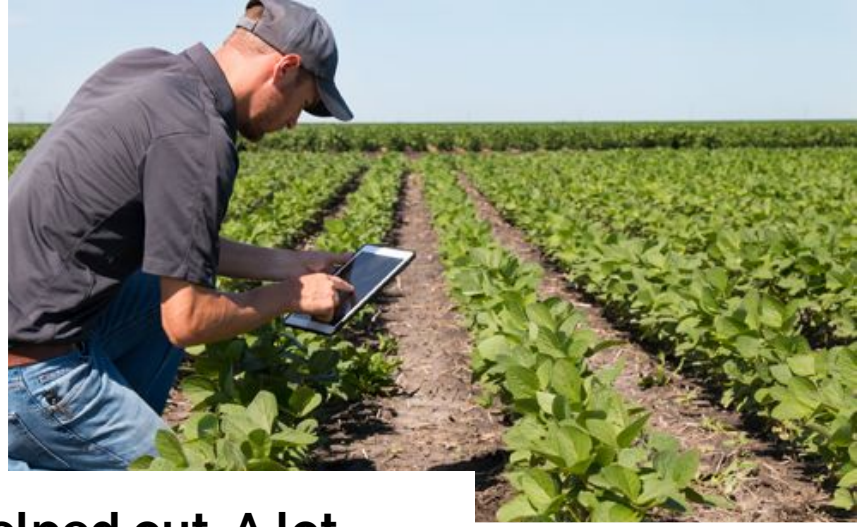
Machine Learning is gullible, insecure and inefficient

Nuno Moniz
PostDoc Fellow @ INESC TEC
Invited Professor @ FCUP

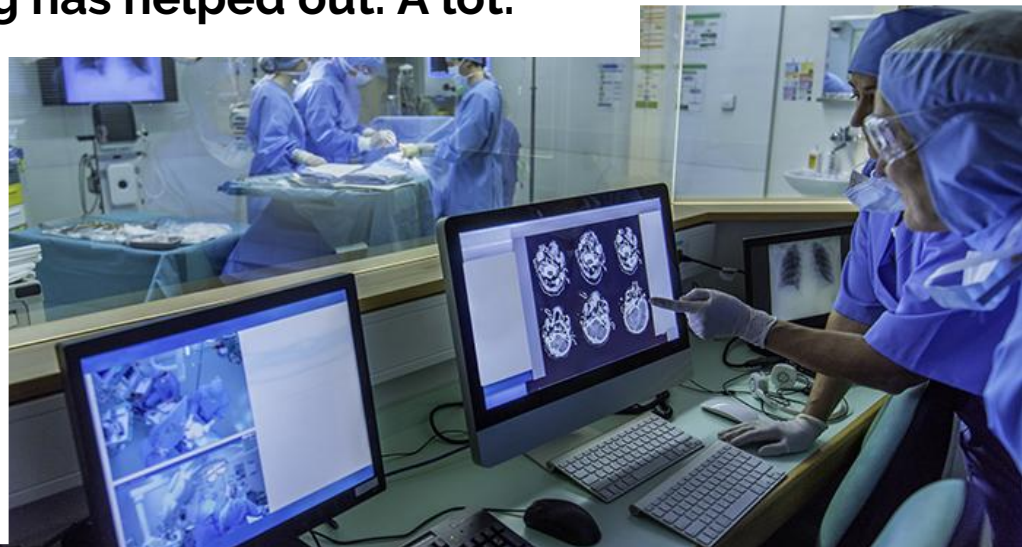
nmmoniz@inesctec.pt

10th December 2018

DSPT
DATA SCIENCE PORTUGAL



Machine Learning has helped out. A lot.



It has also brought a lot of surprises recently

AI “is more profound than… electricity or fire.”

Sundar Pichai @ Recode 19/1/18

“If a typical person can do a mental task with less than one second of thought, we can probably automate it using AI either now or in the near future.”

Andrew Ng @ HBR 9/11/16

But...

“People naively believe that if you take deep learning and scale it 100 times more layers, and add 1000 times more data, a neural net will be able to do anything a human being can do, but that’s just not true.”

