

Information Visualization

eXplainable Artificial Intelligence

Dr. David Koop

Using Visualization in AI/Deep Learning

Visual Analytics in Deep Learning | Interrogative Survey Overview

§4 WHY

Why would one want to use visualization in deep learning?

Interpretability & Explainability
Debugging & Improving Models
Comparing & Selecting Models
Teaching Deep Learning Concepts

§6 WHAT

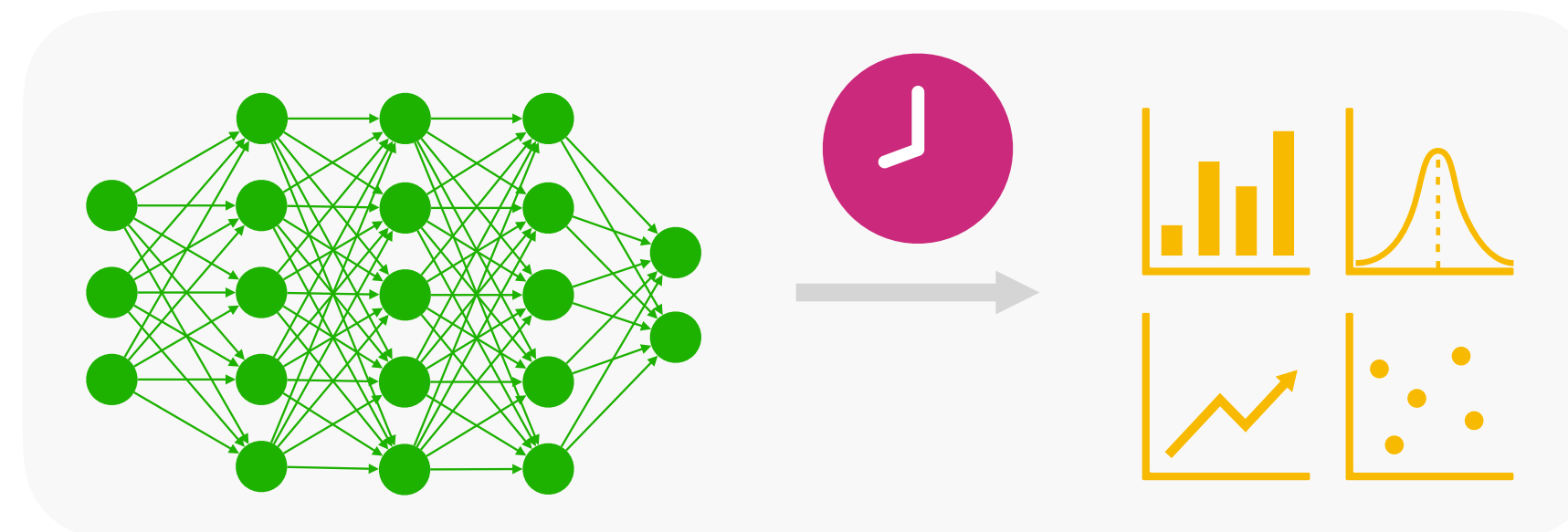
What data, features, and relationships in deep learning can be visualized?

Computational Graph & Network Architecture
Learned Model Parameters
Individual Computational Units
Neurons In High-dimensional Space
Aggregated Information

§8 WHEN

When in the deep learning process is visualization used?

During Training
After Training



§5 WHO

Who would use and benefit from visualizing deep learning?

Model Developers & Builders
Model Users
Non-experts

§7 HOW

How can we visualize deep learning data, features, and relationships?

