Building the AI you trust
Intelligent Sharing of Explanation Experiences by Users for Users

https://isee4xai.com/

Hello@isee4xai.com

Anne.Liret@bt.com(BT)
Our proposal to the XAI challenges

Provide the AI community with an unifying platform, underpinned by case-based reasoning (CBR), in which successful experiences of applying an explanation strategy to an ML task can be captured as cases and retained in a case base for future reuse.

Cases will encode knowledge about the decisions made by a user and the effectiveness of the XAI strategy.

iSEE will recommend how best to explain ML predictions to other users in similar circumstances

- Build an understanding of how organisations judge the quality and success of explanations
- Build a platform and an evaluation conversational module
iSee Story in 2 minutes

Video from the iSee web site (click on the logo, then the triangle)

https://isee4xai.com/
https://youtu.be/XGvv2QeU0k8

Video: iSee in action on a real world use case about radiology diagnosis explanation experience:

https://www.youtube.com/watch?v=k1tNC9XEJZk&ab_channel=iSeechist-era
Why is Evaluating Explanation Complex?

Users have different goals they wish to achieve when receiving an explanation
- What an end-user expects will be very different from a developer for example
- The purpose is subjective

As a result, explanation can take many different forms:
- Either in how it is constructed
- Presented
- Or the level of contextual knowledge it incorporates

Different ideas of what makes an explanation successful
A framework for co-creating

Co-creation of explanation strategies (conversational adaptation) between machine and users

Evaluation of the utility of the explanation from a human-computer interaction and social angle, as well as the perception of it by end user.

Design User

- A domain expert who is responsible for the AI system

End User

- A user of the AI system or other interested party

- Create use case
- Define users, Intents, and AI features
- View use case analytics

- View use case details
- View explanation for a query
- Give feedback on a query

End User Workflow

- Ask question from the user
  - auto fill to construct the query to the KM system

- Get feedback from the end user on their experience

- Presents an explanation to the end user based on their query and use case information
Use case onboarding

**Goal**: Test the Implementation of a specific use case using the ontology

**Design**

- Manage Usecase - iSee (chamath.me)
- Copy of iSee Cockpit Work Flows - diagrams.net
Case

• Used to represent an Explanation Experience
• How Explanation strategy satisfies explanation needs

• **Query**: Problem, AI task and goal assessment, Explanation needs.

• **Case solution**: Explanation strategy
  • Workflow of several evaluation artifacts, represented by BehaviourTrees

• **Case result**: Evaluation, user feedbacks (through the cockpit on how it meets the needs)
A typical path...

- Example-based, post hoc
- Feature importance-based, SHAP
- Global confidence, LIME
- Local accuracy
- Text summarisation
- Saliency maps
Supporting components in iSee platform
Conversational evaluation: Example 1

Adding navigator nodes

Bot: Does this explanation answer your question?
- Yes, I’m happy (success, exit)
- Yes/No, I need a different explanation (go to LIME)
- Yes/No, I have other questions (exit, go to Need)
- I don’t agree with this explanation (disagreement, go to disagreement)

Bot: What do you think?
- Yes, I’m happy (agreement, exit)
- Yes/No, I have other questions (exit, go to Need)
- I don’t agree with this explanation (disagreement, exit, go to disagreement)
## Navigation based Relationships between Explainers

<table>
<thead>
<tr>
<th>Navigation</th>
<th>Traditional BT</th>
<th>Scenarios</th>
<th>Navigation Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence</td>
<td>Any composite with one child</td>
<td>1. User has gone through all explainers in the current intent and other intents 2. User wants to leave</td>
<td>1. Thank you for using jSee. That’s all the explanations I can provide for you today. Would you like to take a questionnaire to evaluate your explanation experience. 2. Do you think we answered your question? Yes/No I’m happy to leave</td>
</tr>
<tr>
<td>Priority</td>
<td>Priority node with two or more children</td>
<td>1. Current parent is priority with right siblings → show other explanation 2. There are other intents → user needs to select a different question/intent</td>
<td>1. Do you think we answered your question? No I would like a different explanation 2. Do you think we answered your question? No I would like to ask a different question</td>
</tr>
<tr>
<td>Variant</td>
<td>Sequence node with two or more children</td>
<td>1. Current explainer has other presentations → show other presentation 2. Current parent is sequence with right sibling → show other explanation</td>
<td>1. Do you think we answered your question? Yes, I would like a different visualisation 2. Do you think we answered your question? Yes, I would like a more explanation</td>
</tr>
<tr>
<td>Complement</td>
<td>Explainer with Sequence node with two or more children</td>
<td>1. Current explainer has other presentations → show other presentation 2. Current parent is sequence with right sibling → show other explanation</td>
<td>1. Do you think we answered your question? Yes, I would like a different visualisation 2. Do you think we answered your question? Yes, I would like more explanation</td>
</tr>
<tr>
<td>Supplement</td>
<td>1. Current parent is sequence with right sibling → show other explanation</td>
<td>Do you think we answered your question? Yes, I would like more information</td>
<td></td>
</tr>
<tr>
<td>Disagreement</td>
<td>1. There are other intents → user needs to select a different question/intent</td>
<td>What would you like to do next? Ask new question</td>
<td></td>
</tr>
</tbody>
</table>
Conversational evaluation: Example 2