François Chollet @ Wired 2/2/18

About this talk

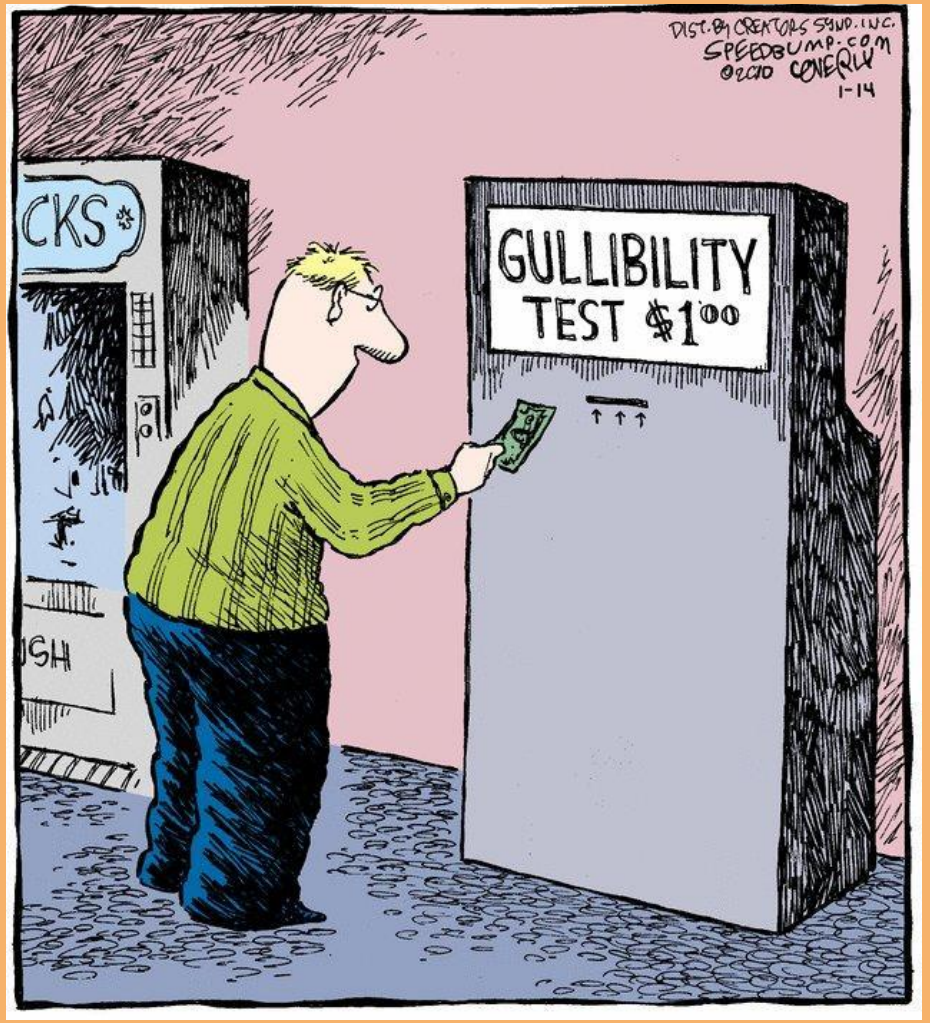
Challenges for Machine Learning

Gullibility

Failure of social intelligence in which a person is easily tricked or manipulated into an ill-advised course of action.

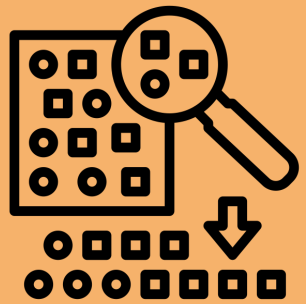
Credulity (closely related)

Tendency to believe unlikely propositions that are unsupported by evidence.



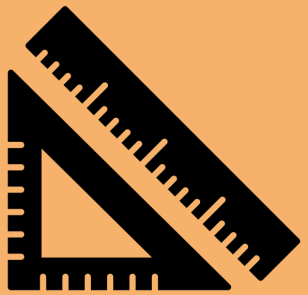
BIAS

[or ‘if the data/algorithm says so…’]

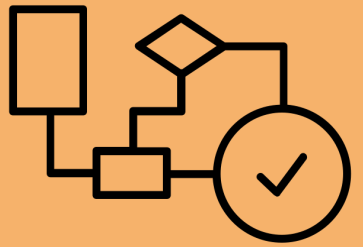


Sample Bias

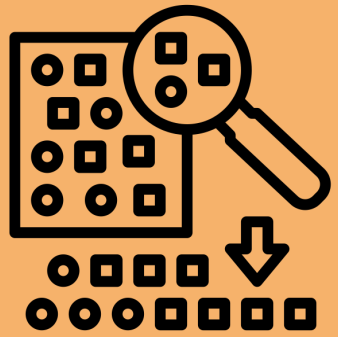
Prejudice Bias



Measurement Bias



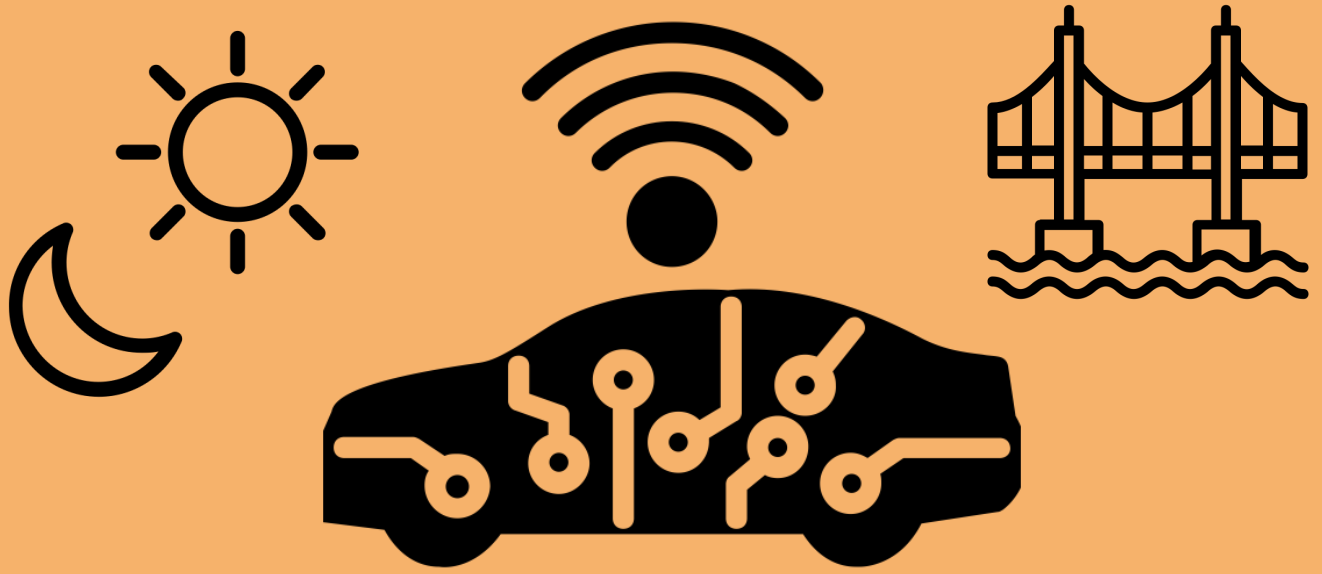
Algorithm Bias



Sample Bias

You will (almost) never train your model with all of the data ...
... therefore you may not accurately represent the entire domain

