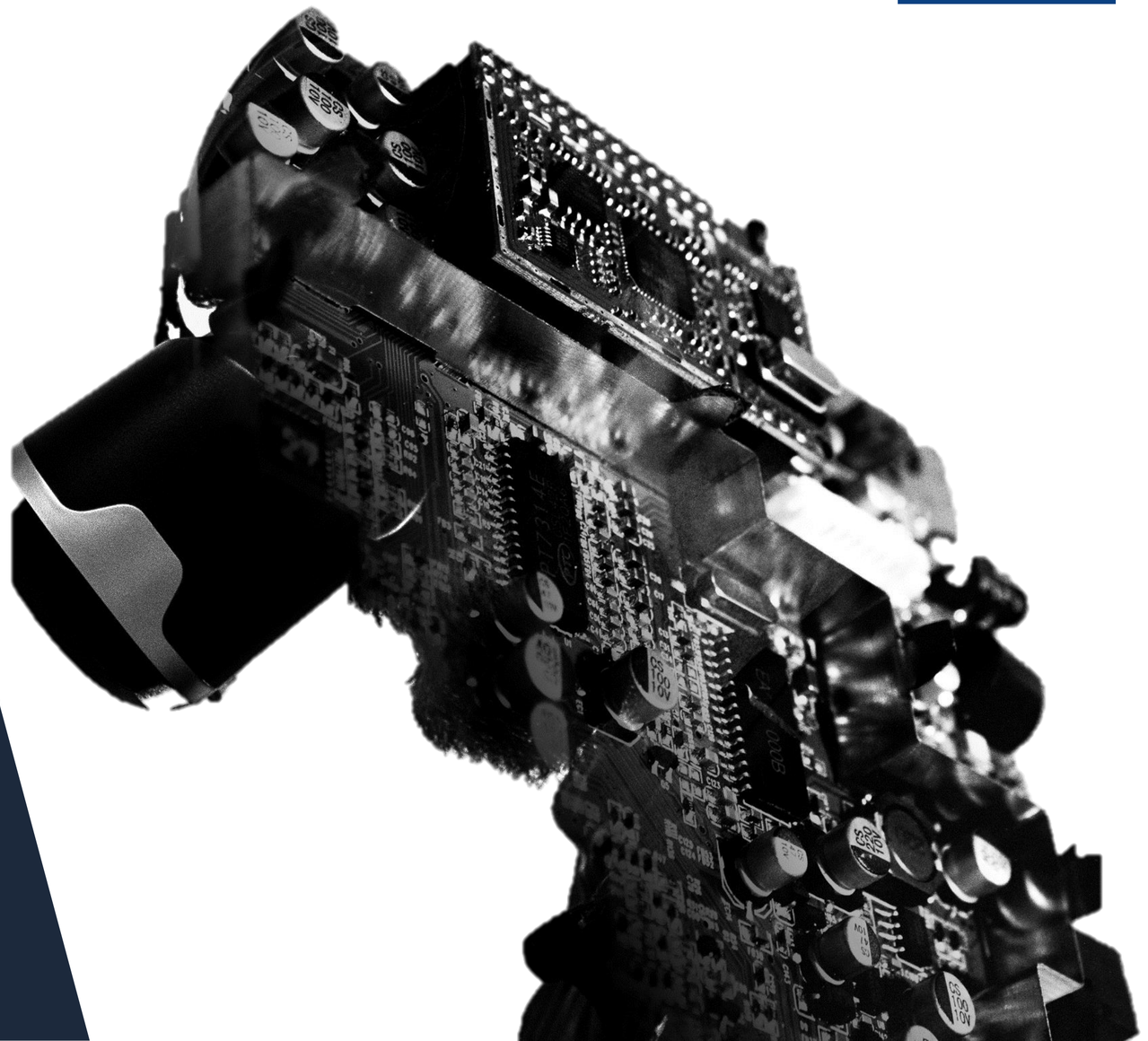
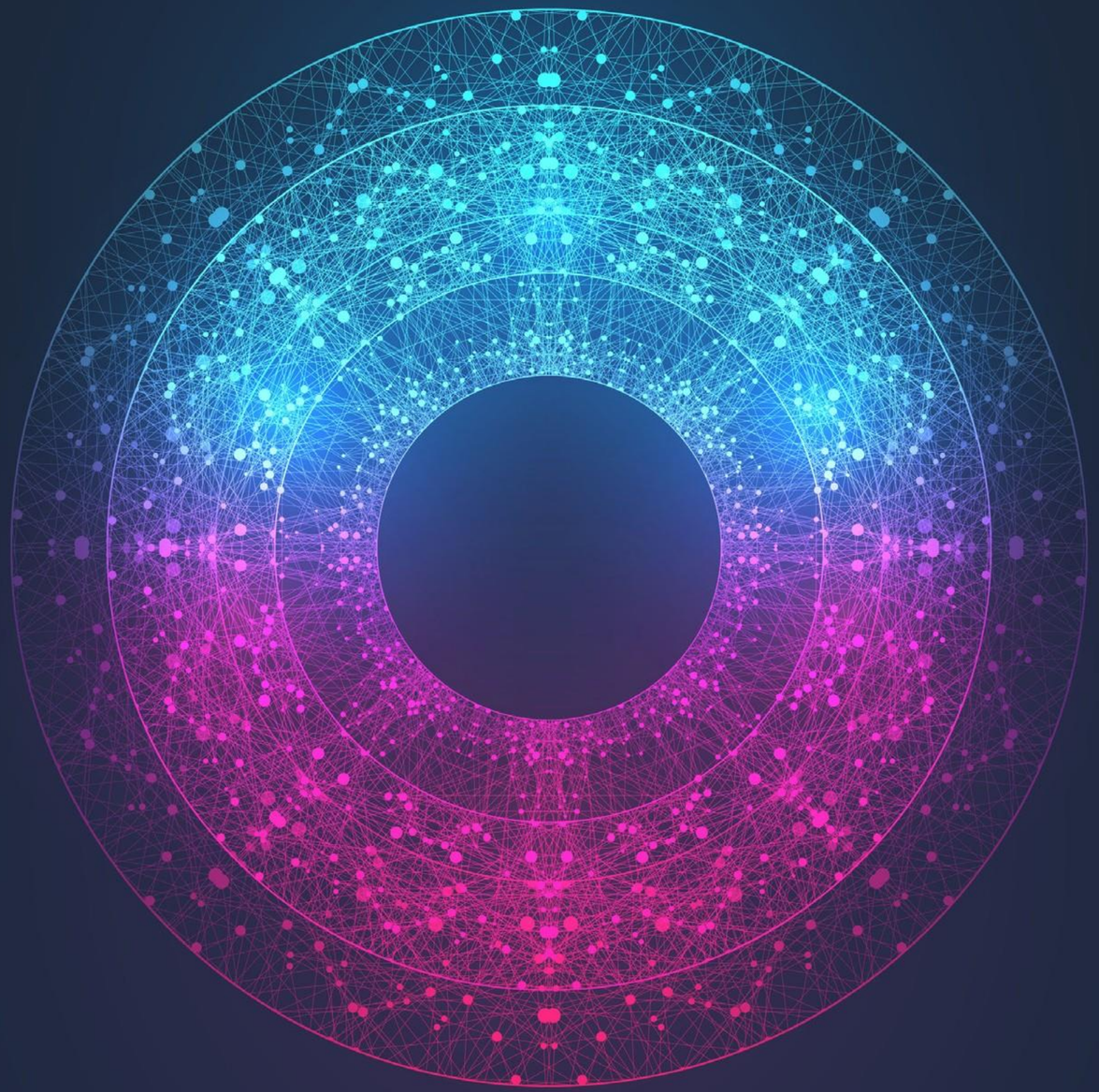
EXPLAINABLE AI IS DEAD!
LONG LIVE EXPLAINABLE AI!

