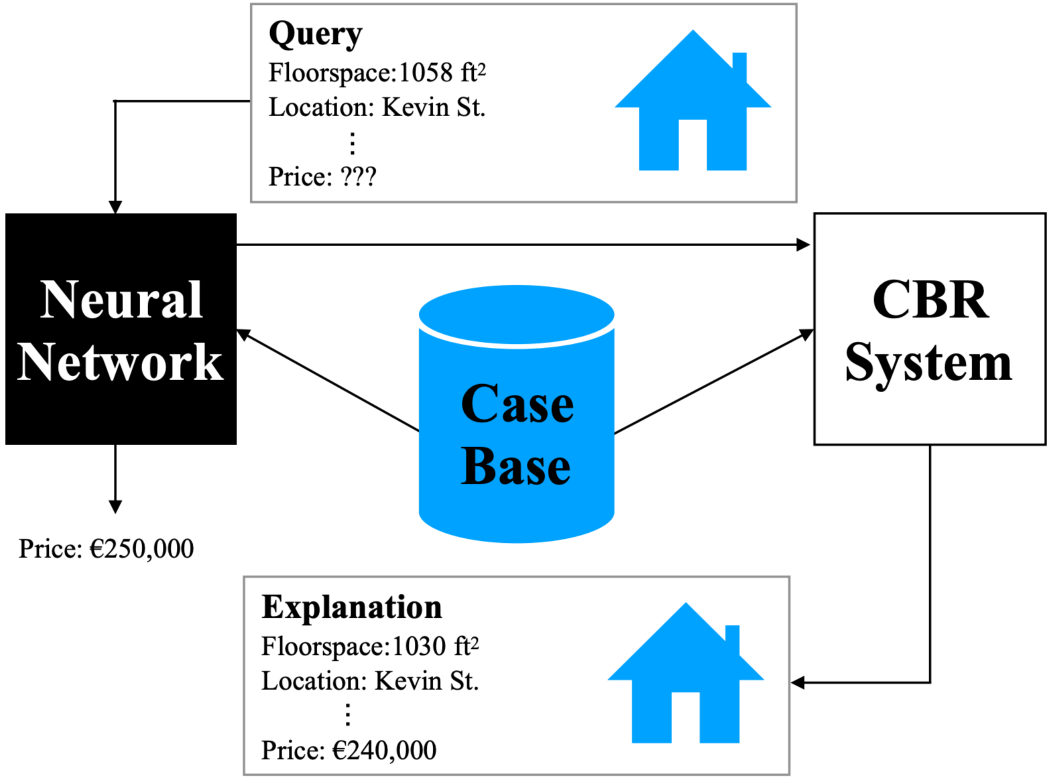
Co-located with ICCBR 2023 in Aberdeen UK



Timeline CBR and XAI Research

CBR as a twin system for explanation

- Model agnostic (post hoc) explanations
  - After-the-fact justifications for a prediction
- Use CBR as a surrogate interpretable model
  - KNN could be used for justification.
  - KNN is very transparent, and its answers should be relevant to the problem.
- But it needs to be faithful to the black box system

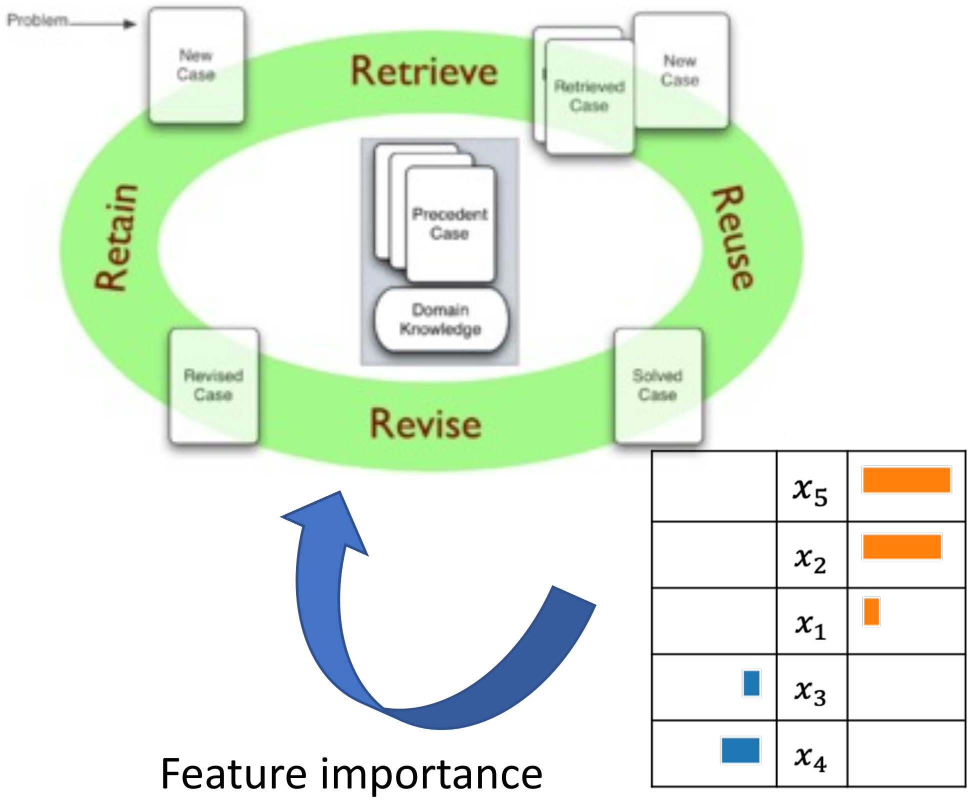


Kenny & Keane, 2019, Twin-Systems to Explain Artificial Neural Networks using Case-Based Reasoning: Comparative Tests of Feature-Weighting Methods in ANN-CBR Twins for XAI, IJCAI.

DisCERN: Explaining with counterfactuals

- Model agnostic & post-hoc
- Use retrieval to identify the neighbourhood
- Not all features are equally important
  - Importance changes locally

- Use adaptation operators (substitution) to generate counterfactuals
  - Use attribution explainers to guide adaptation

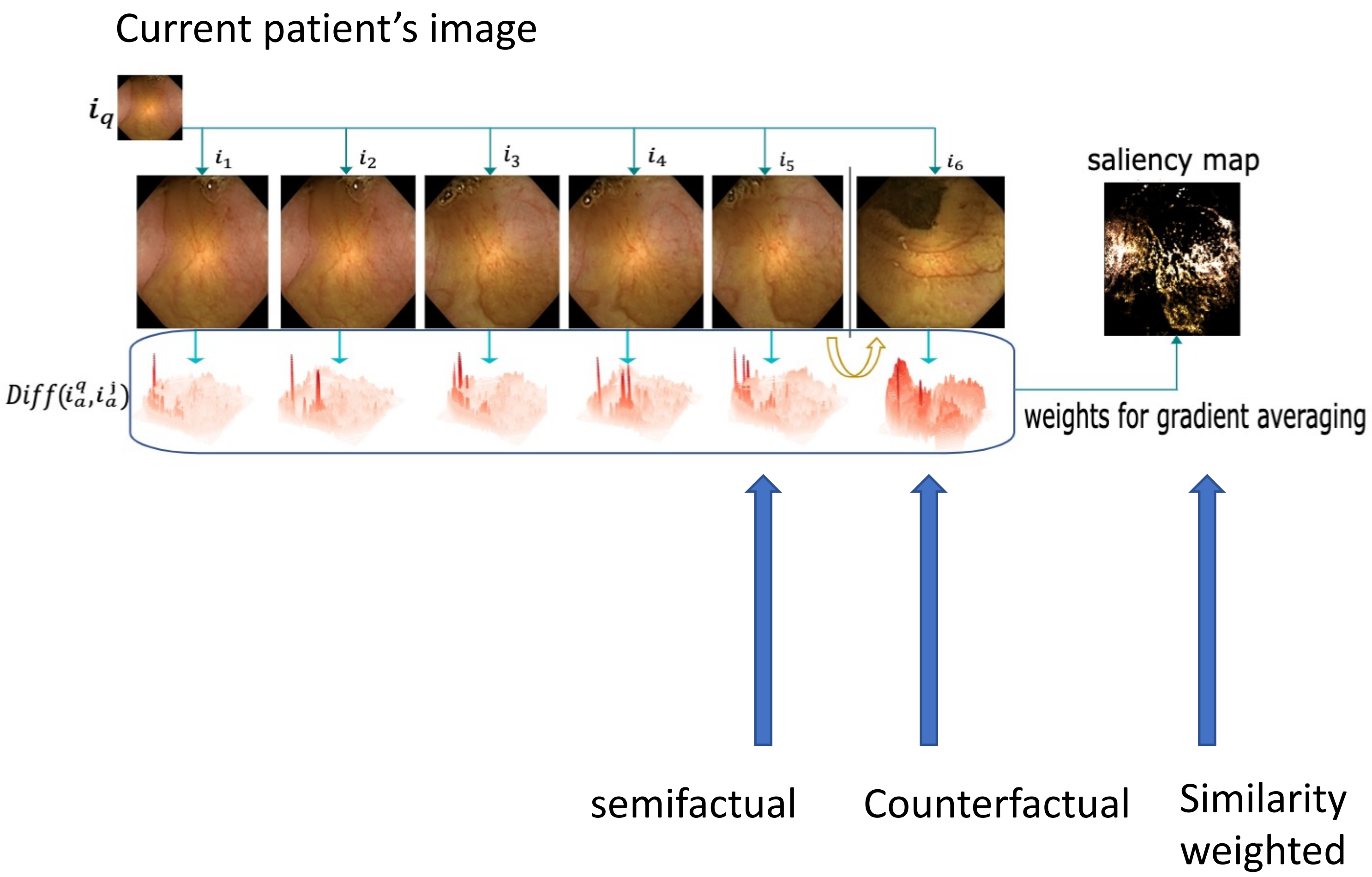


Feature importance
Guided adaptation

Wiratunga, N., Wijekoon, A., Nkisi-Orji, I., Martin, K., Palihawadana, C., & Corsar, D. (2021, November). Discern: Discovering counterfactual explanations using relevance features from neighbourhoods. In 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI) (pp. 1466-1473). IEEE.

Explaining with Semi-factuals

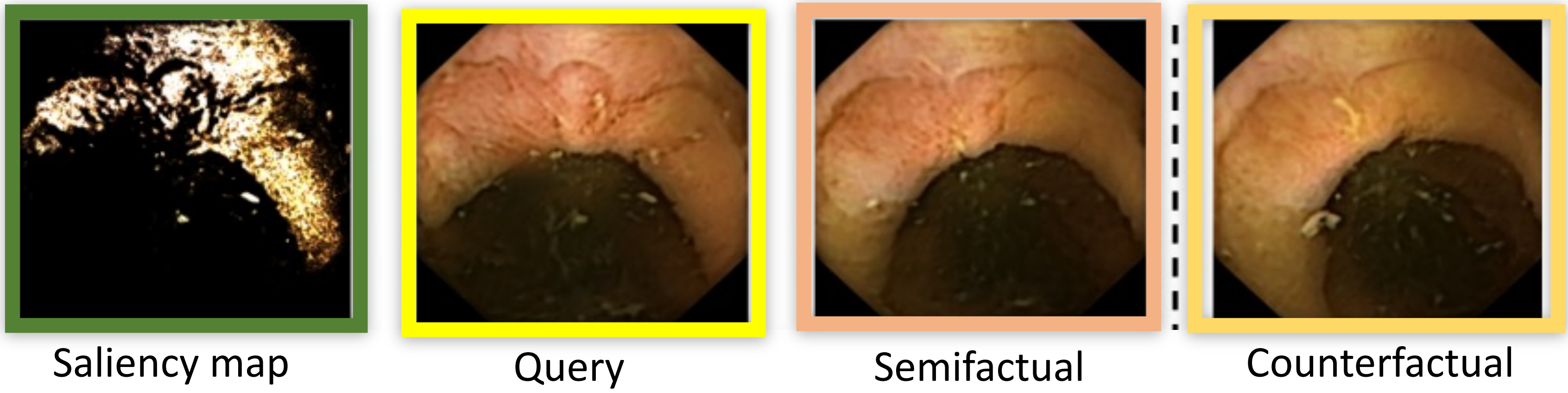
- Applied to medical images to predict abnormalities (e.g. ulcers)
- Generates a set of images to convey disease progression
- Ensures generation is plausible along latent attributes that have a causal relationship (e.g. inflammation)



Vats, A., Mohammed, A., Pedersen, M. and Wiratunga, N. [2023]. This changes to that: combining causal and non-causal explanations to generate disease progression in capsule endoscopy. IEEE (ICASSP 2023)

Explaining with Semi-factuals

Generated Explanation :



" Query image is abnormal due to signs/regions highlighted by the saliency map. The least amount of abnormality required for the prediction to be abnormal is seen in semifactual. However, if the abnormal signs change to as in counterfactual the image would no longer be classified as abnormal "

Vats, A., Mohammed, A., Pedersen, M. and Wiratunga, N. [2023]. This changes to that: combining causal and non-causal explanations to generate disease progression in capsule endoscopy. IEEE (ICASSP 2023)

CBR for sharing explanation experience by users for users

Why iSee? Social and Legal Implications

- EU GDPR'16 regulation subjects have a right to an explanation regarding decisions made using their data.
- Data subjects have a right to contest those decisions.

