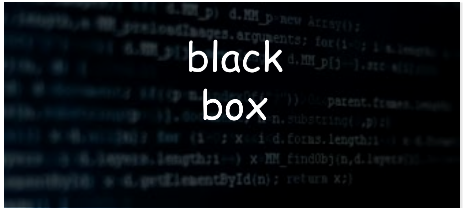
Causes and Explanations in Practical Applications

Hana Chockler

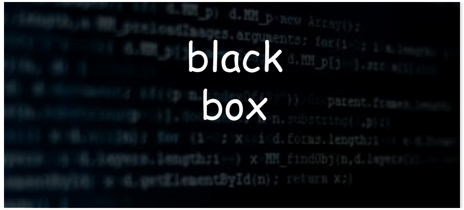
CAUSAL LENS

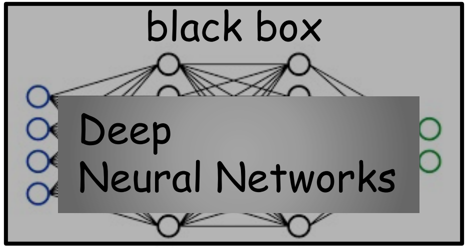
KING'S College LONDON

Modern computerized systems are **huge and difficult to understand**



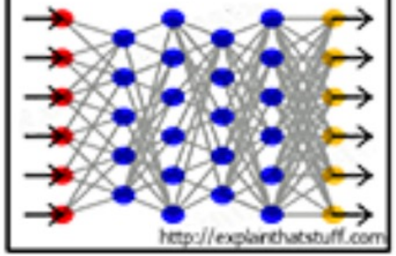
Modern computerized systems are
huge and difficult or even impossible to understand





From DARPA:

AI System



- We are entering a new age of AI applications
- Machine learning is the core technology
- Machine learning models are opaque, non-intuitive, and difficult for people to understand

DoD and non-DoD Applications

Transportation

Security

Medicine

Finance

Legal

Military

User

- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

ico.
Information Commissioner's Office

The UK's independent authority set up to uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals.

Home | Your data matters | For organisations | Make a complaint | Action we've taken

For organisations / Guide to Data Protection / Key DP themes /
Explaining decisions made with Artificial Intelligence

Explaining decisions made with AI

ico. | The Alan Turing Institute

GDPR right to explanation

EUROPEAN COMMISSION

Brussels, 19.2.2020
COM(2020) 65 final

WHITE PAPER

On Artificial Intelligence - A European approach to excellence and trust

Who is the recipient of explanations?

Laypeople

- Why did you do that?
- Can I trust you?

Experts

- Why not something else?
- What if…?

Developers

- When does the system succeed?
- When does it fail?
- How can I correct an error?

Build Causal AI

The new data science workflow

For data scientists

Causal Graph Discovery

Proprietary framework to discover causal drivers from raw data and filter spurious correlations.

Can incorporate domain knowledge & constraints into the causal graph.

Causal AI Model Discovery

Proprietary implementations that take a causal graph & its constraints as input to build a causal model.

Business Interventions & Simulations

Proprietary implementations that recommending actions which lead to the desired outcome.

Proprietary modelling of what-if questions and scenario planning

Causal AI is able to ‘imagine’ counterfactual scenarios

For data scientists

Answer ‘Why’ and ‘What-if’ questions

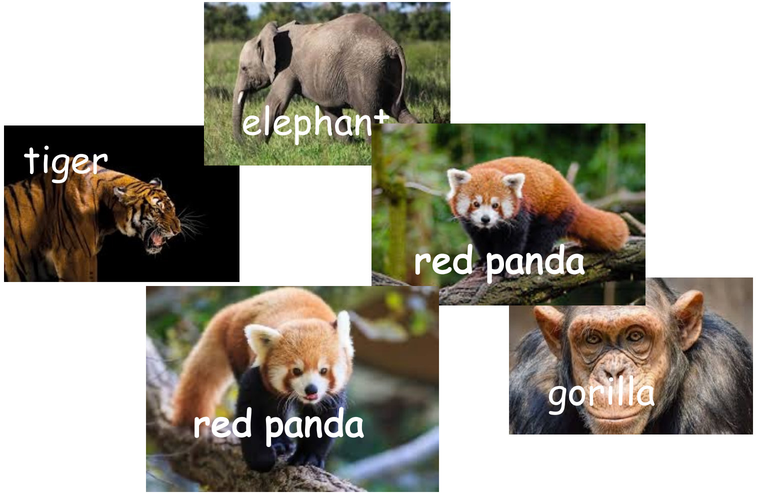
Causal AI can explore potentialities that never actually happened — “counterfactuals” — while maintaining a connection with reality. It can reimagine the past, explaining why events unfolded as they did.

