Existential Questions for Machine Learning

Coralie Busse-Grawitz

bcoralie@ethz.ch

13. August 2019

How reliable is Machine Learning?

ML quality measurement raises existential questions

What is the **origin** of ML models?

Can we **trust** ML?

What does ML tell us about the **truth**?

What is the **purpose** of ML?

ML quality measurement raises existential questions

What is the **origin** of ML models?

Can we **trust** ML?

What does ML tell us about the **truth**?

What is the **purpose** of ML?

To measure ML model quality,
we must understand its origin

Complex question → Data → ML model → Complex answer

ML quality measurement raises existential questions

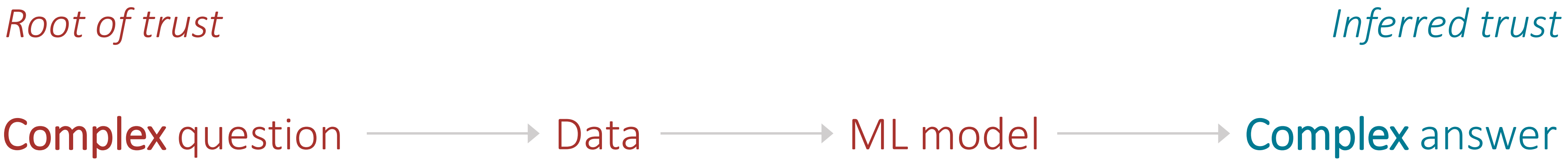
What is the **origin** of ML models?

Can we **trust** ML?

What does ML tell us about the **truth**?

What is the **purpose** of ML?

To measure ML model quality, we evaluate the «stack of trust»

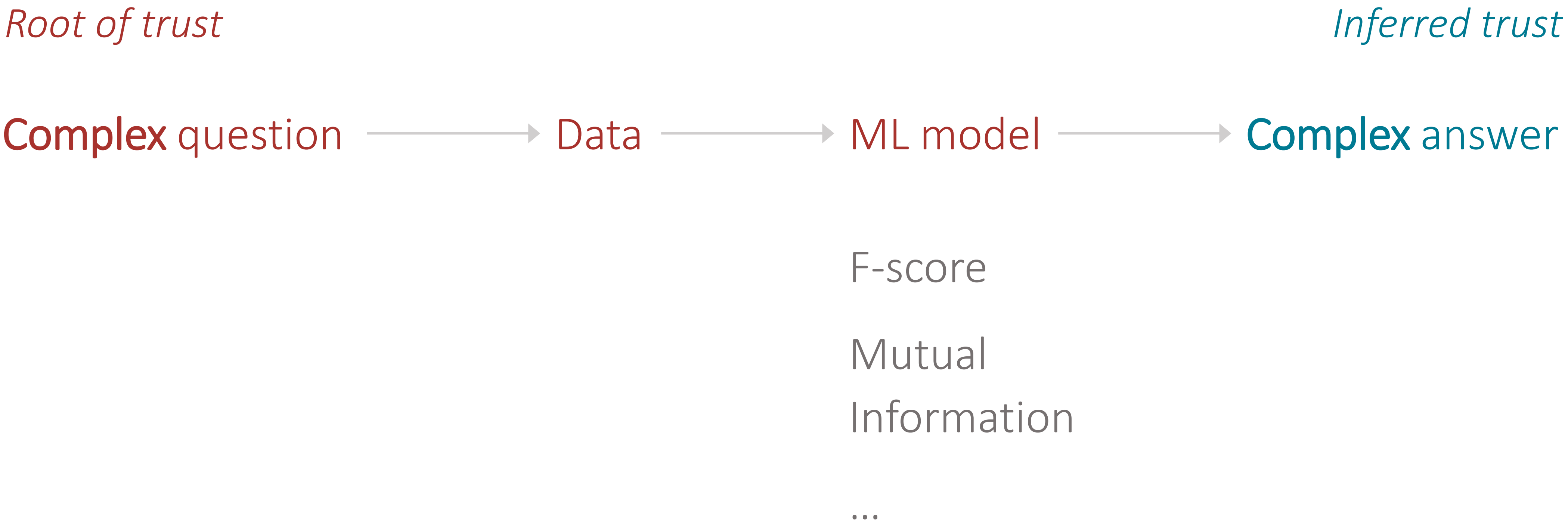


Root of trust

Inferred trust

Complex question → Data → ML model → **Complex** answer

To measure ML model quality, we evaluate the «stack of trust»



