Data Science Meets Social Impact

Wirtschaftsagentur Wien, 4. JUNE 2019

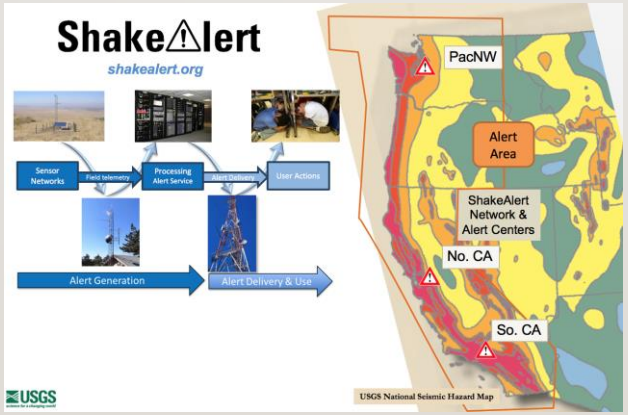
Scott Fertig (with more than a little cribbing from R. Wazir)

Grad school, medical informatics, AI, HPC, family, startups, corporate CIO, move to Vienna, public sector

a life…

uugot.it

VDSG
VIENNA DATA SCIENCE GROUP



isti Instrumental Software Technologies, Inc.

VDSG

data4good - data science for positive social impact

Rapid advances in Artificial Intelligence (AI) and data processing technologies are having a deep impact on society.

We strive for Data Science techniques that are

- Accessible
- Explainable
- Reproducible
- Fair

What's AI?

What's Data Science?

AI:
- The capability of a machine to imitate intelligent human behavior. (MW dictionary)
- A computer system that can perform tasks previously requiring human intelligence.

Data Science:
- Need to define Machine Learning first…

Machine Learning

Instead of programming every step, a machine is programmed to find patterns, and "learns" based on examples. There is a large area of overlap with the field of Statistics.

- Many "flavors", but all are doing some form of pattern recognition which can be applied to new information
- Flexible: if the data changes, the program adapts.
- Can help solve problems that we don't have an easy, ready-made solution for: Fraud detection, earthquake prediction.

Data Science

- Like all technology, can be used for good, merely to sell more stuff, or the not-so-good…

Using methods from Machine Learning and Statistics to solve problems and make decisions based on data

Good Data, Better Governance:

Be a Data Scientist

It is WWII, and you are a statistical consultant to the US military.

They have the following data, collected from American planes returned from engagements over Europe, and they want you to figure out for them which parts of the airplane need more armoring.

What is your advice to them?

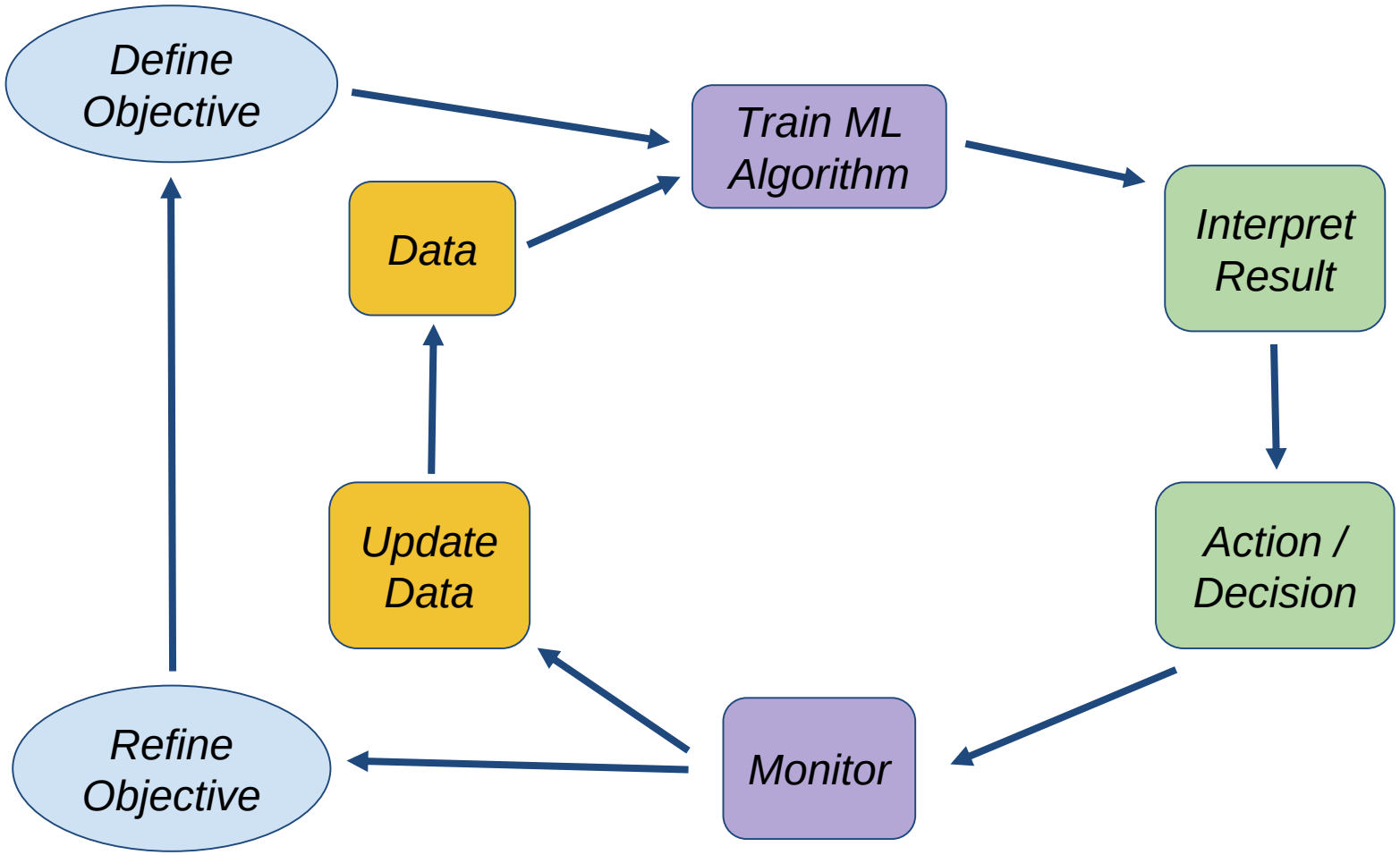
Protect the engines!