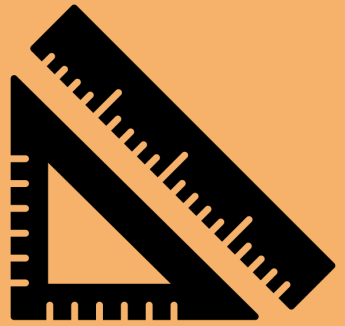
How to mitigate sample bias?



Prejudice Bias

Common when choosing samples or inserting data
Mathematics cannot overcome problems related to prejudice
You'll have to do it yourselves.

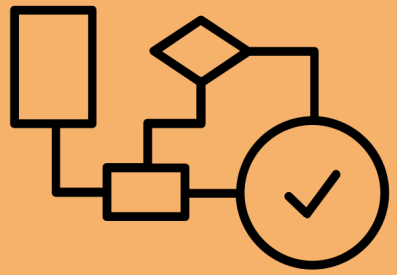
How to mitigate prejudice bias?



Measurement Bias

Relying on specific devices for observations or measurements may provoke a systematic distortion
This will induce constant bias to the data

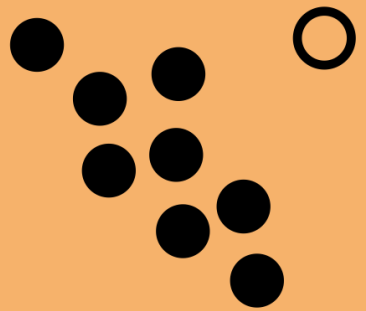
How to mitigate measurement bias?



Algorithm Bias

Algorithms are based on specific assumptions
Average-Behaviour Obsession
Bias-Variance Tradeoff

How to mitigate algorithm bias?



Still, there's a larger problem...

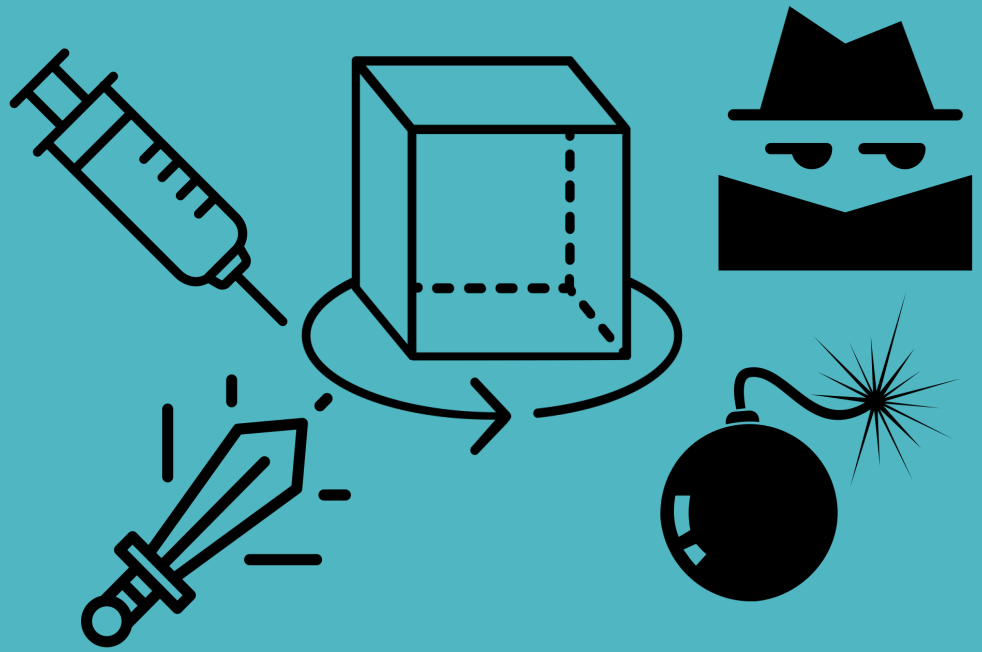
Models are little more than opinions embedded in mathematics

They codify the past, but they do not invent the future

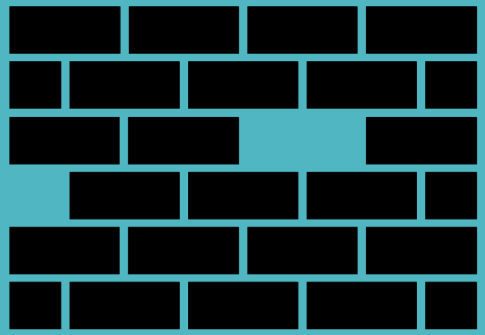
Regardless, they are prone to creating their own reality (data/algorithms)

Can models incorporate auto-reflection and auto-regulation?
Or will this always be dependent of us?

Are Your Models Safe?