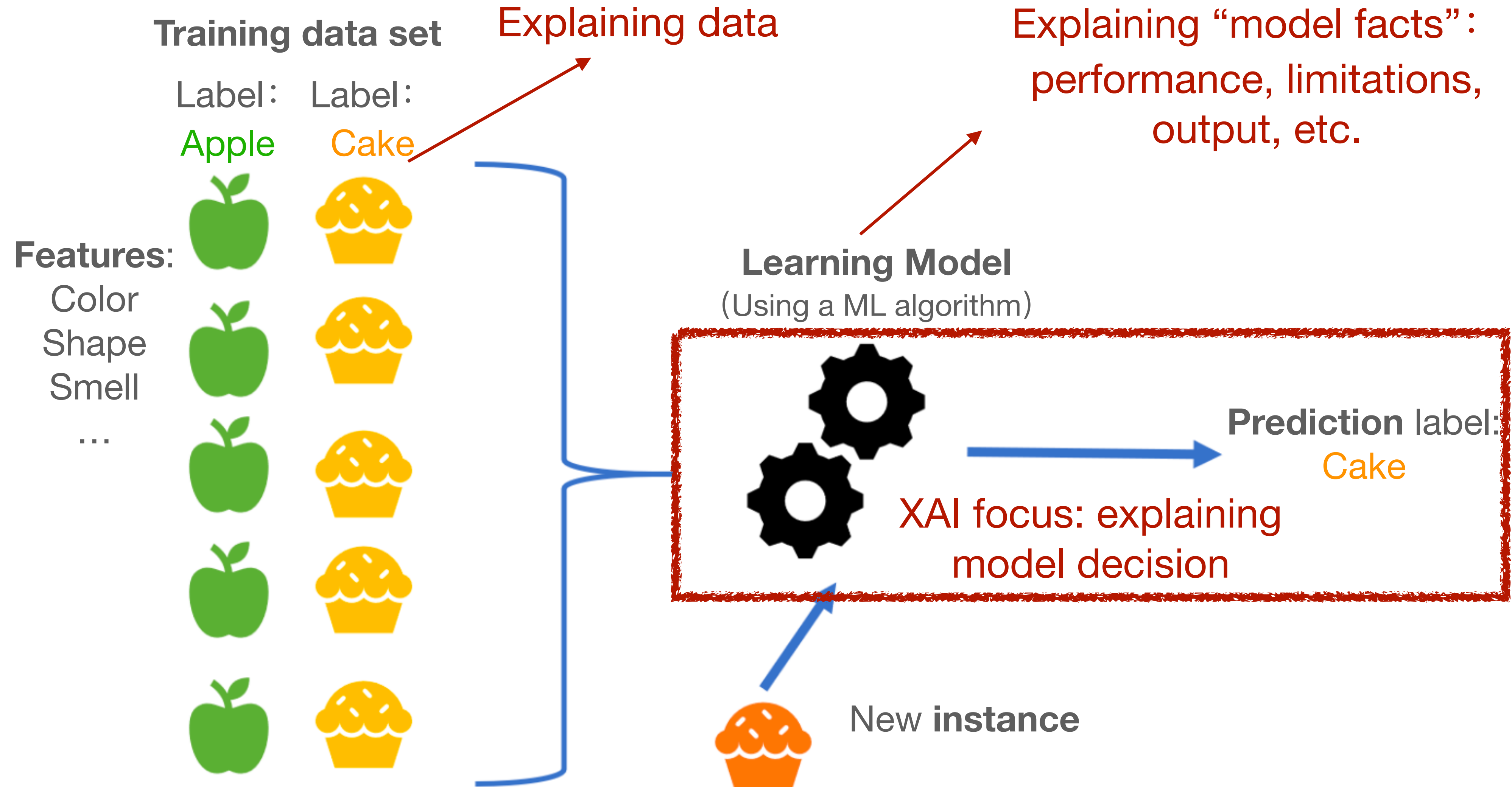
Node-link Diagrams for Network Architecture
Dimensionality Reduction & Scatter Plots
Line Charts for Temporal Metrics
Instance-based Analysis & Exploration
Interactive Experimentation
