A.I. MEETS ETHICS

April 24th 2019

GIST

Welcome by GIST

Debbie Esmans

Tanguy Coenen

Jan Adriaenssens

Director City of Things - IMEC

Rob Heyman

Senior Researcher – VUB-SMIT/IMEC

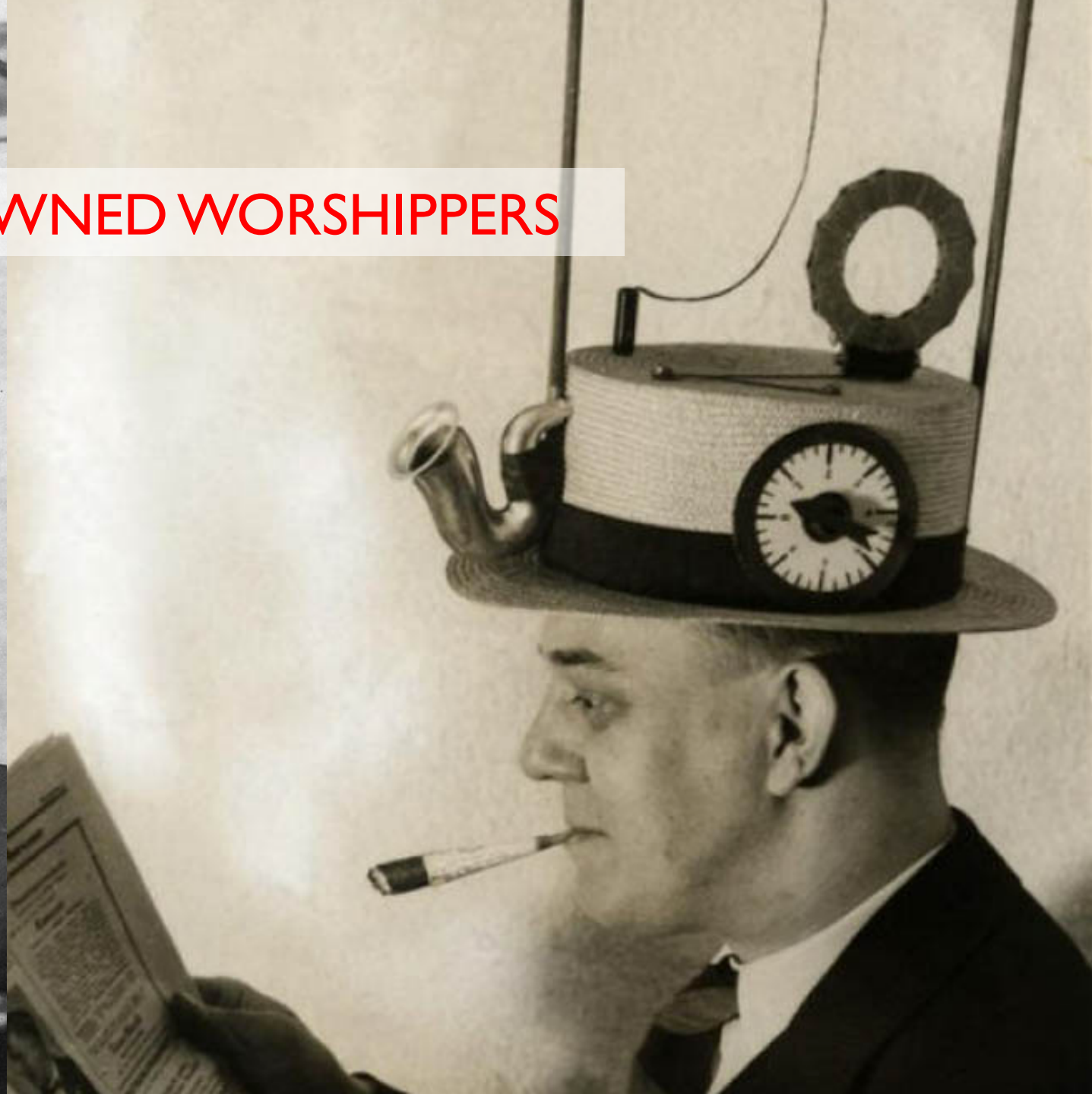
AI AND SOCIETY: A BIRD'S-EYE VIEW ON ETHICS AND AI

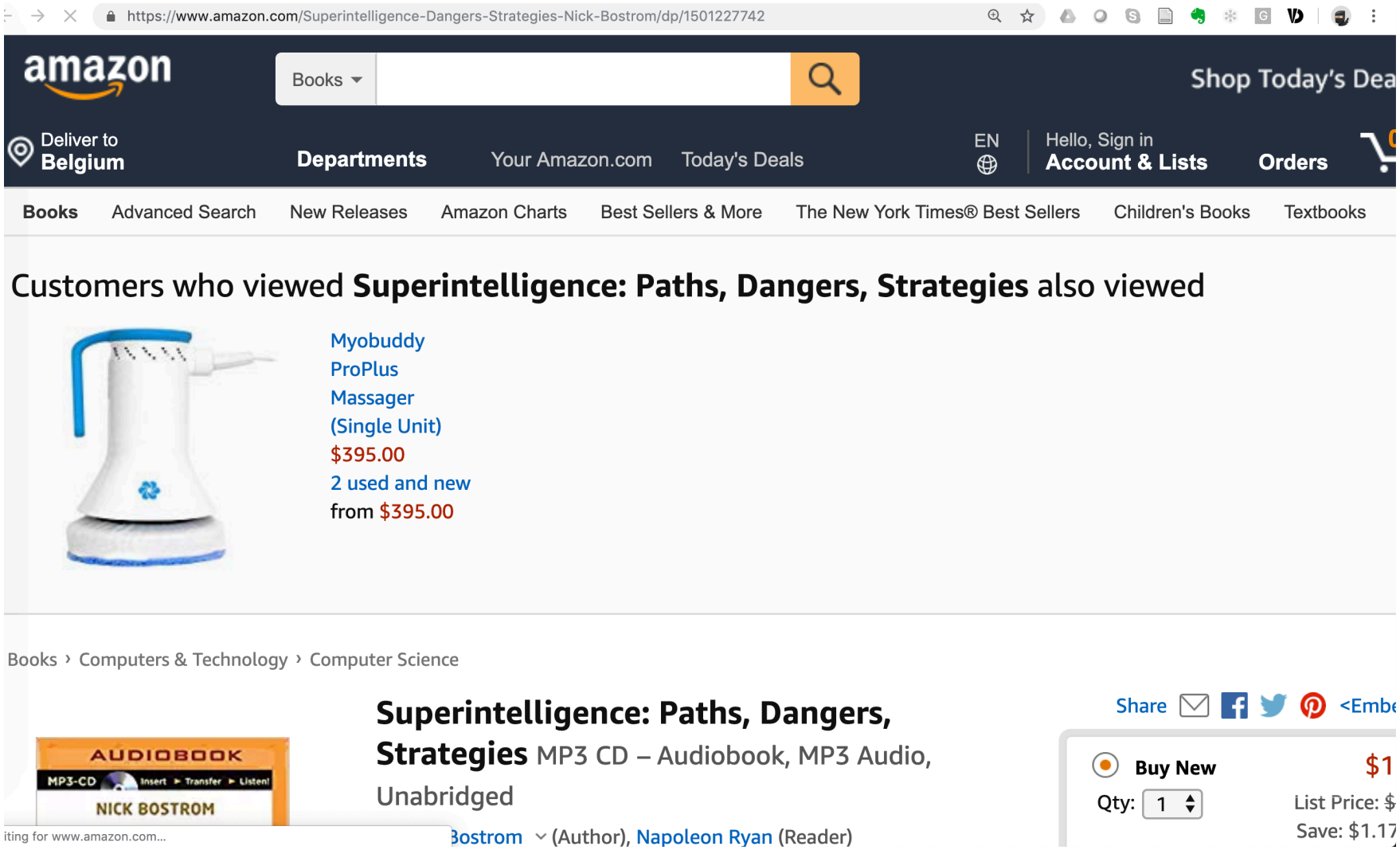
ROB HEYMAN, IMEC-SMIT VUB

WHAT I WANT YOU TO REMEMBER FROM THIS TALK
CONTENTS

- AI can be general or specific
- Ethics can be used for anything by anyone
- Who needs ethics in AI? And why?

THE DROWNED WORSHIPPERS





"Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. (…)" (AI HLEG)

KINDS OF AI

GENERAL AI

- Is able to do what humans do in general
- Might become better, bigger, faster than humans
- Nick Bostrom: Superintelligence: Paths, Dangers, Strategies

SPECIFIC AI

- Is better, bigger, faster at one specific thing than humans
- Brian Christian, Tom Griffiths: Algorithms to Live By: The Computer Science of Human Decisions
  - When should you take the next parking spot, pick a restaurant, a partner? What is the quickest way to sort books in a library?

ETHICS

DISCLAIMER: I AM NOT SUMMARISING A WHOLE FIELD, NOR AM I AN ETHICIST

- **Standardisation**
- Well-founded standards of right and wrong that prescribe what humans ought to do, usually in terms of rights, obligations, benefits to society, fairness, or specific virtues.
- What AI should be as a standard
  - Do not mention the trolley problem

- **Problematisation**
- The study and development of one's ethical standards
- What do we have, what do we need in terms of standards for AI based on what we know?

THE AI ETHICS GUIDELINES GLOBAL INVENTORY

ALGORITHM WATCH

- Similar principles: transparency, equality/non-discrimination, accountability and safety
  - Some have additional on AI being socially beneficial and protecting human rights
- Frameworks are developed by coalitions, organisations can sign sign up
  - Voluntary commitments
- Majority of documents published in 2018-19
- High level principles that do not detail how things are done

UK PRESS COVERAGE ABOUT AI IS DOMINATED BY INDUSTRY (60%)

AN 'ARENA' FOR THE FUTURE OF OUR SOCIETIES

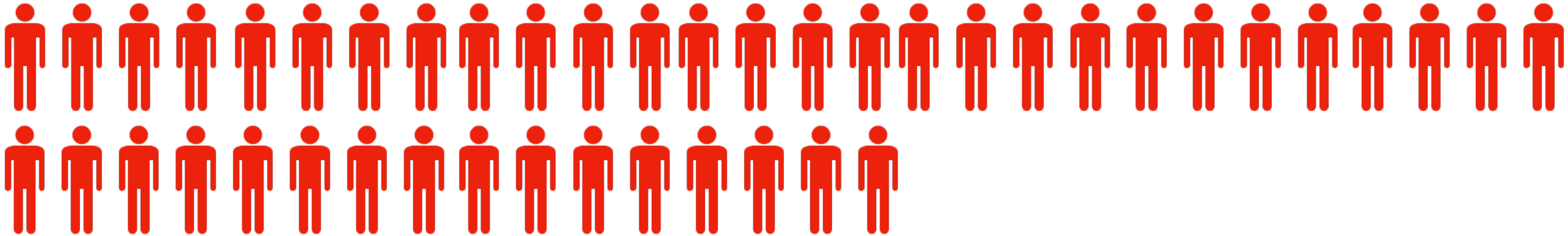
"By amplifying industry's self-interested claims about AI," said one of the researchers, "media coverage presents AI as a solution to a range of problems that will disrupt nearly all areas of our lives, often without acknowledging ongoing debates concerning AI's potential effects. In this way, coverage also positions AI mostly as a private commercial concern and undercuts the role and potential of public action in addressing this emerging public issue."