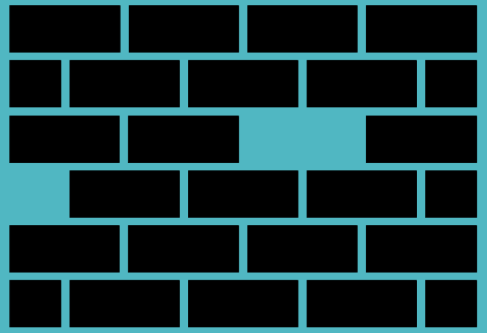
Secure Machine Learning



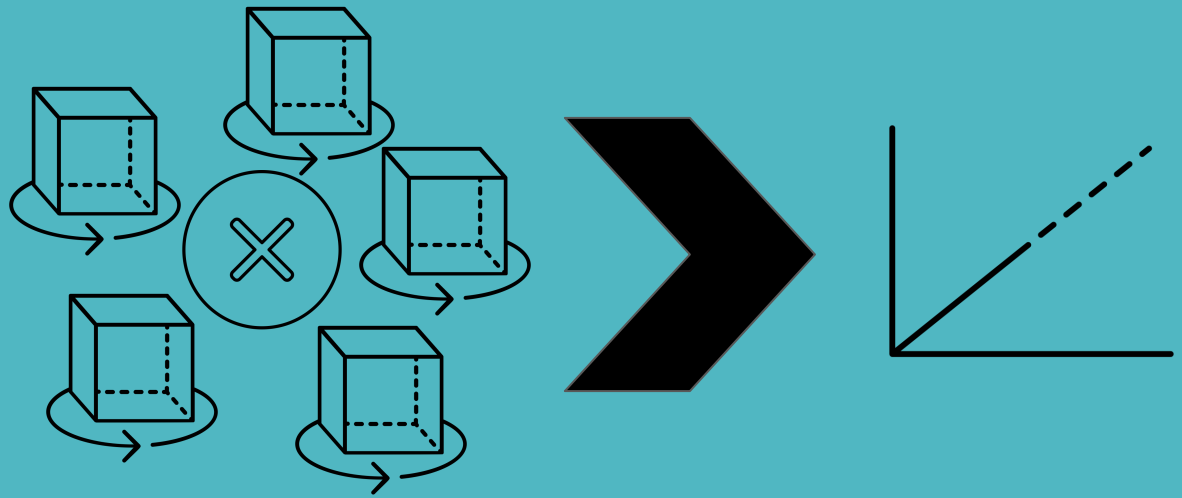
Integrity

Robustness

Privacy

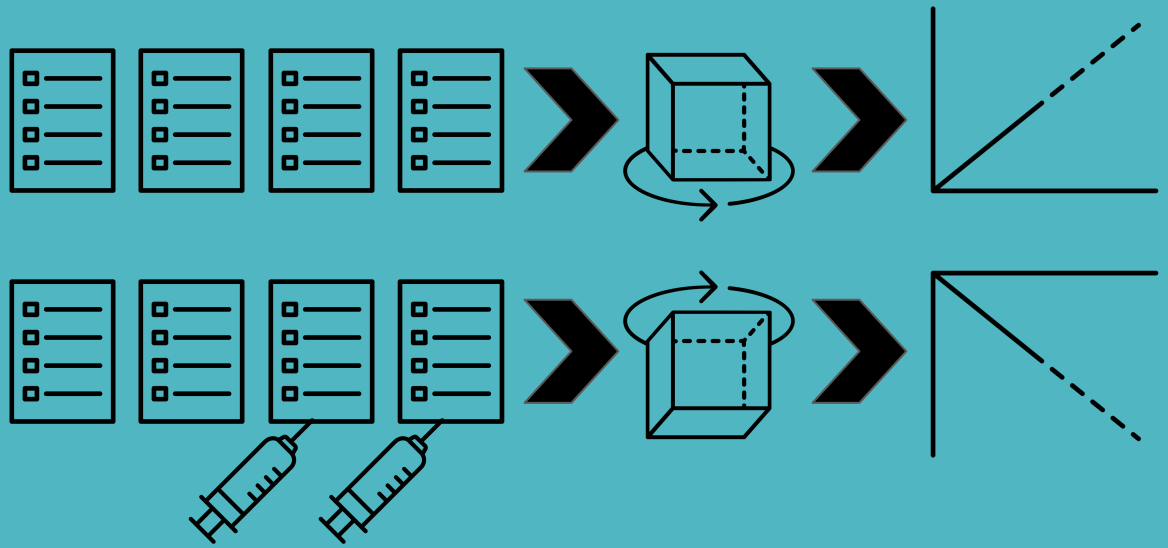


Integrity: Learning architectures capable of withstanding “Denial-of-Model” situations, maintaining expected performance



… messing with the mechanics

Robustness: Ability to sustain attempts of model biasing with data injection, using strategies to guarantee data coherence with domain