Tim Miller

School of Electrical Engineering and Computer Science
The University of Queensland, Australia
timothy.miller@uq.edu.au
@tmiller_uq



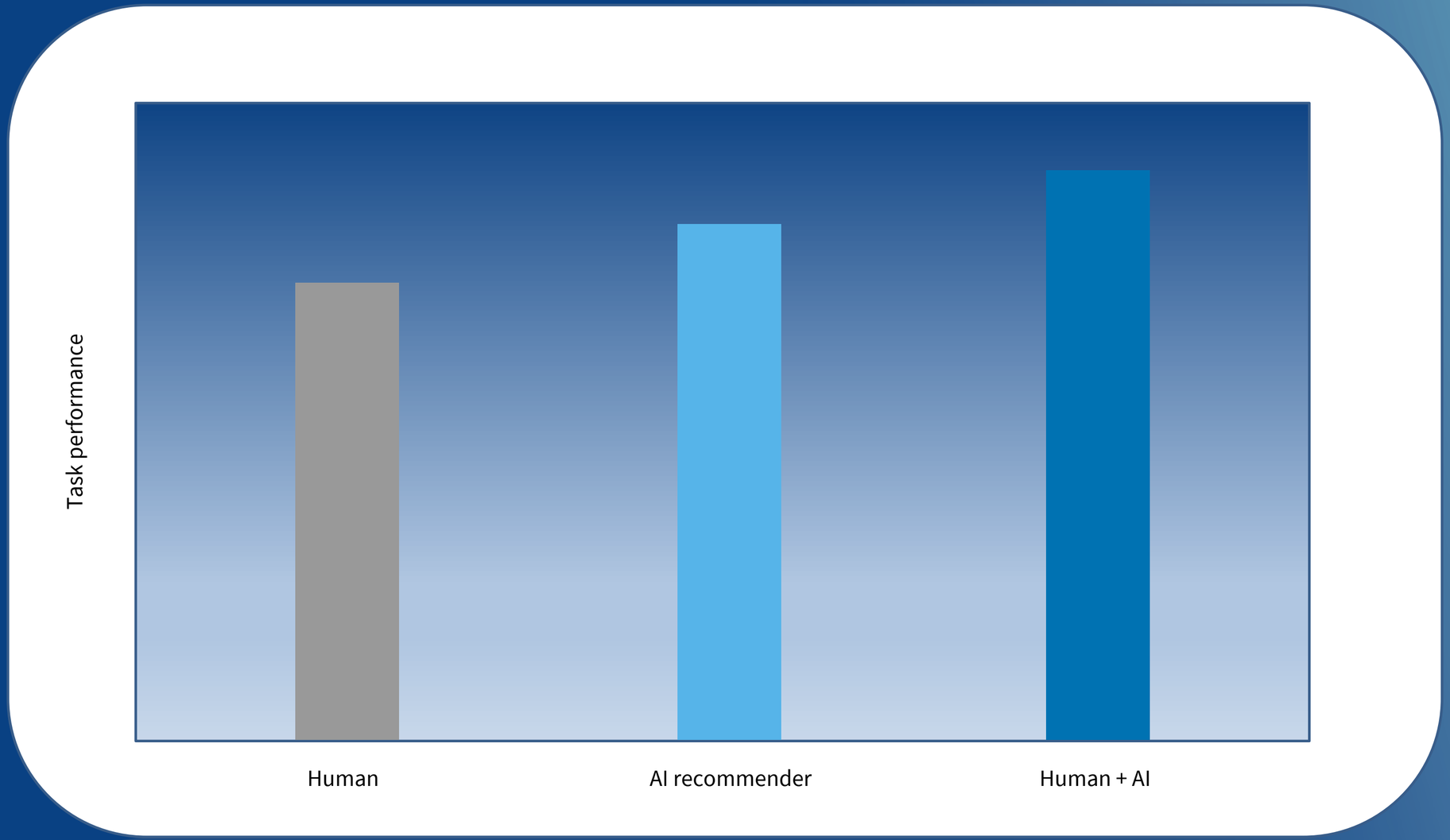
A QUICK SURVEY



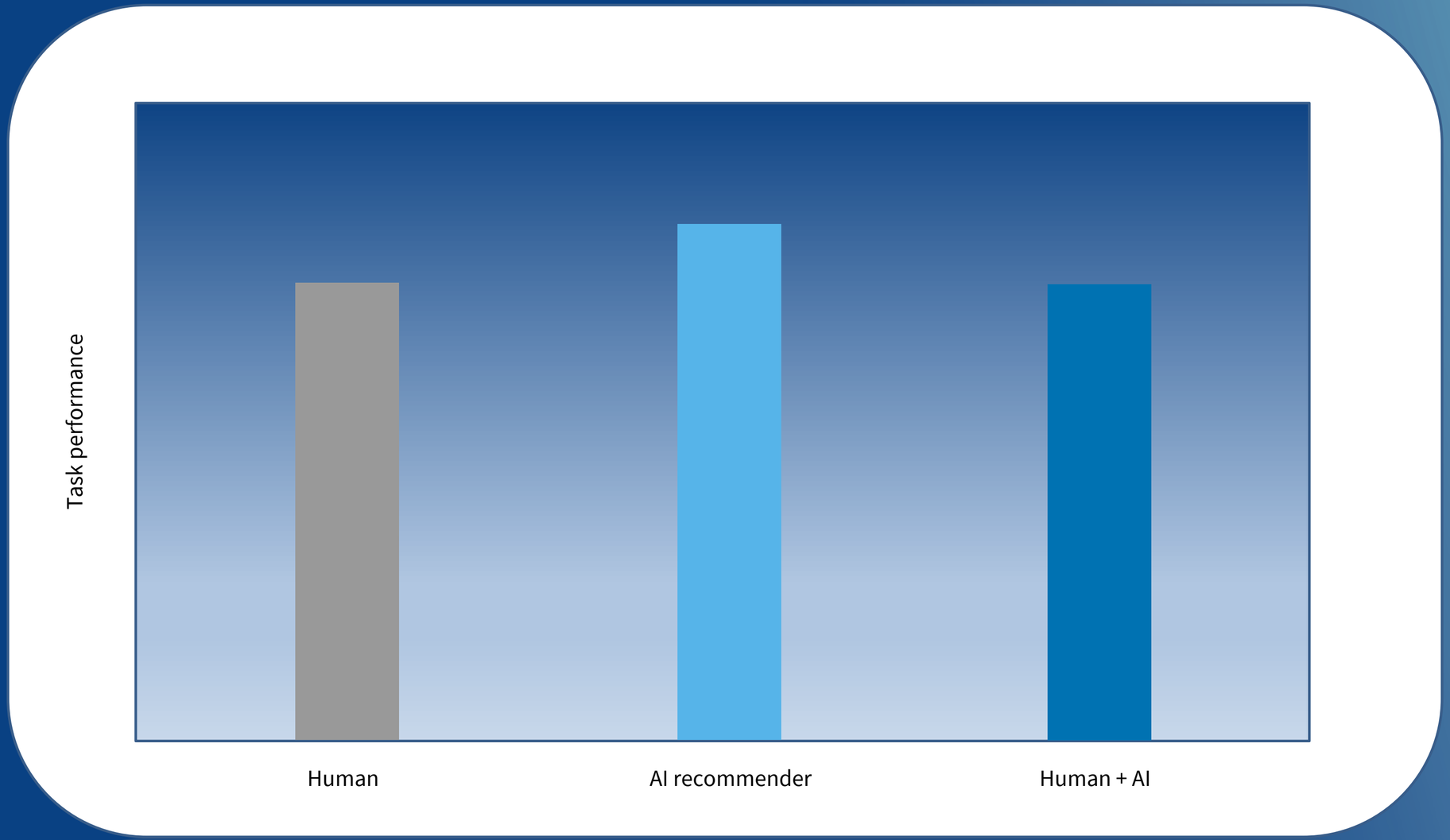
RECOMMENDATION-DRIVEN
EXPLAINABLE AI

HYPOTHESIS-DRIVEN
EVALUATIVE AI

DECISION AIDS



DECISION AIDS

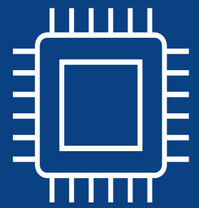


ASSUMPTIONS: DECISION SUPPORT SYSTEMS