(John Naughton: Don't believe the hype: the media are unwittingly selling us an AI fantasy)

ETHICS WASHING MADE IN EUROPE (THOMAS METZINGER)
HLEG COMPOSITION

- Ethicists

- Others (policy, industry):



CONCLUSION: WE STILL HAVE TIME
AND NEXT STEPS

- AI is still interpretive flexible, it can be many things to different people
  - Industry defines, society follows
- Standards, guidelines and ethics
  - smoke screen to whitewash future societal challenges OR
  - Ethics as one way of opening up the debate
- AI in society will be a problem of many hands
  - Improve representation
- Next steps
  - Create openness about challenges and AI failures
  - Invite societal actors
  - Increase transparency to allow feedback
  - See involvement and ethics as an ongoing effort

SOURCES

- Ethics washing made in Europe. Retrieved April 23, 2019, from https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html
- Launch of our "AI Ethics Guidelines Global Inventory." (2019, April 9). Retrieved April 24, 2019, from AlgorithmWatch website: https://algorithmwatch.org/en/launch-of-our-ai-ethics-guidelines-global-inventory/
- Bostrom, N. (2016). *Superintelligence: paths, dangers, strategies*. Oxford, United Kingdom ; New York, NY: Oxford University Press.
- Naughton, J. (2019, January 13). Don't believe the hype: the media are unwittingly selling us an AI fantasy | John Naughton. *The Guardian*. Retrieved from https://www.theguardian.com/commentisfree/2019/jan/13/dont-believe-the-hype-media-are-selling-us-an-ai-fantasy

Mieke De Ketelaere

Customer Intelligence Director - SAS

Ethical AI through the eyes of an engineer

Mieke De Ketelaere (SAS, IMEC)

About this session

Covered: Translation of AI principles to reach FATE in an understandable way

Not covered: Liability

Ethical AI in Business

BUSINESS

How do we approach ethical AI in our company?

DATA SCIENTIST

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} \ \mathcal{L}(f, g, \pi_x) + \Omega(g) \qquad (1)$$

LEGAL

...

3. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.

4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.
...

Ethical AI in Business

BUSINESS

AI TRANSLATOR

DATA SCIENTIST

LEGAL

Ethical Guidelines for Trustworthy AI

About this session: Observations and Objections

Fair

Accountable

Transparent

Explainable

- “I get my algorithms cheaper outside Belgium”
- “We can’t do anything due to GDPR”
- “AI is a black box which I do not trust”
- “Companies don’t want to share data with us”
- “We have invested in the best AI developers”

The Process: From Data to Action

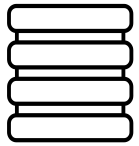
01 Listen

02 Understand

03 Act

"Don't be intimidated by jargon. For example, a model is just a fancy word for "recipe." (C. Kozyrkov)"

The Process: From Data to Action



$Y=\beta_0+ \beta_1x_1+ \beta_2x_2+ \beta_3x_3 + \ldots$

CLV = -1500 -20,5*Age
+ 1,3*Income
+ …

CLV(Mieke) = 789

Loan request: OK!
(Backend)

Automated approval
(Frontend)

Request → Data → Algorithm → Model → Prediction → Decision → Action






Request → Ingredients → Appliances → Recipes → Dish → Observation → Eating

LISTEN **UNDERSTAND** **ACT**

“Imagine trying to start a restaurant by hiring folks who’ve been building microwave parts their whole lives but have never cooked a thing… what could go wrong?” (C. Kozyrkov)

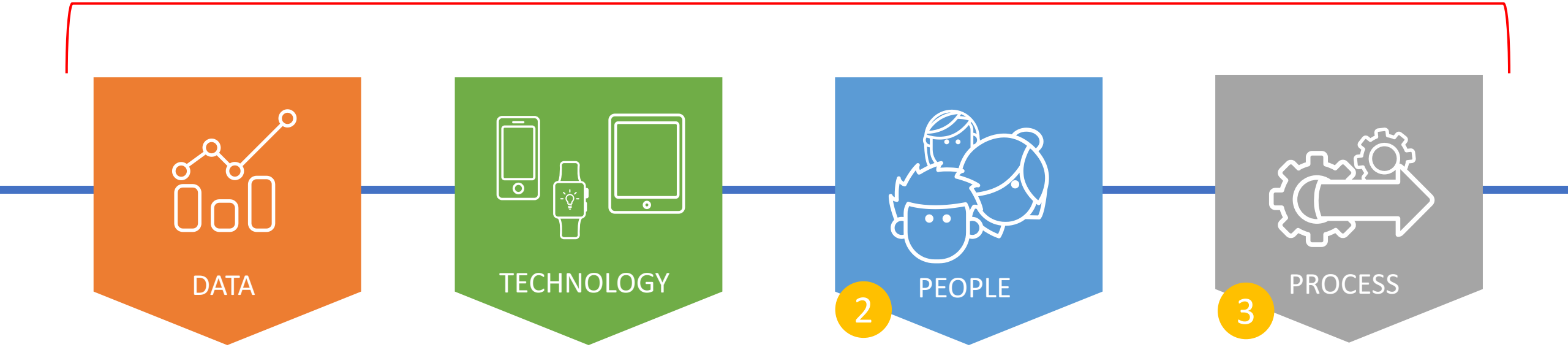
“We have invested in the best AI developers”

The complete people and proces layer is as important as the technology

“If a researcher is your first hire, you probably won’t have the right environment to make good use of them.” (C. Kozyrkov)

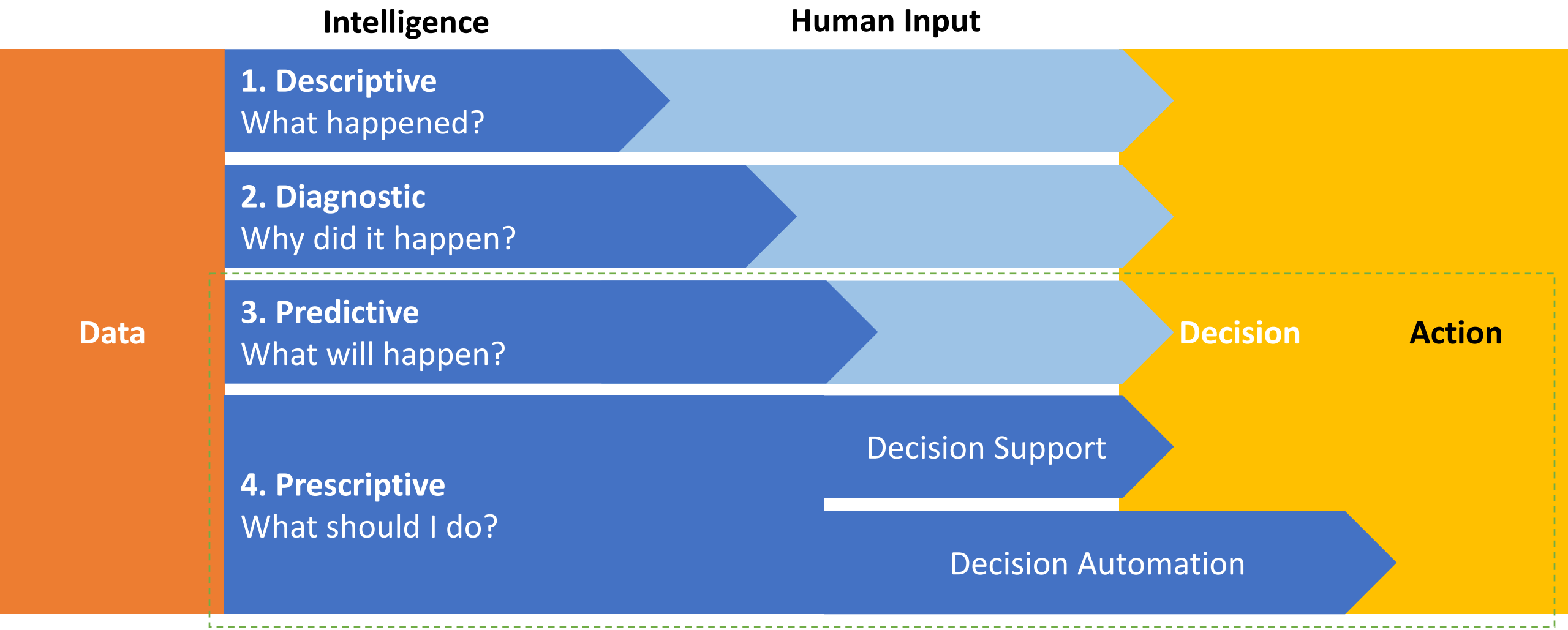
Accountable AI: Organisational Readiness

1 VISION: What is your objective? What is the problem to solve? Expected value?



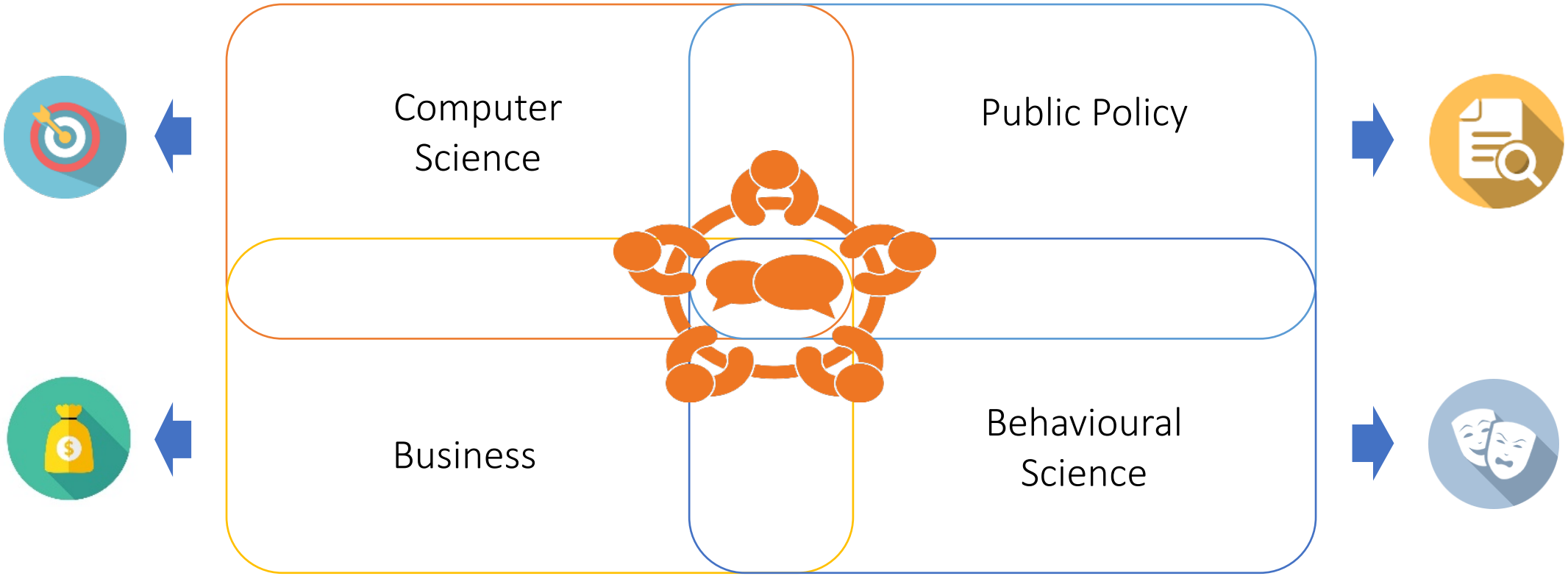
All topics are important to tackle **ACCOUNTABILITY**

1 - The Vision: Understanding the Problem



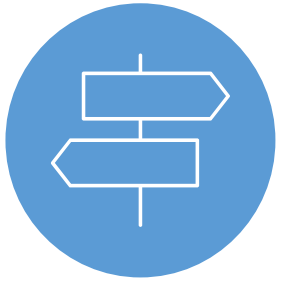
ACCOUNTABILITY is key in all steps of the process and all levels of intelligence.