The General Data protection Regulation (GDPR)

- Limits to decision making based solely on automated processing and profiling (Art 22)
- Right to be provided with meaningful information about the logic involved in the decision (Art 13(2)f. & 15(1) h)

[Paul Nemitz, Principal Advisor, European Commission, Talk at IBM Research, 2018]

the need to develop a future regulatory framework - European Commission in 2020

New approach to regulating AI to build public trust – UK Gov White paper in 2023

iSee Platform: Why do we need it?

- The Idea: Capture, share and re-use experiences of AI explanations with other users who have similar explanation needs.
- Provide the AI community with a unifying open-source platform
  - Underpinned by CBR
- Enabling users to interact with, experiment with, and evaluate explanations
  - Design users and End-users
- Gather XAI best practices
  - Route to compliance

iSee

Intelligent Sharing of Explanation Experience by Users for Users

RGU

BT

Know your users - How do Humans Explain?

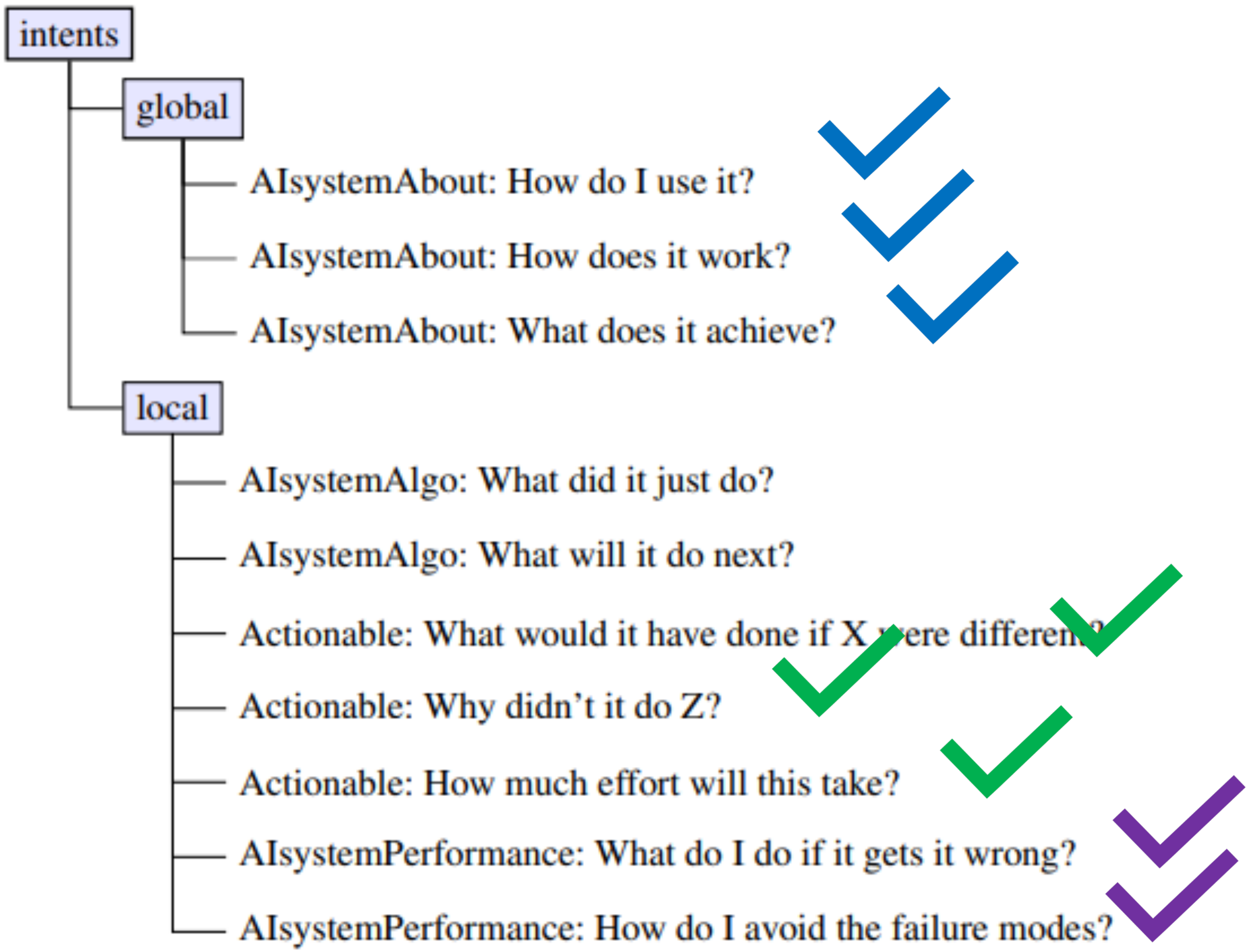
- Good explanation is Coherent
  - Parts of the explanation fit together
  - Compatible with existing beliefs, consistent with evidence
- Good explanation is Complete
  - No gaps in the explanation
- Good explanation is Articulate
  - Preference for complex explanations (multiple causal paths; explanation length)
- Good explanation has Alternatives

Zemla et al, 2017, Evaluating everyday explanations, Psychonomic Society

(ELI5; www.reddit.com/r/explainlikeimfive)
~7M unique visitors per month, Using 3 explanations per question

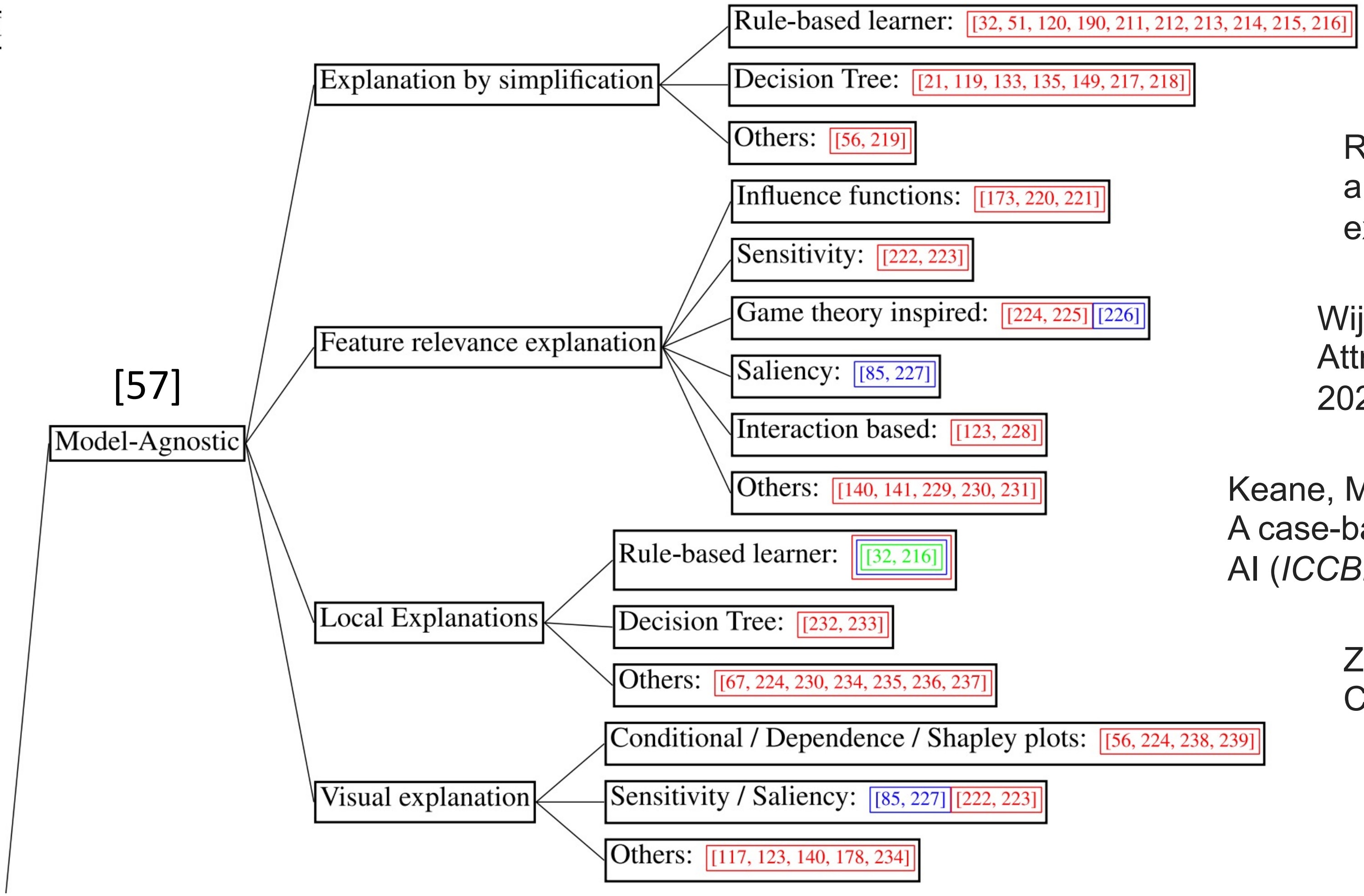
Human Explanation Intents

- Intent taxonomy
- To better understand
  - how the AI system functions
  - how to action a change of circumstances to drive a different outcome
  - How things are causally related



Adapted from Hoffmann, et. al. (2018)

Intents are met by relevant Explainers



Recio-García, J. A., et al. CBR-LIME: a case-based reasoning approach to provide specific local interpretable model-agnostic explanations. *ICCBR 2020*

Wijekoon, A., et al. How Close Is Too Close? The Role of Feature Attributions in Discovering Counterfactual Explanations. ICCBR 2022

