AN INITIATIVE OF
THE NETHERLANDS
RED CROSS

RECENT NEWS ON ARTIFICIAL INTELLIGENCE

510 AN INITIATIVE OF THE NETHERLANDS RED CROSS

One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority

In a major ethical leap for the tech world, Chinese start-ups have built algorithms that the government uses to track members of a largely Muslim minority group.

By Paul Mozur

April 14, 2019

FEBRUARY 17, 2020

Fear of Big Brother guides EU rules on AI

by Alex Pigman

A beginner's guide to the AI apocalypse: Wall-E syndrome

by TRISTAN GREENE — 29 days ago in ARTIFICIAL INTELLIGENCE

Governments Use AI Facial Recognition Tech: Is It Legal?

Posted by Justin Shamlou | Feb 19, 2020

Why algorithms can be racist and sexist

A computer can make a decision faster. That doesn't make it fair.

By Rebecca Heilweil | Feb 18, 2020, 12:20pm EST

RECENT NEWS ON ARTIFICIAL INTELLIGENCE

AN INITIATIVE OF THE NETHERLANDS RED CROSS

EU proposes rules for artificial intelligence to limit risks

By KELVIN CHAN February 19, 2020

‘Really scary what I saw in Hong Kong’: EU to urge caution on use of facial recognition amid fears of Big Brother-style society

EU seeks ‘responsible’ AI to dispel Big Brother fears

Published 1 day ago on February 20, 2020
By **Agence France-Presse**

RECENT NEWS ON ARTIFICIAL INTELLIGENCE

Google AI tool will no longer use gendered labels like 'woman' or 'man' in photos of people

Google's Cloud Vision API will tag images as 'person' to thwart bias

By Kim Lyons | Feb 20, 2020, 12:43pm EST

Only a better understanding of AI can help end our fear of robots

February 20, 2020

By **Vignesh Radhakrishnan**

TODAY'S TOPIC

RESPONSIBLE ARTIFICIAL INTELLIGENCE

EXISTING FRAMEWORKS / INITIATIVES

510 AN INITIATIVE OF THE NETHERLANDS RED CROSS

Microsoft — **FATE: Fairness, Accountability, Transparency, and Ethics in AI**

The Alan Turing Institute — **Fairness, transparency, privacy**

European Parliament — **A governance framework for algorithmic accountability and transparency**

DeNederlandscheBank EUROSYSTEEM — **General principles for the use of Artificial Intelligence in the financial sector**

Université de Montréal — **Montréal Declaration for Responsible Development of Artificial Intelligence**

facebook — **Integrity**

Google AI — **Responsible AI Practices**

EXISTING FRAMEWORKS / INITIATIVES

510 AN INITIATIVE OF THE NETHERLANDS RED CROSS

Microsoft	FATE: Fairness, Accountability, Transparency, and Ethics in AI	ACADEMIC
The Alan Turing Institute	Fairness, transparency, privacy	ACADEMIC
European Parliament	A governance framework for algorithmic accountability and transparency	THEORETICAL
DeNederlandscheBank EUROSYSTEEM	General principles for the use of Artificial Intelligence in the financial sector	THEORETICAL
Université de Montréal	Montréal Declaration for Responsible Development of Artificial Intelligence	PLEDGE
facebook	Integrity	ACADEMIC
Google AI	Responsible AI Practices	TOOLS

SUPPLY vs DEMAND

AN INITIATIVE OF THE NETHERLANDS RED CROSS

What is available ?