1 - The Vision: A Multidisciplinary Debate



1 -The Vision: A Multidisciplinary Debate

KPIs



Identify decisions

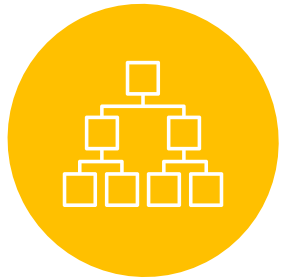
Define decisions

Business/Legal context

Units and roles

Data

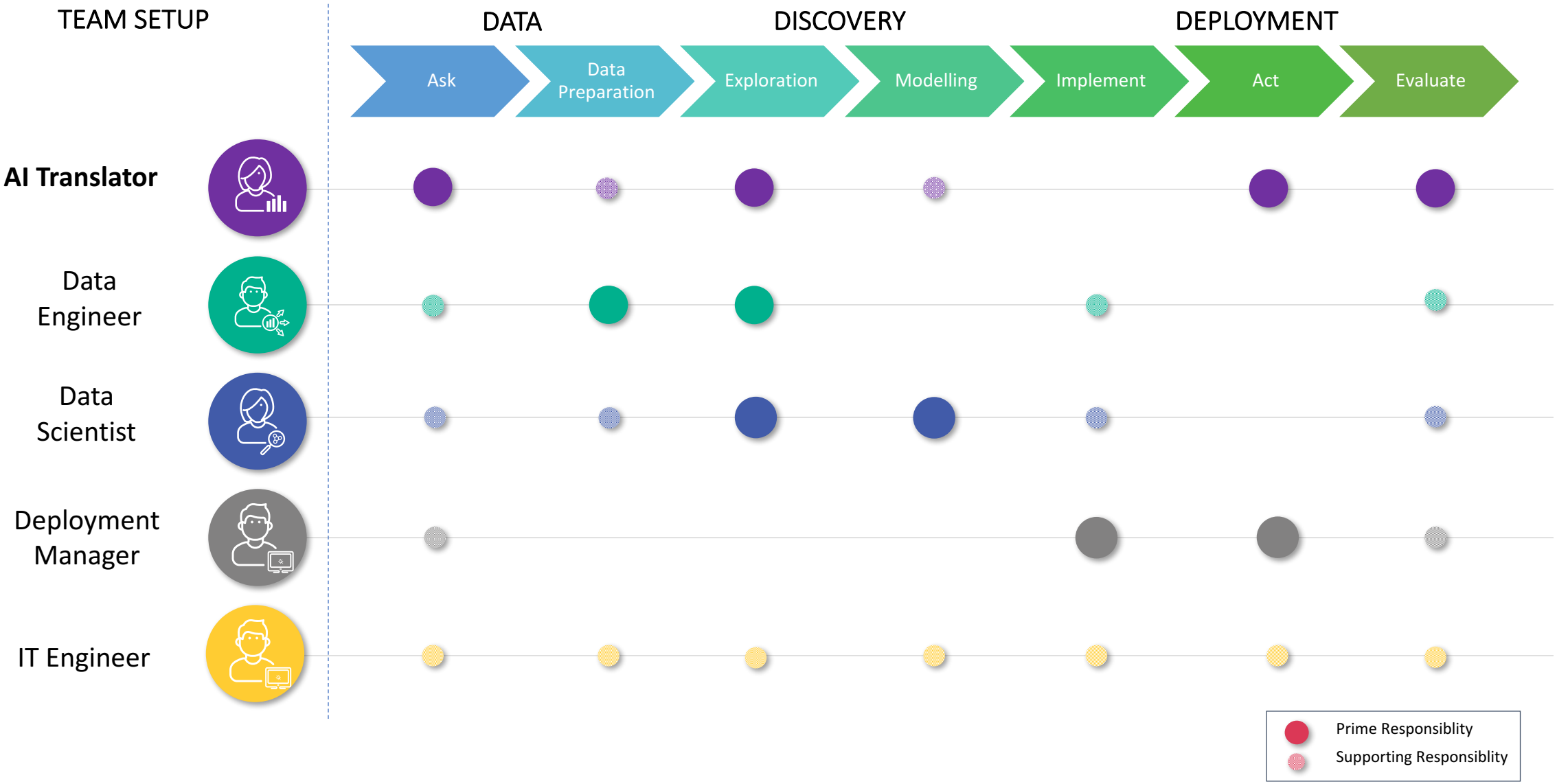
Knowledge



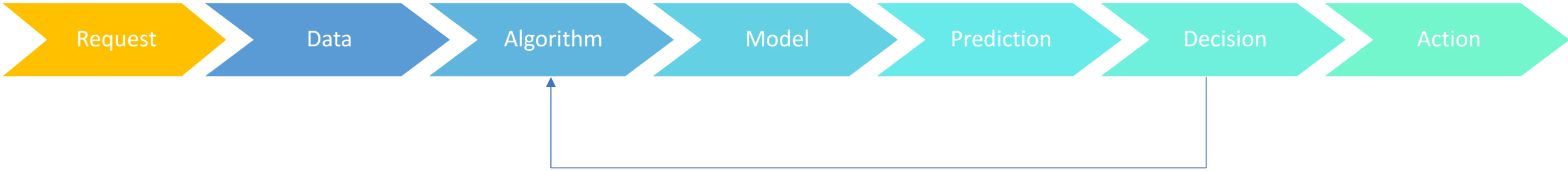
Decision Specs

ADVICE: Use a business canvas for the planning your AI projects

2 – The People: The Roles



3 - The Process: From Data to Action

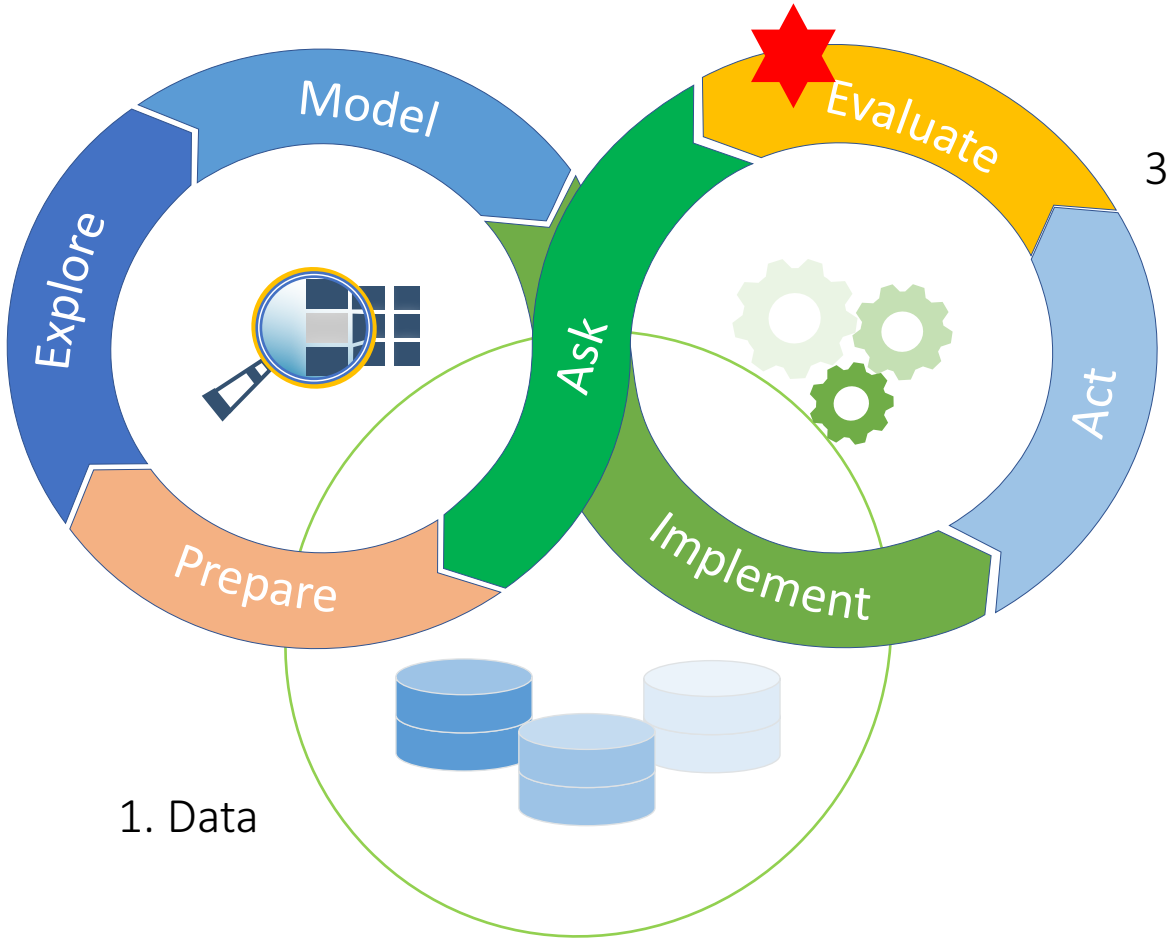


"Communicate, in a clear and proactive manner, information to stakeholders about the AI system's capabilities and limitations, enabling realistic expectation setting, and about the manner in which the requirements are implemented.

Be **transparent** about the fact that they are dealing with an AI system."

3 – The Process: The Lifecycle of AI Systems

2. Discovery



3. Deployment

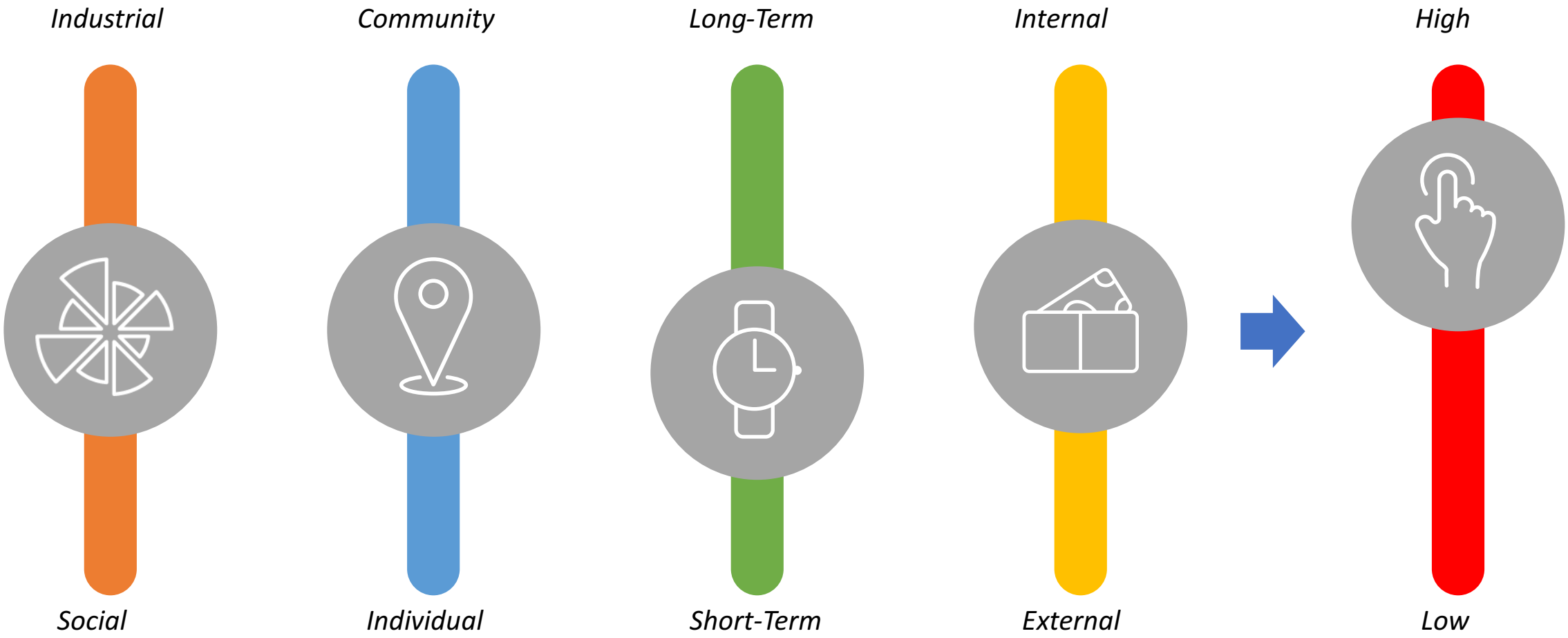
1. Data

Benefit of the Lifecycle

- Consistency
- Efficiency
- Agility
- Reusability
- Transparency
- Monitoring

NOTE: The AI lifecycle is the optimal treatment after the **PPP syndrome (POC, Pilot, Prototype)**.