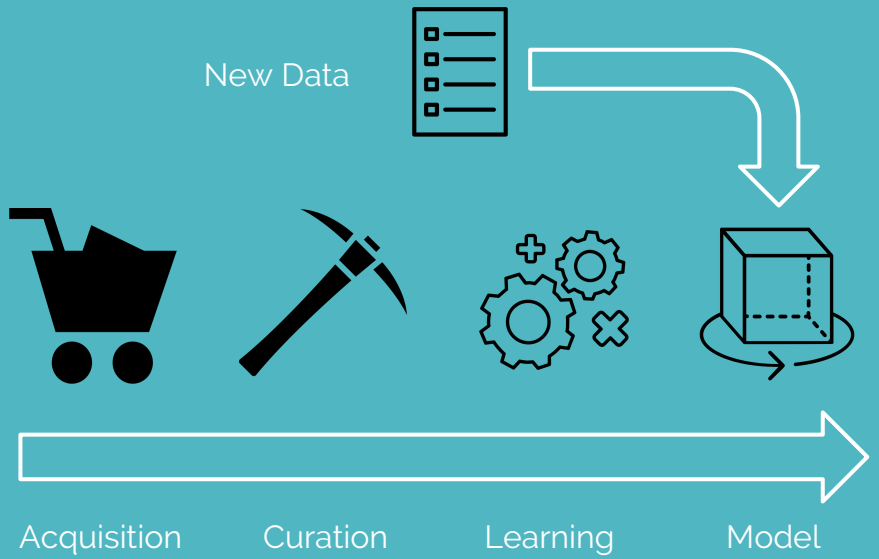


… messing with the fuel

Privacy: Ensure both data and model privacy, guaranteeing that users' data is kept anonymous and that no single model explains the entire domain

… messing with the ECU

Really. What Can Go Wrong?



Biased Acquisition of Data

Reverse Anonymization

Data Spoofing

Reverse Model Engineering

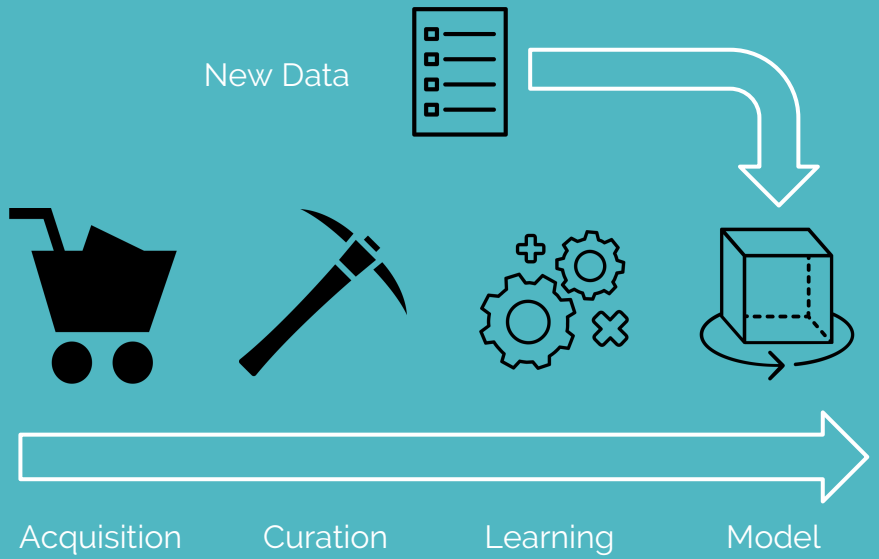
Business Understanding

Model Shutdown

...

Moniz, N., Portela, B., Vinagre, J. (2019).
A Survey on Secure Machine Learning [in preparation]

Ok. How Can It Be Avoided?

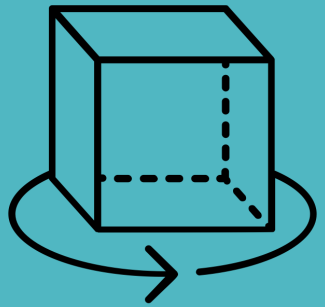


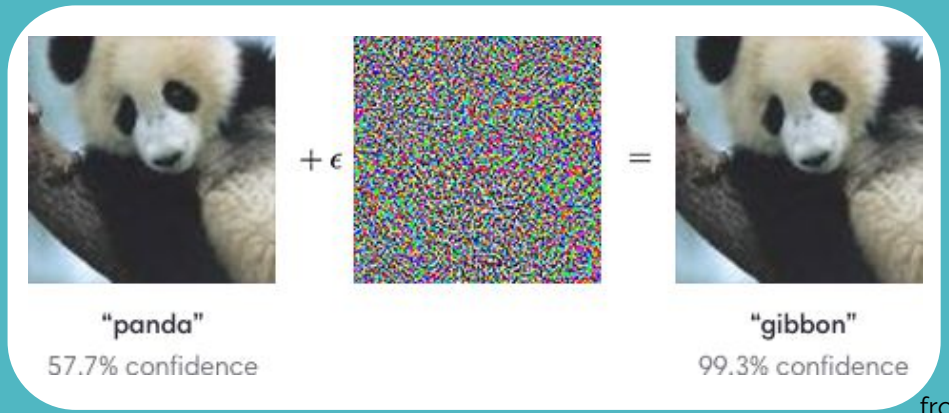
- Data Validation
- Differential Privacy
- Blockchain
- Adversarial Training
- Model Abstraction
- Neural Networks
- Distributed Learning
- ...

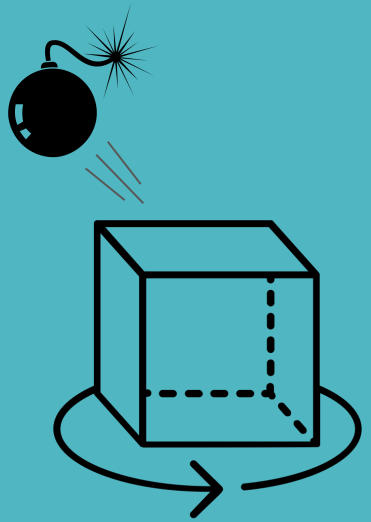
Moniz, N., Portela, B., Vinagre, J. (2019).
A Survey on Secure Machine Learning [in preparation]

Coming up next (?)