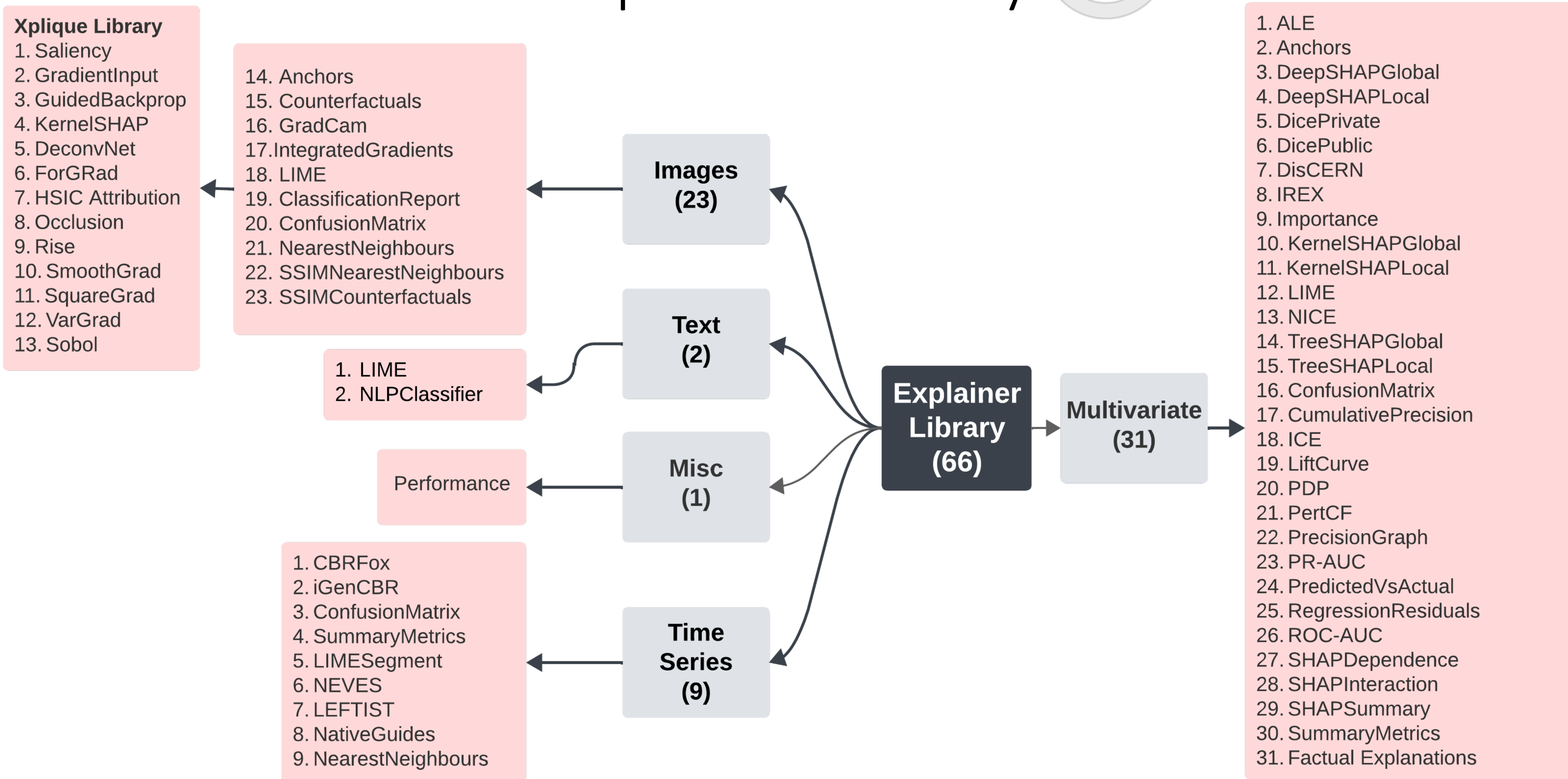
Keane, M. T., & Smyth, B. Good counterfactuals and where to find them: A case-based technique for generating counterfactuals for explainable AI (*ICCBR 2020)*

Zhao, Z., Leake, D., Ye, X., & Crandall, D. (2022). Generating Counterfactual Images: Towards a C2C-VAE Approach.

Kenny, E. M., & Keane, M. T. On generating plausible counterfactual and semi-factual explanations for deep learning. (AAAI 2021)

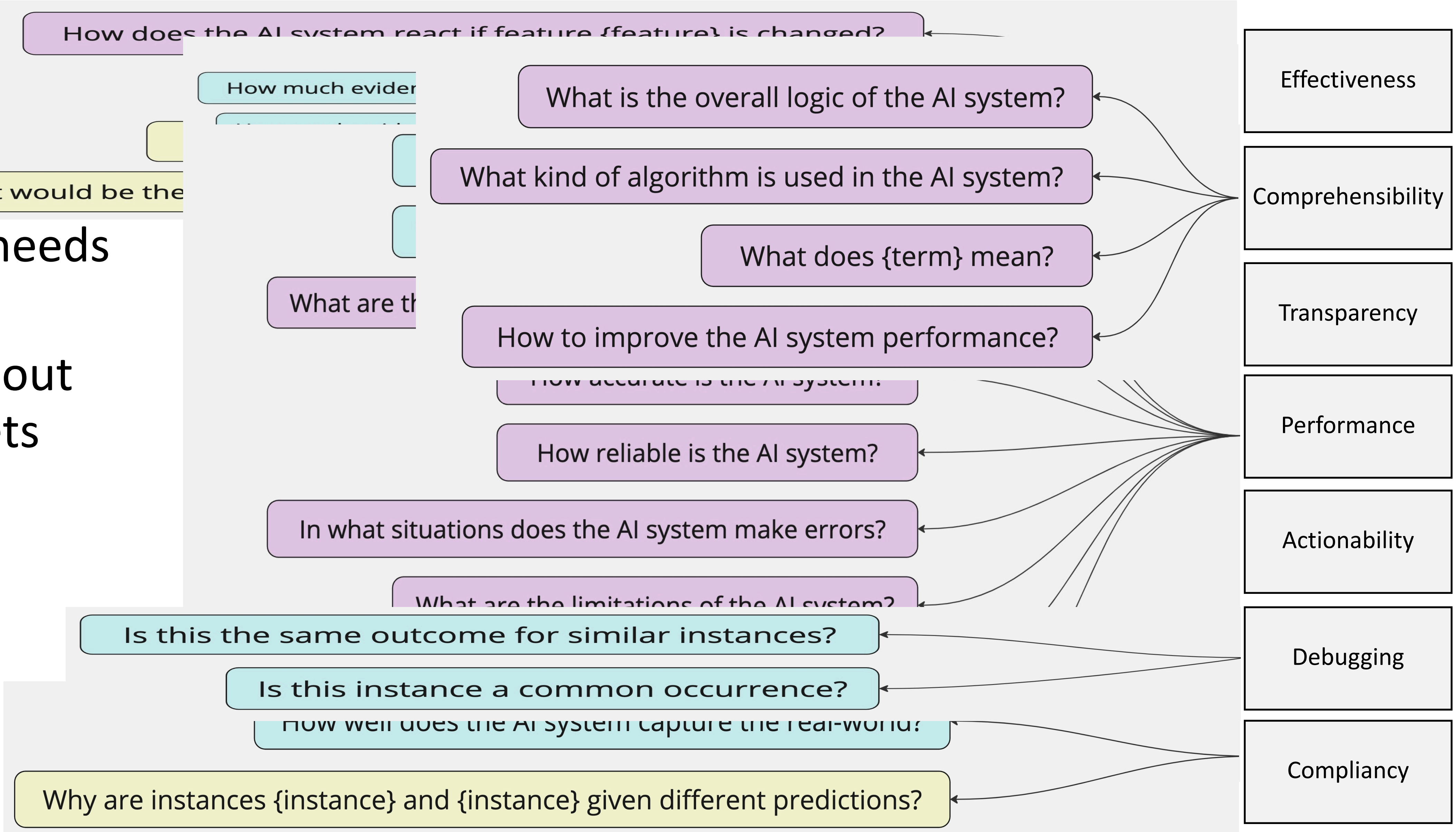
Extract from Arrieta, A. B., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, 82-115.

iSee Explainer Library



Explanation intents taxonomy

- Explanation needs as questions
- Questions about specific targets
  - dataset
  - model
  - prediction



Map Explanation need to explainer types

Explanation Type

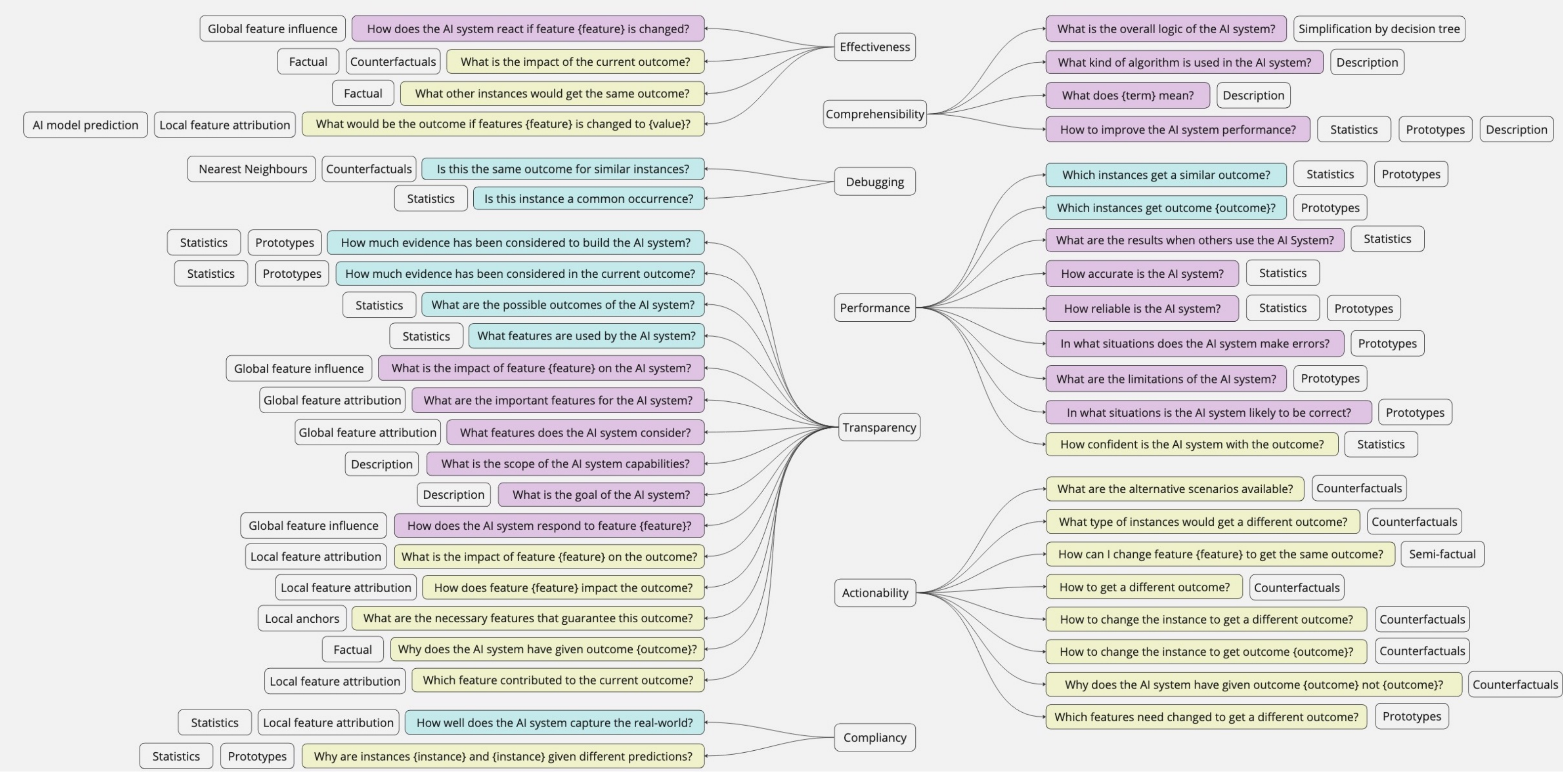
Nearest Neighbours | Counterfactuals | Statistics | Prototypes | Description | Global feature influence | AI model prediction

Simplification by decision tree | Local feature attribution | Global feature attribution | Factual | Semi-factual | Local anchors

Intents

- Effectiveness
- Comprehensibility
- Transparency
- Performance
- Actionability
- Debugging
- Compliancy

- Explanation needs as questions
- Questions about specific targets
  - dataset
  - model
  - Prediction
- Explainer types mapped to intents via questions

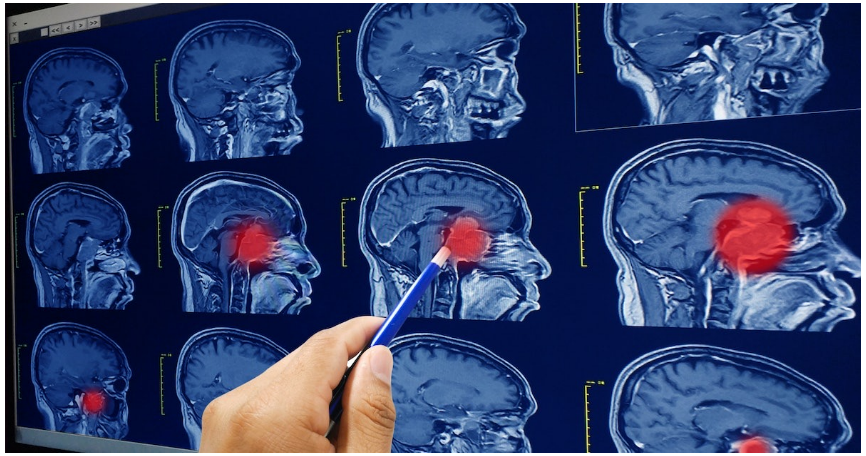


iSee Users

IAAA-MX

brainguard RESEARCH

Prediction of extreme natural events



Brain health monitoring with PET

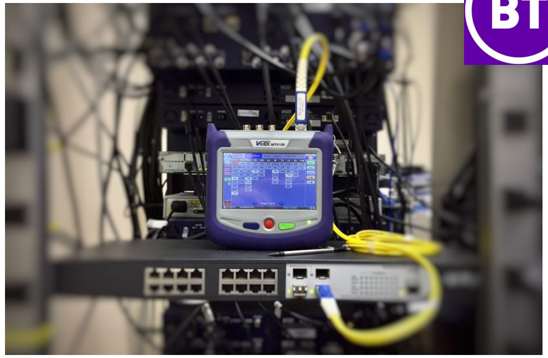
Jiva.ai

Radiology fracture detection

BOSCH

Anomaly detection in production lines

BT



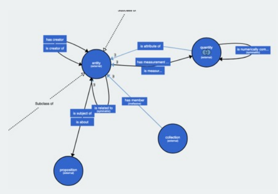
“next” actions recommendation in complex telecom service workflows

https://isee4xai.com/usecases/

CBR in iSee

What is an Explanation Experience?