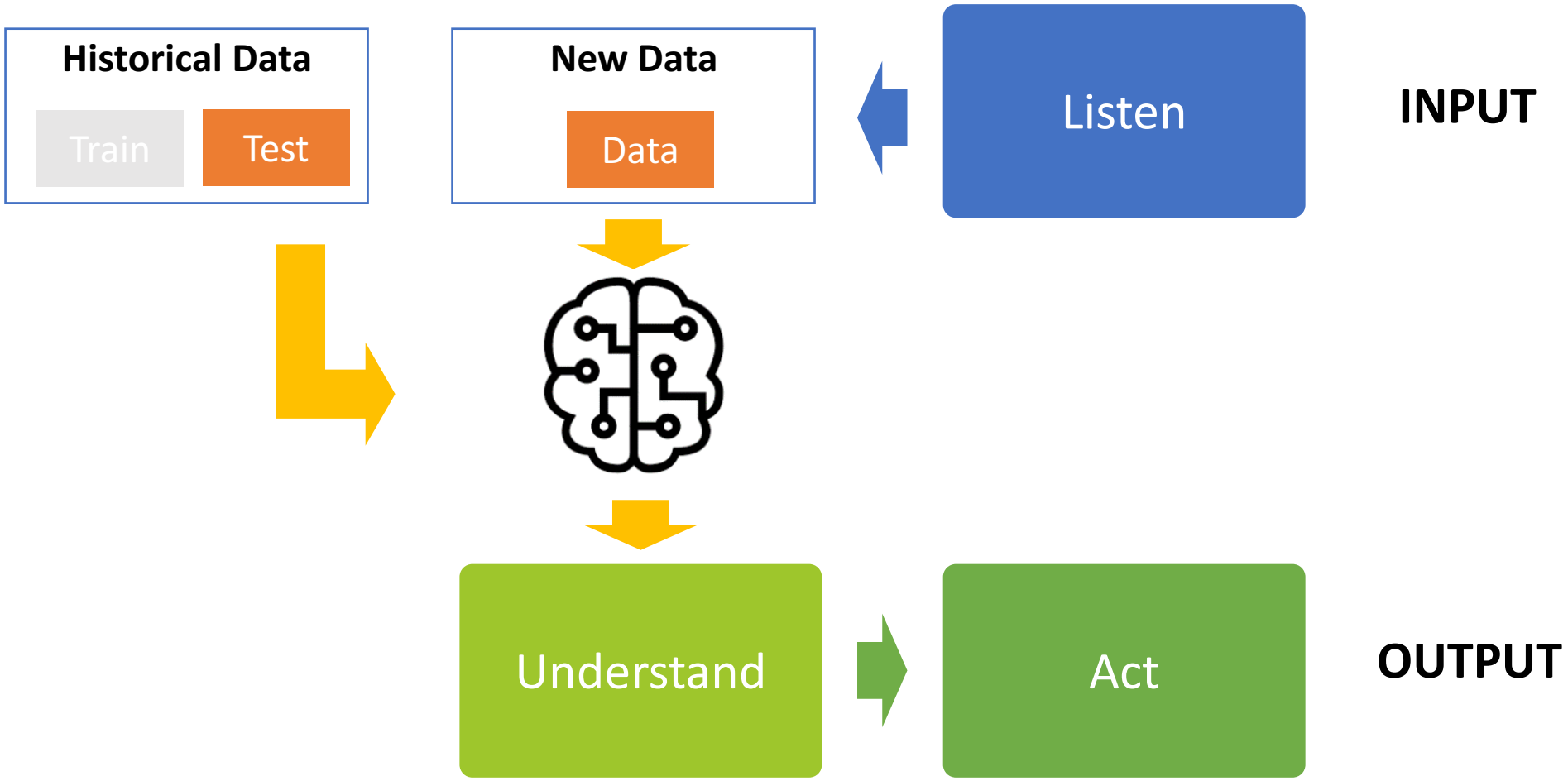
3 – The Process: The Lifecycle of AI Systems



“I get my algorithms cheaper outside BE”

The impact of cognitive bias in AI can kill your brand

Known Issues



Human versus Artificial Intelligence

	HUMAN INTELLIGENCE	ARTIFICIAL INTELLIGENCE
Speed of processing		✓
Copy		✓
Accuracy on Big Data		✓
Modus		✓
Size of storage	✓	
Energy	✓	

Human versus Artificial Intelligence

	HUMAN INTELLIGENCE	ARTIFICIAL INTELLIGENCE
Learn Experience	✓	
Adopt Change	✓	
Parallel Processing	✓	
Type of Input Data	✓	
Creativity	✓	
Empathy	✓	

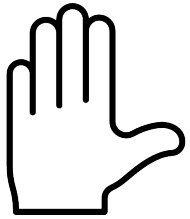
Human versus Artificial Intelligence

	HUMAN INTELLIGENCE	ARTIFICIAL INTELLIGENCE
BIAS		

Cognitive Bias introduces Computational Bias

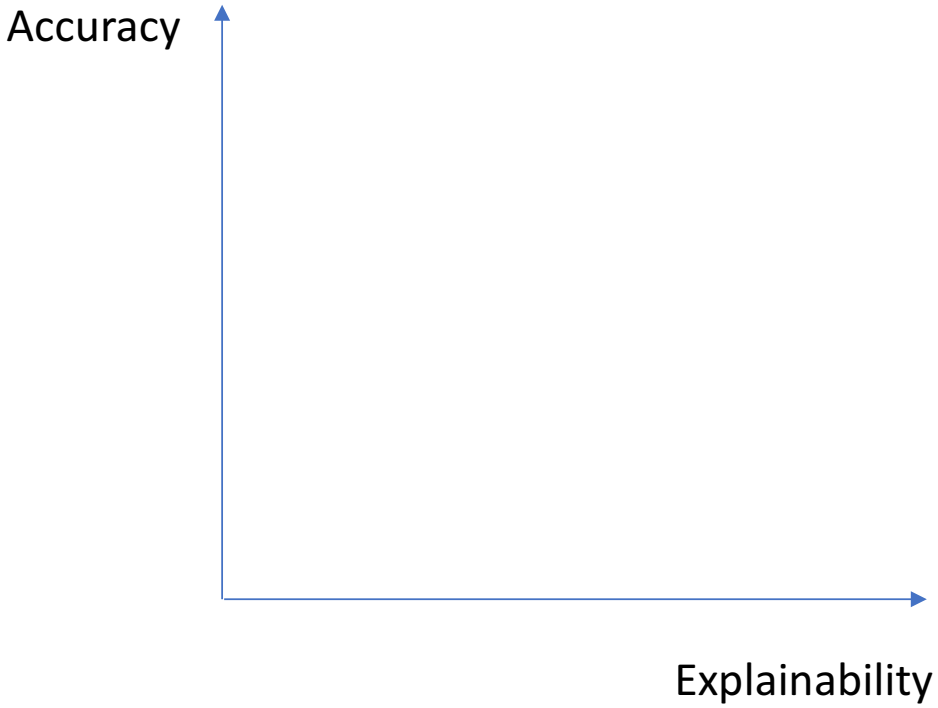
Availability Bias

Enchoring Bias



Confirmation Bias

Overconfidence Bias



EXAMPLES

Anti-Bias Framework

	Diagnosis	Treatment	Prevention
Listen	✓	✓	✓
Decide	✓	✓	✓
Act	✓	✓	✓

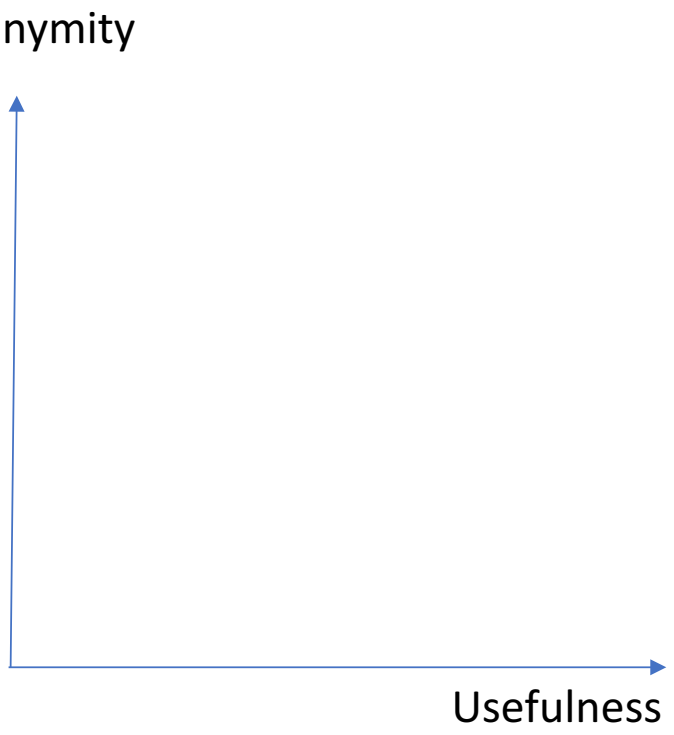
“We can’t do anything due to GDPR”

Synthetic data and differential privacy can solve the problem

Tradeoff between anonymity and usefulness

Perfect Anonymity

- **Pro**: 100% Data Privacy
- **Con**: Data loss and distortions can compromise conclusions

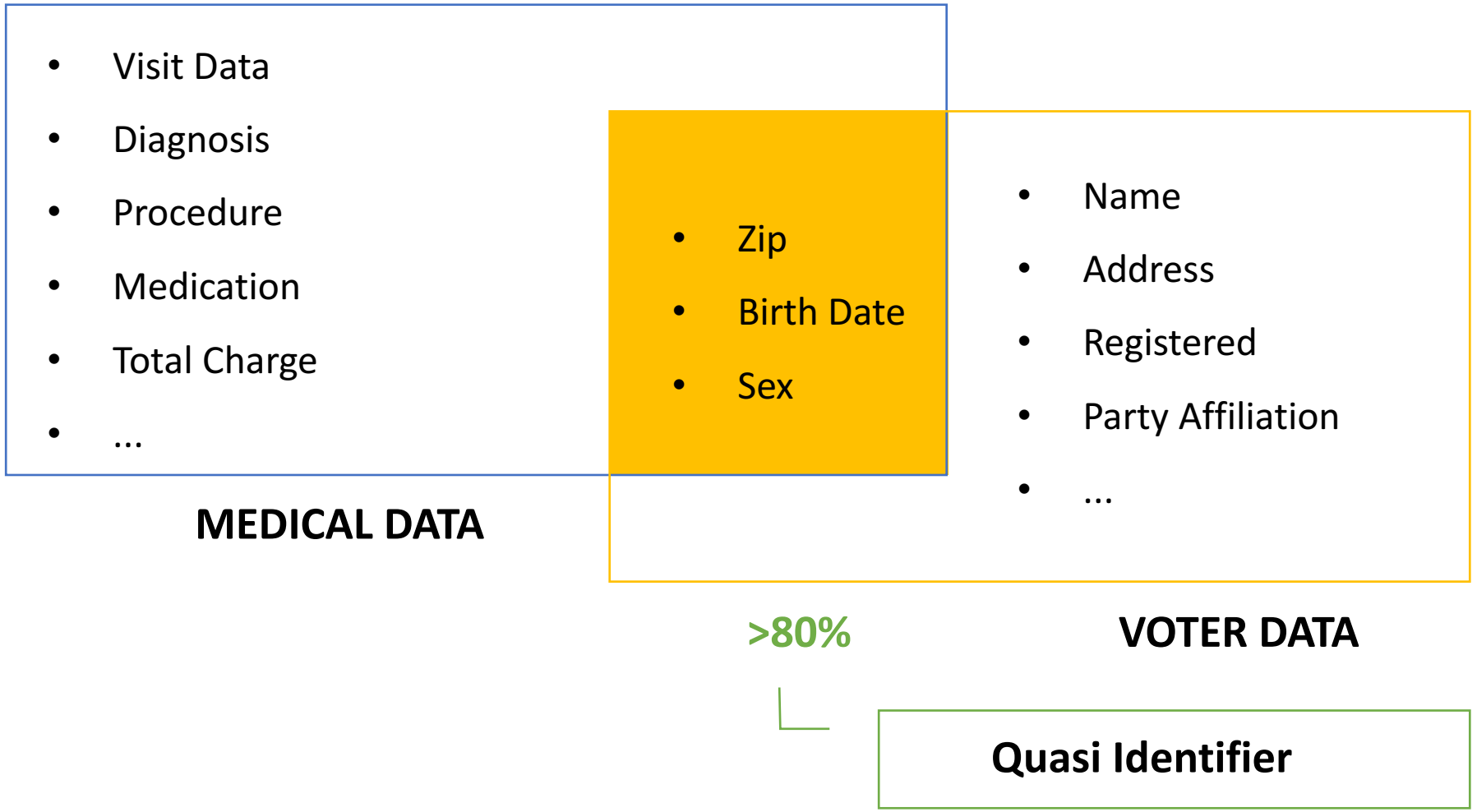


Perfect Usefulness

- **Pro**: A maximum of data based insights is possible
- **Con**: Disclosure of individuals is possible

Privacy concerns

The Problem with Data Anonymisation



Tackling Data Privacy: Depersonalisation Approaches

Depersonalisation

Standard Approach

Adaption of real data using data manipulation techniques such as replacement, suppression, etc

Issues with standard depersonalisation approach

Name	Age	Disease
Alice	30	No
Bob	30	No
Thomas	40	Yes
Ann	50	No
Frank	50	Yes

Question: What is the youngest age of a person with the disease?