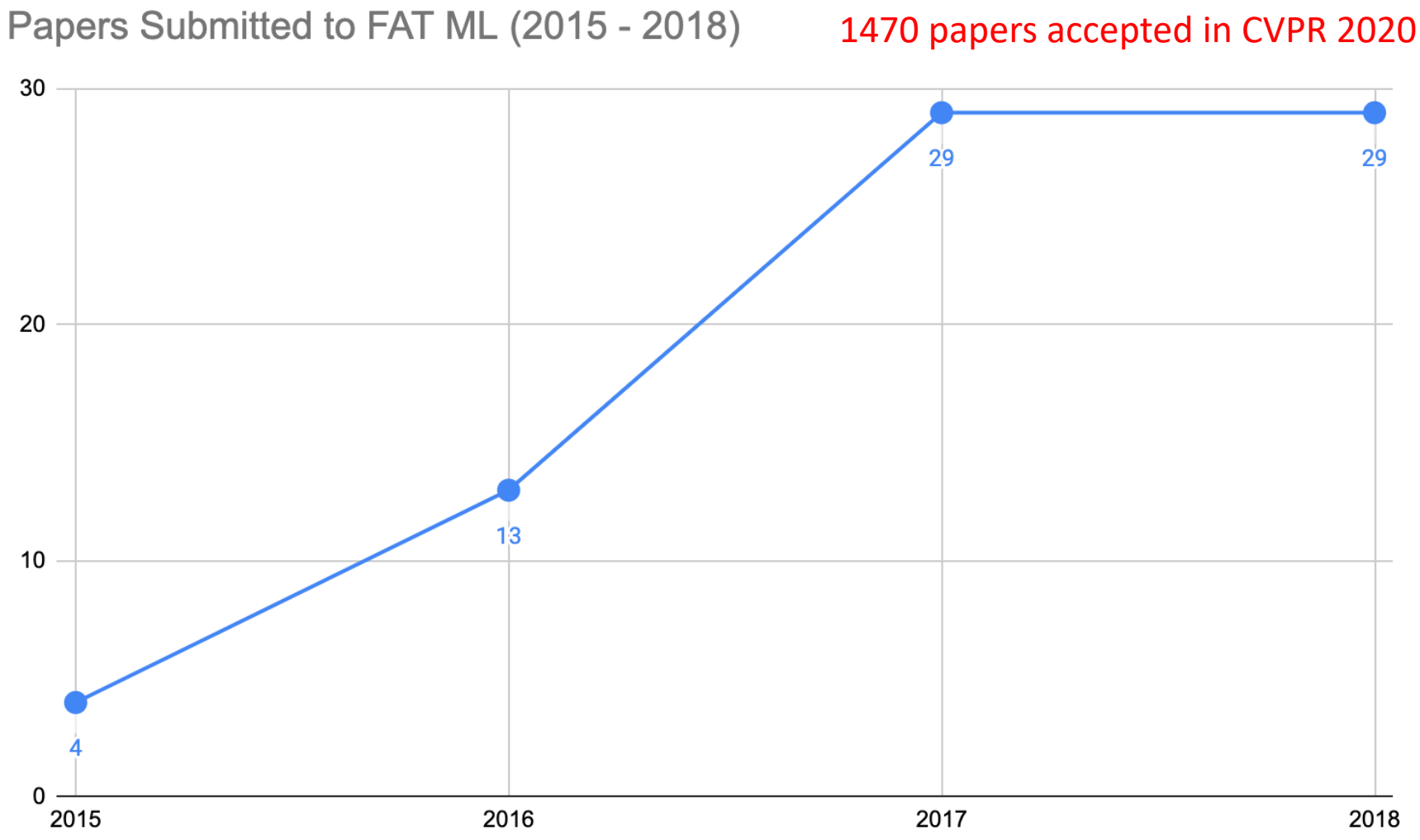
- **ACADEMIC**
- **THEORETICAL**
- **PLEDGE**
- **TOOLS**

What is needed ?

- **PRACTICAL**

ACADEMIC ACTIVITY

AN INITIATIVE OF THE NETHERLANDS RED CROSS



1470 papers accepted in CVPR 2020

UNIVERSITY OF AMSTERDAM

Mandatory course since 2019:

Fairness, Accountability, Confidentiality and Transparency in AI

by

Dr. Maarten de Rijke

from the

Information and Language Processing Systems (ILPS-UvA)