What happens when models start attacking other models for competitive gain?







from OpenAI

Let’s talk about efficiency

20 Watts

Sunway TaihuLight (3rd)

15.4 Megawatts
0,000013%

20 Terawatts
0,0000000001%

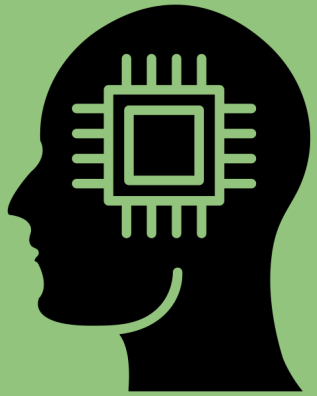
Let’s talk about efficiency

20 Watts

Sunway TaihuLight (3rd)

15.4 Megawatts
0,000013%

(a fraction)



20 Terawatts
0,0000000001%

AI's dirty secret: Energy-guzzling machines may fuel global warming

NewScientist

Advances in artificial intelligence could lead to massive growth in energy use as smart machines push into every corner of our lives

Christie Hemm Klok / New York Times / Redux / eyevine

How to estimate energy consumption?

We apologize, but a straightforward cross-platform approach to estimate energy consumption for different types of algorithms is not available at the moment.

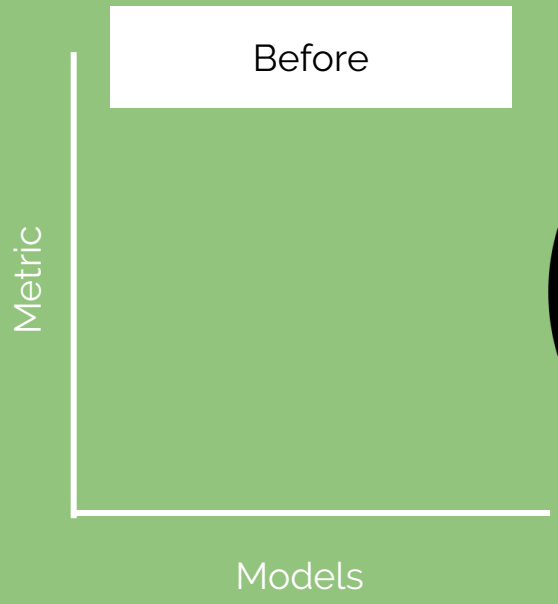
García Martín, E. et al. (2018). How to Measure Energy Consumption in Machine Learning Algorithms. In Green Data Mining Workshop (ECMLPKDD 2018). LNCS.

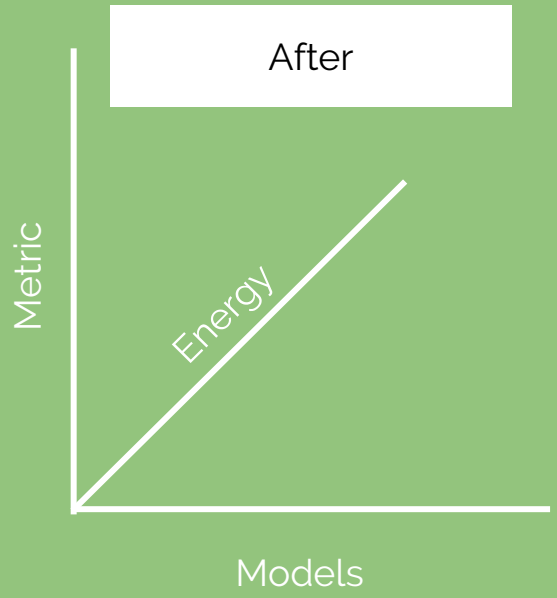
So: how to estimate energy consumption?

Type	Technique	Level
Analytical	Simulation	Architecture
Empirical	PMC	Instruction

García Martín, E. et al. (2018). How to Measure Energy Consumption in Machine Learning Algorithms. In Green Data Mining Workshop (ECMLPKDD 2018). LNCS.

Green Machine Learning?





What we need (for starters)