Issues with standard depersonalisation approach

K-Anonymity Approach

Name	Age	Disease
Alice	30	No
Bob	30	No
Carl	40	No
Thomas	40	Yes
Ann	50	No
Frank	50	Yes

Question: What is the youngest age of a person with the disease?

Issues with standard depersonalisation approach

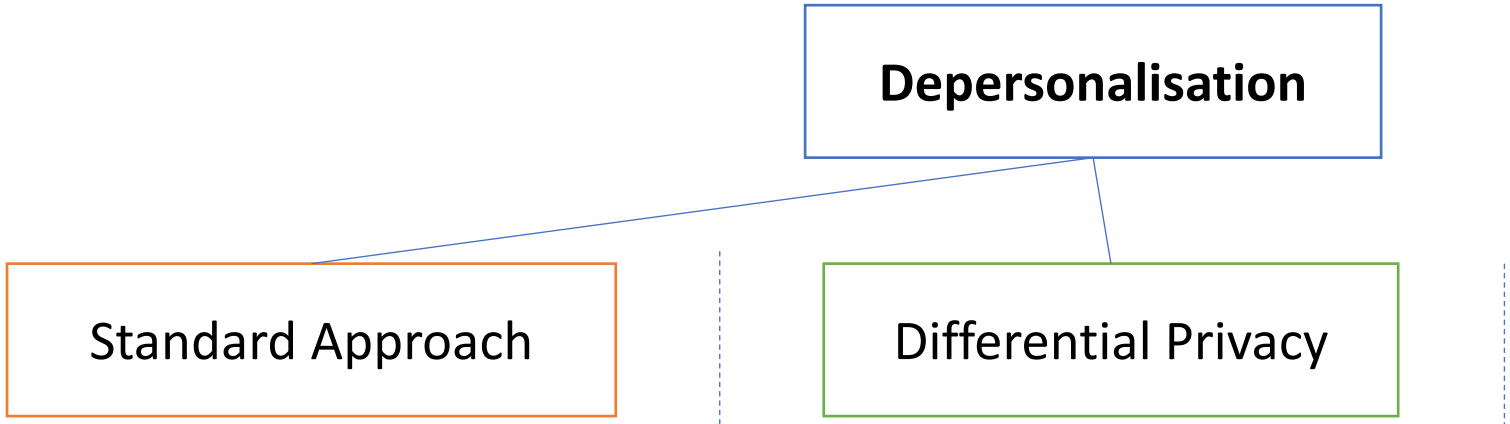
The problem of composition

Name	Age	Disease
Alice	30	No
Bob	30	No
Carl	40	No
Thomas	40	Yes
Ann	50	No
Frank	50	Yes

Name	Weight	Disease
Alice	60	No
Bob	90	No
Carl	90	No
Thomas	100	Yes
Ann	60	No
Frank	100	Yes

Question: What is the youngest age and minimal weight of a person with the disease?

Tackling Data Privacy: Depersonalisation Approaches



Adaption of real data using data manipulation techniques such as replacement, suppression, etc

A mechanism which reports an **approximate answer**, typically generated **randomly** on the basis of the **true answer** and of some **probability distribution.** Mainly used to gain insights from **large datasets**

Introducing: Differential Privacy

Adding some probalistic noise

Name	Age	Disease
Alice	30	No
Bob	30	No
Carl	40	No
Thomas	40	Yes
Ann	50	No
Frank	50	Yes

Name	Weight	Disease
Alice	60	No
Bob	90	No
Carl	90	No
Thomas	100	Yes
Ann	60	No
Frank	100	Yes

Question: What is the youngest age and minimal weight of a person with the disease?

Answer:

Noisy answers – Age

- 40 with probability 1/2
- 30 with probability 1/4
- 50 with probability 1/4

Noisy answers - Weight

- 100 with probability 4/7
- 90 with probability 2/7
- 60 with probability 1/7

Introducing: Differential Privacy

- A good mechanism which provides a good **trade-off** between **privacy** and **utility**.

POTENTIAL SOLUTION TO TACKLE PRIVACY

Google

Uber

QuickType suggestions • Emoji suggestions • Lookup Hints • Safari Energy Draining Domains • Safari Autoplay Intent Detection (macOS High Sierra) • Safari Crashing Domains (iOS 11) • Health Type Usage (iOS 10.2)

Tackling Data Privacy: Depersonalisation Approaches

Depersonalisation

Standard Approach

Adaption of real data using data manipulation techniques such as replacement, suppression, etc

Differential Privacy

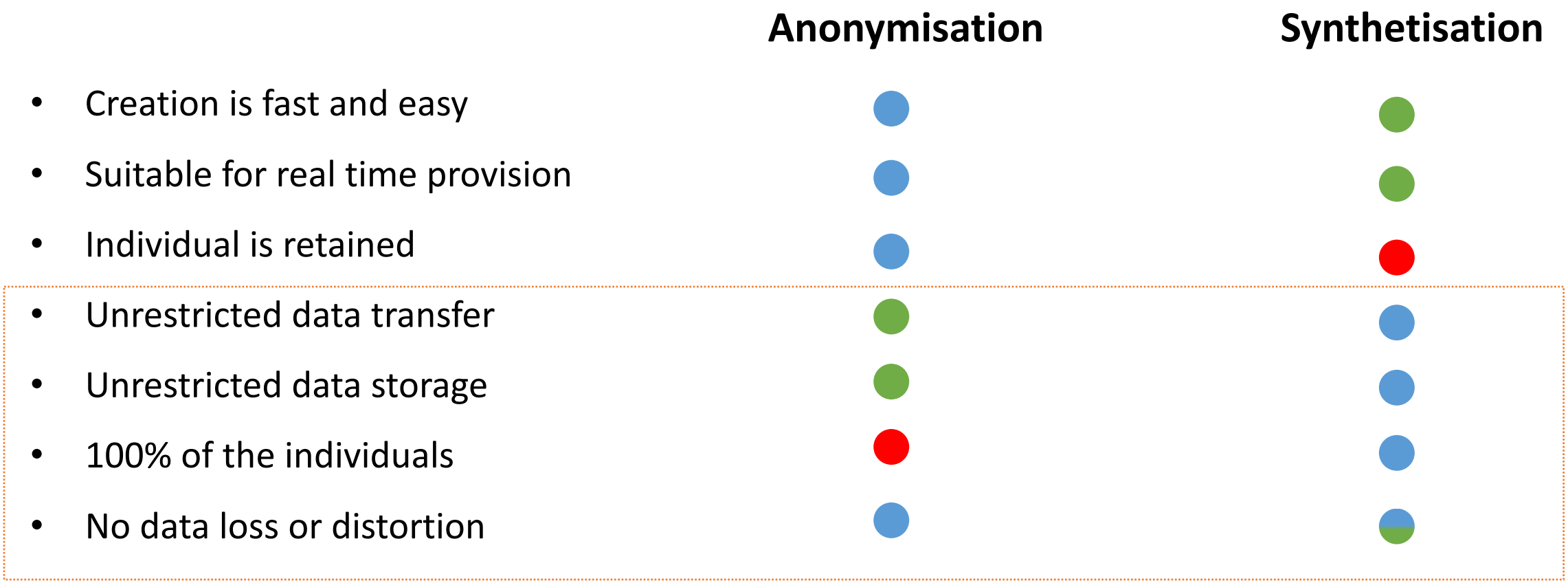
A mechanism which reports an **approximate answer**, typically generated **randomly** on the basis of the **true answer** and of some **probability distribution.** Mainly used to gain insights from **large datasets**

Synthetic Data

Synthetic data is information that's **artificially manufactured** rather than generated by real-world events. It is used as a **stand-in for test datasets** of production or operational data, to **validate mathematical models** and, increasingly, to **train machine learning models.**

Introducing: Synthetic Data

Introducing: Synthetic Data



POTENTIAL SOLUTION TO TACKLE PRIVACY

Introducing: Synthetic Data

http://www.thispersondoesnotexist.com/

Introducing Synthetic Data

Medical Data

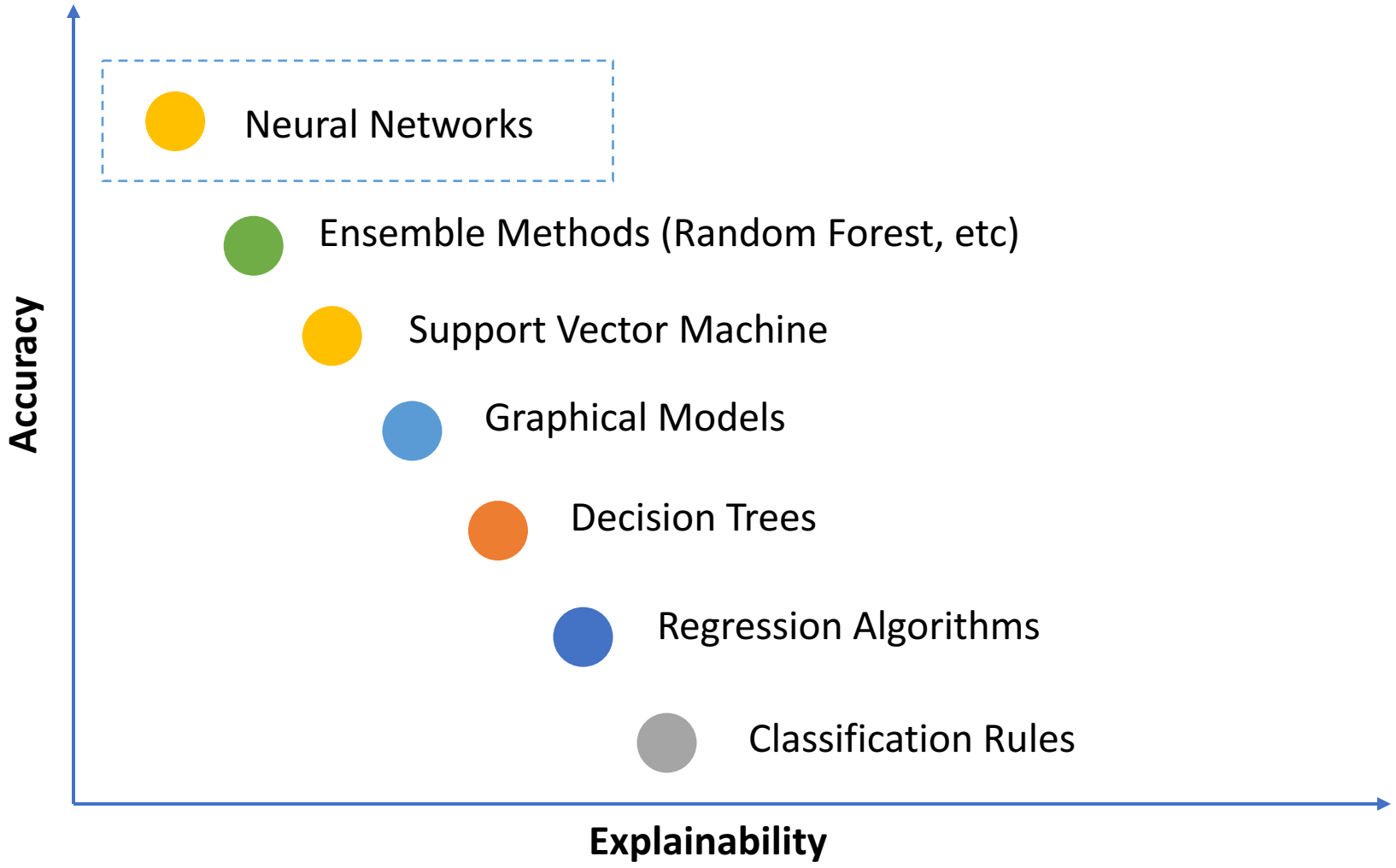
Financial Data

“AI is a black box which I do not trust”

LIME and other techniques can help the explainability of AI

“If you refuse to trust decision-making to something whose process you don’t understand, then you should fire all your human workers, because no one knows how the brain (with its hundred billion neurons!) makes decisions.” (C. Kozyrkov)