Root of trust

Inferred trust

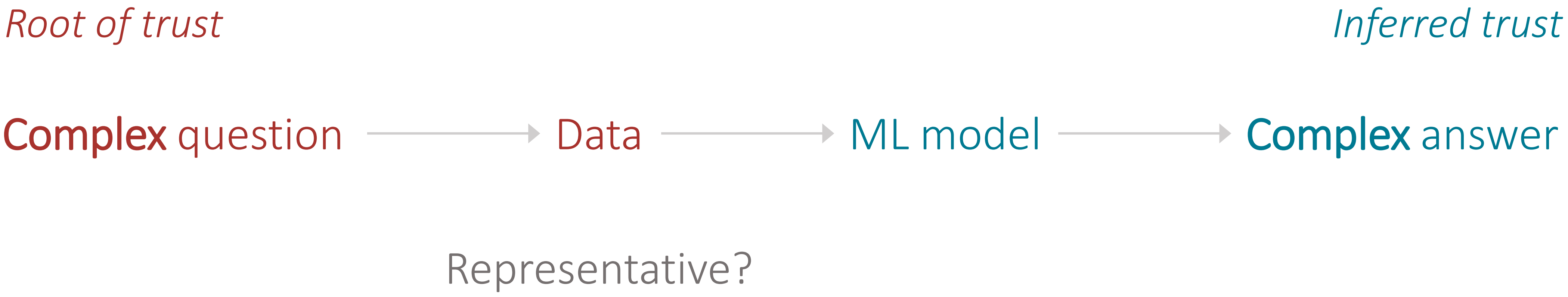
Complex question → Data → ML model → **Complex** answer

F-score

Mutual Information

…

To measure ML model quality,
we evaluate the «stack of trust»



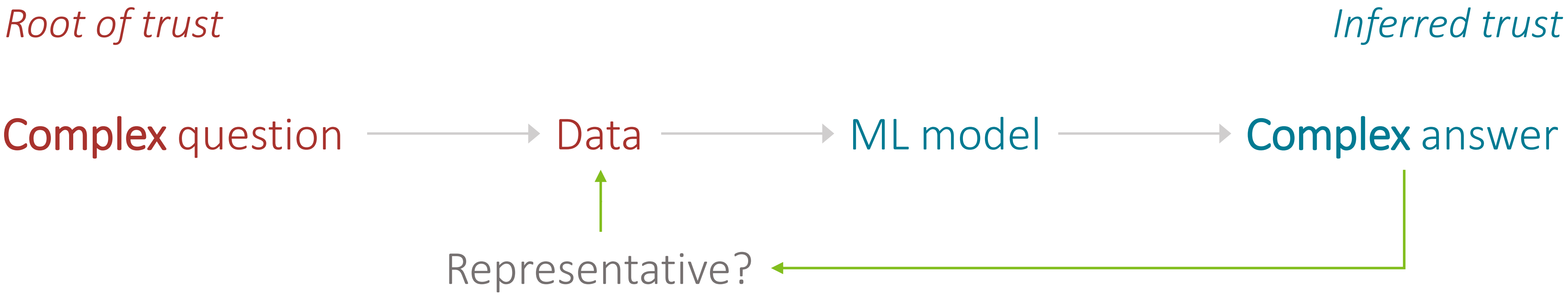
Root of trust

Inferred trust

Complex question → Data → ML model → **Complex** answer

Representative?

To measure ML model quality, we evaluate the «stack of trust»



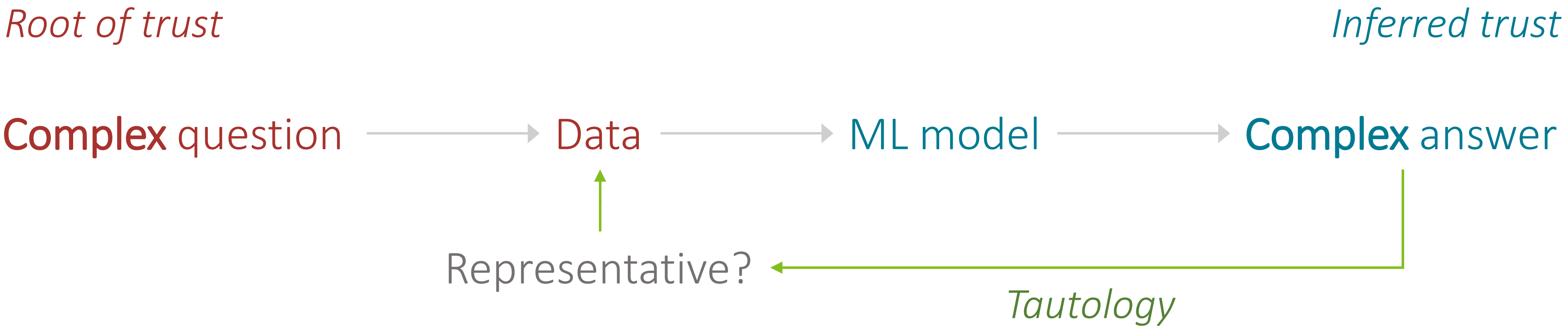
Root of trust

Inferred trust

Complex question → Data → ML model → **Complex** answer

Representative?