More FAT papers accepted in ACM FAccT 2018 - 2020

DEFINE RESPONSIBLITY

Blame the…

1. Scientist
2. Developer
3. CXO
4. Investors
5. Publisher

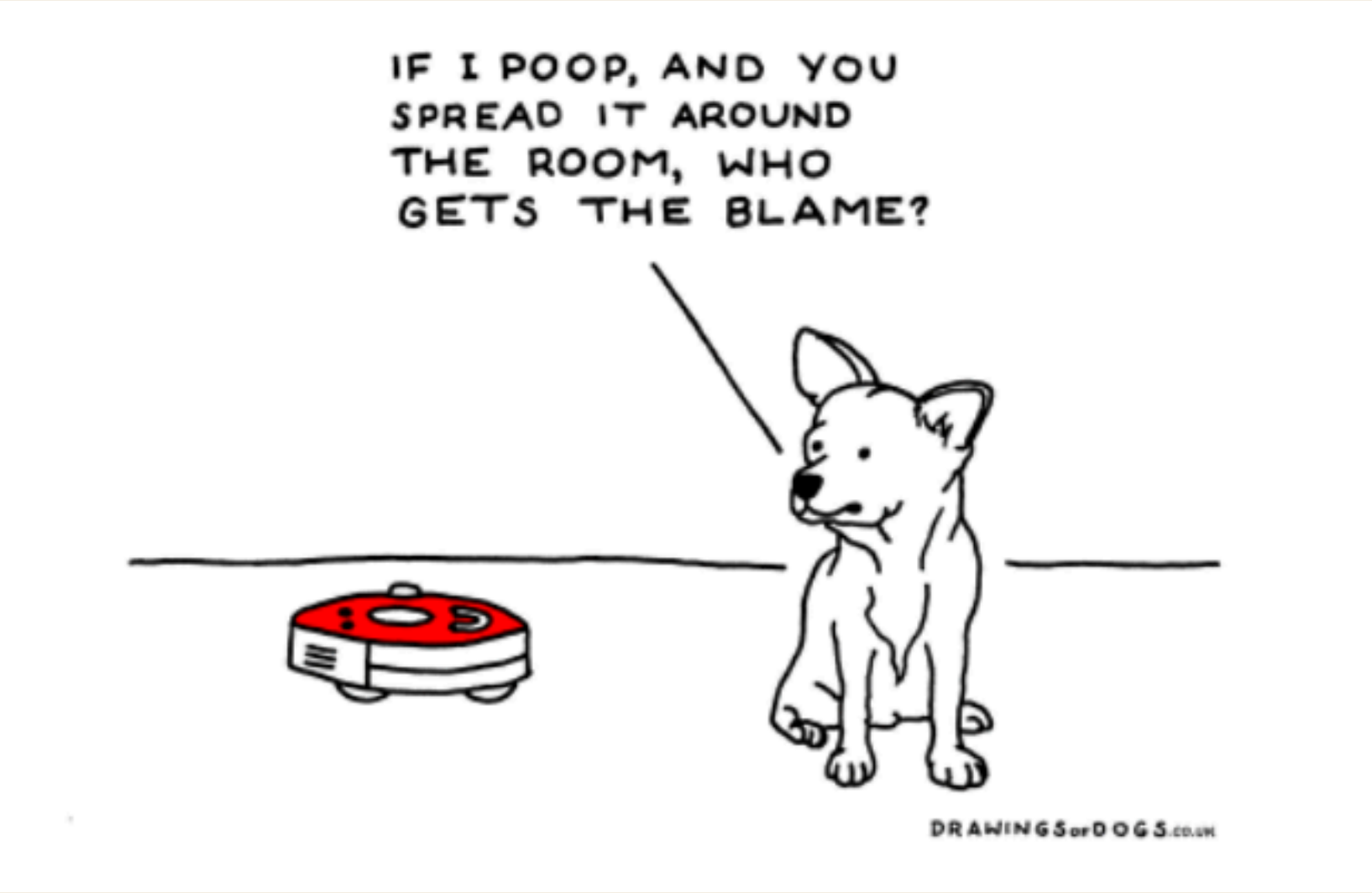
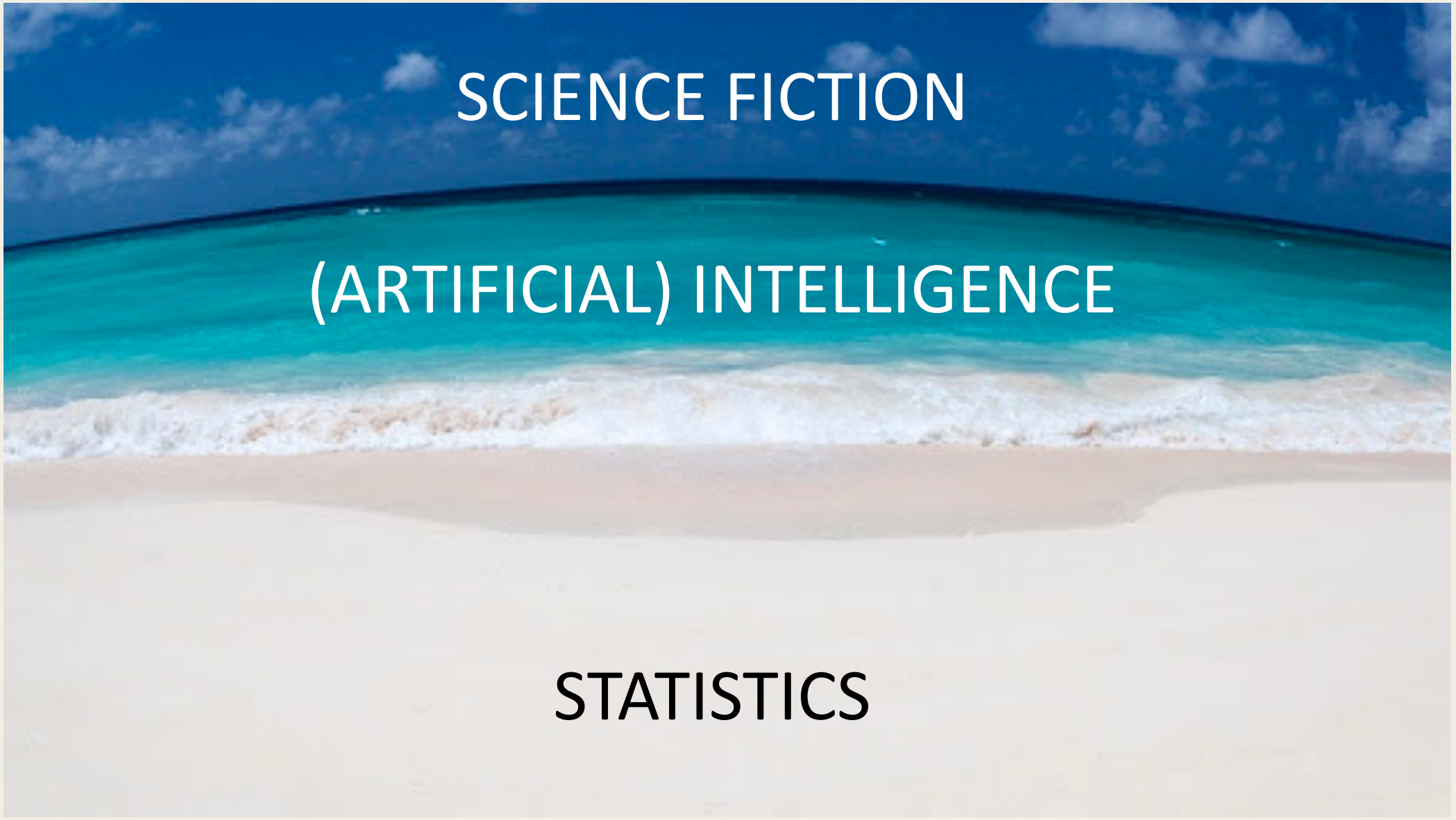