Algorithms for Attribution & Feature Visualization

§9 WHERE

Where has deep learning visualization been used?

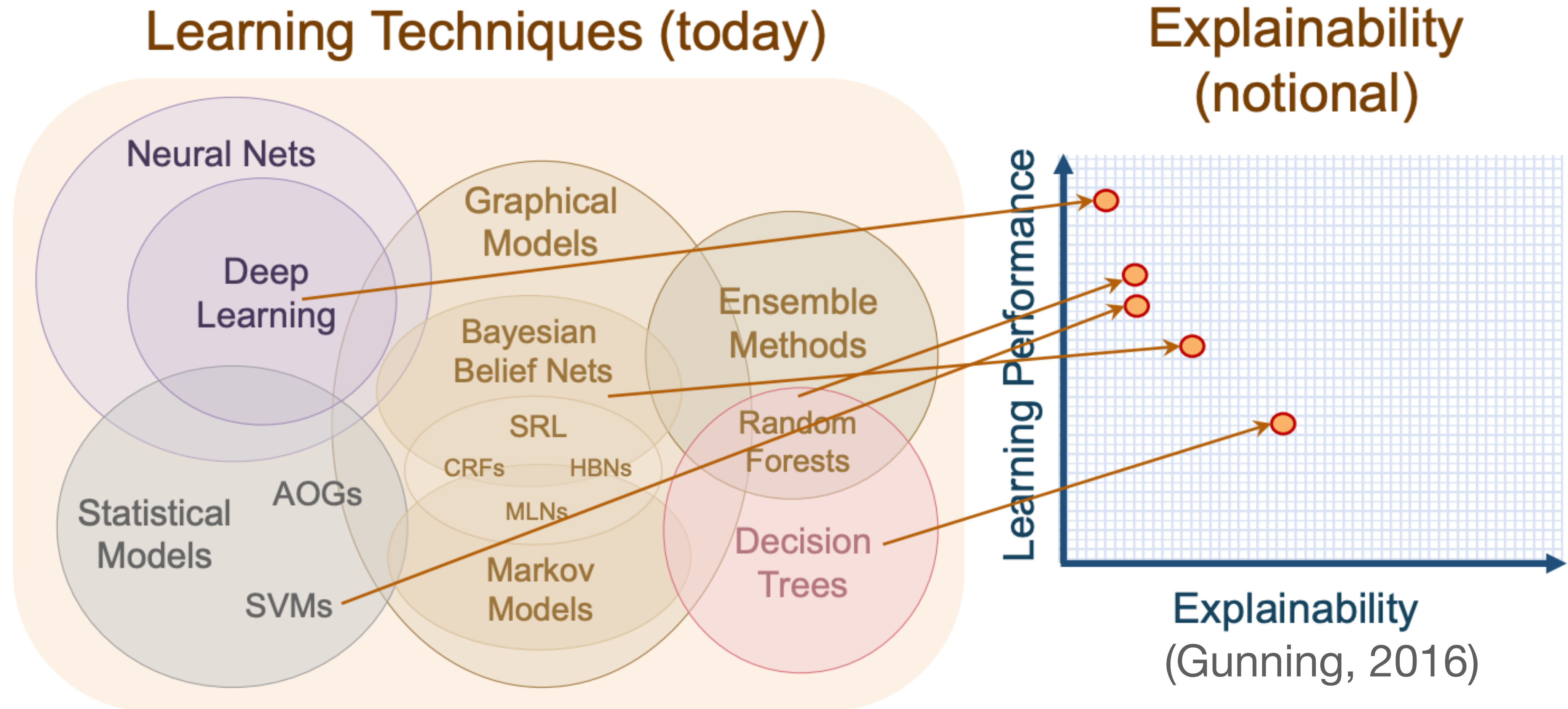
Application Domains & Models
A Vibrant Research Community