To measure ML model quality, we evaluate the «stack of trust»



Root of trust

Inferred trust

Complex question → Data → ML model → **Complex** answer

Representative?

Tautology

ML quality measurement raises existential questions

What is the **origin** of ML models?

Can we **trust** ML?

What does ML tell us about the **truth**?

What is the **purpose** of ML?

ML models the dataset,
but does not find truth.

ML quality measurement raises existential questions

What is the **origin** of ML models?

Can we **trust** ML?

What does ML tell us about the **truth**?

What is the **purpose** of ML?

The purpose of ML
is to understand it.

The purpose of ML is to understand it.[1]

[1] At least if you need to trust all output

Two aspects of understanding ML are tracebacks and robustness certificates

Trace back decisions

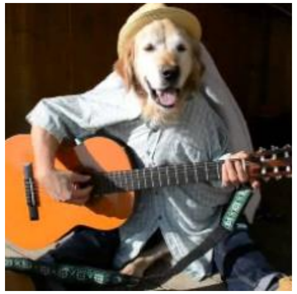
⇒ SP-LIME [1]

Certify robustness

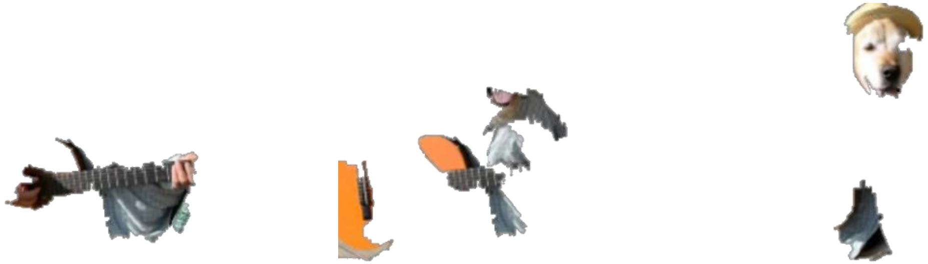
⇒ DeepPoly [2]