- Why are our factory’s solar panels failing with heat spots? What’s the root cause?
- How would growth stocks have been impacted this past quarter had consumer price inflation risen by 1%?
- What would have happened to quarterly sales had we not increased advertising spend?

Deep Neural Networks for image classification

Training phase

For developers



DNN for classifying animals

?

After sufficient training...

Deep Neural Networks for image classification

Classification phase

DNN for classifying animals

red panda

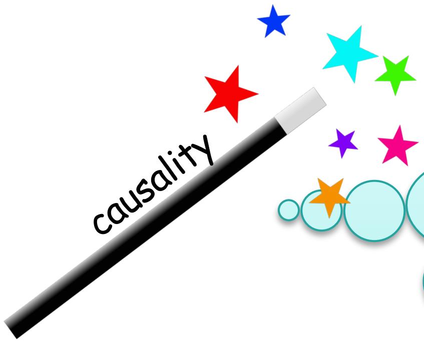
How can we trust DNN's output?

Explanations for Deep Neural Network's decisions

DNN for classifying animals

red panda

Because of this part: 



Explanation:
minimal, sufficient,
non-trivial subset
of highest-ranked
causes

How to detect misclassification?

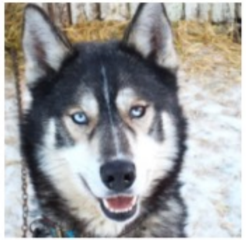
Example: wolves vs huskies

Training phase:

Pictures of wolves and huskies

DNN

Classification phase:



(husky)

DNN

~~wolf~~

© "Why Should I Trust You?"

Subtle misclassification – uncovered by explanations

DNN for classifying images

seems ok

cowboy hat

Because of this part: 

Explanation uncovered misclassification!

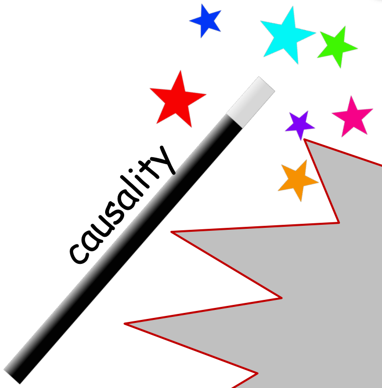
Explanations for DNN's decisions –Based on ranking of causes

DNN for classifying animals

red panda

Because of this part:

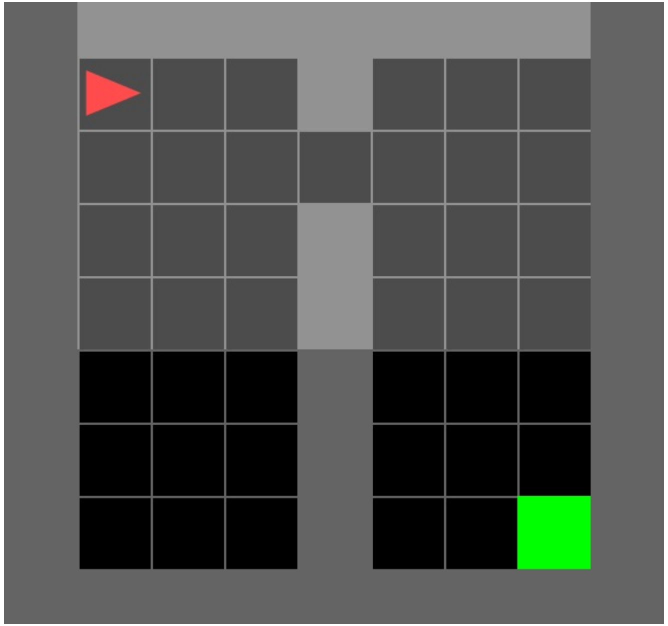
How to compute a minimal and sufficient subset of causes efficiently?