eXplainable Artificial Intelligence

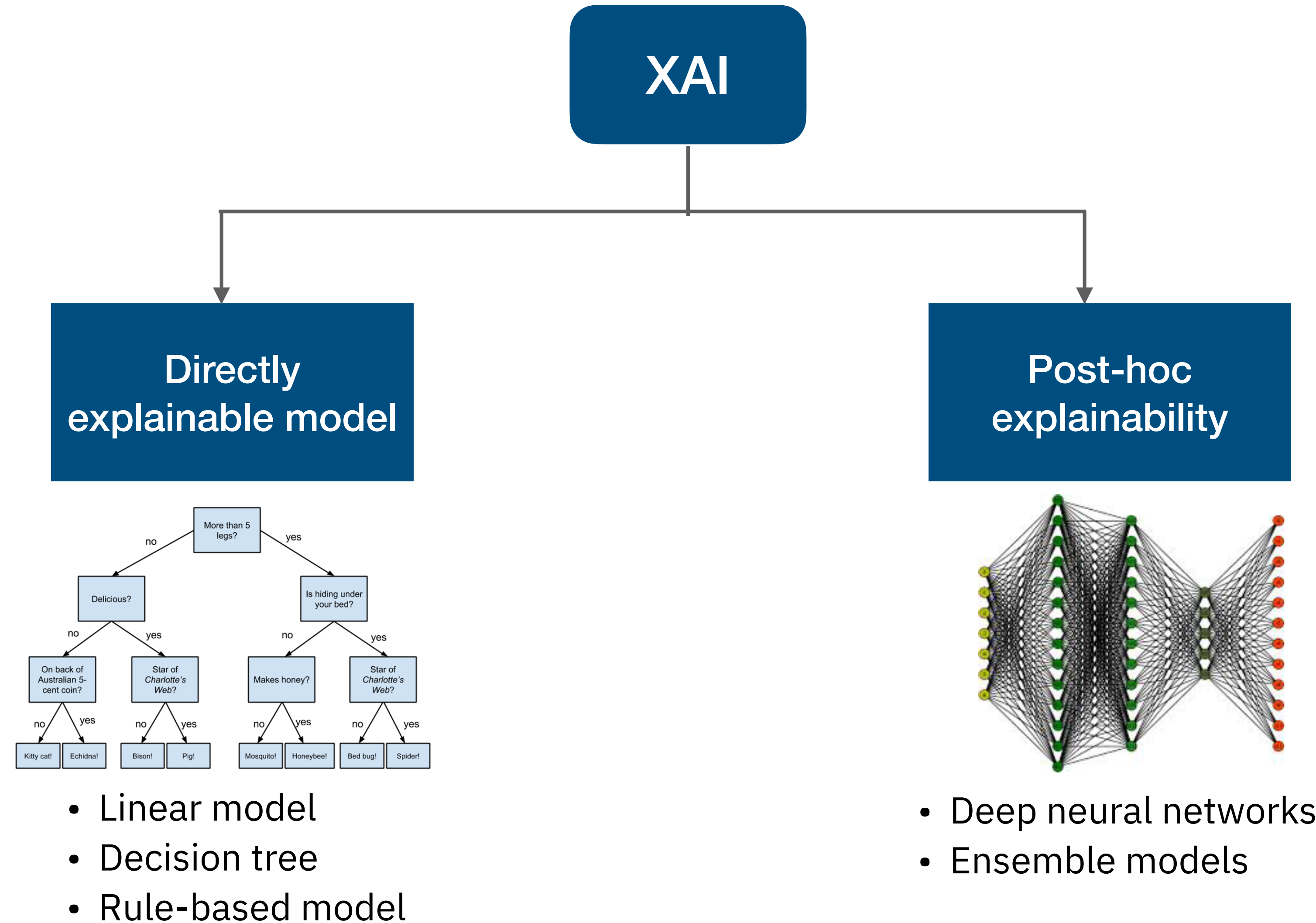


[Liao et al.]