ASSUMPTIONS: DECISION SUPPORT SYSTEMS

ASSUMPTIONS: DECISION SUPPORT SYSTEMS

(DIS)TRUST AND (UNDER-)RELIANCE



TRUSTWORTHY

NOT TRUSTWORTHY

TRUSTED

DISTRUSTED

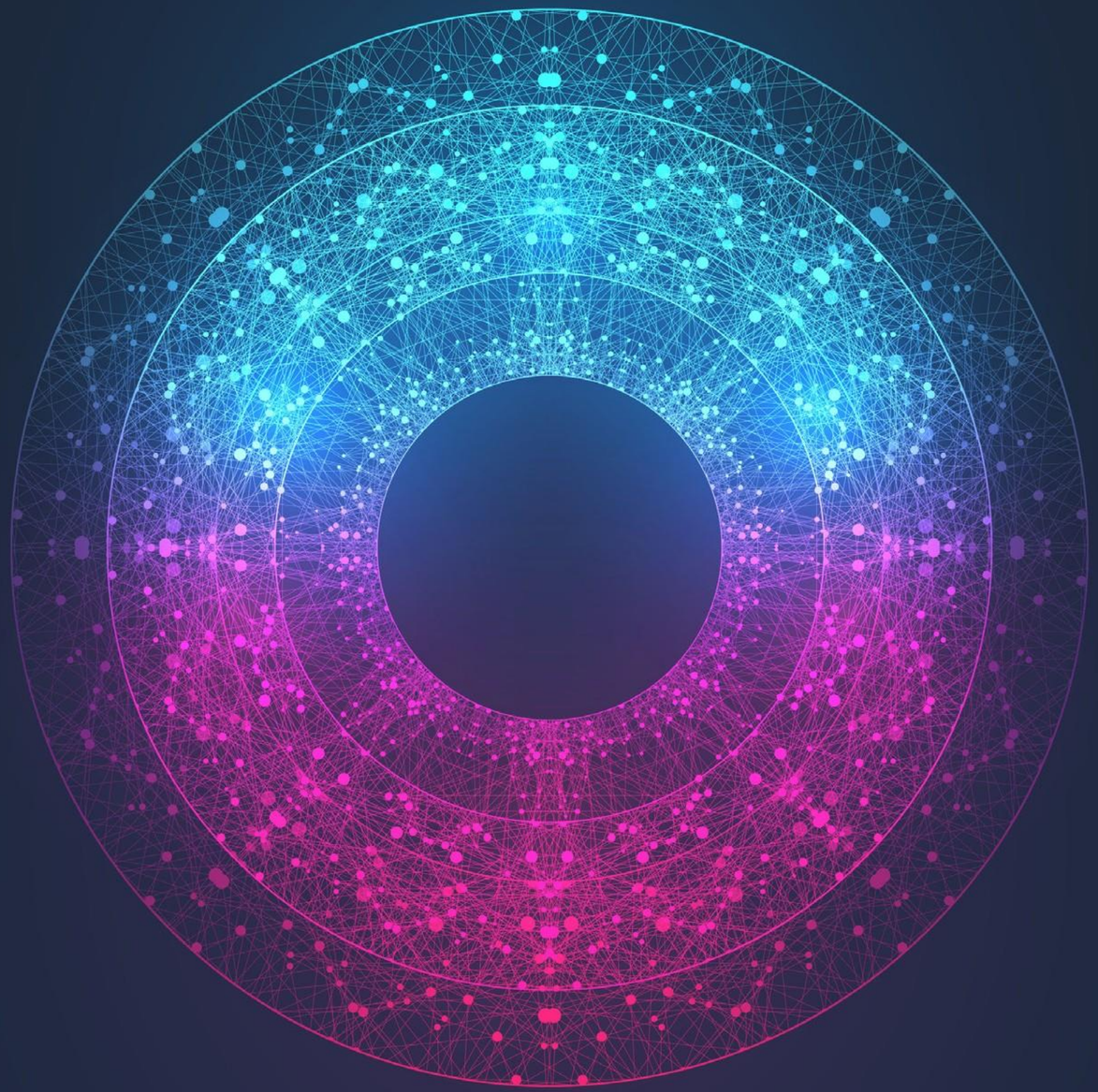
(DIS)TRUST AND (UNDER-)RELIANCE

	TRUSTED	**DISTRUSTED**
TRUSTWORTHY	WARRANTED TRUST/RELIANCE	UNWARRANTED DISTRUST/ UNDER-RELIANCE
NOT TRUSTWORTHY	UNWARRANTED TRUST/ RELIANCE	WARRANTED DISTRUST/ UNDER-RELIANCE