Adding navigator nodes

Bot: Does this explanation answer your question?
• Yes, I’m happy (agreement, exit)
• Yes, I need a more explanation (go to ALE)
• No, I need different explanation (go to DisCERN)
• Yes/No, I have other questions (exit, go to Need)
• I don’t agree with this explanation (disagreement, go to disagreement)

Bot: What do you think?
• Yes, I’m happy (agreement, exit)
• Yes/No, I need a different explanation (go to DisCERN)
• Yes/No, I have other questions (exit, go to Need)
• I don’t agree with this explanation (disagreement, go to disagreement)
Conversational evaluation : Example2

Adding navigator nodes

Bot: Does this explanation answer your question?
- Yes, I'm happy (agreement, exit)
- Yes, I need a more explanation (go to ALE)
- Yes/No, I have other questions (exit, go to Need)
- I don't agree with this explanation (disagreement, go to disagreement)

Bot: Do you think we answered your question?
- Yes, I'm happy (agreement, exit)
- Yes/No, I have other questions (exit, go to Need)
- I don't agree with this explanation (disagreement, go to disagreement)

Bot: What do you think?
- Yes, I'm happy (agreement, exit)
- Yes, I need a more explanation (go to SHAP)
- Yes/No, I have other questions (exit, go to Need)
- I don't agree with this explanation (disagreement, go to disagreement)
Example 2: Telecom Engineering Domain

Agents are responsible for supporting engineers in the field by organising task intervention and allow progress of complex tasks.
Example 2: Telecom Engineering Domain

An AI system to support desk agents in deciding on Next Action and help Triage process.

An AI system to predict potential Failure cause to field engineers ahead of starting the Fibre tasks.

What is the explanation used for (goal)
Example 2: Telecom Engineering Domain

A planner may be interested in
- Why does the AI believe this action is required?
- What are the other possible actions supporting the top one?

An expert may be interested in
- Key facts and relevance weights
- Numerical information presented in NLG text
Example 2: Telecom Engineering Domain

An engineer may be interested in:
- A summary of the action to be done (Aerial cable required)
- Actions made on similar past task to allow progression and closure
- Some contextual knowledge about Hazards (tree, pole renew)
- How likely is the algorithm to be correct?
- Confidence measures
- Human-in-the-loop

Pole validation

External Event: 34%
Dangerous animal: 10%
Time constraint: 20%
Personas and Intents (telecom use case)

- **Planner**
  - Explanation is decision-support
  - About / Education
    - “Why has this recommendation been made ?”
    - “What are other notes that led to the same action ?”
- **Mental Model**
  - "Why do you think this is the next network task?"
- **Performance**
  - “How do I know the system works effectively?”
  - How confident is the AI model ? what are the other possibilities?

- **Developers**
- **Explanation as an auditing tool**
- **Debugging**
  - "How has this prediction been made?"
  - "How has this explanation been generated?"
  - Which inputs led to this result?
BT Telecom – Planner (desk based agents - admin), engineers

• (sequence) Confidence Score: how confident is the AI system about the prediction?
  • Technique: Euclidean distance to similarity between the query and the recommended case
  • Explanation type: Statistical explanation

• Would you like alternative explanation? Yes/No (*)

• (sequence) Confidence score of k nearest neighbours
  • Technique: Euclidean distance to similarity between the query and the nn (Bruno)
  • Explanation type: NN explanation + Statistical explanation

• (priority) Would you like some thing else? Yes/No (*)

• (sequence) Feature attribution explanation word level of the recommended case
  • Technique: tf-idf vector comparison (Bruno)
  • Why recommended based on words

• (sequence) Sentence attribution explanation
  • Technique: tf-idff vector comparison but based on words form previous (Bruno)
  • “Reason” why they are similar: reason being a sentence
Example 1: Medicine and Health Domain

A doctor may be interested in:
- Why does the AI believe there is a fracture?
- Saliency Mapping
- Graphical Representation
- Little contextual knowledge required

A patient may be interested in:
- How likely is the algorithm to be correct?
- Confidence Measures
- Numerical information presented in NLG text
- Some contextual knowledge required

Who is the explanation for?