[1] Ribeiro et al, 2016
[2] Singh et al, 2019

SP-LIME [1] selects representative classification examples on a budget



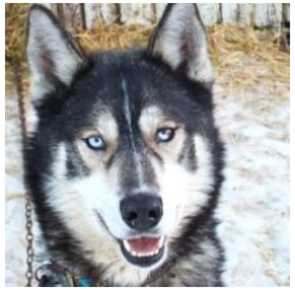
⇒



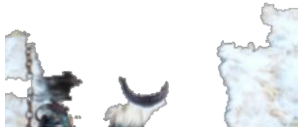
E-Guitar

Acoustic Guitar

Labrador



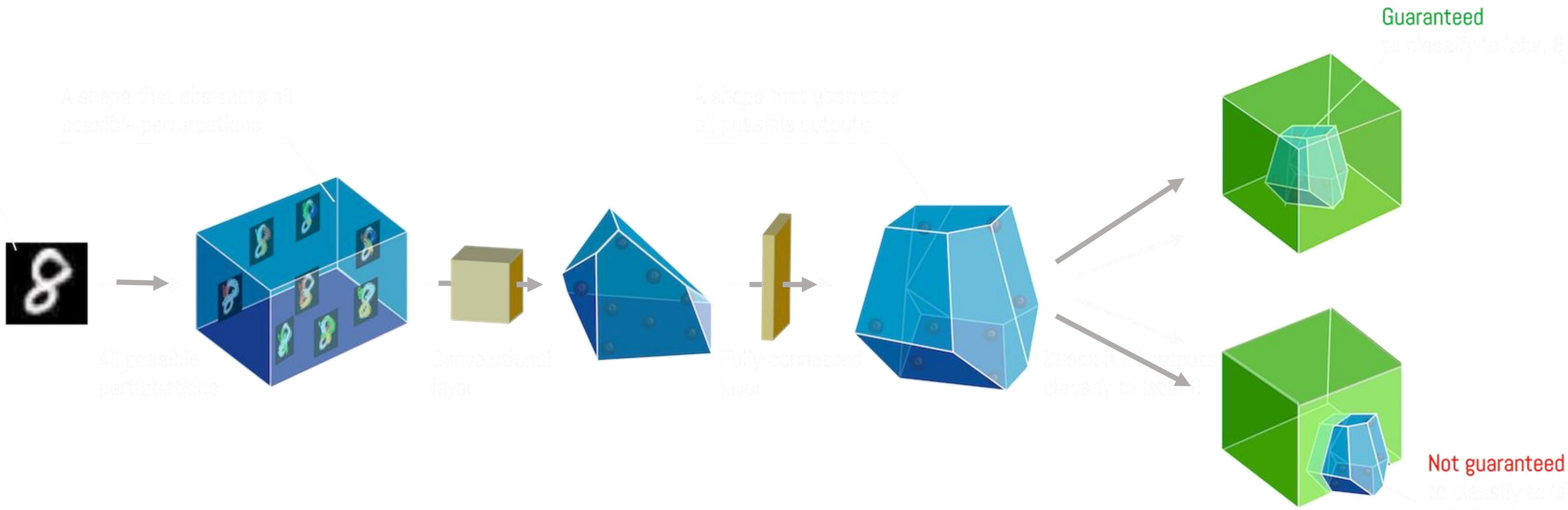
⇒



Wolf

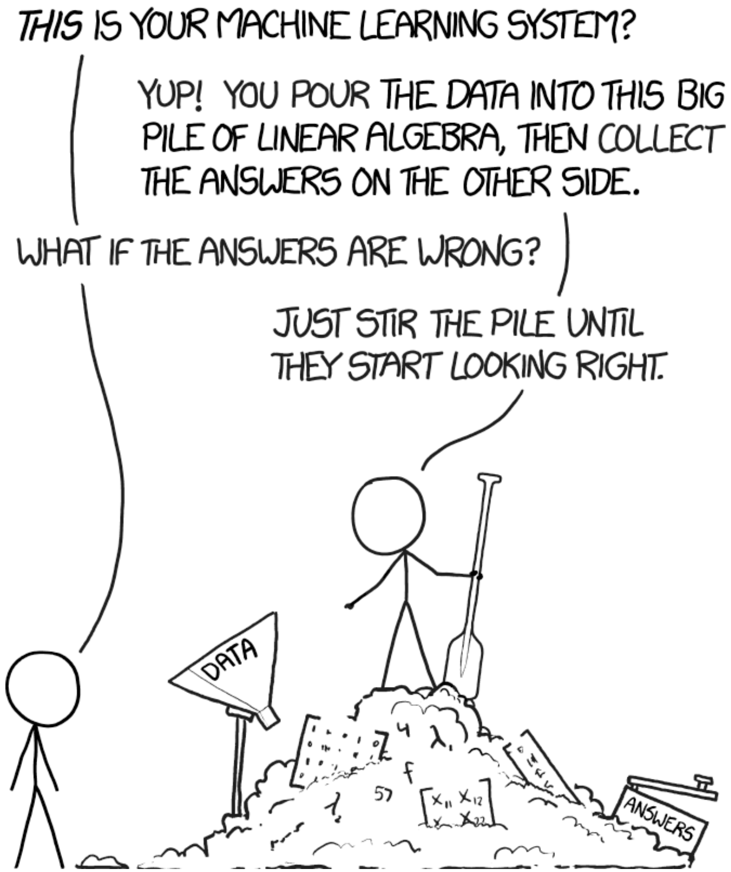
[1] Ribeiro et al, 2016

DeepPoly [2] transforms floating-point polyhedra to prove robustness under complex pertubations


A shape that abstracts all possible perturbations
All possible perturbations
Convolutional layer
A shape that abstracts all possible outputs
Fully-connected layer
Check if all outputs classify to label 8
Guaranteed
to classify to label 8
Not guaranteed
to classify to la

[2] Singh et al, 2019   safeai.ethz.ch

These existential questions are possible discussion topics


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

xkcd.com/1838/

What is the **origin** of ML models?

Can we **trust** ML?

What does ML tell us about the **truth**?

What is the **purpose** of ML?

What **directions** should research take?

Coralie Busse-Grawitz

bcoralie@ethz.ch

Sources

[1] Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "Why should i trust you?: Explaining the predictions of any classifier." *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 2016.

[2] Singh, Gagandeep, et al. "Boosting Robustness Certification of Neural Networks." (2018).