- Problem description
  - User seeking an explanation
  - Application domain
  - AI Model that generated output being explained
  - Explanation needs
- Solution
  - The explanation strategy with explainers used
- Outcome
  - The user's evaluation of the explanation strategy



iSee Ontology

User Group

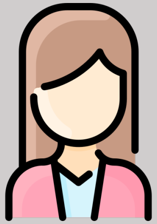
AI System and Domain

Explanation Needs

Explanation Strategy

Evaluation Outcomes

one-explainer-(does not)-fit-all



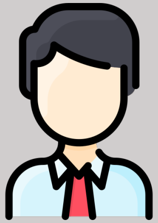
Automated Loan Approval System

Loan Applicant whose loan application was rejected

How can I get my loan approved?

Counterfactual & Semi-factual Explainers

Now I know what to do next time I apply for the loan!



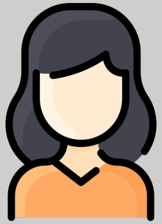
Automated Loan Approval System

Training Loan Officer learning about different scenarios

Why was this loan rejected?

Local Feature Attribution Explainer

I need more evidence on why the loan was rejected!



Automated Loan Approval System

Auditor carrying out a regulation check for fairness

What attributes are used by the automated system?

Global Feature Attribution Explainer

I am concerned on how much age affects the loan outcome!