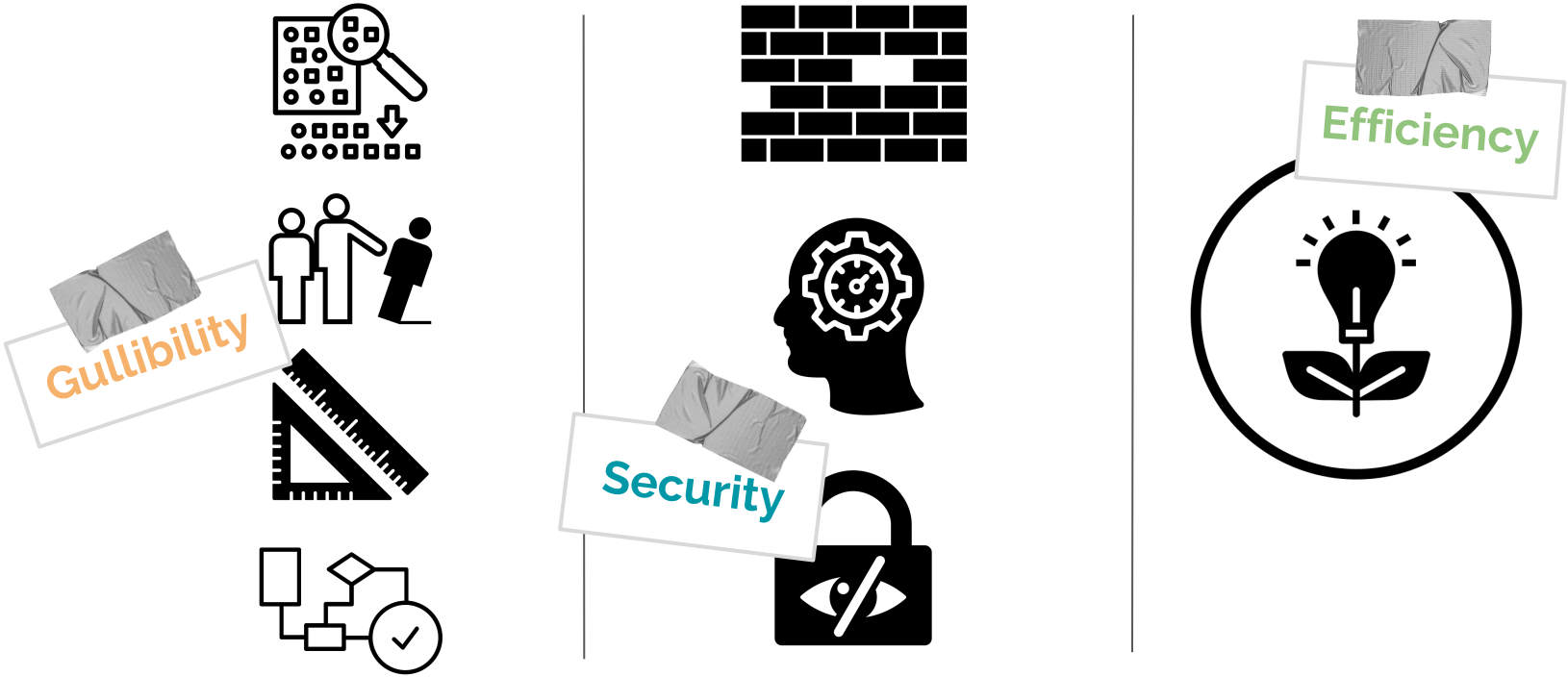
Standard cross-platform approach to measure consumption of energy in ML

Broad benchmark of data-domain-algorithm performance in terms of energy efficiency

Impact of parametrization in energy consumption

Establishing Performance / Energy-Efficiency tradeoff

[Recap]



Going Forward: Reliability / Explainability

How can we trust ML?

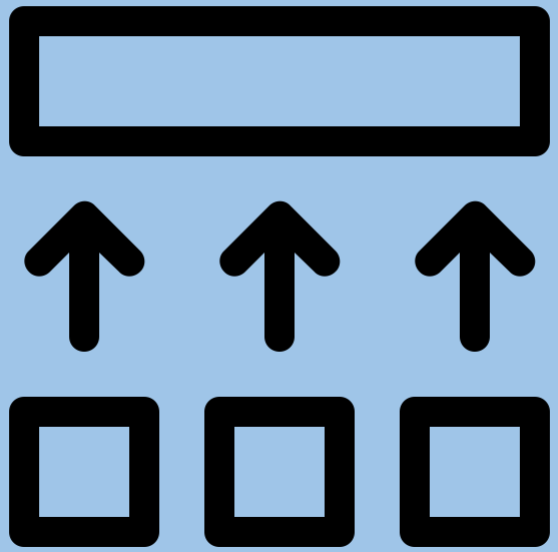
Where did that prediction come from?

Going Forward: Accountability

Where did it all go wrong?

Who's to blame?

Going Forward: Generalization

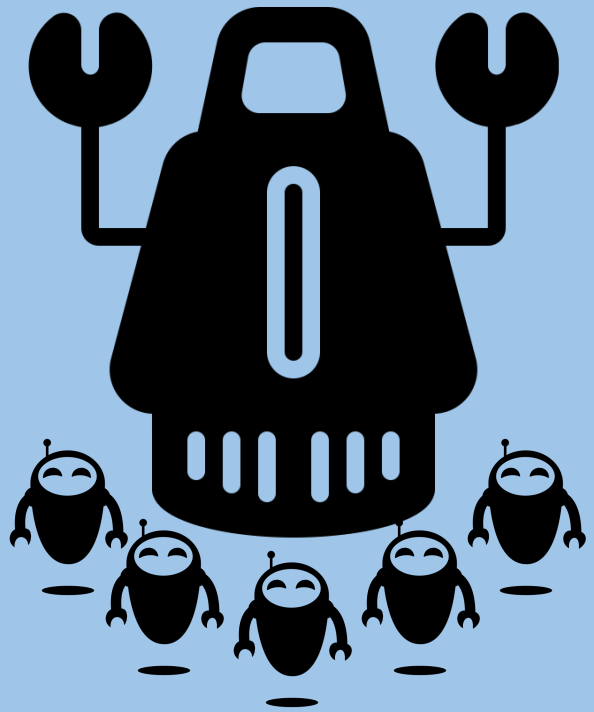


ML does not generalize concepts

ML is not capable of abstract reasoning

(yet?)

Going Forward: AutoML



Learn to learn

Nevermind DL for a minute: which tools are better for this data

Your solutionism level?

- 100B neurons per human (only 15% are active)
- More connections in the human body than stars in the galaxy

AI and ML have and will do even more great things...
But we can do better.