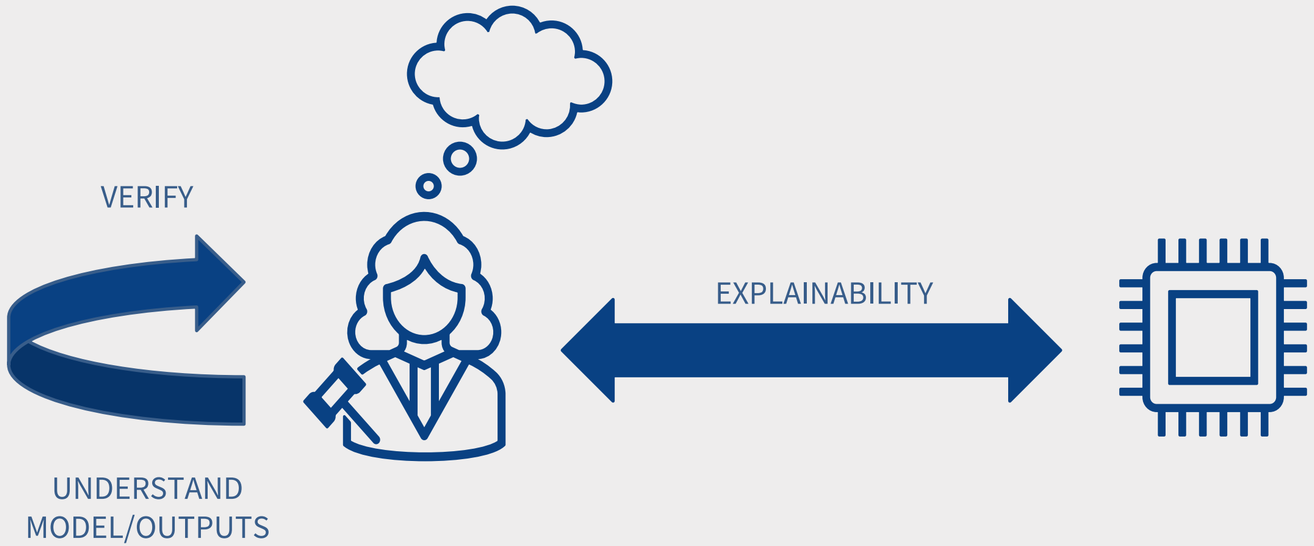
(DIS)TRUST AND (UNDER-)RELIANCE

	TRUSTED	DISTRUSTED
TRUSTWORTHY	**WARRANTED TRUST/RELIANCE**	~~UNWARRANTED DISTRUST/ UNDER-RELIANCE~~
NOT TRUSTWORTHY	~~UNWARRANTED TRUST/ RELIANCE~~	**WARRANTED DISTRUST/ UNDER-RELIANCE**