iSee CBR cycle

Design User

1 Knowledge capture for case description

2 Retrieve explanation strategies from similar past situations

5 Revise explanation strategies

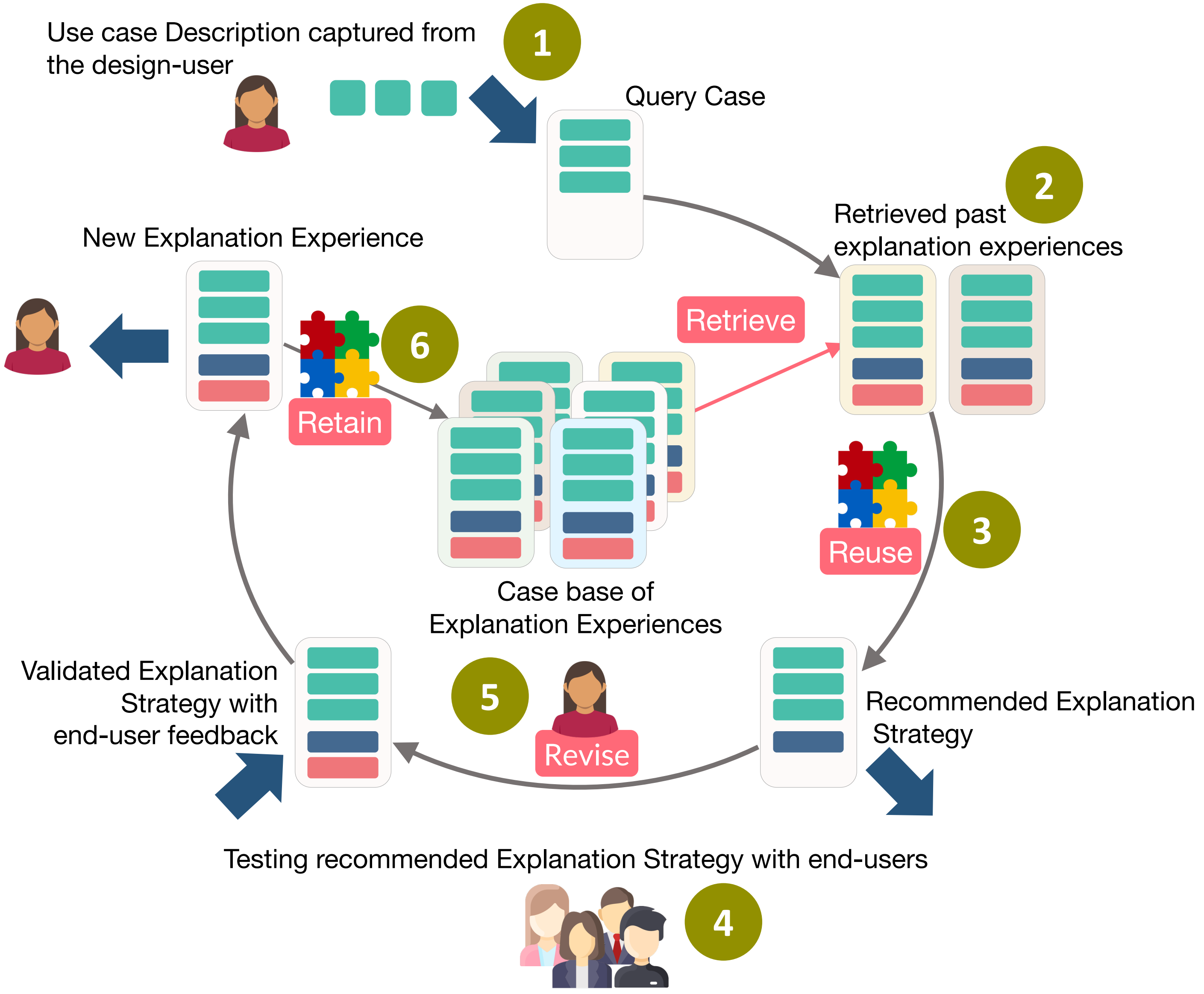
End Users

4 Conversational feedback gathering for collaborative case revision and retention

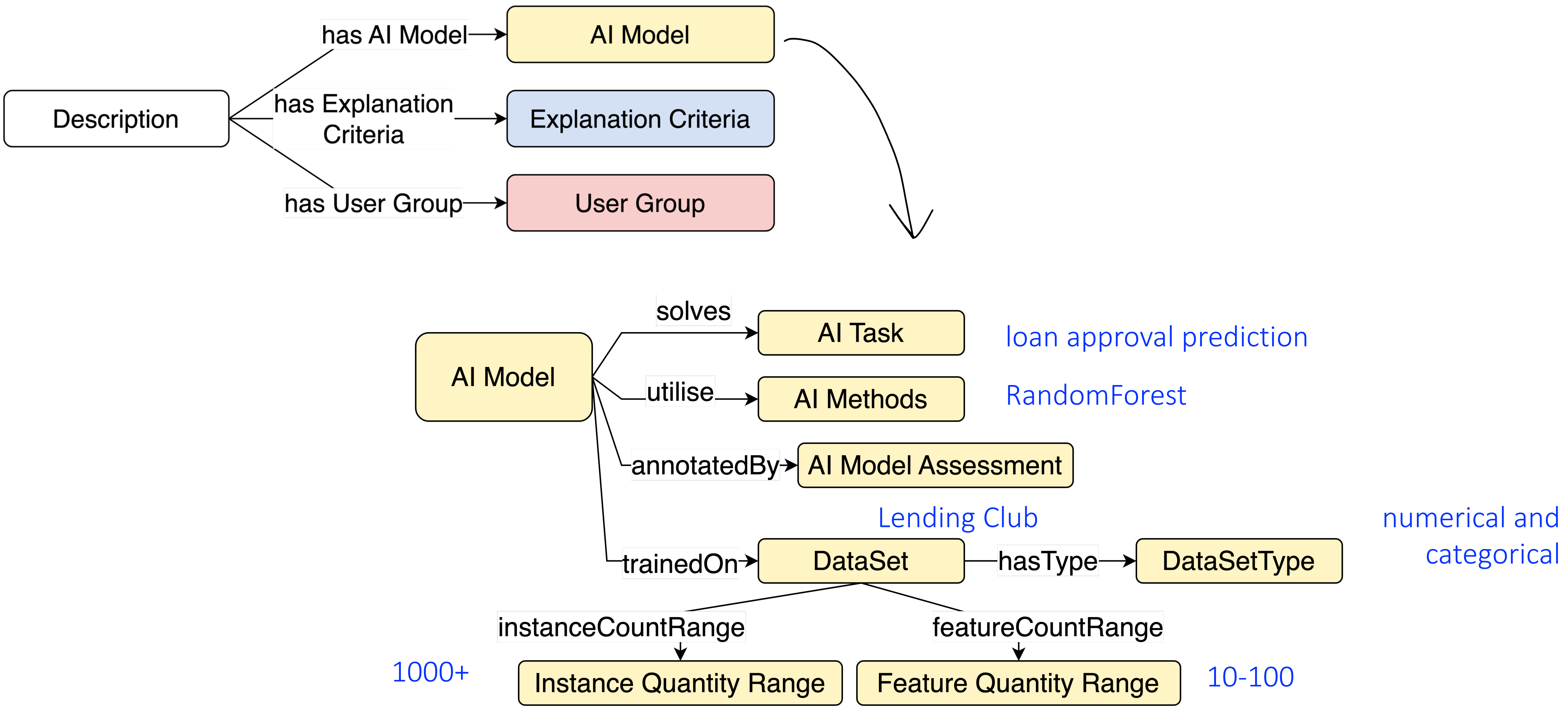
iSee

3 Transformational and Constructive Adaptation Algorithms

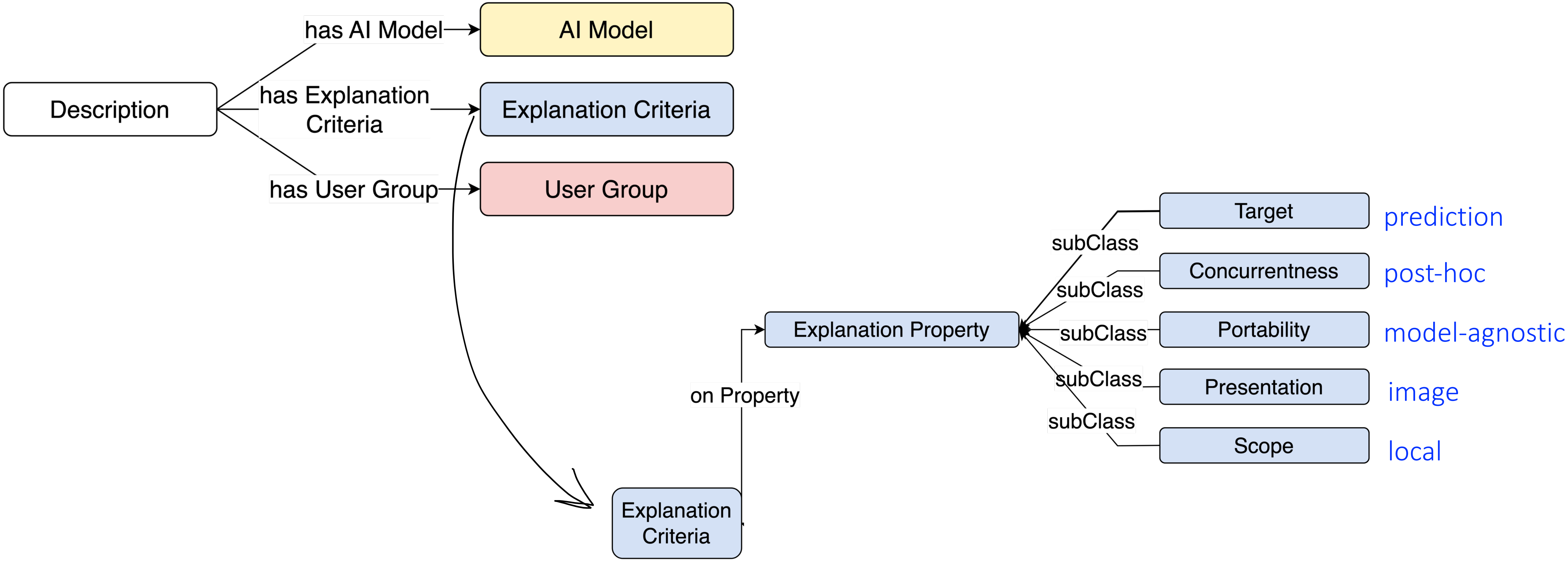
6 Retain Algorithms



What does a case represent? the problem part

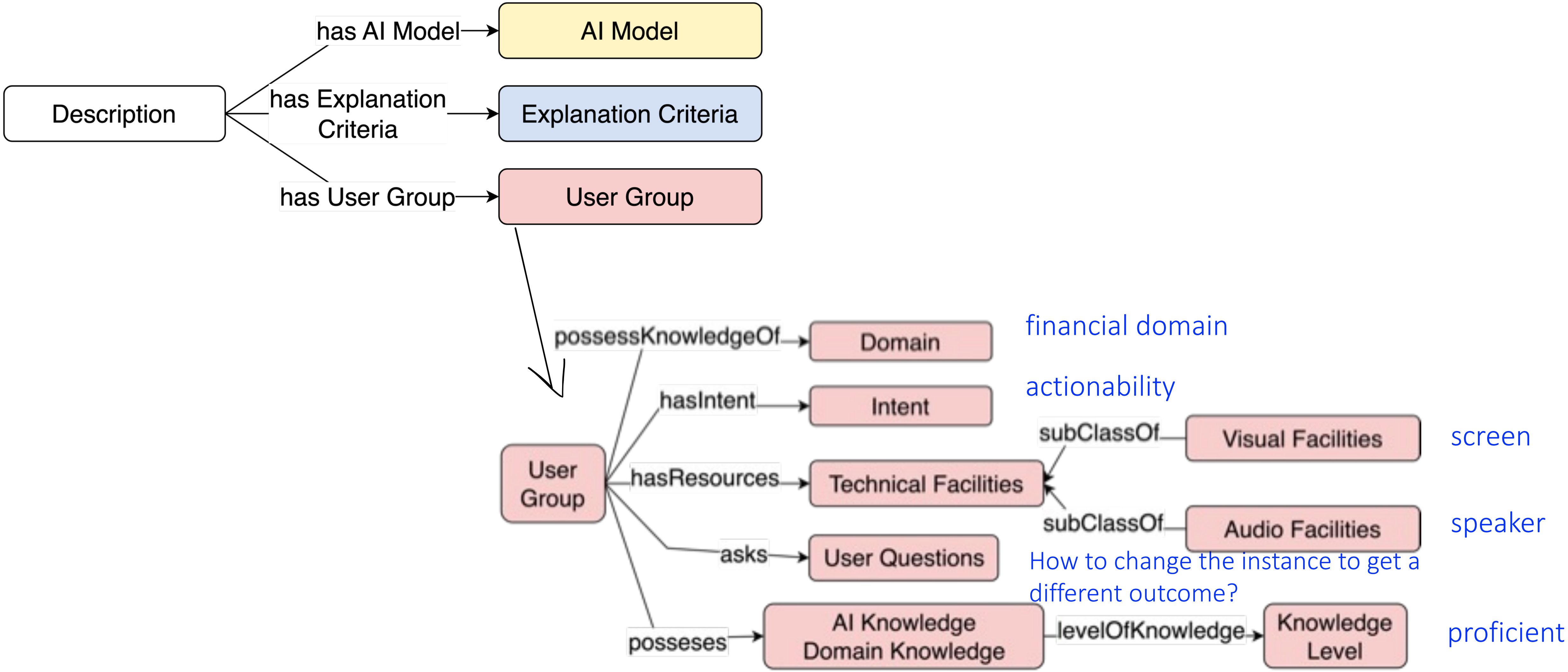


What does a case represent? the problem part



https://purl.org/heals/eo

What does a case represent? the problem part



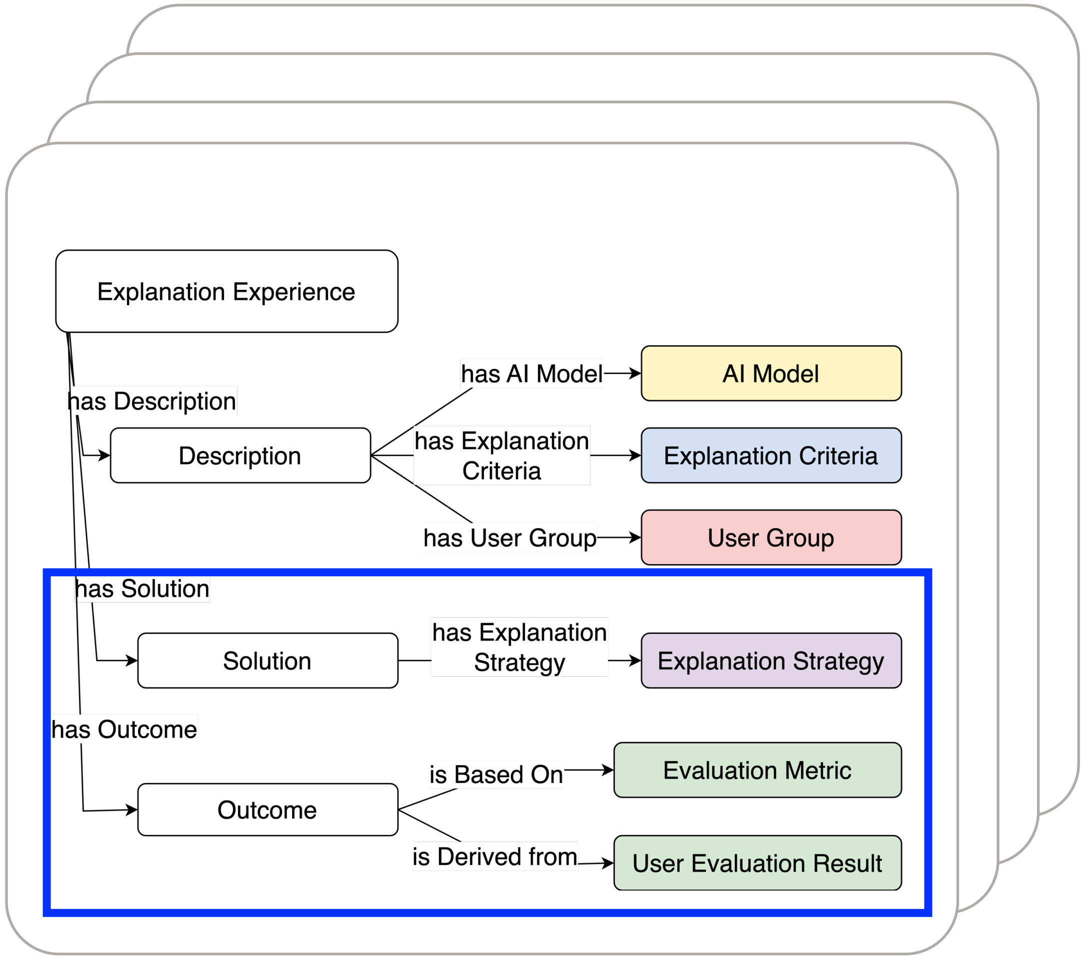
Case retrieval to recommend candidate explanation strategies

Ontology	Case Attribute	Ontology Component	Similarity Metric	Solution
AI Model	AI Task	Class	Wu&Palmer [22]	-
	AI Method	Class	Wu&Palmer [22]	-
	Dataset Type	Individual	Exact Match	-
Explanation Criteria	Portability	Individual	Exact Match	-
	Scope	Individual	Exact Match	-
	Target	Individual	Exact Match	-
	Presentation	Class	Exact Match	-
	Concurrentness	Individual	Exact Match	-
User Group	Intent	Individual	Exact Match	-
	TechnicalFacilities	Individual Set	Query Intersection	-
	AIKnowledgeLevel	Individual	Exact Match	-
	DomainKnowledgeLevel	Individual	Exact Match	-
	User Questions	Individual Set	Query Intersection	-
Behaviour Tree	Explanation Strategy	N/A	N/A	✓

CloodCBR

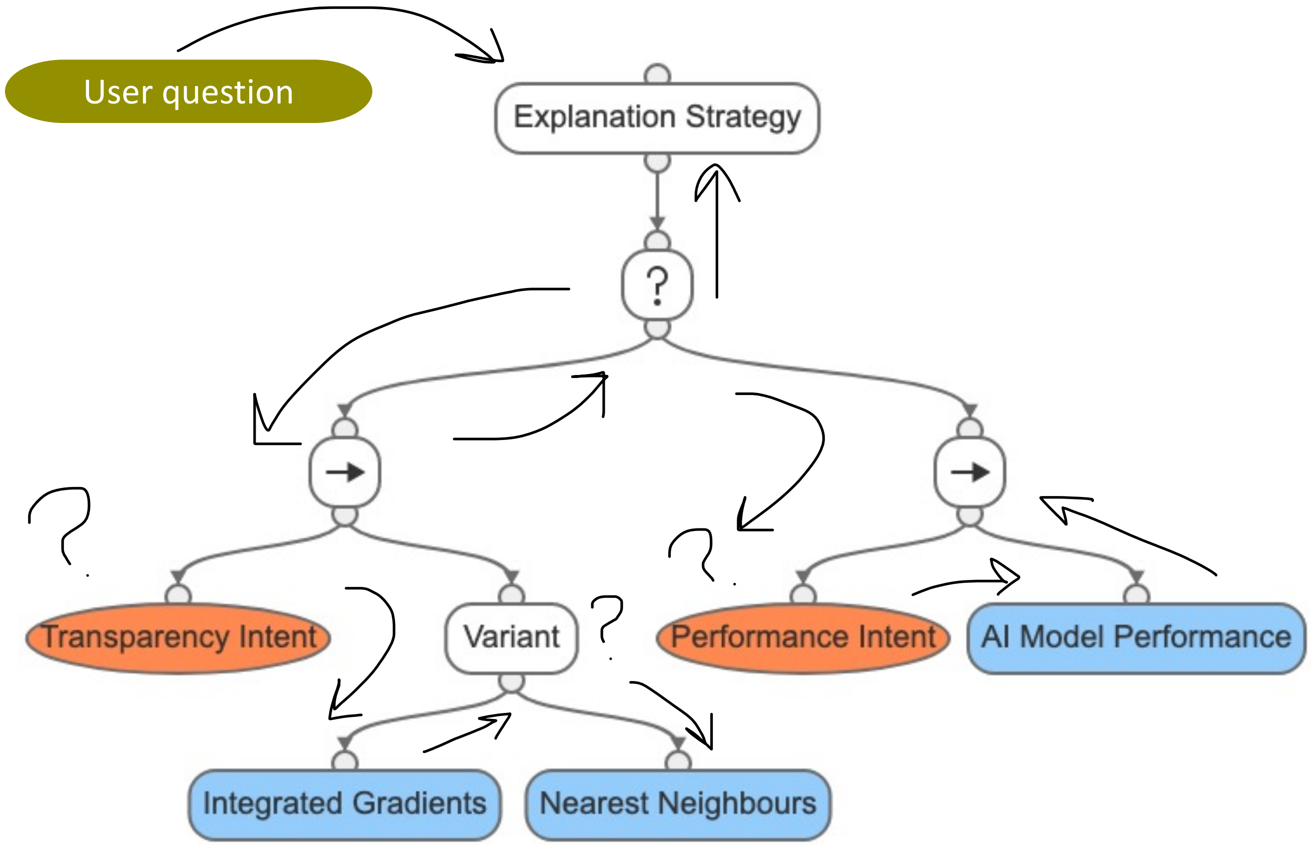
Case retrieval to recommend candidate explanation strategies