95% accurate
Example 2: Anomaly Detection Domain

The use of AI models within cybersecurity response aims at helping cyber SOC members automatically respond to cyberattacks both faster and more efficiently.

Goal is to improve task performance, business impact.

Expert Analyst may be interested in:
- Why the result
- Counterfactuals, probability, post-hoc
- External knowledge required
- Why a certain action was suggested
- General model behaviour
- Local AI system behaviour

Manager may be interested in:
- Global understanding of AI system
- How effective it may be if performed
- Graphical representation

What makes an explanation a success
User experience of explanation

If FAILURE, comeback and try another from the catalogue
Do you wish to contribute with methods for the library or use cases?

- The goal of iSee project is to apply the platform to real problems such as Anomaly Detection in production line, Cyber Threat Detection, Telecom Diagnosis, Radiology diagnosis, Natural Event prediction.

- Since explanation goals and utility is very inherent to the use case and the user profile, we are looking for inputs from both users of AI or explanation system, practitioners of explanation methods.

Contact us at hello@isee4xai.com
Use case onboard path

Version 4th July 2022
UC Progress line

**Kick-off**
- **Maturity**
- **Business needs**

**Understanding (survey)**
- **Models**
- **AI/Human goal**
- **Explainers**
- **User intent**
- **Explanation strategy**
- **Questions for evaluation**

**Case structure**

**Ontology Mapping**
- **Model available**: upload it (pkl)
- **Explanation strategy exists**: upload it (beh tree?)
- **No explainer**: skip to retrieval step
- **Reusable explainer available**: add to isee catalog with the WP2 team

**Add to Cockpit**

**Explanation retrieval**
- **Cockpit**: retrieve strategy « image »
- **Build or Review the evaluation questions script?**
- **Connect to iSee conversation layer**

**Feedback**
- **AI model in iSee**: Using Chatbot, query the model through iSee and answers questions to provide feedbacks
- **AI model not in iSee**: Download Chatbot version connecting the external model to the iSee API and answers questions. Transfer the feedback to iSee team or through iSee API/private git repo.
UC Progress line with co-creation

Kick-off
- Maturity
- Business needs
- What makes AI trustworthy,
- What makes explanation a success,

Understanding (survey)
- Models
- AI/Human goal
- Explainers
- User intent
- Explanation strategy
- Questions for evaluation

Case structure

Ontology Mapping
- Model available: upload it (pkl)
- Explanation strategy exists: upload it (beh tree?)
- No explainer: skip to retrieval step
- Reusable explainer available: add to iSee catalog with the WP2 team

Add to Cockpit

Explanation retrieval
- Cockpit: retrieve strategy « image »
- Build or Review the evaluation questions script?

Add to Cockpit

Feedback
- AI model in iSee:
  - Using Chatbot, query the model through iSee and answers questions to provide feedbacks
- AI model not in iSee:
  - Download Chatbot version connecting the external model to the iSee API and answers questions.
  - Transfer the feedback to iSee team or through iSee API/private git repo.?

Co-creation designer

Collaborative evaluation – end user
Each UC can bring multiple cases into Cockpit / ontology instantiation / json formatted case / platform

V1: case retrieval for the invent validation goal -> implement UC with this goal (no explainer provided in input)
V2: case revise/ for the Xplainers strategy recommendation goal -> implement UC with this goal (explainer provided in input)

**Consumer of case-based explanation strategy and explainers**
Use cases have an AI (black box) system (with/without training data)
situation1 [AI system + explanation goals, no explainer]
situation2 [dataset + AI system, no explainer]

**Producer of case-based explanation strategy or explainers**
Use cases have an AI (black box) system (with/without training data)
situation3 [dataset + embedded AI-explainer]
situation4 [dataset + AI system, dedicated external explainer]

iSee helps retrieving one of the existing explainer (retrieve and reuse from the iSee case base).
Will be using iSee to evaluate or certify their models.

Helps populate the case base with new explainer without experiences.
Not using iSee to evaluate or certify their models.
iSee helps finding reusable explanation methods and build strategies