SELF-EXPLANATION IN DECISION MAKING



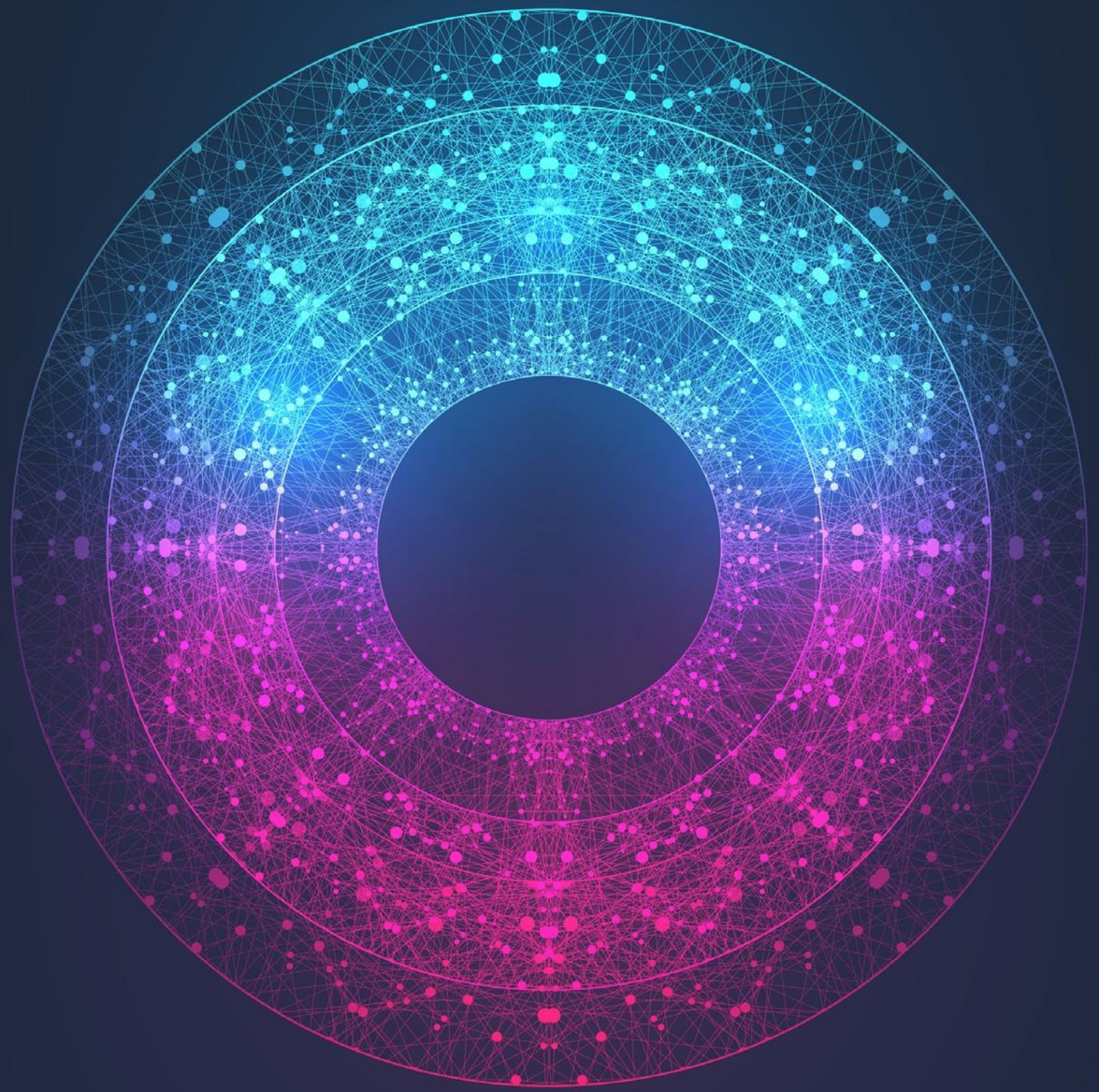
SELF EXPLANATION



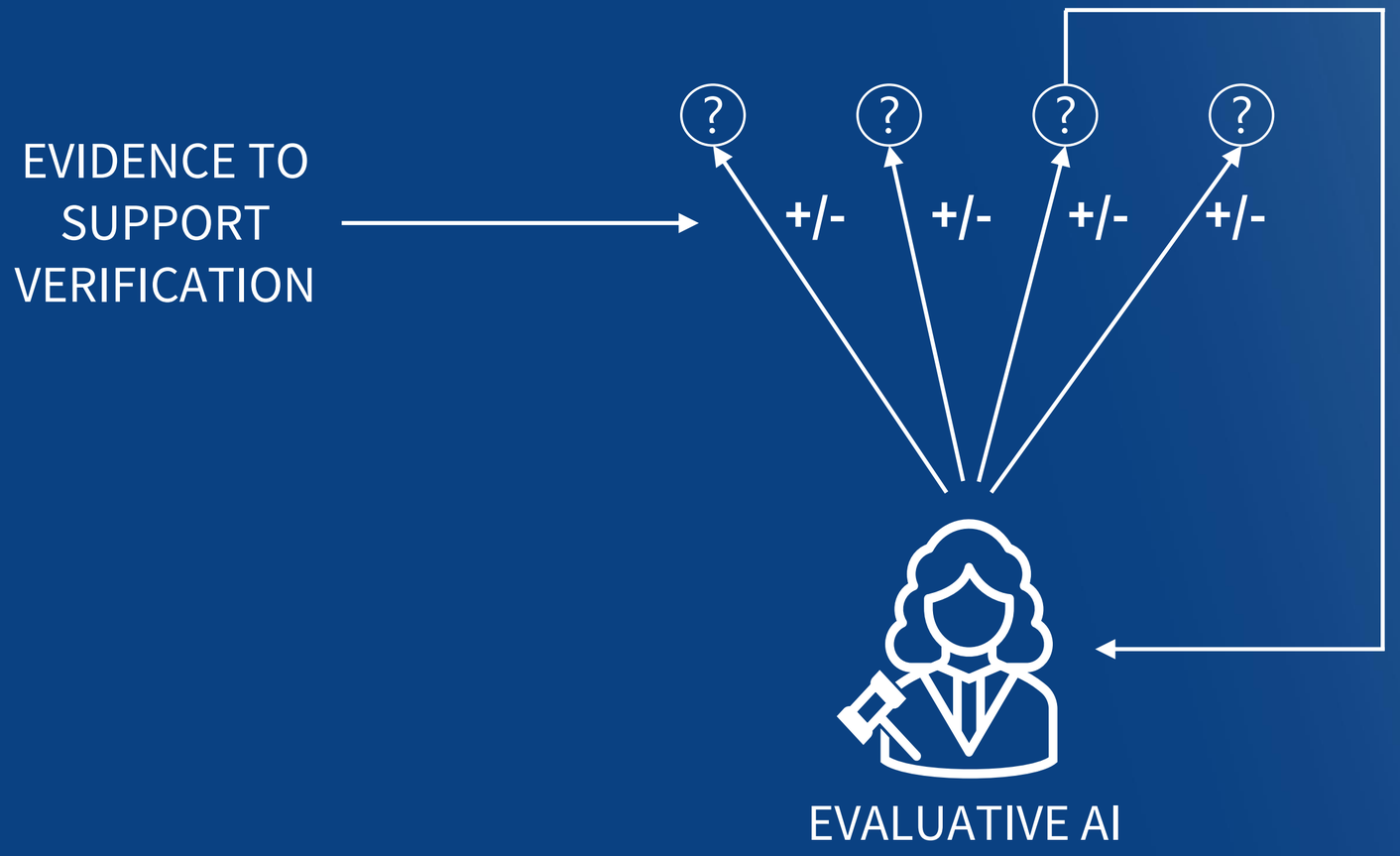
ABDUCTIVE REASONING AND VERIFICATION

PROCESS	REQUIREMENTS
1. Observe event	Design interfaces to determine what has happened
	Design interfaces to highlight unusual events
2. Generate hypotheses	Help to construct (likely) hypotheses
3. Judge plausibility	Help to explore how causes affect outputs
	Find evidence to support and refute hypotheses
4. Resolve explanation	Identify and record important information
5. Extend explanation	Support hypothesis revision
	Support interactive exploration

DECISION SUPPORT = VERIFICATION



EVALUATIVE AI



EVALUATIVE AI

Lesion

Notes

Patient reports itchiness and bleeding.
Lesion has changed colour.

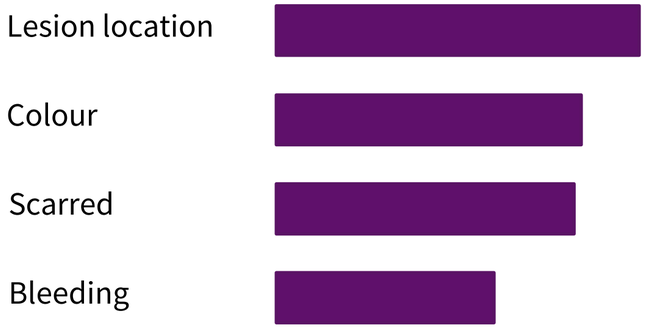
Lesion location

- ○ Head
- ○ Face
- ◉ Back
- ○ Front Torso
- ○ Upper arm
- ○ Hand/Lower Arm
- ○ Upper Leg
- ○ Foot/Lower Leg