- iSee Case base currently holds 17 seed cases from the literature
  - filtered list of cases from a literature review of 50 peer-reviewed papers
  - Casebase to mature with Retain



https://cloodcbr.com/ CloodCBR

What does a case represent? the solution part

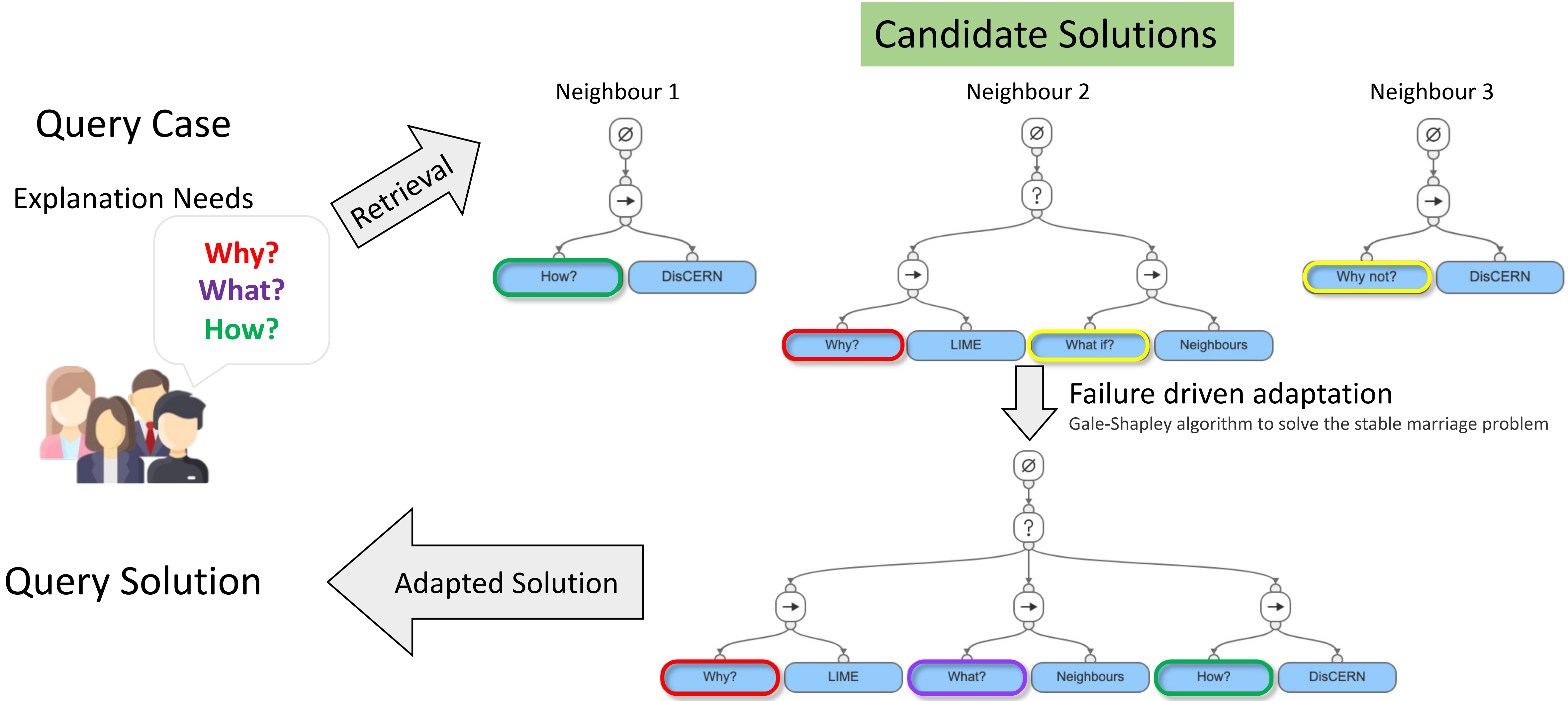


"If user indicate transparency intent, execute Integrated Gradients Explainer and present feature attributions"

"if they would like to verify with a different explainer execute Nearest Neighbours Explainer and present examples"

"else if user indicate performance intent, present AI model performance metrics to the user"

Transformational reuse to adapt solutions

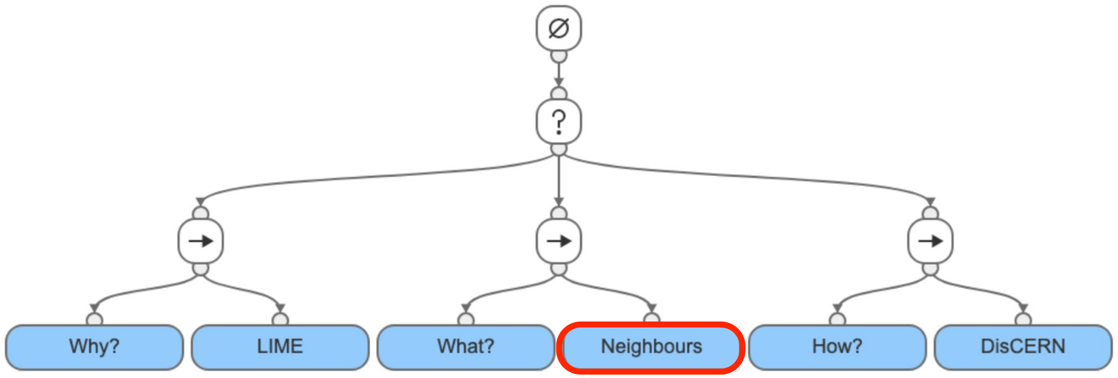


Nkisi-Orji, I., et al. (2023, July). Failure-Driven Transformational Case Reuse of Explanation Strategies in CloodCBR. In International Conference on Case-Based Reasoning (pp. 279-293). Springer Nature Switzerland.

