

# **Explainable AI is Dead, Long Live Explainable AI!**

Hypothesis-driven Decision Support using Evaluative AI

*Tim Miller (2023) - Conference FAcT*

**Charles Vin**

Sorbonne Université - 21216136

# Quick summary

Arg for a Paradigm Shift in (X)AI for Decision Support

→ 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.

See "Automation bias"  $\Rightarrow$  Problems after deployment:  
AI systems ignored OR over-reliance related problems.

# Over/Under-reliance

## Causes

- Over-reliance: lack of cognitive engagement ;
- Under-reliance: Algorithmic aversion.

When adding current XAI tools for more explanation  
⇒ Confirmation bias (called *fixation* in the paper).

# Over/Under-reliance

## Solutions

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

# How we make 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 (2012)

- 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
- Judgement & Possibilities: 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

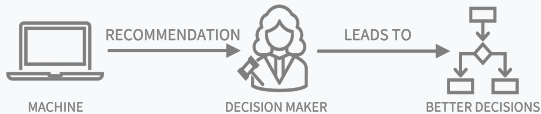


Figure: A model of giving recommendations for decision support.

Decision makers can carefully consider recommendations ;

→ Leading to better decisions ;

× empirical evidence suggests this is not the case.

× Options

$1/n$  Possibilities & Judgement

× Trade-offs

× Understandable

# Giving recommendations with explanatory information



Figure: A model of XAI for decision support.

Giving reasons or explanations for decisions

→ Mitigates the problem of distrust ;

→ Leading to better decisions ;

× Empirical evidence suggests people do not pay careful attention to the reasons/explanations.

× Options

$1/n$  Possibilities & Judgement

✓/× Trade-offs

✓ Understandable

# Giving recommendations with cognitive forcing



Figure: A model of cognitive forcing.

Withholding recommendations & giving an explanation ;

- forces people to engage ;
- limit over-reliance ;
- Better decisions ;
- × still a "recommend and defend" approach.
- × Least preferred method by participants

✓/× Options

$1/n$  Possibilities & Judgement

✓/× Trade-offs

✓ Understandable

# The evaluative AI framework

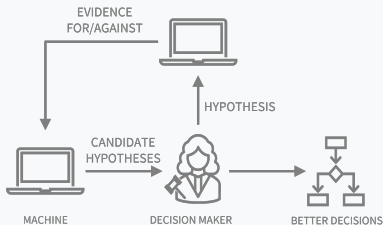


Figure: A model of Evaluative AI.

- Align with decision making processes
- Keep the decision maker in control
- Ask users to rely on evidence instead of recommendations.

# The evaluative AI framework

## Properties

**Lesion**

**Notes**

Patient reports itchiness and bleeding.  
Lesion has changed colour.

**Lesion location**

☐ Head ☐ Upper arm  
☐ Face ☐ Hand/Lower Arm  
☒ Back ☐ Upper Leg  
☐ Front Torso ☐ Foot/Lower Leg

**Your hypothesis**

Melanoma  
Melanocytic Nevus  
**Basal Cell Carcinoma**  
Actinic Keratosis  
Benign Keratosis  
Dermatofibroma  
Vascular Lesion

**Evidence for**

Lesion location   
Colour   
Scarred   
Bleeding

**Evidence against**

Asymmetric shape   
Changed colour   
Itchiness

- ✓ Options
- ✓ Possibilities & Judgement
- ✓ Trade-offs
- ✓ Understandable

Figure: A simple prototype of a diagnostic interface using evaluative AI.

# The evaluative AI framework

## Zoom on properties

- ✓ Options
  - Show the most likely options (with or without probabilities) ;
  - Not a single recommendation.
- ✓ Possibilities & Judgement
  - The machine provide feedback on humain judgement only.
- ✓ Trade-offs
  - Offer real trade-offs between *any* set of two options ;
  - Evaluative AI provides evidence both for and against each option, irrelevant of the judged likelihood of that option ;
  - *Option awareness* in the literature.

# The evaluative AI framework

## Differences with cognitive forcing

- **Control** Permit to explore hypothesis, not a single recommendation ;
- Built on the way we makes decision (identify, compare, choose)

# Long live XAI

- Evaluative AI is designed only toward decision making, Evaluative AI  $\subset$  XAI ;
- XAI is still needed and more adapted to many situation (eg. making decision at scale) ;
- Recommendation based models : base of any XAI techniques ;
- Many existing XAI tools are already adapted to Evaluative AI
  - Constrastive explanation ;
  - Feature importance (eg. SHAP).



# Limitation

- Why would people pay attention to evidence this time?
  - Evaluative AI
    - Better control ;
    - Process built on the way we makes naturally decision
      - people would naturally follow
      - ≠ Contrary to recommendation-driven approches ;
  - × Proof?
- Cognitive load remain a problem
  - Evaluative AI still reduce the quantity of information the decision maker needs (only revelant information are presented)
  - × Still the less prefered solution by decision makers

# Limitation

- More introduction around automation bias needed ;
-

# Bibliography

- Tim Miller. Explainable ai is dead, long live explainable ai! hypothesis-driven decision support using evaluative ai. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, pages 333–342, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701924. doi: 10.1145/3593013.3594001. URL <https://doi.org/10.1145/3593013.3594001>.
- J. Frank Yates and Georges A. Potworowski. 198 Evidence-Based Decision Management. In *The Oxford Handbook of Evidence-Based Management*. Oxford University Press, 06 2012. ISBN 9780199763986. doi: 10.1093/oxfordhb/9780199763986.013.0012. URL <https://doi.org/10.1093/oxfordhb/9780199763986.013.0012>.

Beamer template from here 