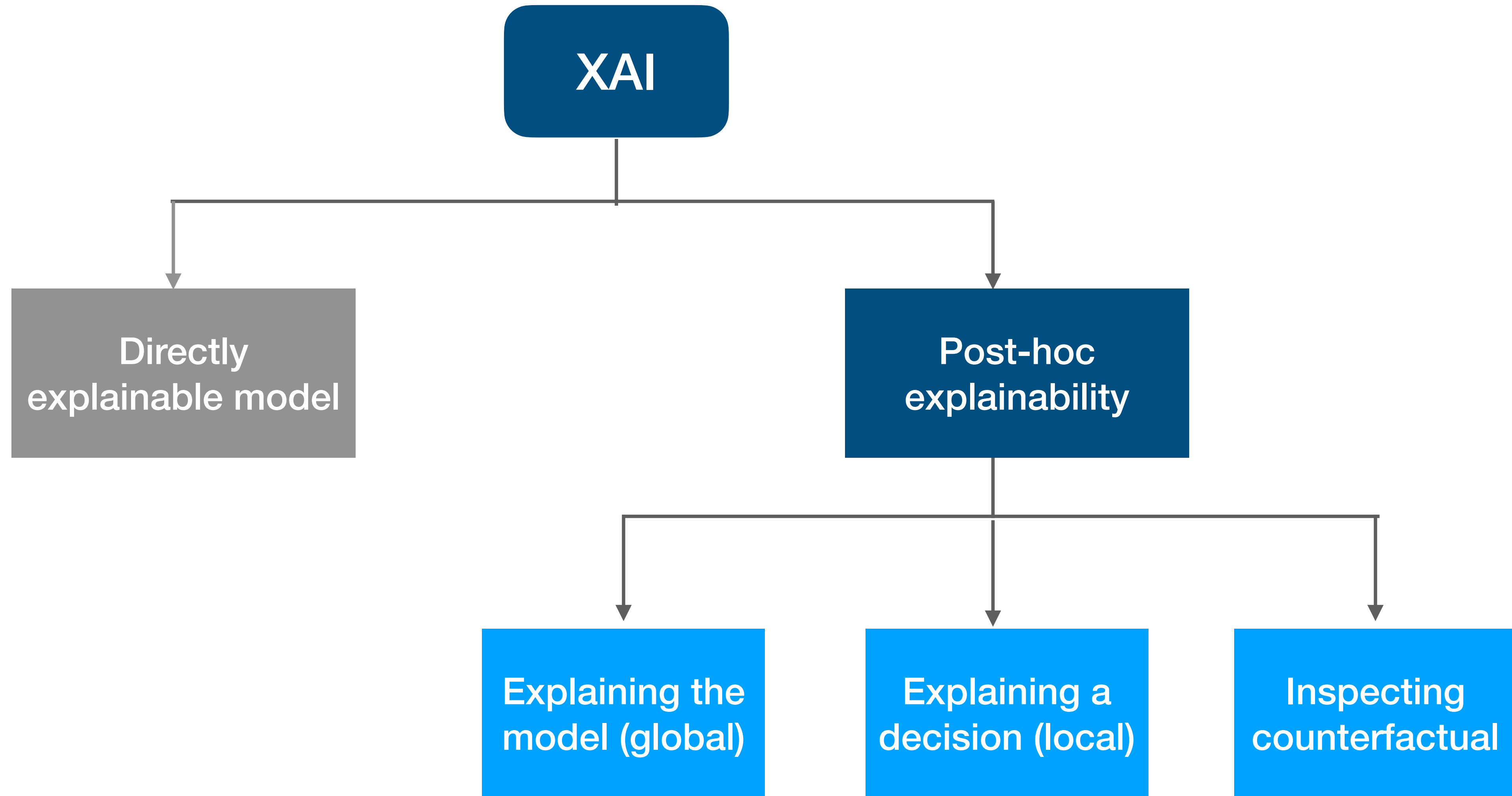
XAI Tradeoff



XAI Possibilities



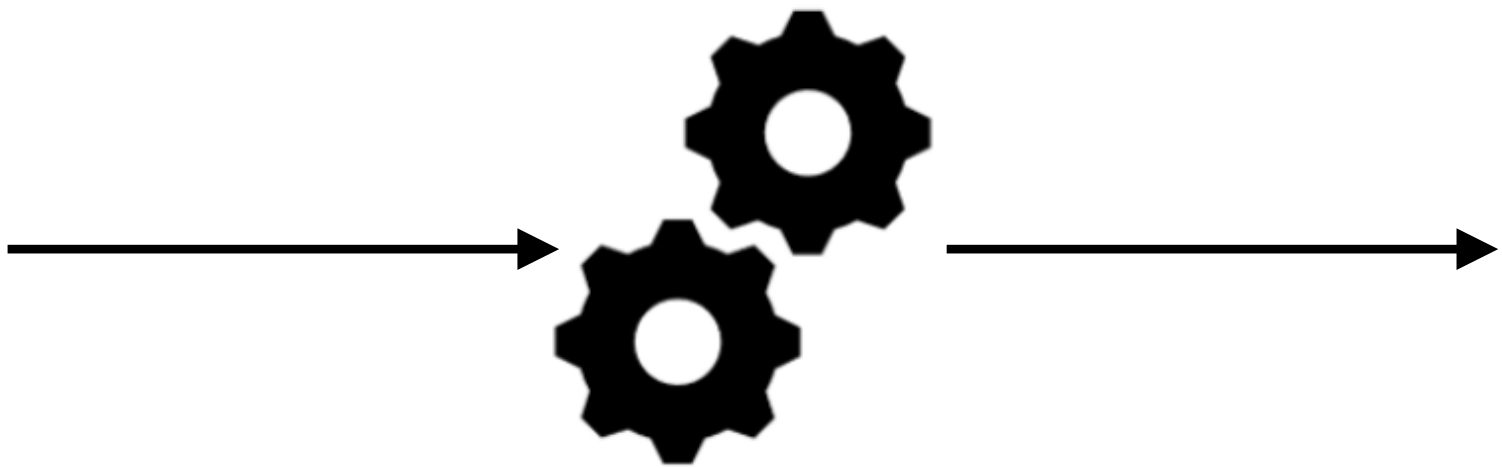
XAI Post-Hoc Techniques



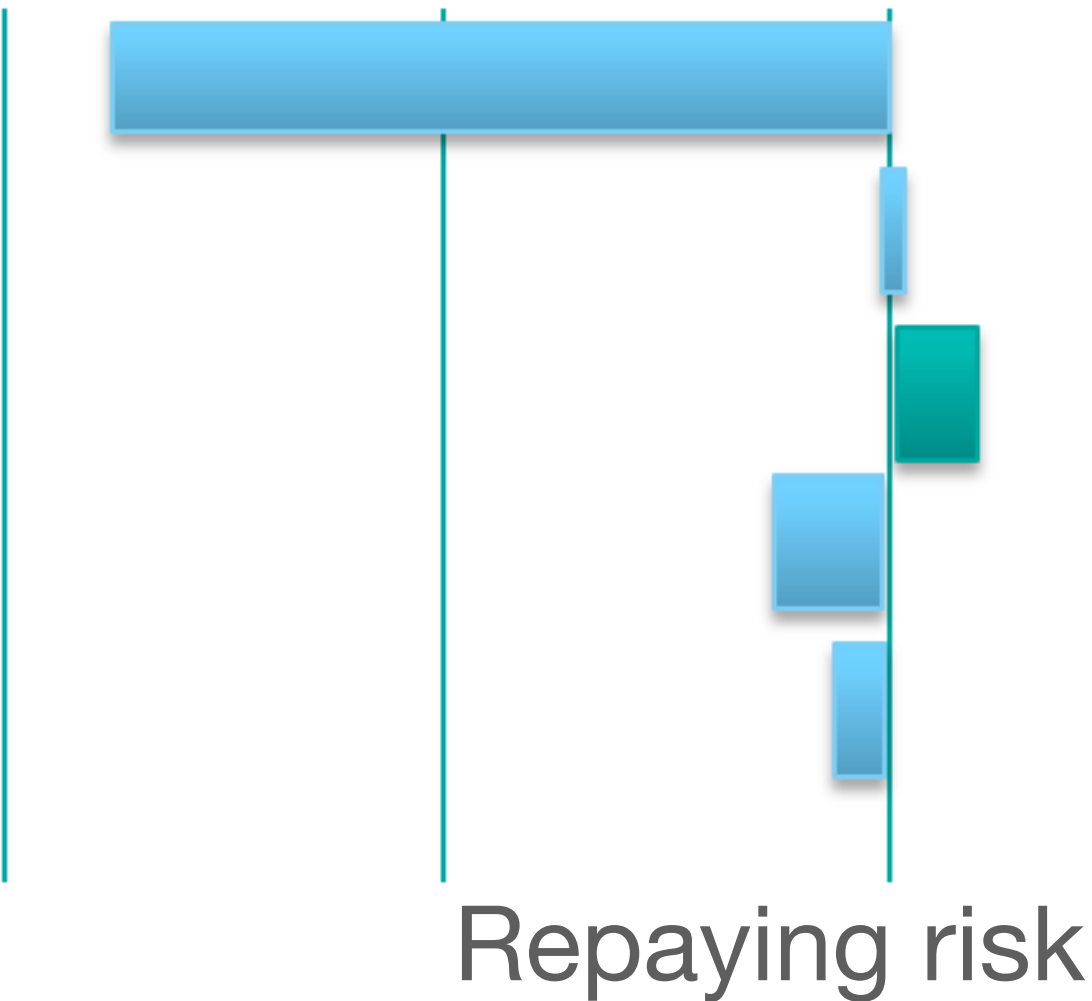
Feature Importance in a Decision

Customer: Jason

Assets score: 88
No. Of satisfactory trades: 0
Mo. since account open: 3
No. of inquiries: 1
Debt percentage: 10%



Risk of failing to repay: low



- Assets score
- No. Of satisfactory trades
- Mo. since account open
- No. Of inquiries
- Debt percentage



Loan officer

Why is Jason predicted of low risk?
Can I trust this prediction?

Counterfactual to Explain a Decision

Customer: Ana

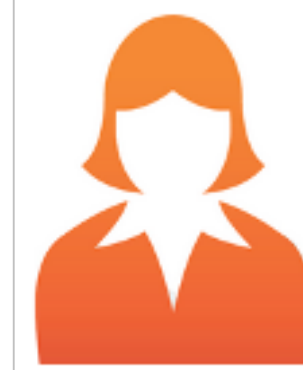
Assets score: 65
No. Of satisfactory trades: 1
Mo. since account open: 12
No. of inquiries: 4
Debt percentage: 50%



**Risk of failing
to repay: high**

Sue

Assets score: 66
No. Of satisfactory trades: 1
Mo. since account open: 12
No. of inquiries: 3
Debt percentage: 28%
Repaid on time



Bank customer

Why was my loan application rejected?
How can I improve in the future?

XAI Evaluation

Inherent “goodness” metrics

- Fidelity/faithfulness
- Stability
- Compactness
- ...

Faithfulness

Correlation between the feature importance assigned by the interpretability algorithm and the effect of features on model accuracy.



