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.
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*

Complex question → Data → ML model → Complex answer

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

**Root of trust**

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

**Inferred trust**

F-score

Mutual Information

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

Root of trust

Complex question → Data → ML model → Complex answer

Representative?

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

**Root of trust**

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

Representative? → **Tautology**

**Inferred trust**
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]

[2] Singh et al, 2019
SP-LIME [1] selects representative classification examples on a budget


[2] Singh et al, 2019

safeai.ethz.ch
These existential questions are possible **discussion topics**

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

xkcd.com/1838/
Sources