Intelligent Machines

Microsoft's neo-Nazi sexbot was a great lesson for makers of AI assistants

Outrage over AI that 'identifies gay faces' as Google experts say the machine relies on patterns in how people take selfies and NOT on their facial features

Thank you!

(and some credits)

For the artwork

The Noun Project: Creative Stall, Arthur Shlain, Ester Barbato, Rutmer Zijlstra, Creaticca Creative Agency, Luis Prado, Wahyu Unggul Sejati, Sergey Novosyolov, Vectors Markets, Ruslan Dezign, ProSymbols, karremovic, Marek Polakovic, faisalovers, Rose Alice Design, Paisley, AlfredoCreates.com/icons & Flaticondesign.com, Ilaria Bernareggi, Nerea Martínez Orduña, JohnnyZi, Saeful Muslim, Oksana Latysheva, Chameleon Design, Atif Arshad, Noura Mbarki, Samy Menai, sachin modgekar, Scott Lewis, Ben Davis, Rigo Peter, H Alberto Gongora, Bakunetsu Kaito, Becris, Dan Hetteix, Jonathan Gibson.

For their inspiring work/presentations and/or our conversations

Ricardo Baeza-Yates, Paula Branco, Pavel Brazdil, Vítor Cerqueira, João Gama, Eva García Martinez, Alípio M. Jorge, Daniel Loureiro, Margaret Mitchell, Hélder Oliveira, Mariana Oliveira, Arian Pasquali, Bernardo Portela, Mario Rasetti, Rita P. Ribeiro, Carlos Soares, Luís Torgo, João Vinagre, ...

Machine Learning is gullible, insecure and inefficient

Nuno Moniz
PostDoc Fellow @ INESC TEC
Invited Professor @ FCUP

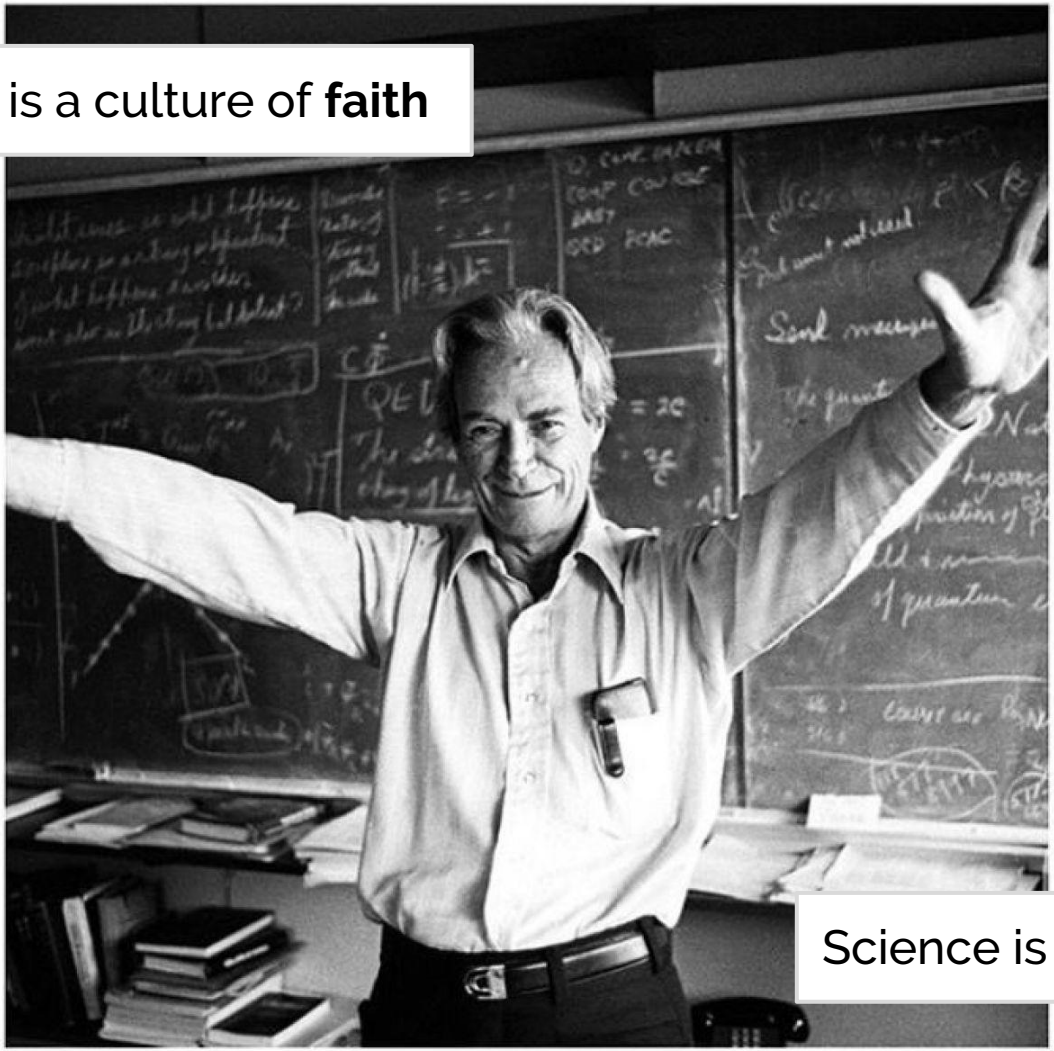
nmmoniz@inesctec.pt

10th December 2018

DSPT
DATA SCIENCE PORTUGAL

Preface

Religion is a culture of **faith**



Science is a culture of **doubt**

Richard Feynman