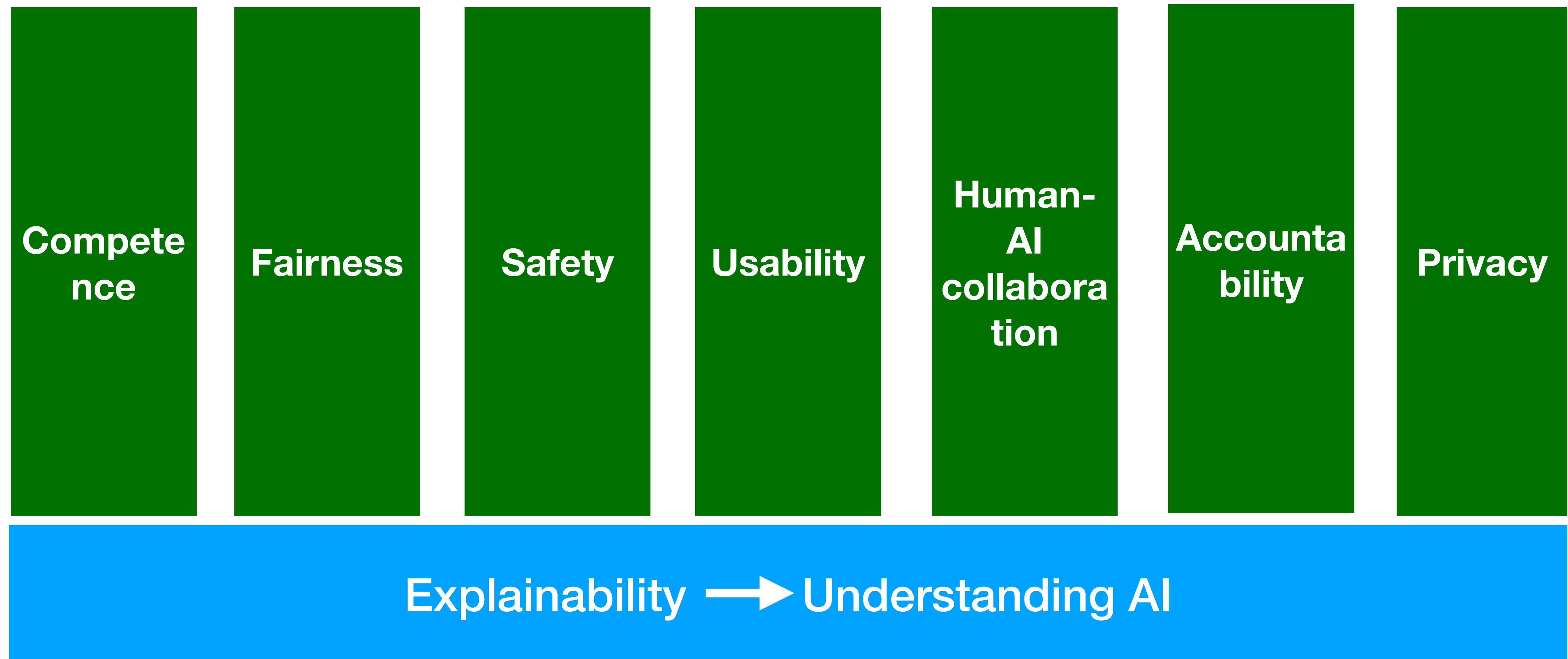
User-dependent measures

- Comprehensibility
- Explanation satisfaction
- ...

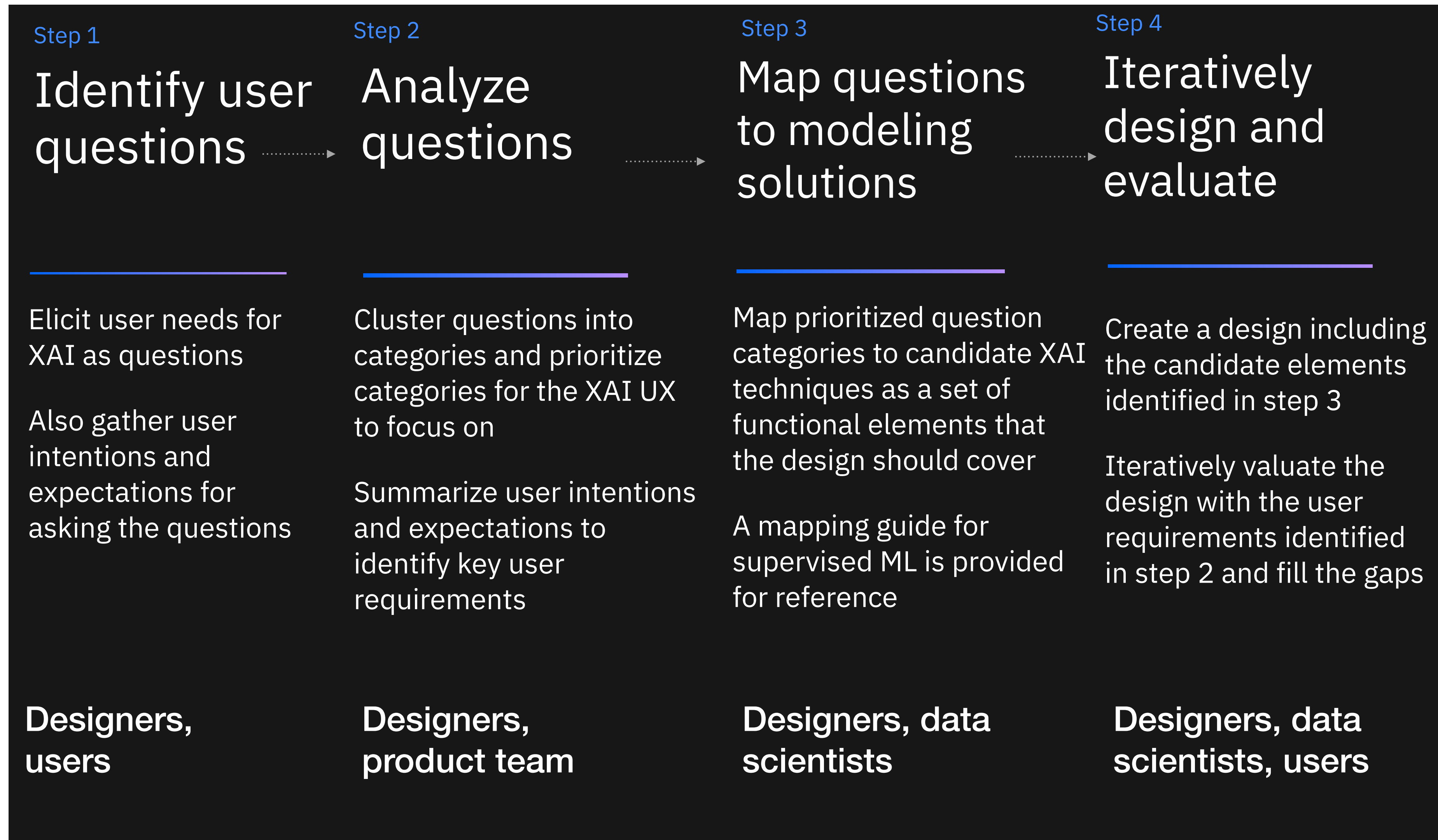
Task oriented measures

- Task performance
- Impact on AI interaction
 - Trust (calibration) in model
- Task or AI system satisfaction