The Challenge: Accuracy versus Explainability



Motives for AI Explainability

1 Producer/Model Maker

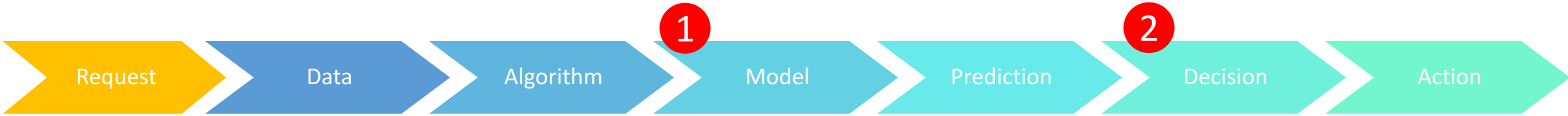
Data scientist

- Debugging and improving
- Exploring and discovering latent or hidden feature interactions
- Model variability
- Model comparison
- Domain knowledge
- Bring transparency to enable trust

2 Consumer/Model Breaker

Business Owner/Customer/Auditor/Data Engineer

- Explain the model/algorithm
- Explain key features driving the KPI
- Verify and validate the accountability
- Identify blind spots to prevent attacks
- Ability to share the explanations
- Comply with data protection regulations



1

Model Explainability: Occlusion Sensitivity Analysis

Example: Insurance Fraud

- c0: safe driving
- c1: texting – right
- c2: talking on the phone – right
- c3: texting – left
- c4: talking on the phone – left
- c5: operating the radio
- c6: drinking
- c7: reaching behind
- c8: hair and makeup
- c9: talking to passenger

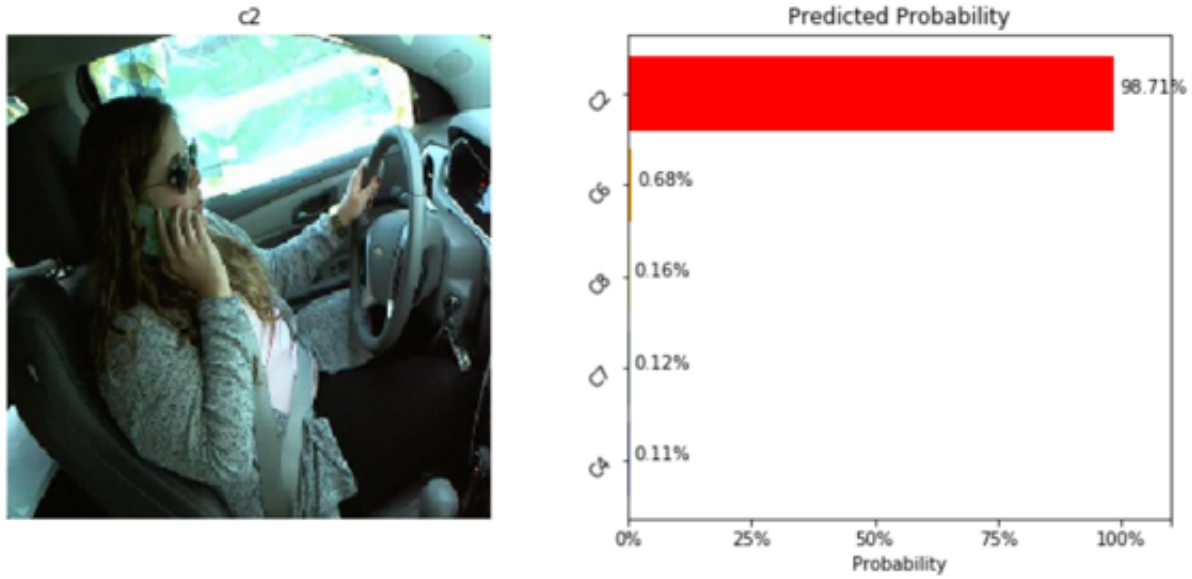
1

Model Explainability: Occlusion Sensitivity Analysis

Return a correct prediction

c0: safe driving c1: texting - right c2: talking on the phone - right c3: texting - left c4: talking on the phone - left c5: operating the radio c6: drinking c7: reaching behind c8: hair and makeup c9: talking to passenger

```
In [36]: model2.plot_predict_res(type='C',image_id=0)
```

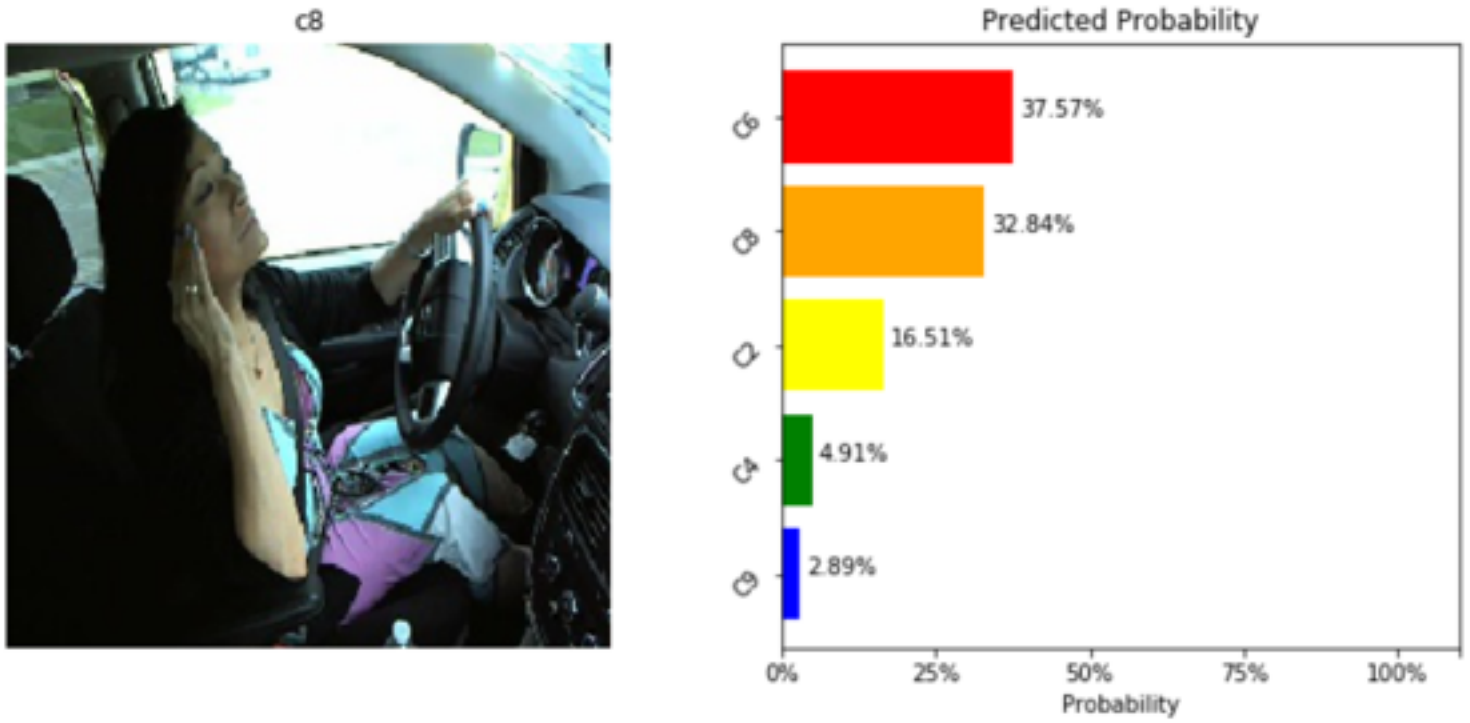


Model Explainability: Occlusion Sensitivity Analysis

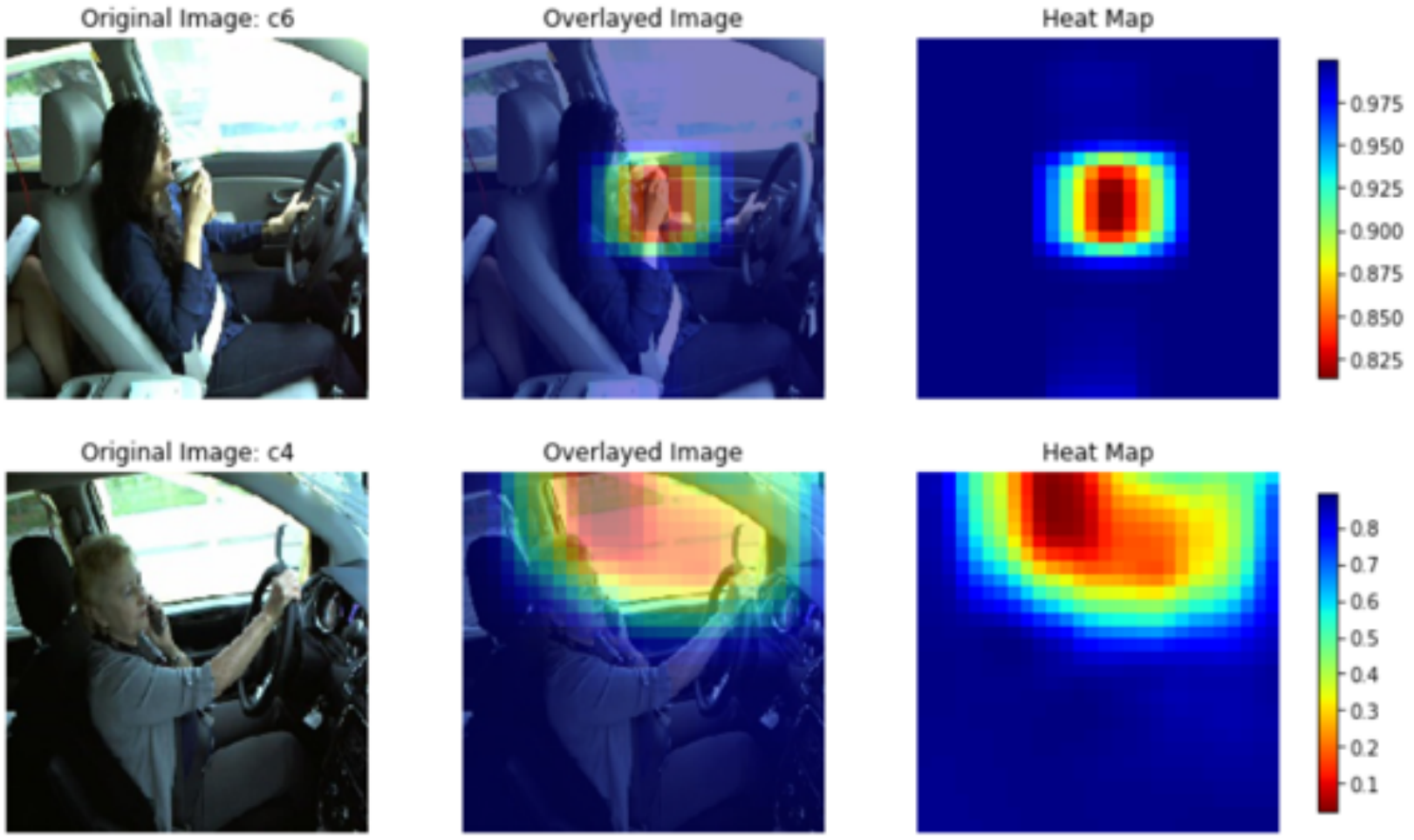
1

Return a misclassified image

```
In [37]: model2.plot_predict_res(type='M',image_id=0)
```