Section of plane	**Bullet holes per square foot**
Engine	1.11
Fuselage	1.73
Fuel system	1.55
Rest of the plane	1.8

Taken from: "How Not To Be Wrong" by Jordan Ellenberg, p.6. The Penguin Press 2014

Good Data, Better Governance:

Social Inclusion with Data Science



Good Data, Better Governance:

Social Inclusion through Data Science

Analyzing or Predicting Demand	Identify potential recipients/beneficiaries. Help in reaching target group.
Optimizing Resource Allocation / Predicting Outcome	Estimate the results of your service for each case, so you can prioritize by impact or urgency.
Analyzing Impact	Does your program lead to the desired results? To what extent?
Early Warning Systems	Find out why/when recipients or volunteers are likely to quit a program
Fraud Detection/Abuse	Detecting misuse of funds and/or services
Enriching Data	Enrich existing data by combining with other sources, and using AI to tag it

Analyzing or Predicting Demand: **DataKind** ***Poverty Detection Based on Satellite Imagery***	Objectives: Automate a data-driven process for targeting villages to receive cash transfers. Use satellite imagery and machine learning to identify, on a village-by-village level, the proportion of thatch and metal roofed homes, with thatch as a proxy for extreme poverty. http://www.datakind.org/projects/using-the-simple-to-be-radical/

Good Data, Better Governance:

Garbage In, Garbage Out

Algorithmic decision-making is widespread, and extending its field of operations: from criminal justice, credit, and insurance, to medical diagnosis and treatment, human resource management, and social work.