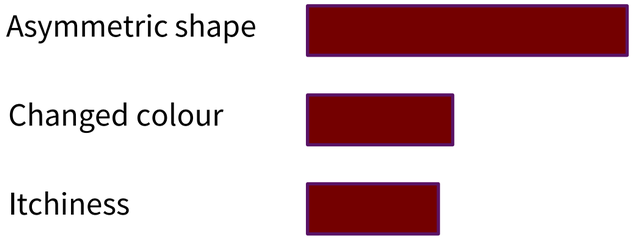
Your hypothesis

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

Evidence for



Evidence against



EVALUATIVE AI

Lesion

Notes

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

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

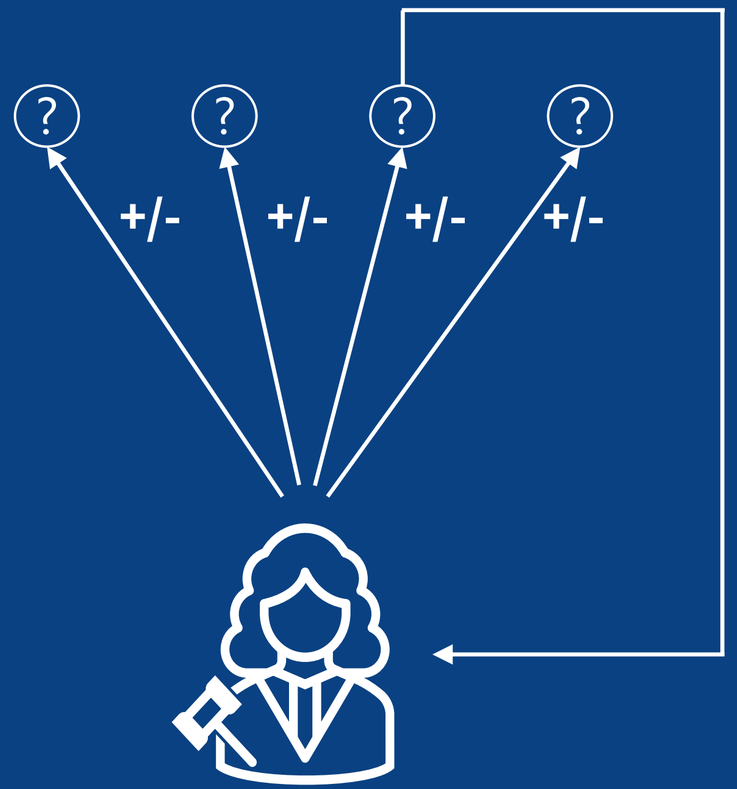