Image Source: https://towardsdatascience.com/ethics-of-ai-a-comprehensive-primer-1bfd039124b0

DEFINE ARTIFICIAL INTELLIGENCE



WHERE TO START?

AN INITIATIVE OF THE NETHERLANDS RED CROSS

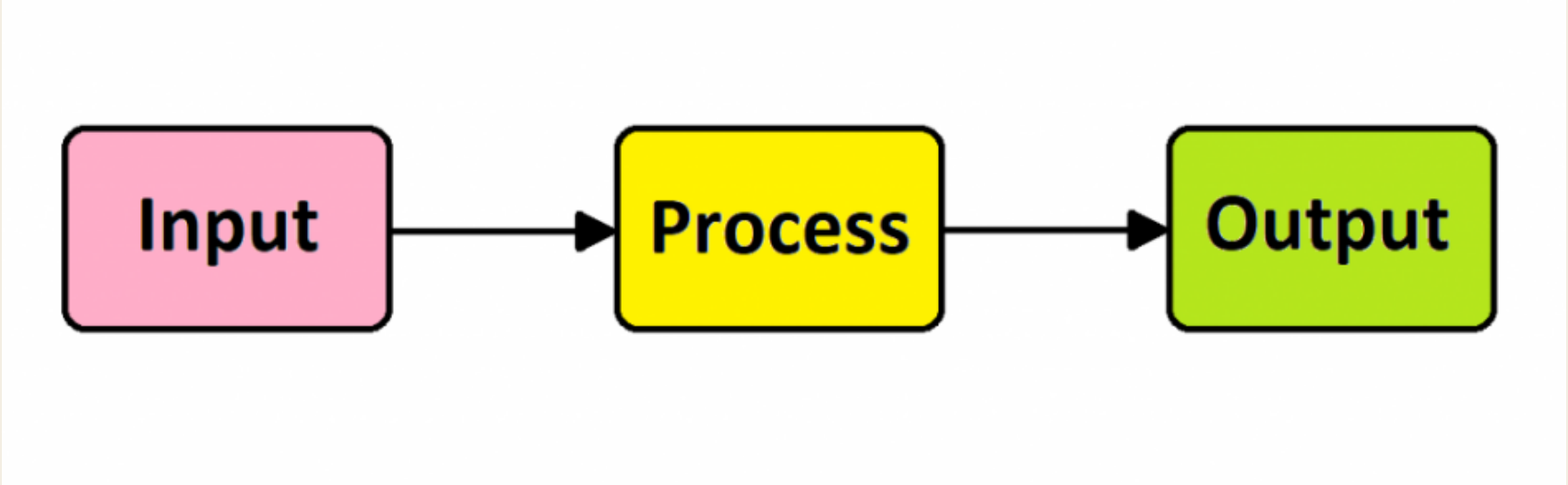
Distil from existing practices

Pros,	Cons,
Tangible starting point	Non-exhaustive reach
Enactable by practitioners	Ever-changing methods
Bottom-up approach	No defined metrics