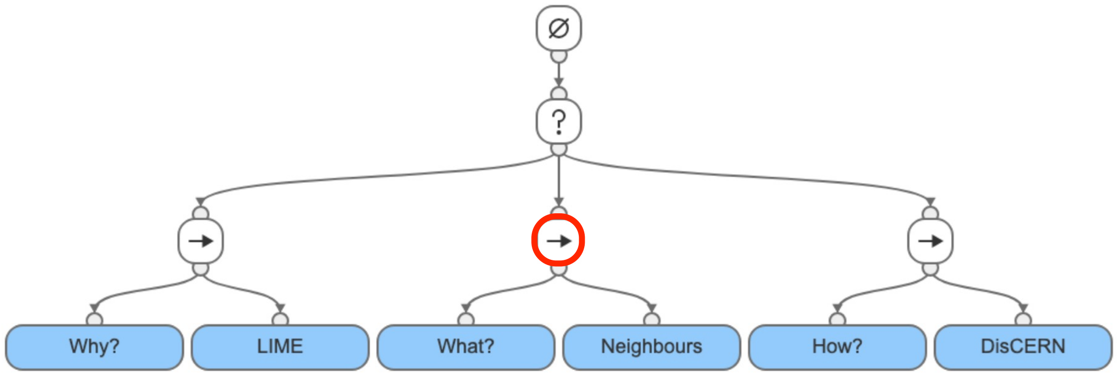
CloodCBR
https://cloodcbr.com/

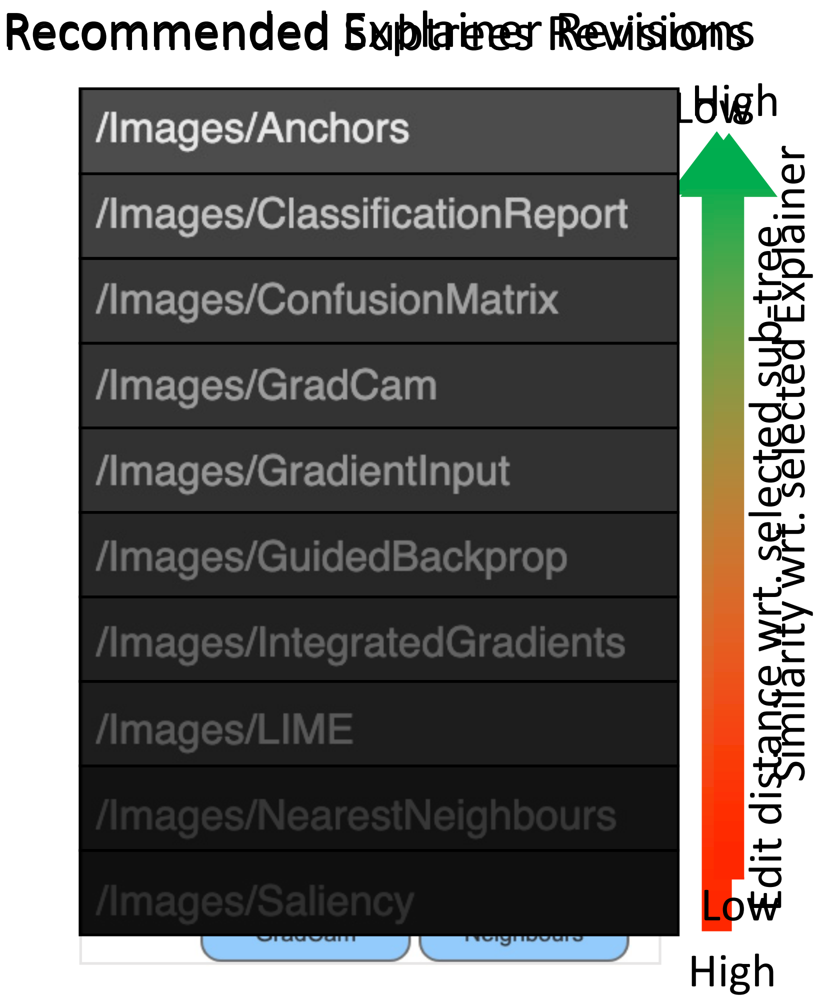
Constructive adaption for revision

Revise Explainers



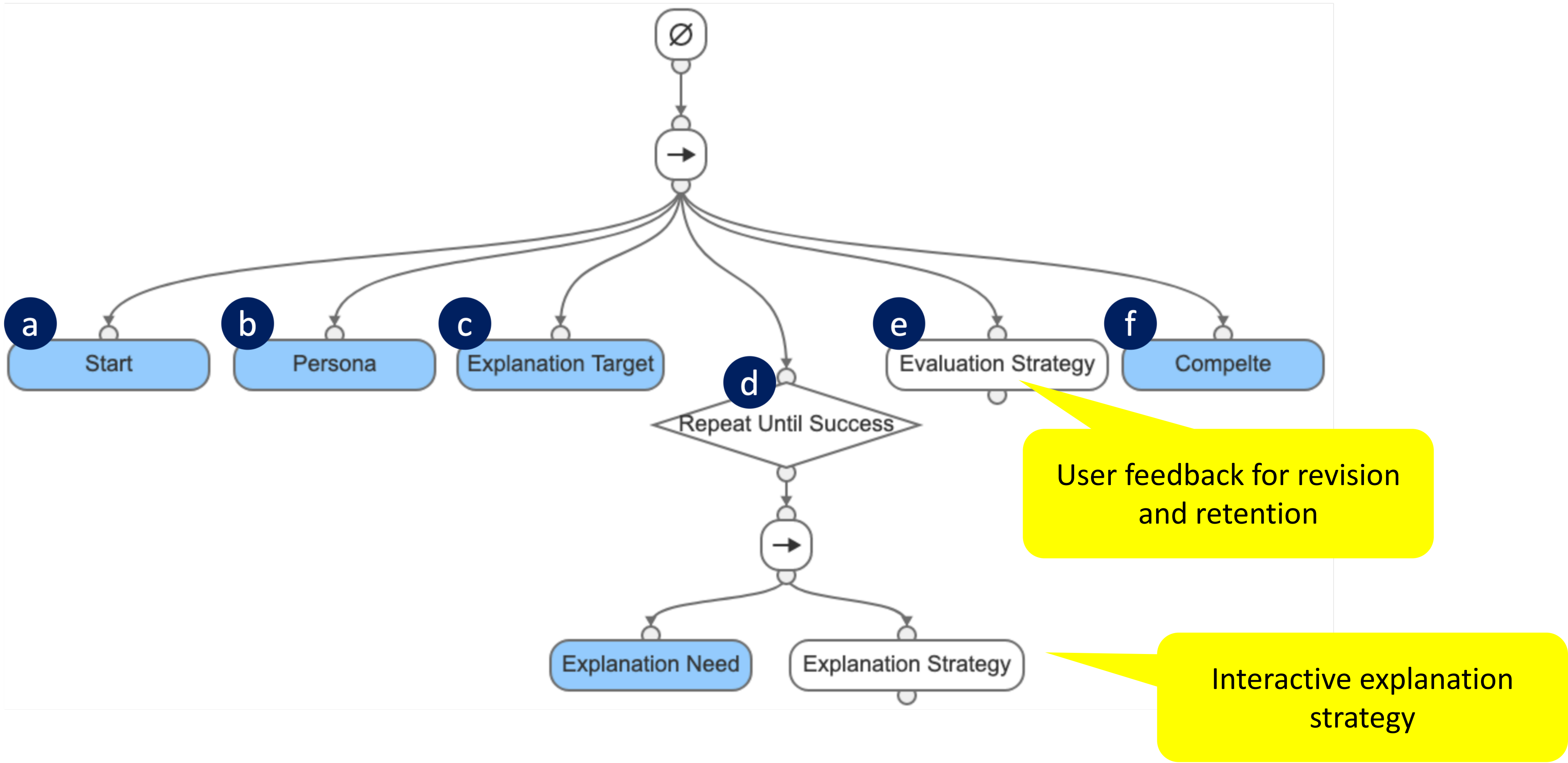
Revise Subtrees





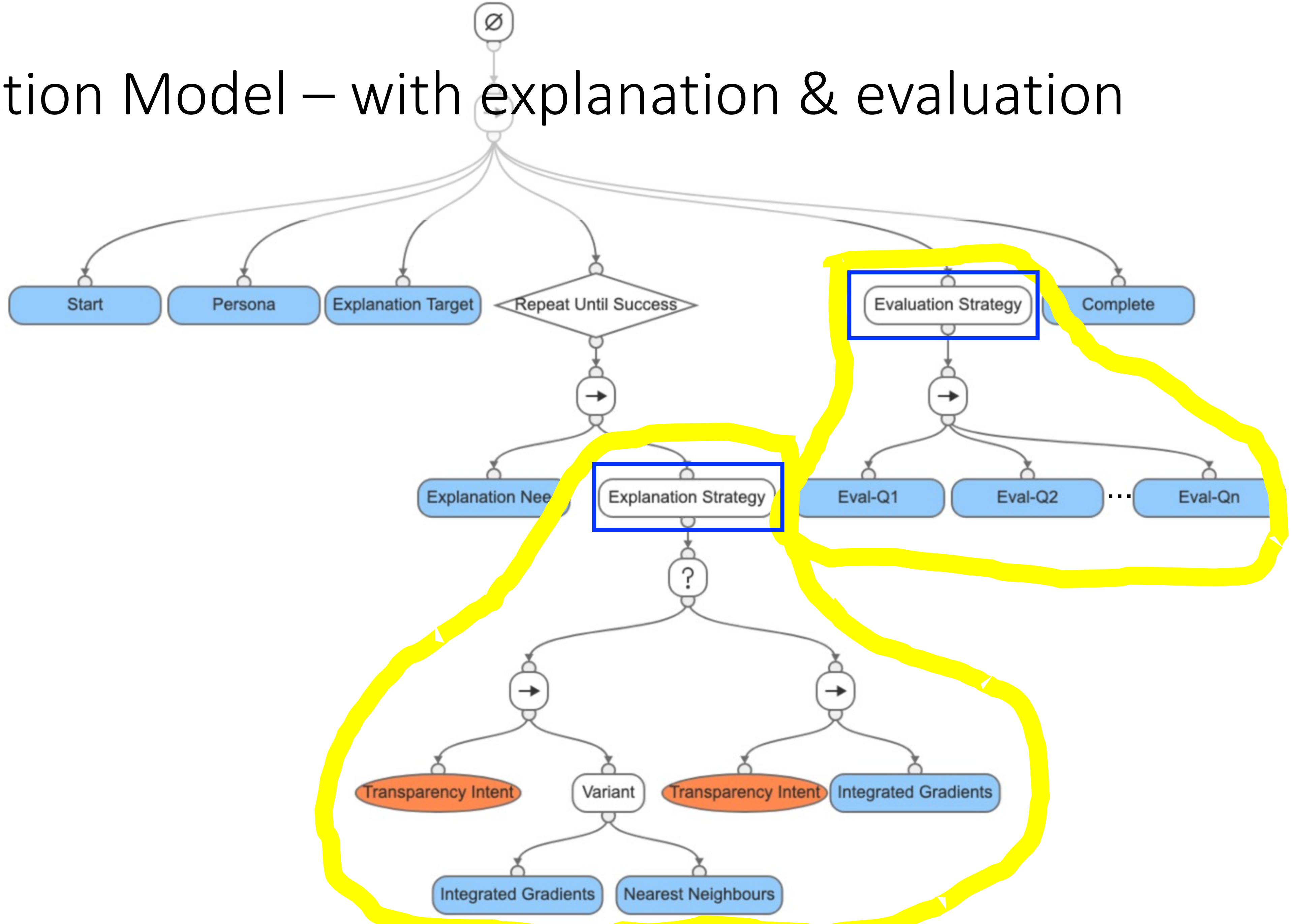
Caro-Martinez. M, et al. (2023) Unveiling the Potential of Semantically-backed Behaviour Trees for Retrieving Explanation Experiences: A Comparative Study, Under Review for Elsevier Journal Expert Systems with Applications

How to create explanation experiences? iSee Interaction Model

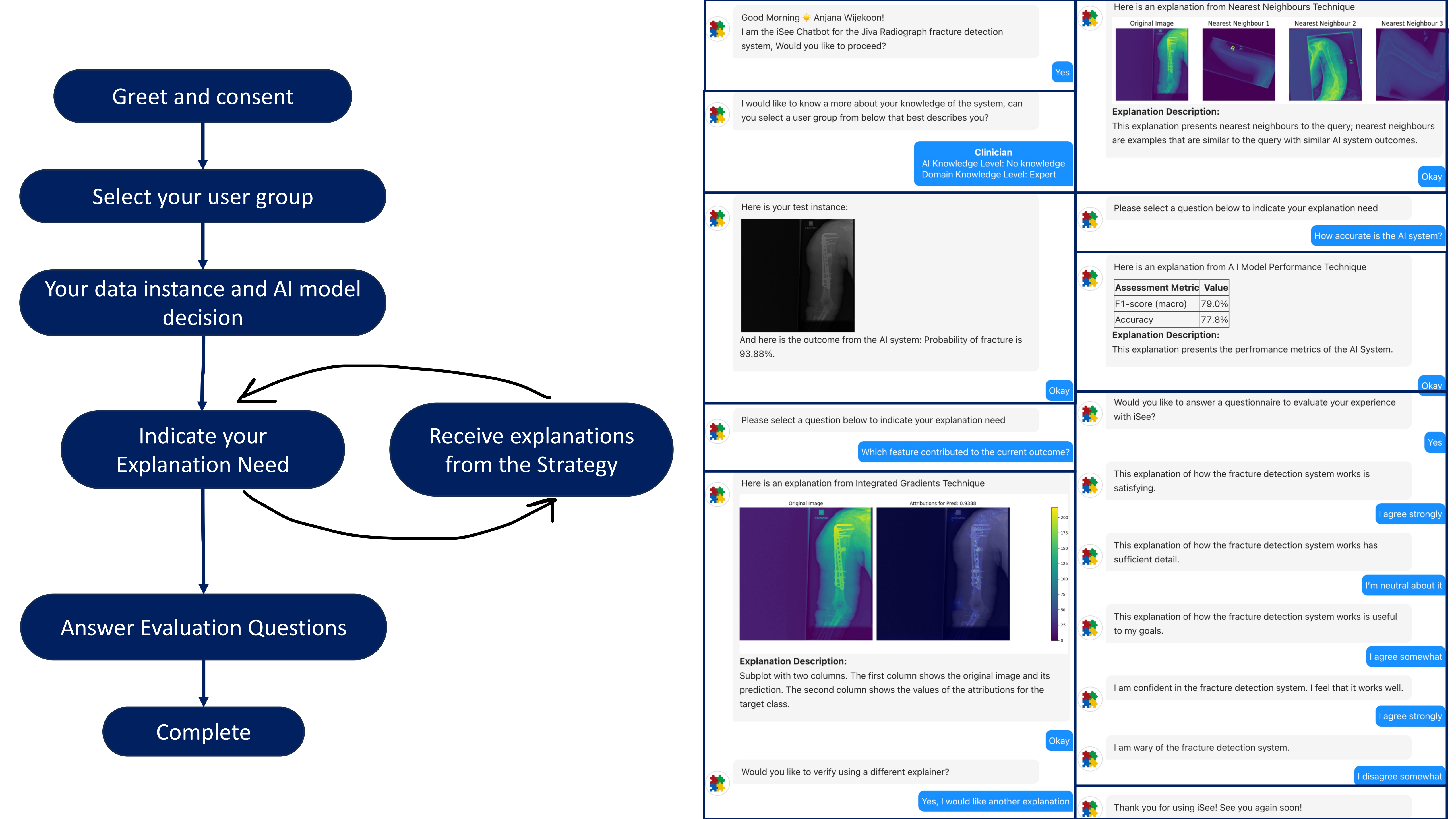


Wijekoon, A., Corsar, D., & Wiratunga, N. (2022). Behaviour Trees for Conversational Explanation Experiences. arXiv preprint arXiv:2211.06402.

iSee Interaction Model – with explanation & evaluation strategy



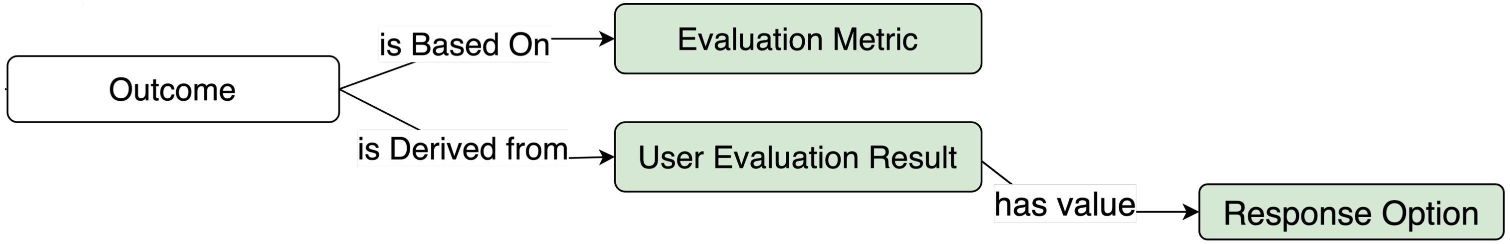
Wijekoon, A., Corsar, D., & Wiratunga, N. (2022). Behaviour Trees for Conversational Explanation Experiences. arXiv preprint arXiv:2211.06402.