XAI as the Foundation for Responsible AI



Question-Driven XAI UX



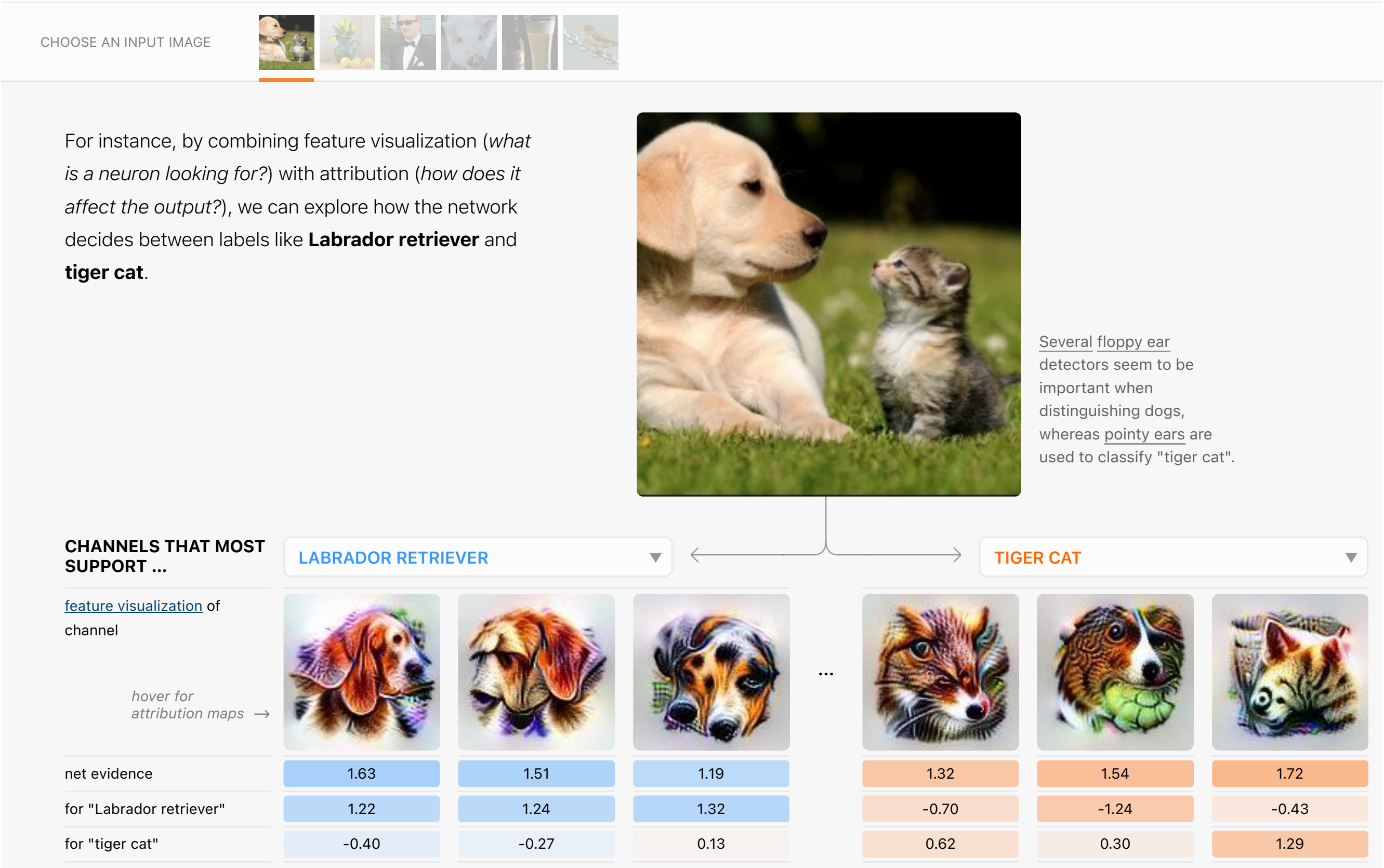
Projects

- Time to make progress on projects
- End of semester quickly approaching
- Questions?

Today's Paper

The Building Blocks of Interpretability

Interpretability techniques are normally studied in isolation.
We explore the powerful interfaces that arise when you combine them—
and the rich structure of this combinatorial space.



[C. Olah et al.]