EVALUATION

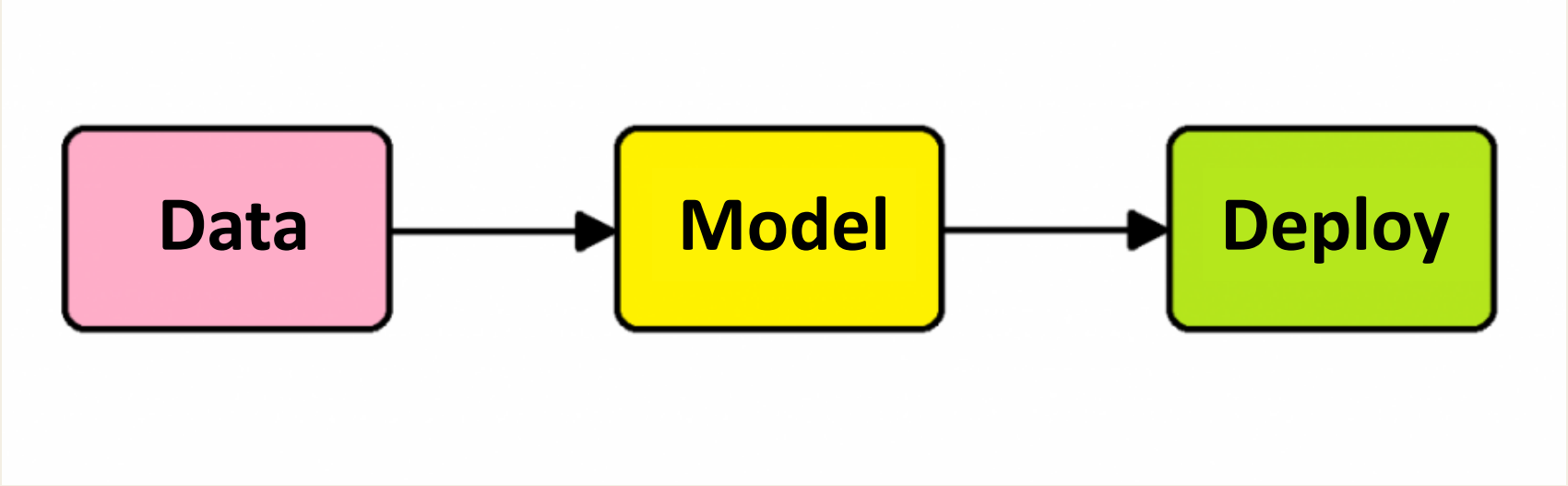
SETTLE FOR AN APPROXIMATION

FUNDAMENTAL COMPONENTS



SETTLE FOR AN APPROXIMATION

FUNDAMENTAL COMPONENTS in A.I.



DATA

510 AN INITIATIVE OF THE NETHERLANDS RED CROSS

Types of data,

1. Numerical
2. Categorical
3. Images
4. Text
5. Location
6. Activity
7. Measurements
8. Video

Sources of data,

1. Datasets
2. Logs
3. Reports
4. Forms
5. Recordings
6. Photo Gallery
7. Consumption History
8. Declarations

Grouped by,

1. Person
2. Organization
3. Region
4. Purpose
5. Legal

CONSTRAINTS

1. Be generalizable to cover all projects.

2. Not be time intensive to apply. (15-30 mins)

3. Defined metrics which capture our goal.

DATA

14 Yes/No questions regarding data,

- Is the source of data/labels known?
- Is the dataset publicly available?
- Is the dataset published in a conference or journal?
- Is the data anonymized?
- Are infographics/graphs used to convey the analyses on the dataset?

MODEL

17 Yes/No questions regarding model,

- Is a baseline model available for reference?
- Have traditional models been evaluated for reference?
- Is the model based on published research?
- Is the trained model publicly shared?
- Is the model published in a conference or journal?

DEPLOY

13 Yes/No questions regarding deployment,

- Is there a human in the decision making process?
- Are the decisions made by the model analysed at regular intervals?
- Is the confidence score considered in the decision of the model?
- Is the prediction confidence shared in a decision report?
- Is the inference data stored for future review, analyses or modelling?

METRICS

AN INITIATIVE OF THE NETHERLANDS RED CROSS

Fairness	Decisions are **unbiased** to identity features such as gender, race, nationality, sexual orientation, religion, political opinion, skin colour, education, address and age.
Accountability	Decisions should be traceable, reproducible and **liable**. All decisions and their consequences must be owned and explained by the decision maker or administrator of the decision maker.
Confidentiality	Personal information which can be used to identify individuals or communities must be **secure** and above intentional or accidental misuse.
Transparency	Understanding why a decision is made is key to establishing **trust** in the decision process. Accompanying decisions with explanations and analysis reports help understand the outcome.

APPLIED

3 Projects (2 within 510)

Project Name	FACT Score
ADA - Caladrius	**41.98%**
MCF Hackathon	**32.10%**
Philippines Typhoon	**65.43%**

Less than an hour to apply.

A hackathon project gets a lower **score** which makes sense. It was **easy to use** by the evaluators without any instructions.

The flood project evaluation was *mostly* unguided.

UNTIL YOU HELP US…

Image Source: https://www.brusselstimes.com/all-news/magazine/95950/challenges-opportunities-and-ethical-dilemmas-how-artificial-intelligence-could-transform-the-european-economic-and-political-landscape/

HOW CAN YOU HELP US?

AN INITIATIVE OF THE NETHERLANDS RED CROSS

Project Document – Word file on Teams

Evaluation Form – Excel file on Teams

Scoring Optimization – Word or Excel

Evaluate your project – Copy the evaluation form

FEEDBACK and SUGGESTIONS are welcome ☺

RESPONSIBLE A.I. TEAM

510 AN INITIATIVE OF THE NETHERLANDS RED CROSS

1. KAMAL AHMED
2. JOACHIM RAMAKERS
3. GULFARAZ RAHMAN
4. …YOU!

UNTIL YOU HELP US...

Image Source: https://xkcd.com/1838/

THANK YOU