EXPLAINABLE AI IS DEAD!
LONG LIVE EXPLAINABLE AI!

Tim Miller
School of Electrical Engineering and Computer Science
The University of Queensland, Australia
timothy.miller@uq.edu.au
@tmiller_uq
A QUICK SURVEY
RECOMMENDATION-DRIVEN EXPLAINABLE AI

HYPOTHESIS-DRIVEN EVALUATIVE AI
DECISION AIDS

Task performance

Human  AI recommender  Human + AI
DECISION AIDS

Task performance

- Human
- AI recommender
- Human + AI
ASSUMPTIONS: DECISION SUPPORT SYSTEMS
ASSUMPTIONS: DECISION SUPPORT SYSTEMS
ASSUMPTIONS: DECISION SUPPORT SYSTEMS

ANSWER

ASK ‘WHY?’

EXPLAIN

ENABLES

LEADS TO
(DIS)TRUST AND (UNDER-)RELIANCE

TRUSTWORTHY

NOT TRUSTWORTHY

TRUSTED

DISTRUSTED
<table>
<thead>
<tr>
<th>Trustworthiness</th>
<th>Trusted</th>
<th>Distrusted</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trustworthy</strong></td>
<td>Warranted Trust/Reliance</td>
<td>Unwarranted Distrust/Under-Reliance</td>
</tr>
<tr>
<td><strong>Not Trustworthy</strong></td>
<td>Unwarranted Trust/Reliance</td>
<td>Warranted Distrust/Under-Reliance</td>
</tr>
</tbody>
</table>

**DIS)TRUST AND (UNDER-)RELIANCE**
(DIS)TRUST AND (UNDER-)RELIANCE

<table>
<thead>
<tr>
<th></th>
<th>TRUSTED</th>
<th>DISTRUSTED</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUSTWORTHY</td>
<td>WARRANTED TRUST/RELIANCE</td>
<td>UNWARRANTED DISTRUST/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UNDER-RELIANCE</td>
</tr>
<tr>
<td>NOT TRUSTWORTHY</td>
<td>UNWARRANTED TRUST/</td>
<td>Warranted DISTRUST/</td>
</tr>
<tr>
<td></td>
<td>RELIANCE</td>
<td>UNDER-RELIANCE</td>
</tr>
</tbody>
</table>
SELF-EXPLANATION IN DECISION MAKING
SELF EXPLANATION

VERIFY

UNDERSTAND MODEL/OUTPUTS

EXPLAINABILITY
## ABDUCTIVE REASONING AND VERIFICATION

<table>
<thead>
<tr>
<th>PROCESS</th>
<th>REQUIREMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Observe event</td>
<td>Design interfaces to determine what has happened</td>
</tr>
<tr>
<td></td>
<td>Design interfaces to highlight unusual events</td>
</tr>
<tr>
<td>2. Generate hypotheses</td>
<td>Help to construct (likely) hypotheses</td>
</tr>
<tr>
<td>3. Judge plausibility</td>
<td>Help to explore how causes affect outputs</td>
</tr>
<tr>
<td></td>
<td>Find evidence to support and refute hypotheses</td>
</tr>
<tr>
<td>4. Resolve explanation</td>
<td>Identify and record important information</td>
</tr>
<tr>
<td>5. Extend explanation</td>
<td>Support hypothesis revision</td>
</tr>
<tr>
<td></td>
<td>Support interactive exploration</td>
</tr>
</tbody>
</table>
DECISION SUPPORT = VERIFICATION
EVALUATIVE AI

EVIDENCE TO SUPPORT VERIFICATION

EVALUATIVE AI
EVALUATIVE AI

Patient reports itchiness and bleeding. Lesion has changed colour.

Lesion location
- Head
- Face
- Back
- Front Torso
- Upper arm
- Hand/Lower Arm
- Upper Leg
- Foot/Lower Leg

Your hypothesis
- Basal Cell Carcinoma
- Melanoma
- Melanocytic Nevus
- Actinic Keratosis
- Benign Keratosis
- Dermatofibroma
- Vascular Lesion

Evidence for
- Lesion location
- Colour
- Scarred
- Bleeding

Evidence against
- Asymmetric shape
- Changed colour
- Itchiness
Patient reports itchiness and bleeding. Lesion has changed colour.

**Lesion location**
- Head
- Face
- Back
- Front Torso
- Upper arm
- Hand/Lower Arm
- Upper Leg
- Foot/Lower Leg

**Evidence for**
- Asymmetric shape
- Changed colour
- Itchiness
- Bleeding
- Colour

**Evidence against**
- Scarred
- Legion location
EVALUATIVE AI

CONTRASTIVE EXPLANATION
EVALUATIVE AI AND CAUSALITY
CAUSALITY AND VERIFICATION

JUDEA PEARL
WINNER OF THE TURING AWARD
AND DANA MACKENZIE

THE BOOK OF WHY

THE NEW SCIENCE
OF CAUSE AND EFFECT
CAUSALITY AND VERIFICATION

If you treat causality as if it is only "in the mind" you would end up thinking it is "in the data", which is much more dangerous than thinking it "in Nature", not "in the mind". Besides, it is in Nature.

"Causality is in the mind, not in the data," Deaton quoting Heckman et al. at #DRIRCT @yudapearl

4:35 PM · Nov 18, 2020

Sanjeev Kumar @Axlamdax · Nov 18, 2020
IS EXPLAINABLE AI DEAD?
IS EXPLAINABLE AI DEAD?

LONG LIVE EXPLAINABLE AI!
**KEY TAKEAWAYS**

**EXPLAINABLE AI**

Explainable decision aids don’t really improved decision making (much)

Some false assumptions

- People look to machine recommendations
- People look to machine explanations
- Intuition needs to be overridden

**HOWEVER ....**

Evaluative AI provides the framework

- Build on expertise and expert intuition
- Assist verification --- don’t always ‘recommend and defend’
- Causal models may assist verification better

**Explainable AI is dead! ...**

... Long live explainable AI!