Explainable AI is Dead, Long Live Explainable AI!

Hypothesis-driven Decision Support using Evaluative AI

Charles Vin

Sorbonne Université - 21216136

Quick summary

Paradigm Shift in AI Decision Support
$\rightarrow$ Evaluative AI Concept
Goals:

- Human-centered Approach
- Going Beyond Recommendations
- **Mitigating Over-Reliance**
- Support for Hypothesis Evaluation
- Machine-in-the-Loop Paradigm

Over/Under-reliance

Definitions

- **Over-reliance**: Decision makers accept a machine recommendations, even when it is wrong, but would be rejected if coming from a human.
  - *The machine "must be right" because it's a machine*
- **Under-reliance**: Machine outputs are consistently rejected, even when it is correct, but would be accepted if coming from a human.

$\Rightarrow$ Problems after deployement:
AI systems ignored OR over-reliance related problems.

Over/Under-reliance

Causes

- Over-reliance: Automation bias ;
- Under-reliance: Algorithmic aversion.

When adding XAI tools for more explaination
$\Rightarrow$ Confirmation bias (called fixation in the paper).

Over/Under-reliance

Solutions

- Cognitive forcing
  - Eg. forcing people to give a decision before seeing a recommendation ;
  - Slightly mitigated overreliance, but not enought to lead to a statistically significant differences ;
  - Least prefered method by participant : people not wanting to exert mental energy.
- Changing the XAI framework

What makes a good decisions?

In a simple way:

- Identify options
- Compare options
- Choose an option

In a less simple way: the 10 "cardinal decision issue" outlined by Yates and Potworowski

- Needs, mode, Investment, Options, Possibilities, Judgements, Value, Trade-offs, Acceptability, Implementation

What makes a good decisions support system?

Summed up

- Options: Help to identify options, well as help to narrow down the list of feasible or realistic options
- Possibilities: Help to to identify possible outcomes
- Judgement & Value: Help to judge which outcomes are most likely and what will be the positive and negative impacts
- Trade-offs: Help to make trade-offs on the above criteria for each options
- Understandable: Help to understand how and why the tools works as it does, and when it fails

Does current decision support align with those criteria?

Giving recommendations with no explanatory information

Figure 3

A model of giving recommendations for decision support.

This assumes that decision makers will carefully consider recommendations. However, empirical evidence suggests this is not the case.

- ✗ Options
- $1/n$ Possibilities
- $1/n$ Judgement & Value
- ✗ Trade-offs
- ✗ Understandable

Giving recommendations with explanatory information

Figure 3

A model of giving recommendations for decision support.
This assumes that decision makers will carefully consider recommendations. However, empirical evidence suggests this is not the case.

- ✗ Options
- ✗ Possibilities
- ✗ Judgement & Value
- ✗ Trade-offs
- ✗ Understandable

Giving recommendations with cognitive forcing

Figure 3

A model of giving recommendations for decision support.

This assumes that decision makers will carefully consider recommendations. However, empirical evidence suggests this is not the case.

- ✗ Options
- ✗ Possibilities
- ✗ Judgement & Value
- ✗ Trade-offs
- ✗ Understandable

Bibliography

Template: Here 

A plain frame has no headline

`\Alegreya`	Lorem ipsum dolor sit amet
`\AlegreyaMedium`	**Lorem ipsum dolor sit amet**
`\AlegreyaExtraBold`	**Lorem ipsum dolor sit amet**
`\AlegreyaBlack`	**Lorem ipsum dolor sit amet**
`\AlegreyaSansThin`	Lorem ipsum dolor sit amet
`\AlegreyaSansLight`	Lorem ipsum dolor sit amet
`\AlegreyaSans`	Lorem ipsum dolor sit amet
`\AlegreyaSansMedium`	**Lorem ipsum dolor sit amet**
`\AlegreyaSansExtraBold`	**Lorem ipsum dolor sit amet**
`\AlegreyaSansBlack`	**Lorem ipsum dolor sit amet**

Alert! A *plain* frame does not show the progress bar but still appears in it unless the frame comes after `\End`

A ***STANDOUT*** frame can be used to focus attention

In combination with *plain*,
it makes a nice thank-you slide!

https://github.com/piazzai/arguelles
https://ctan.org/pkg/beamertheme-arguelles