Evidence for

- Asymmetric shape
- Changed colour
- Itchiness
- Bleeding
- Colour

Evidence against

- Scarred
- Legion location

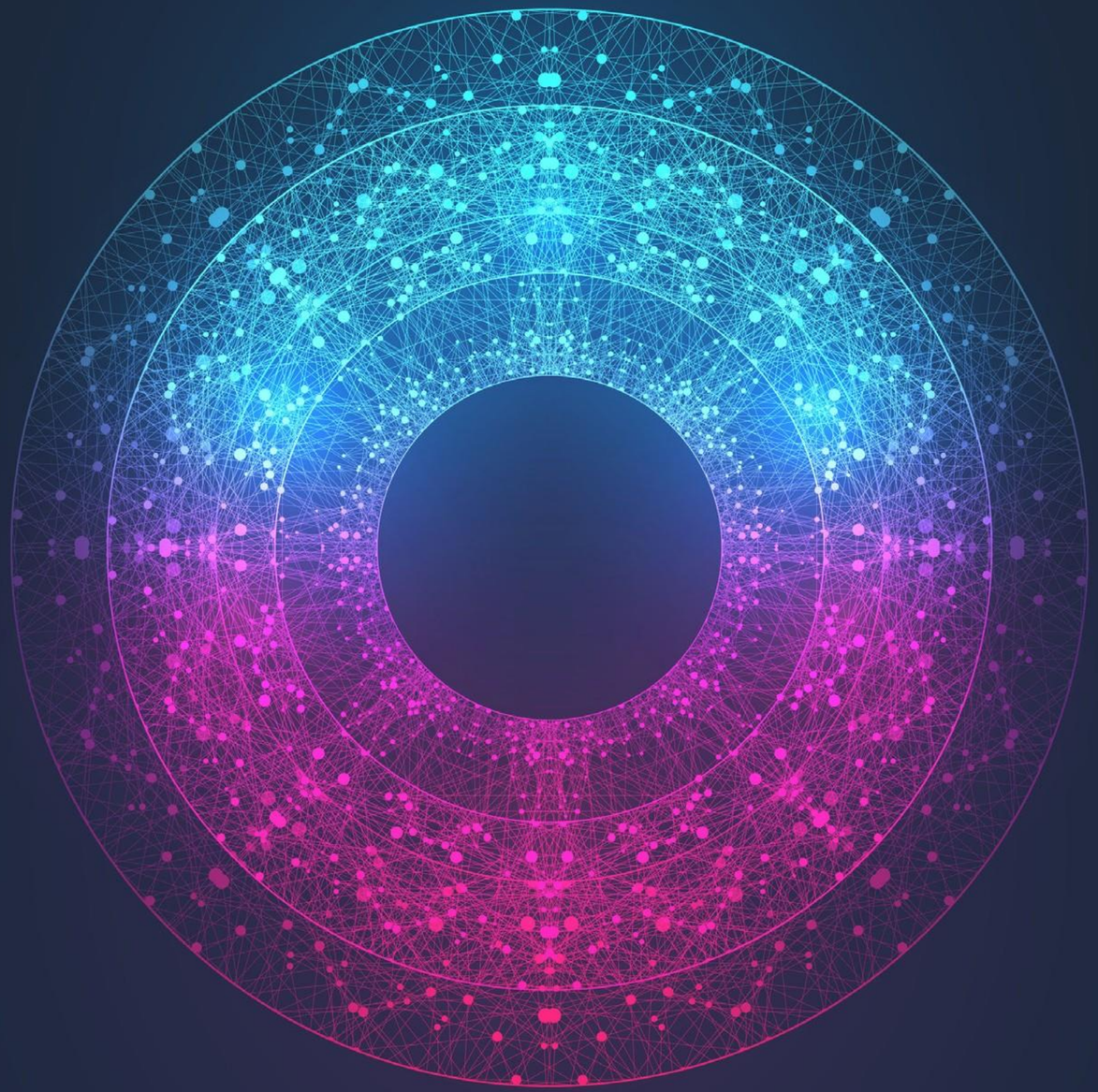
EVALUATIVE AI



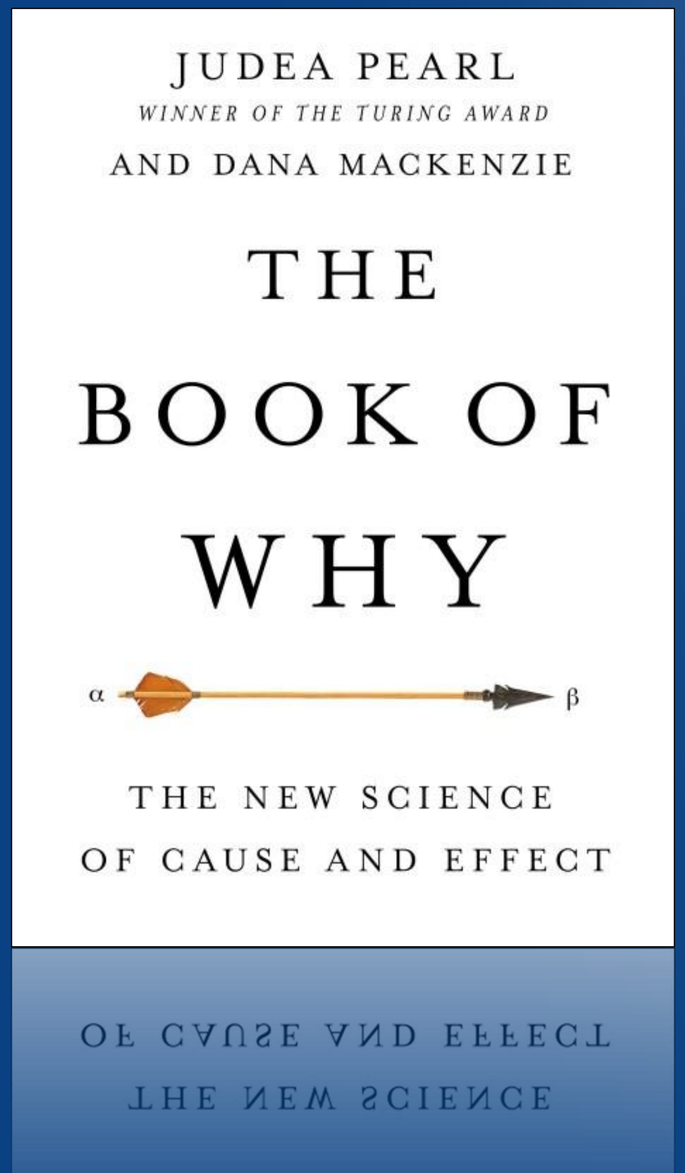
EVALUATIVE AI

CONTRASTIVE EXPLANATION

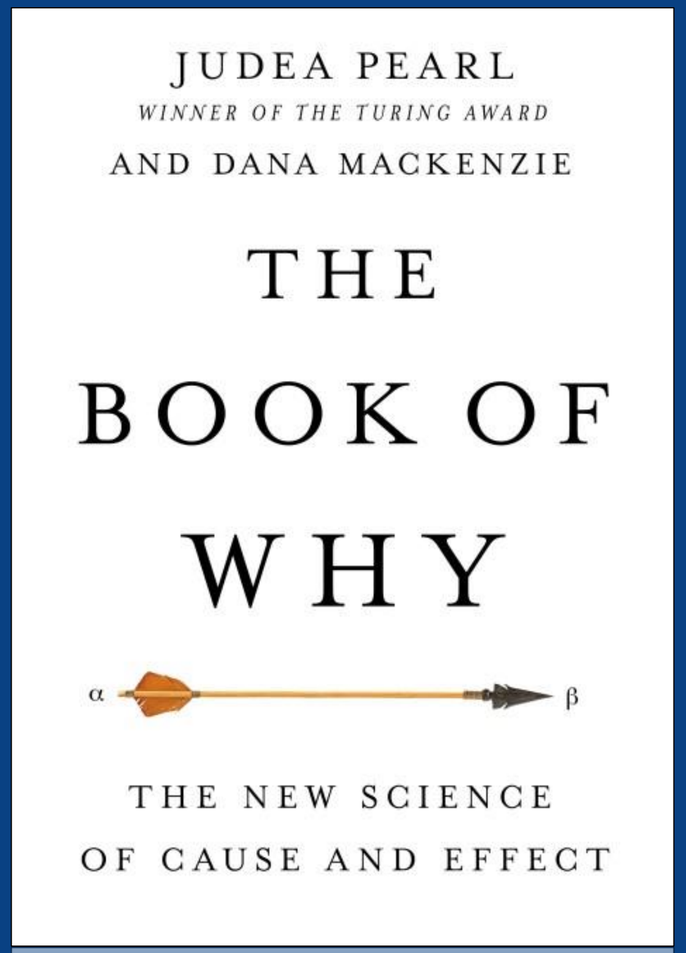