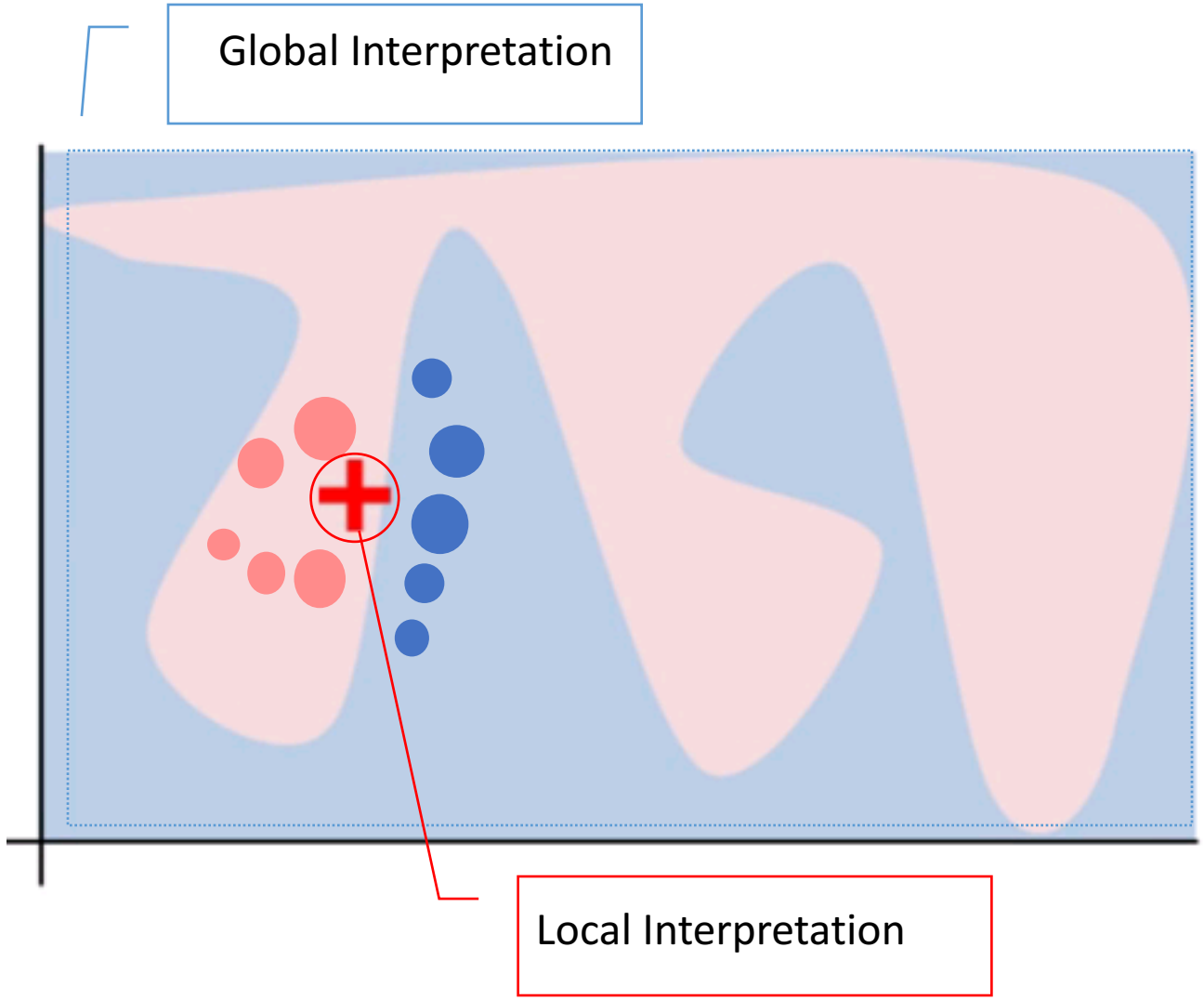


Model Explainability: Occlusion Sensitivity Analysis (OSA)

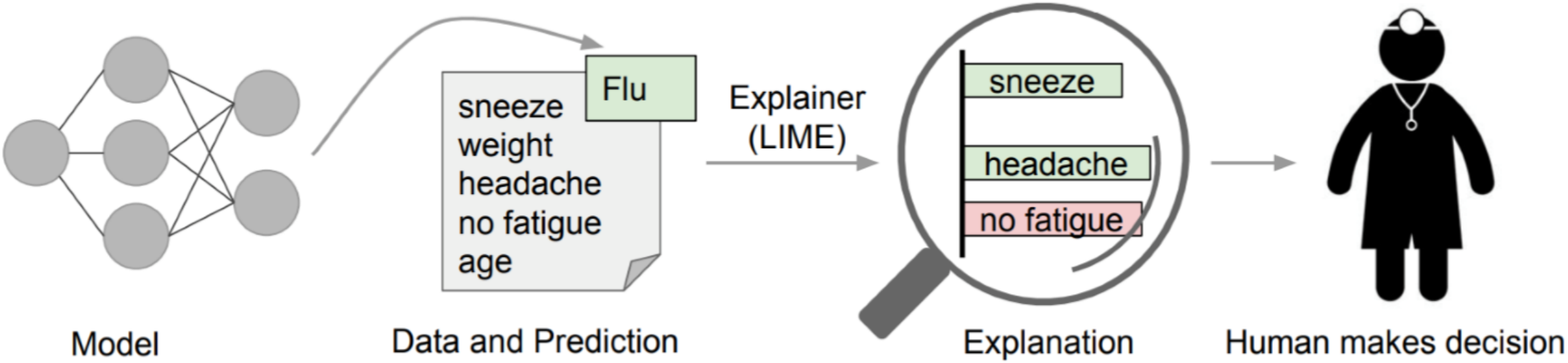


2 Model Explainability: LIME

LIME (Local Interpretable Model-Agnostic Explanations) explains the **conditional interaction** between dependent (response) variables and independent (predictor) variables **for an individual prediction**

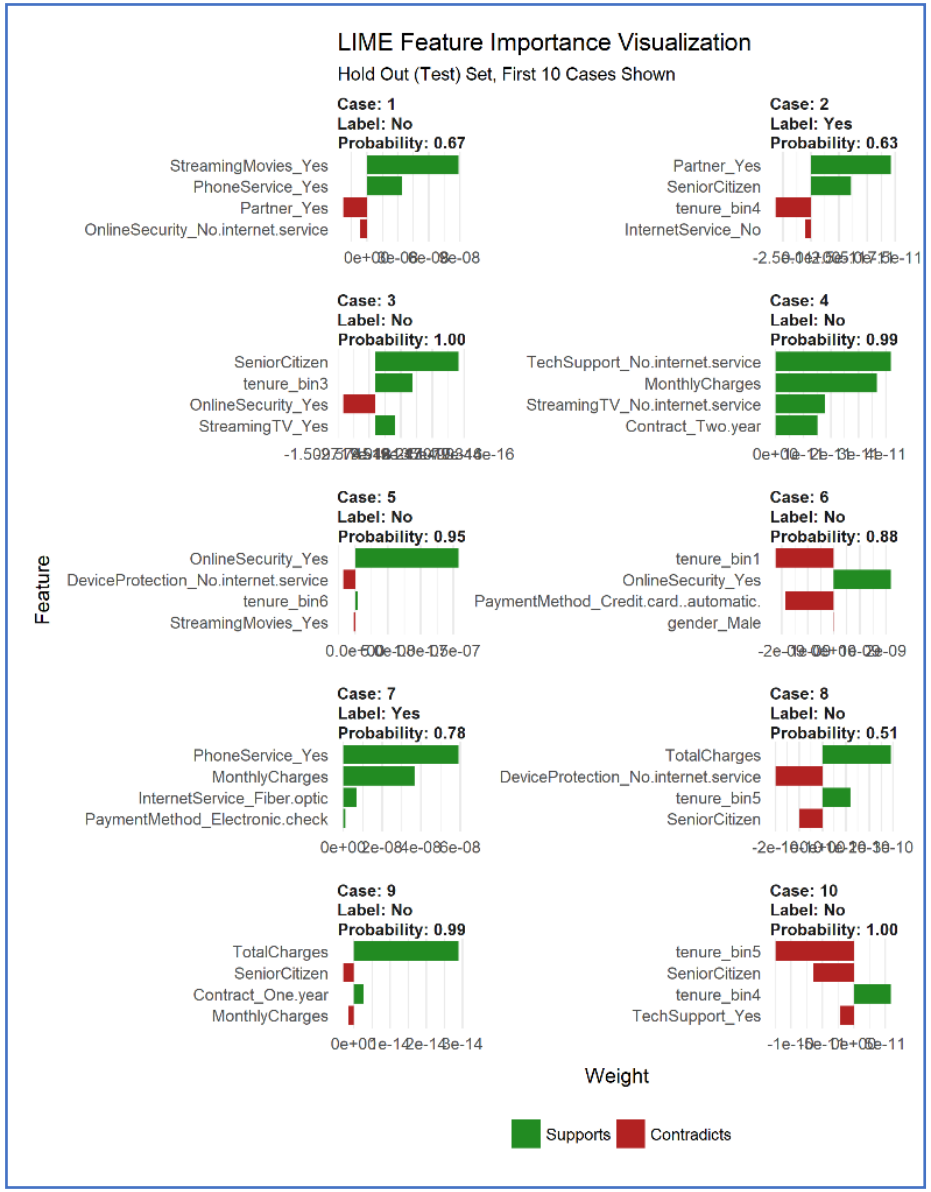


2 Model Explainability: LIME



2 Model Explainability: LIME

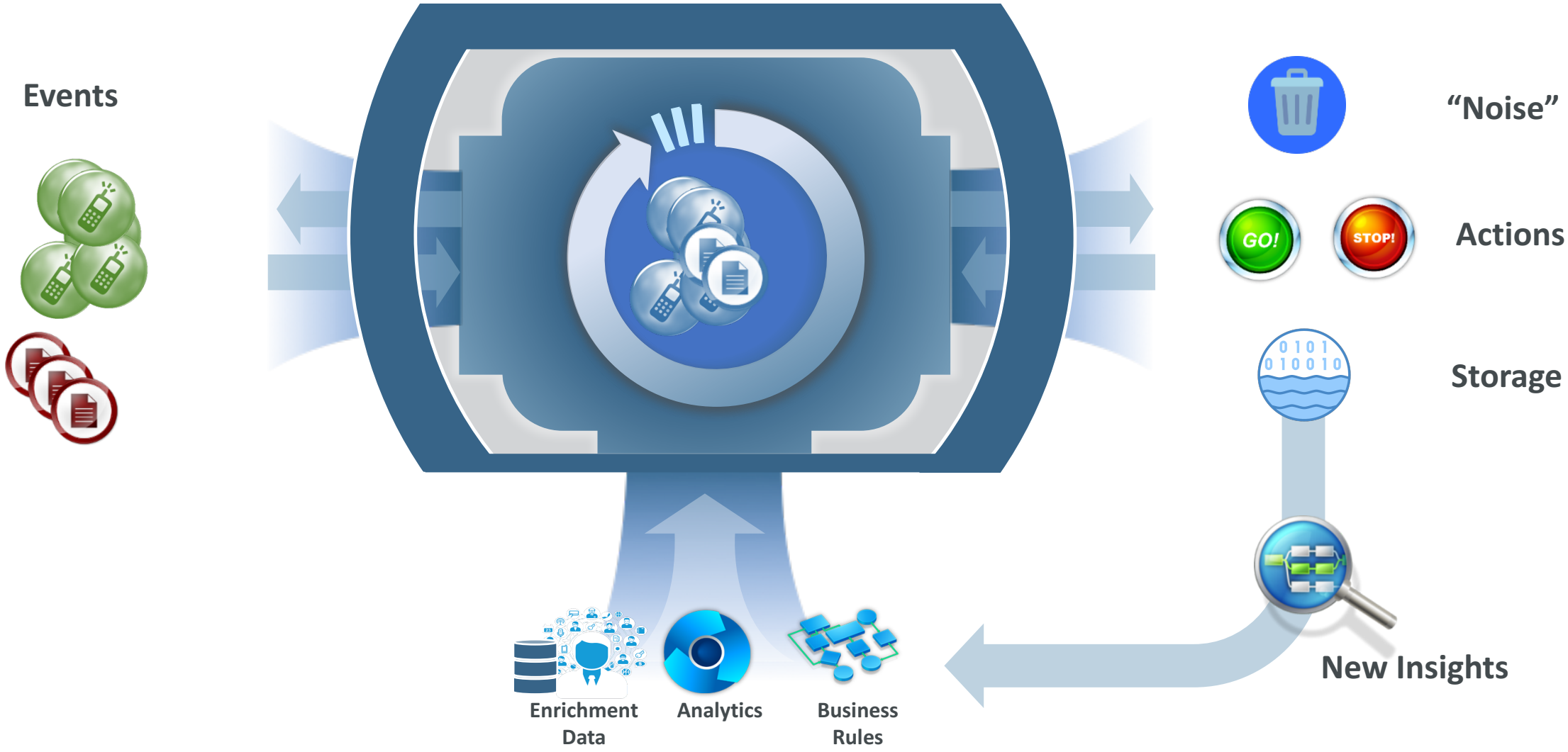
Question: Will my customer churn?