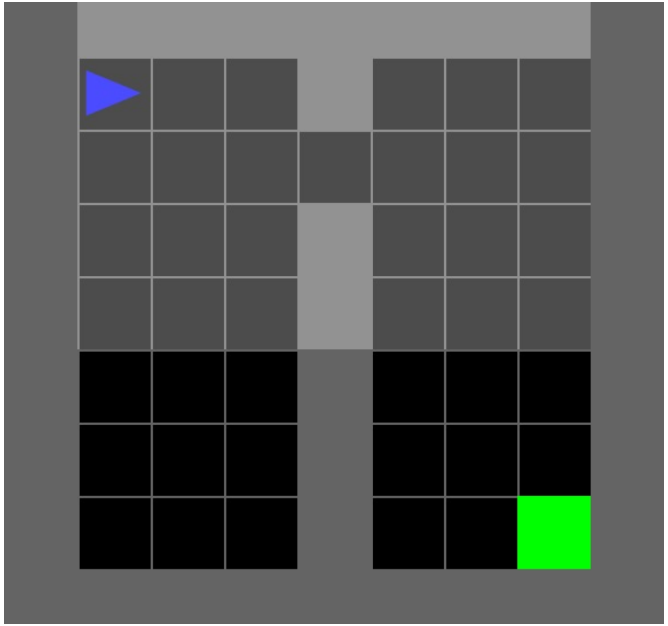
Use causal analysis for ranking

Explanations of Reinforcement Learning Policies

Original RL policy



Policy with only the top-ranked decisions

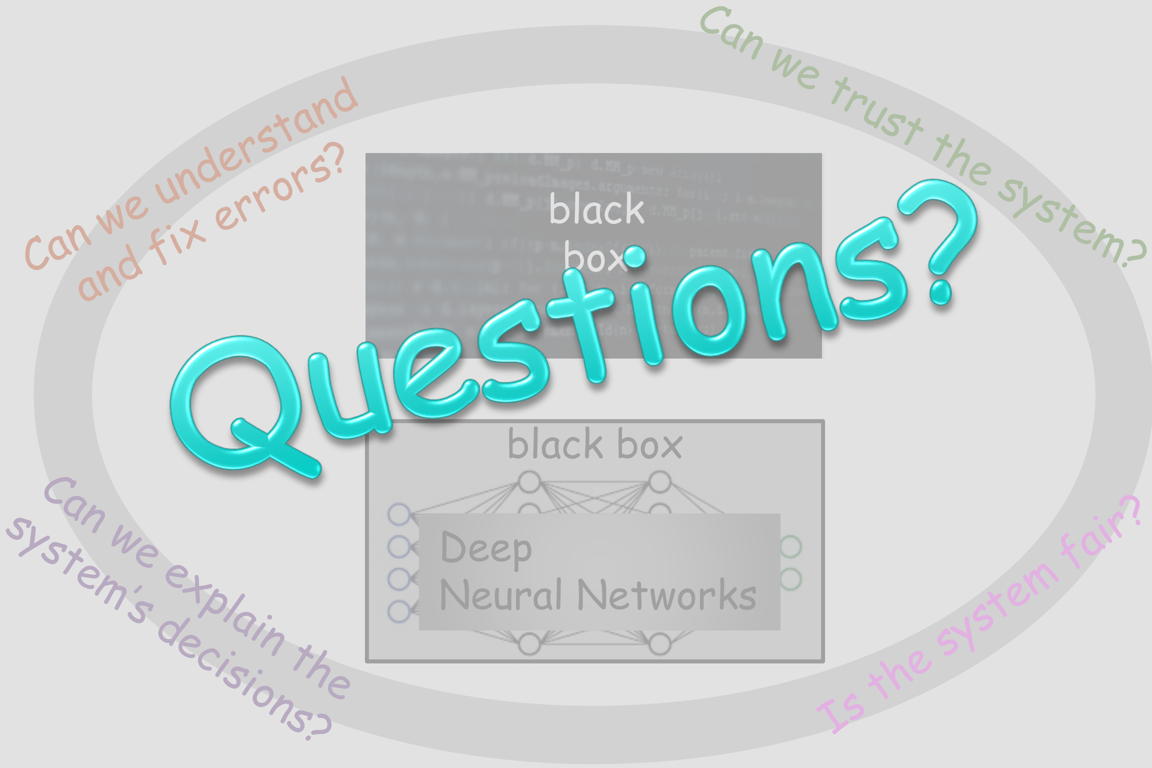


causaLens

We are hiring!

- Research Scientist - Machine Learning and Causality
- Director of Applied Data Science
- VP of Engineering
- Engineering Team Lead

Please email join@causaLens.com for more information


Can we understand and fix errors?
Can we trust the system?
black box
Questions?
black box
Deep Neural Networks
Can we explain the system's decisions?
Is the system fair?