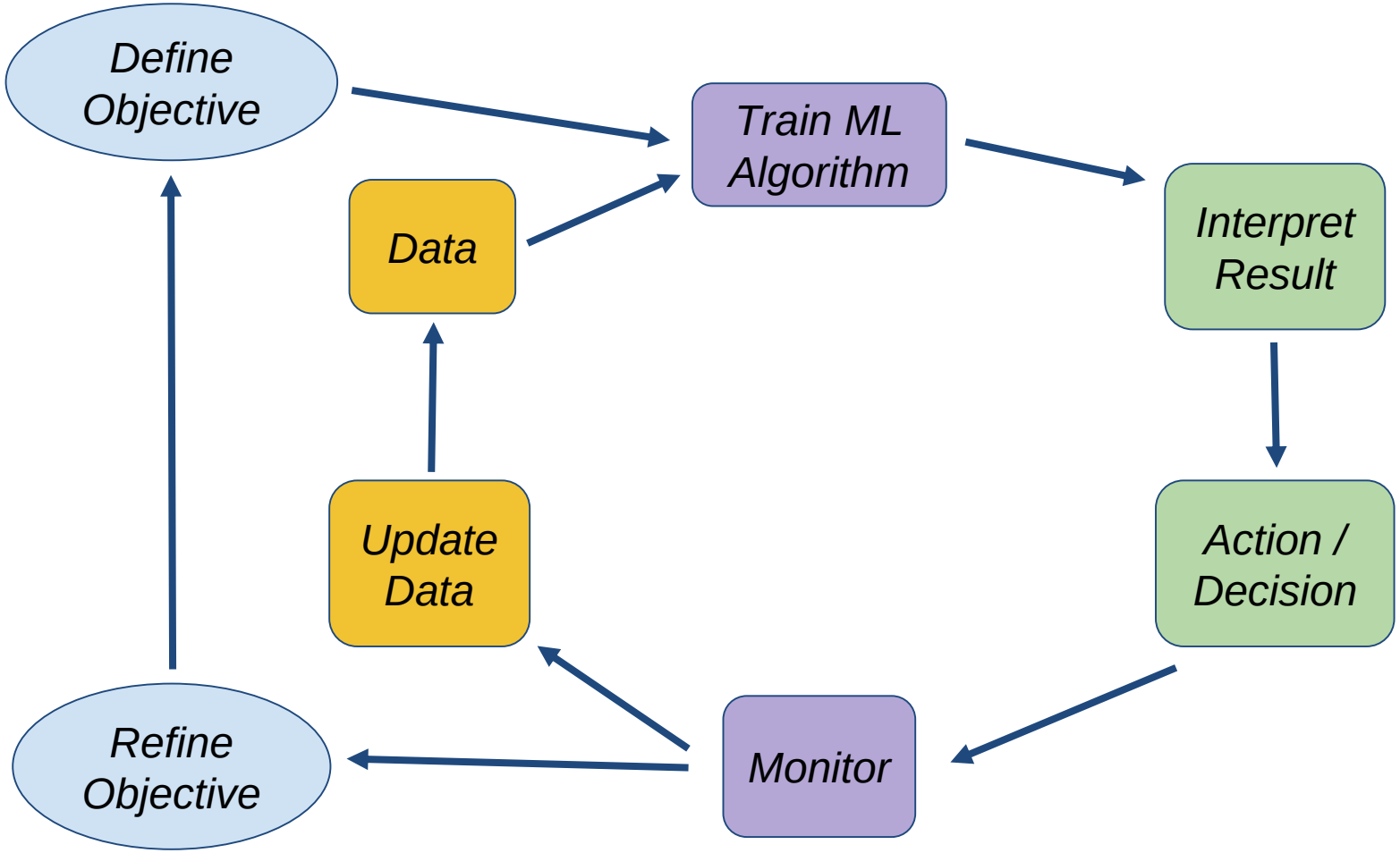
But it's not all milk and honey. Many are raising voices of concern:

- S. U. Noble. *Algorithms of Oppression: How search engines reinforce racism.* NYU Press, 2018.
- C. O'Neil. *Weapons of math destruction: How big data increases inequality and threatens democracy.* Broadway Books, 2016.
- J. Buolamwini. *How I'm fighting bias in algorithms.* TED Talk

How can the algorithms go so wrong?

Good Data, Better Governance:

Garbage In, Garbage Out



Good Data, Better Governance:

Garbage In, Garbage Out

Where the data can go wrong:

- *Insufficient quantity of data*
- *Poor-quality data*

  Missing values, outliers, errors
- *Irrelevant data*

  Too many factors, a lot of them not useful, can lead the algorithm astray
- *Nonrepresentative data*

  Polls (eg BREXIT)

  Medical tests (Clinical trials Gender Gap, The Guardian)
- *Biased data*

  Gender bias in natural language tasks (Man is to Computer Programmer as Woman is to Homemaker, NIPS)

  Recidivism risk algorithms (AI is sending people to jail – and getting it wrong)

AI is sending people to jail—and getting it wrong
Using historical data to train risk assessment tools could mean that machines are copying the mistakes of the past.

“We are not risks,” she said. “We are needs.”

Many AI Issues Boil Down To Ethics

- Machines may be capable of complex calculation and "learn" behavior, but they cannot make qualitative or moral judgments.
- Ultimately there must always be human accountability for the decisions that machines originate.
- The use and manipulation of a massive amount of data creates an information asymmetry. This confers power on those who control it at the potential expense of those who are the subject of it.
- Thank you!