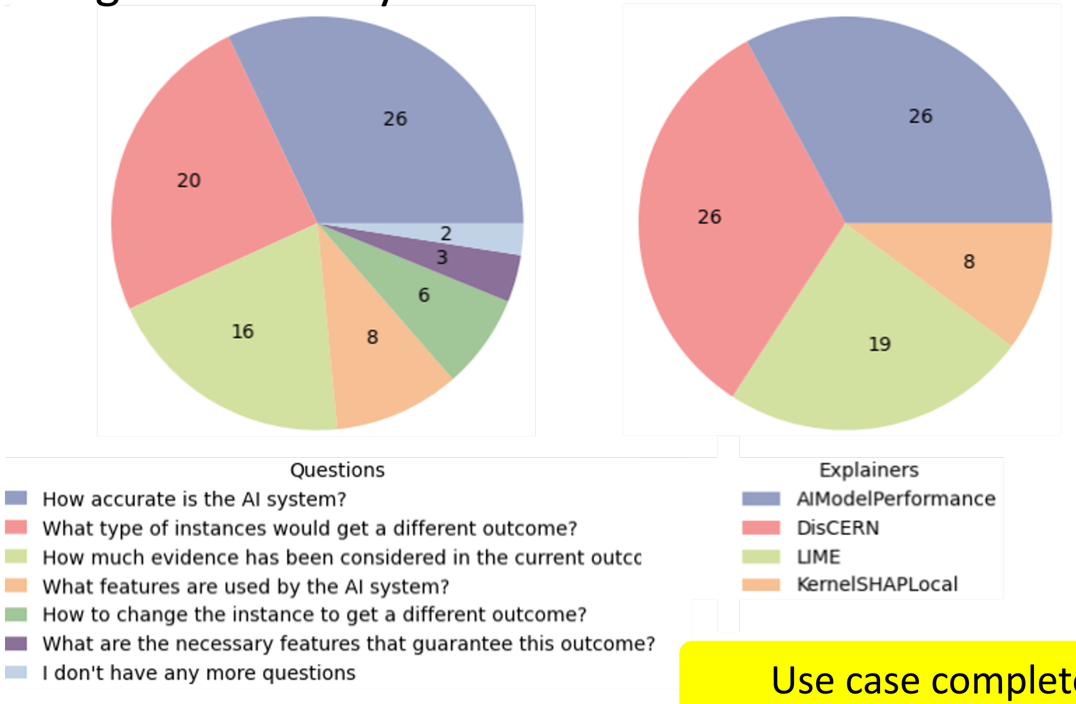
Greet and consent
Select your user group
Your data instance and AI model decision
Indicate your Explanation Need
Receive explanations from the Strategy
Answer Evaluation Questions
Complete
Good Morning Anjana Wijekoon!
I am the iSee Chatbot for the Jiva Radiograph fracture detection system, Would you like to proceed?
Yes
I would like to know a more about your knowledge of the system, can you select a user group from below that best describes you?
Clinician
AI Knowledge Level: No knowledge
Domain Knowledge Level: Expert
Here is your test instance:
And here is the outcome from the AI system: Probability of fracture is 93.88%.
Okay
Please select a question below to indicate your explanation need
Which feature contributed to the current outcome?
Here is an explanation from Integrated Gradients Technique
Original Image
Attributions for Pred: 0.9388
Explanation Description:
Subplot with two columns. The first column shows the original image and its prediction. The second column shows the values of the attributions for the target class.
Okay
Would you like to verify using a different explainer?
Yes, I would like another explanation
Here is an explanation from Nearest Neighbours Technique
Original Image
Nearest Neighbour 1
Nearest Neighbour 2
Nearest Neighbour 3
Explanation Description:
This explanation presents nearest neighbours to the query; nearest neighbours are examples that are similar to the query with similar AI system outcomes.
Okay
Please select a question below to indicate your explanation need
How accurate is the AI system?
Here is an explanation from A I Model Performance Technique
Assessment Metric | Value
F1-score (macro) | 79.0%
Accuracy | 77.8%
Explanation Description:
This explanation presents the perfromance metrics of the AI System.
Okay
Would you like to answer a questionnaire to evaluate your experience with iSee?
Yes
This explanation of how the fracture detection system works is satisfying.
I agree strongly
This explanation of how the fracture detection system works has sufficient detail.
I'm neutral about it
This explanation of how the fracture detection system works is useful to my goals.
I agree somewhat
I am confident in the fracture detection system. I feel that it works well.
I agree strongly
I am wary of the fracture detection system.
I disagree somewhat
Thank you for using iSee! See you again soon!

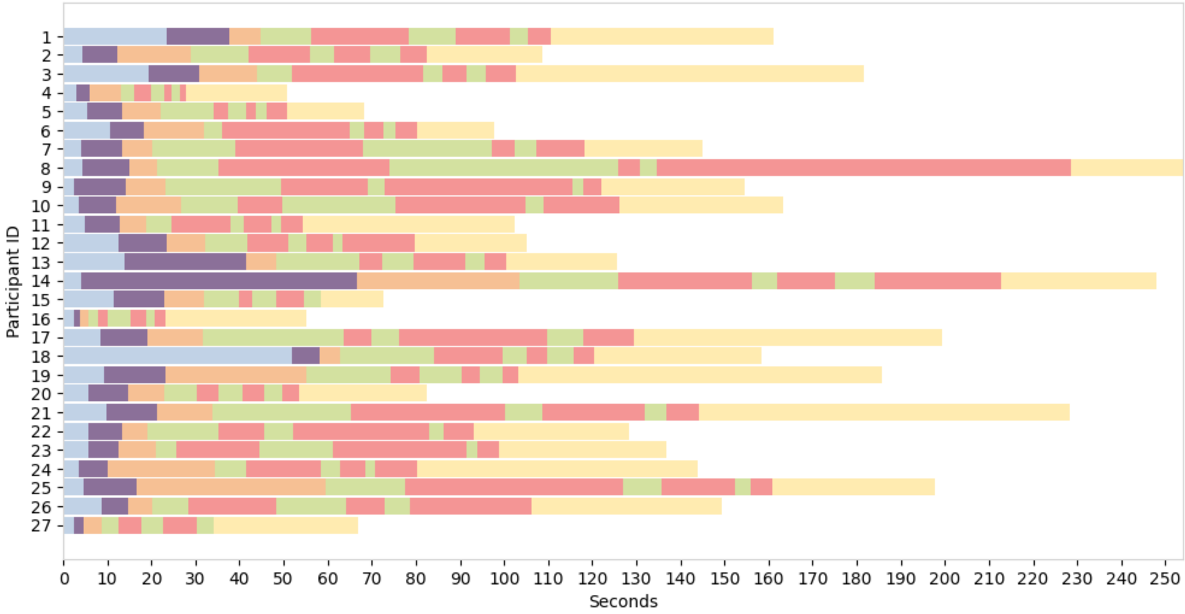
What does a case represent? the outcome part

Feedback to Analytics



Explanation Strategy Quality
High-level Analytics



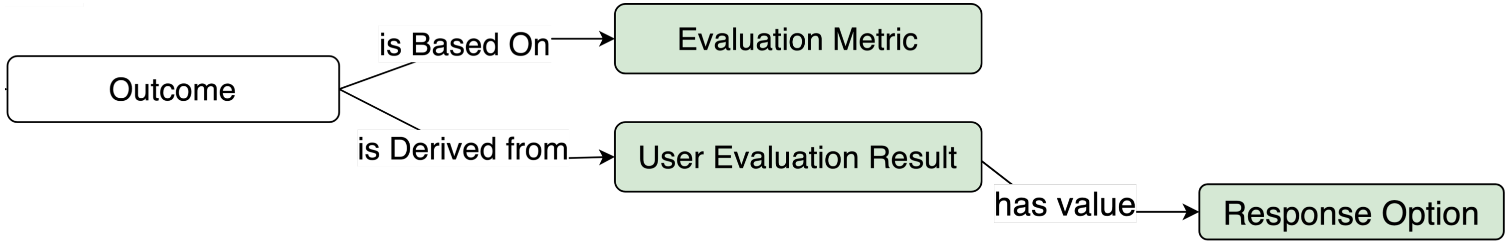


Use case completeness

Cognitive Load

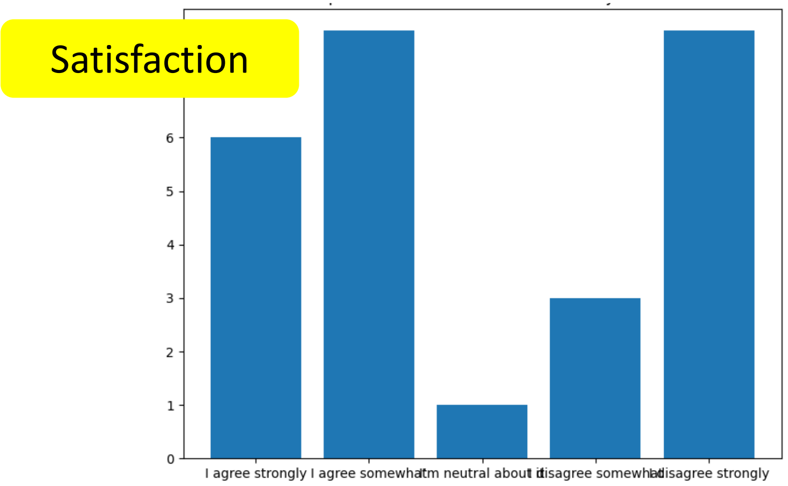
What does a case represent? the outcome part

Feedback to Analytics

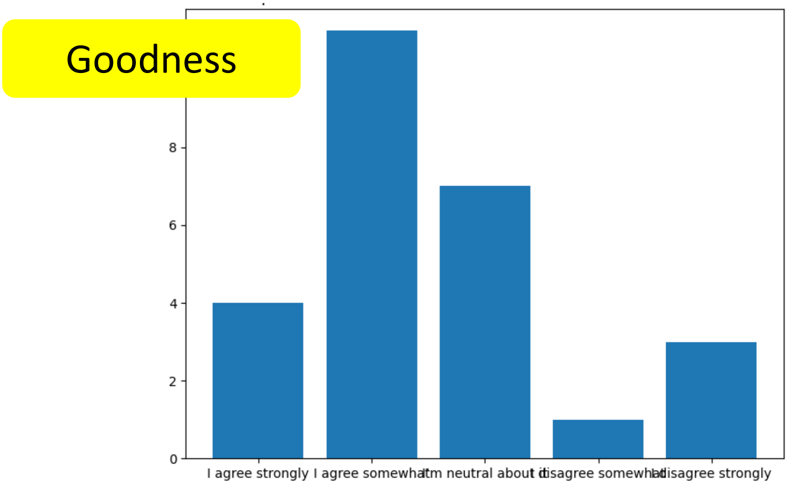


Explanation Strategy Quality Detailed Analytics

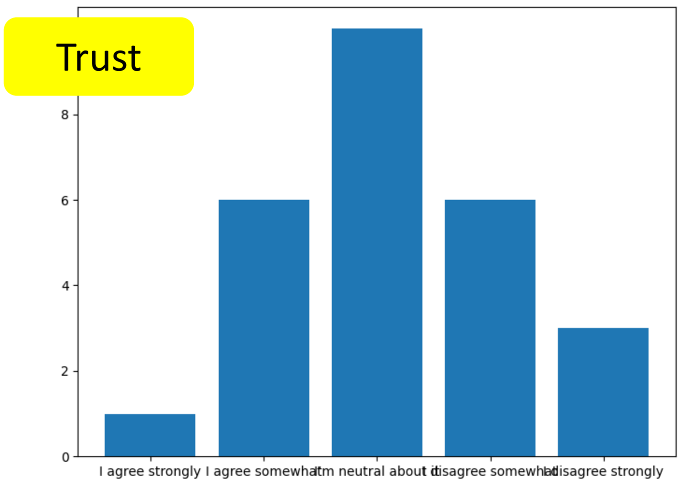
The explanation of the AI Model sufficiently detailed



The explanation lets me know how accurate or reliable the AI model is



The explanation lets me know how trustworthy the AI model is



Conclusions

- CBR is a methodology to share best practice
  - Works with multiple and diverse modalities
  - Join the CloodCBR open source platform development
- iSee platform uses CBR to capture experiences of best practice in XAI.
  - Call for explainers, usecases, evaluation strategies
- CBR provides a path toward building models that have reasoning competence
  - Experiential knowledge
  - Situationally aware
  - Deterministic matching

URL : https://cloodcbr.com/

CloodCBR

URL : https://cockpit-dev.isee4xai.com/usecases
username: test@isee4xai.com
password: Design@iSeeTest

iSee

Artificial Intelligence & Reasoning

It’s a Team effort …

n.wiratunga@rgu.ac.uk