EVALUATIVE AI AND CAUSALITY

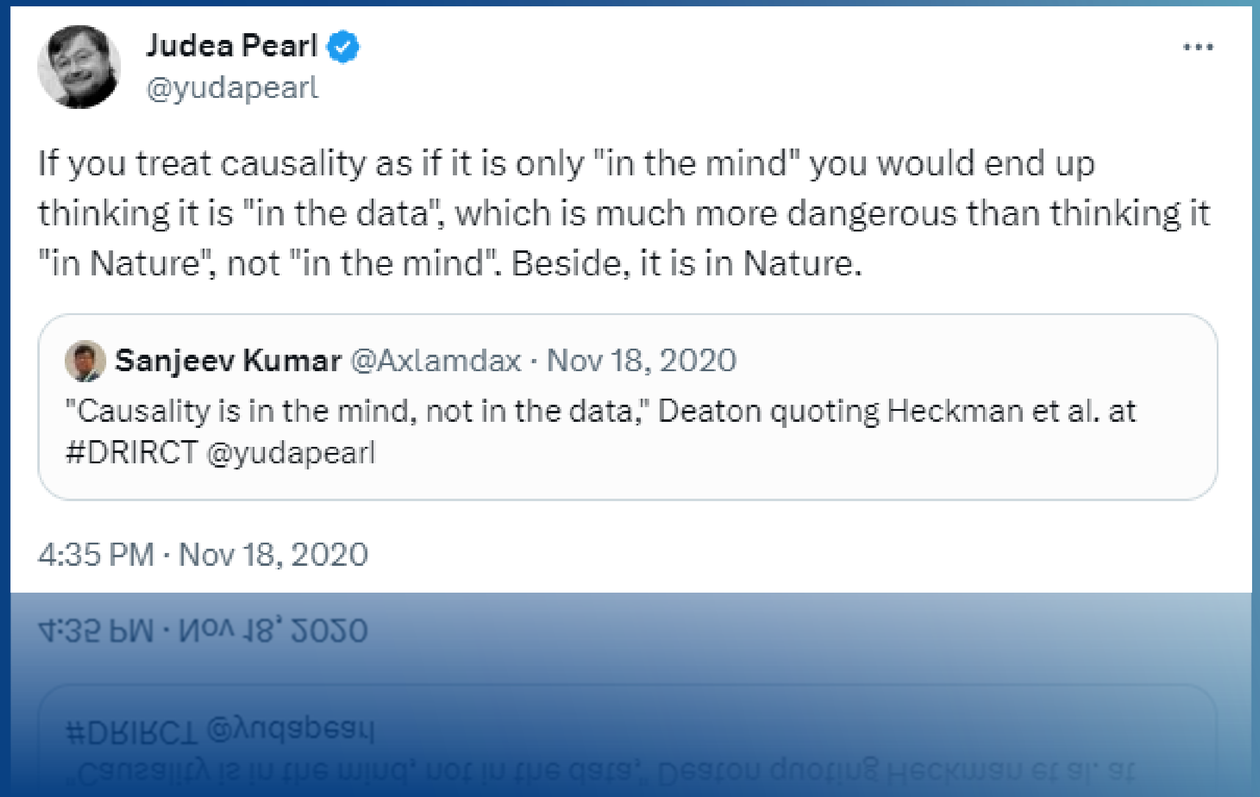


CAUSALITY AND VERIFICATION

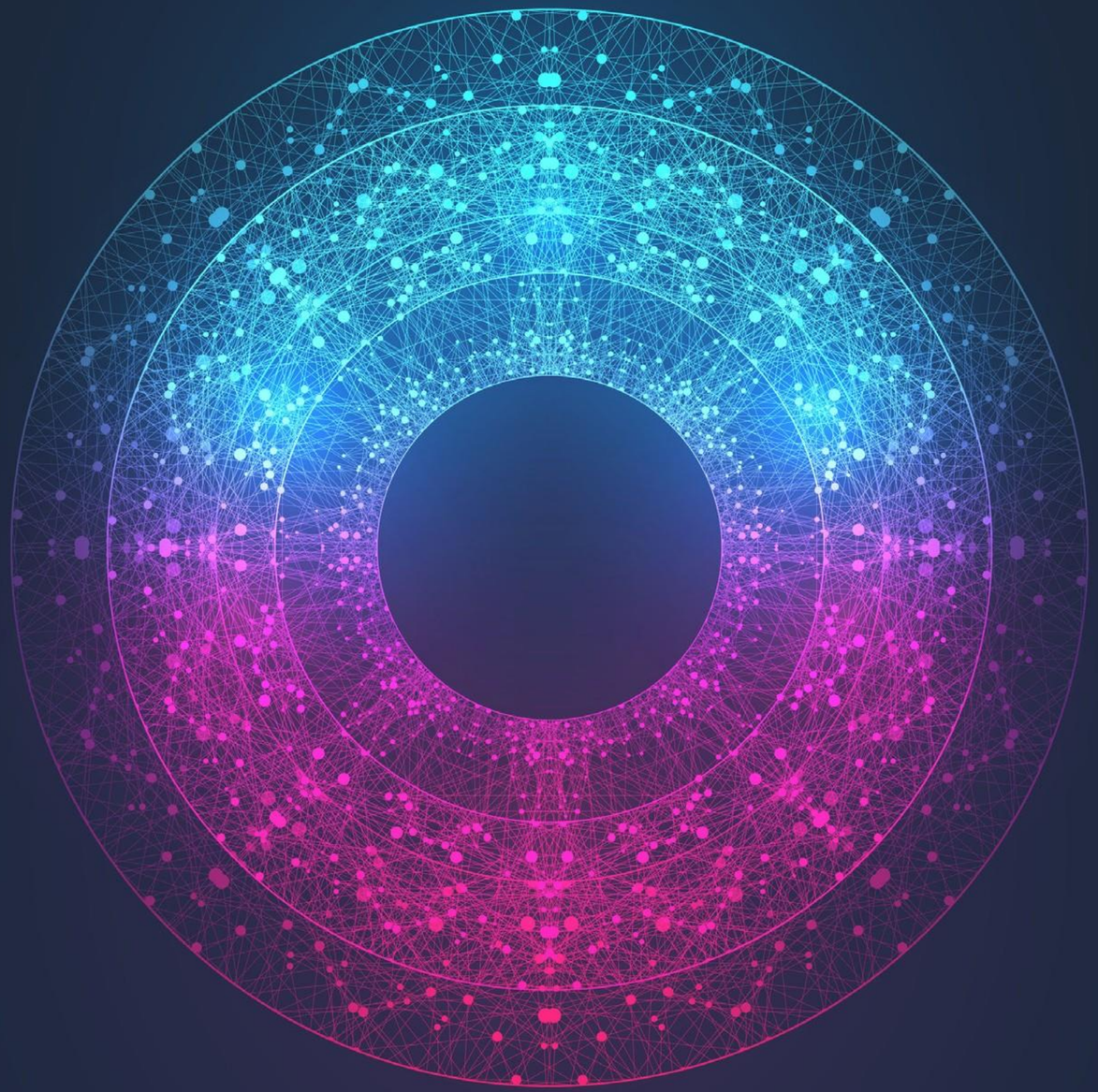


CAUSALITY AND VERIFICATION



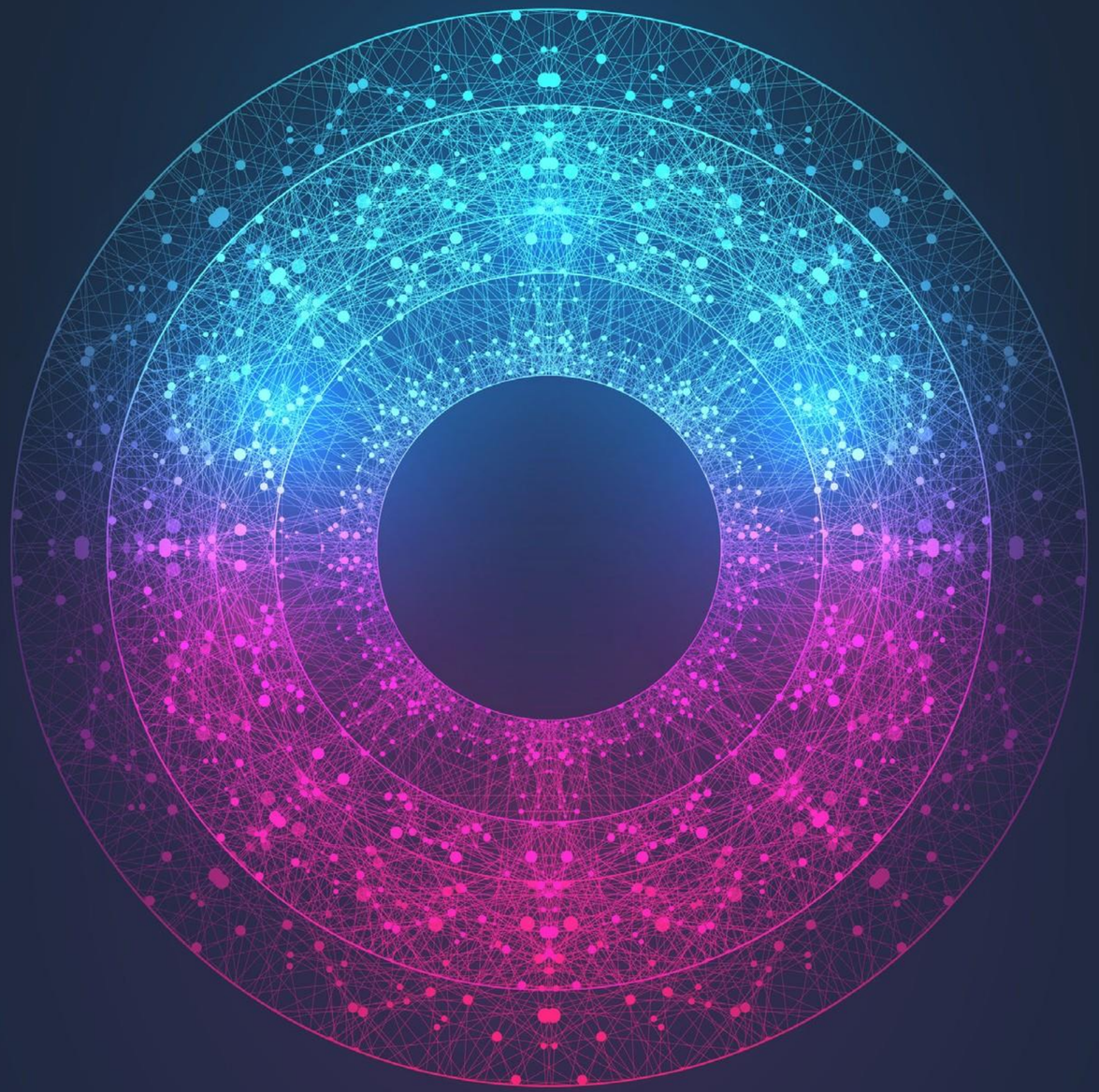


IS EXPLAINABLE AI DEAD?



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!