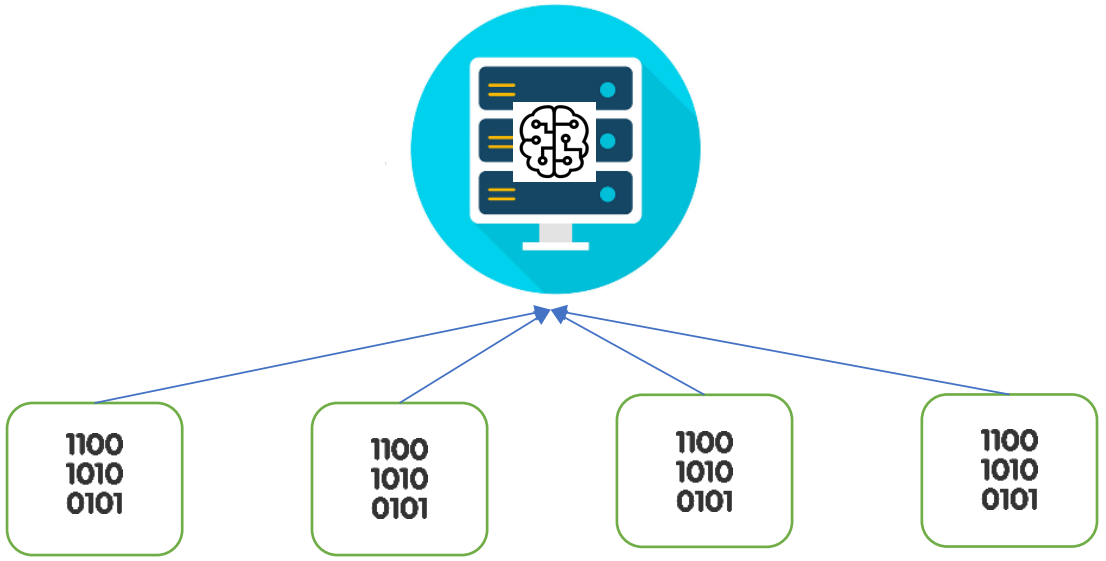
“Companies don’t want to share data with us”

Distributed learning can help latency, competition and privacy issues in one go

Big Data Wave 2.0: IOT



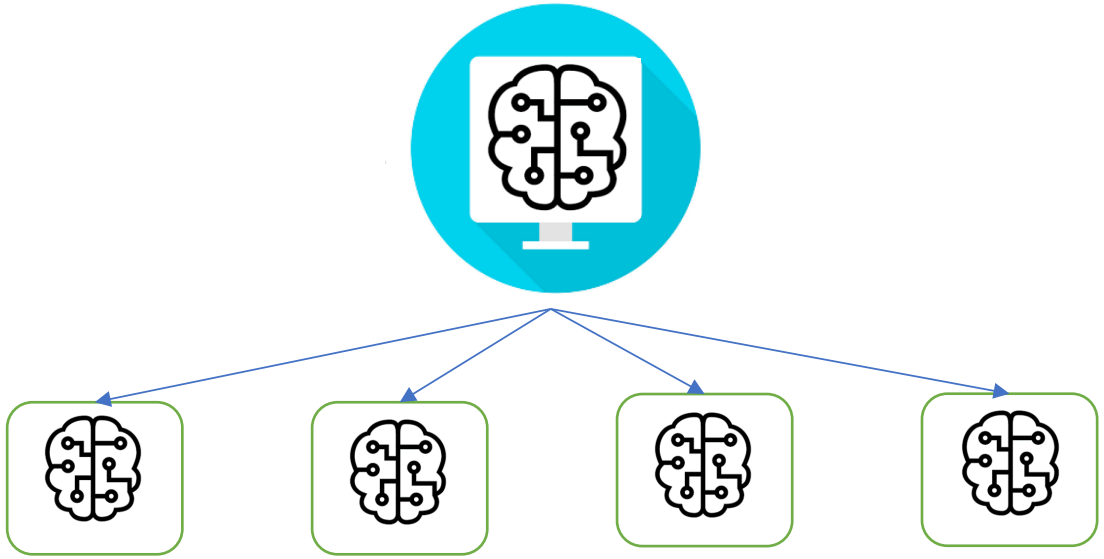
Introducing Federated/Distributed Learning



TRADITIONAL SETUP

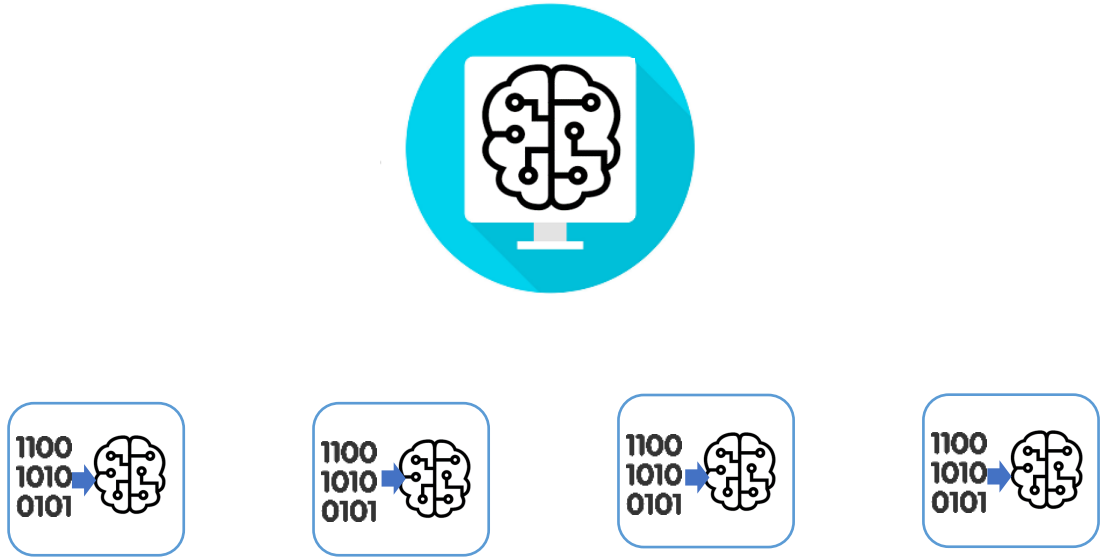
A network of nodes **shares training data** with the server

Introducing Federated/Distributed Learning



Step 1 and 2

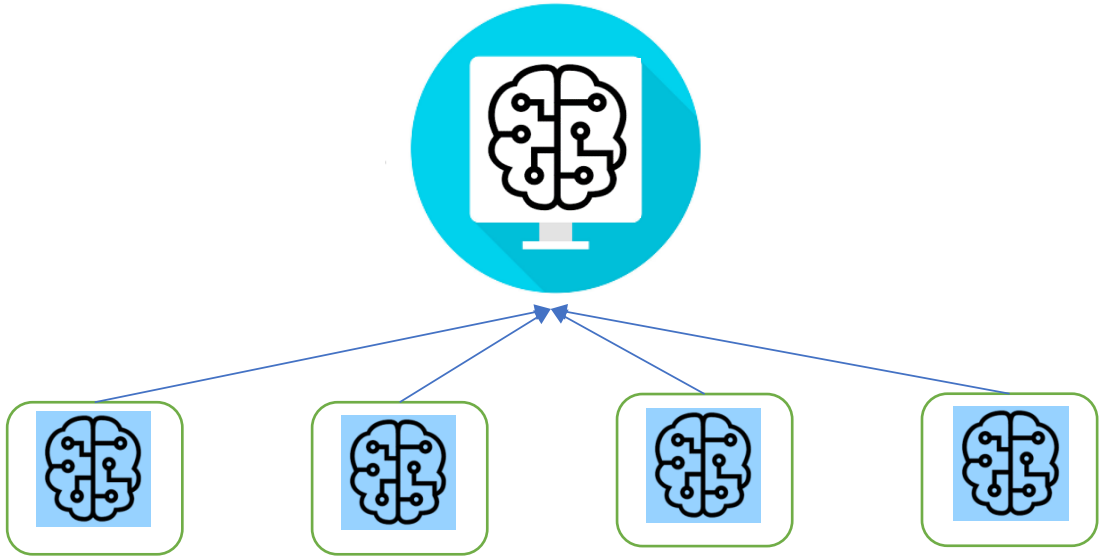
1. The server has an untrained model.
2. It sends a copy of that model to the nodes.



Step 3 and 4

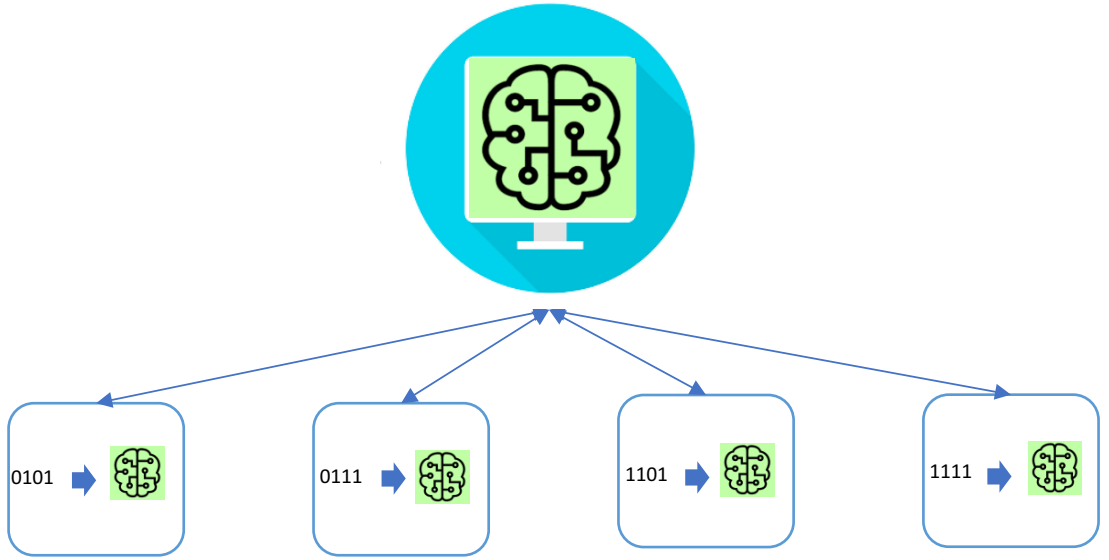
3. The nodes have data on which to train the model.
4. Each node trains the model to fit the data it has.

Introducing Federated/Distributed Learning



Step 5 and 6

5. Each node sends a copy of the model back to the server
6. The server combines the models by taking and average. The process is repeated many times.



Result

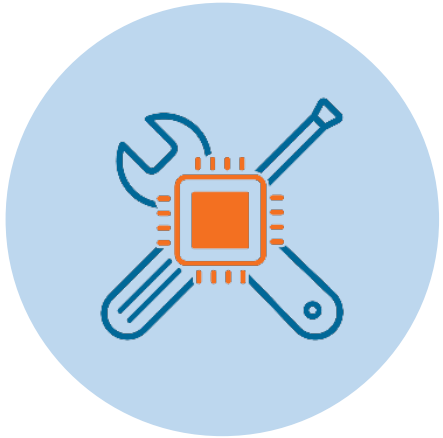
The server now has a model that captures the patterns in the training data on all the nodes. But at no point the nodes shared their training data.

Introducing Federated/Distributed Learning

Federated learning is typically used ...

- Bandwidth or power consumptions are a concern
- High-cost of data transfer
- Latency problems
- Smarter problems
- Privacy needs to be assured

Introducing Federated Learning: Examples



Predictive Maintenance /Industrial IOT

Healthcare (wearables, prognostics, etc)

Smartphones

The AI Translator: Ultimate Goal

Understanding fear

+

Installing trust

Thank you!

Sir Anthony Seldon

Author of “The Fourth Education Revolution: Will Artificial Intelligence liberate or infantilise humanity?

Co-founder of The Institute for Ethical AI in Education (UK)

THE UNIVERSITY OF BUCKINGHAM

Opportunities and Risks for AI in education

Sir Anthony Seldon
Vice-Chancellor

Wednesday 24th April 2019

What AI or 4.0 entails

- Artificial intelligence
- Machine learning
- Virtual reality
- Augmented reality
- Mixed reality
- Trans-humanism

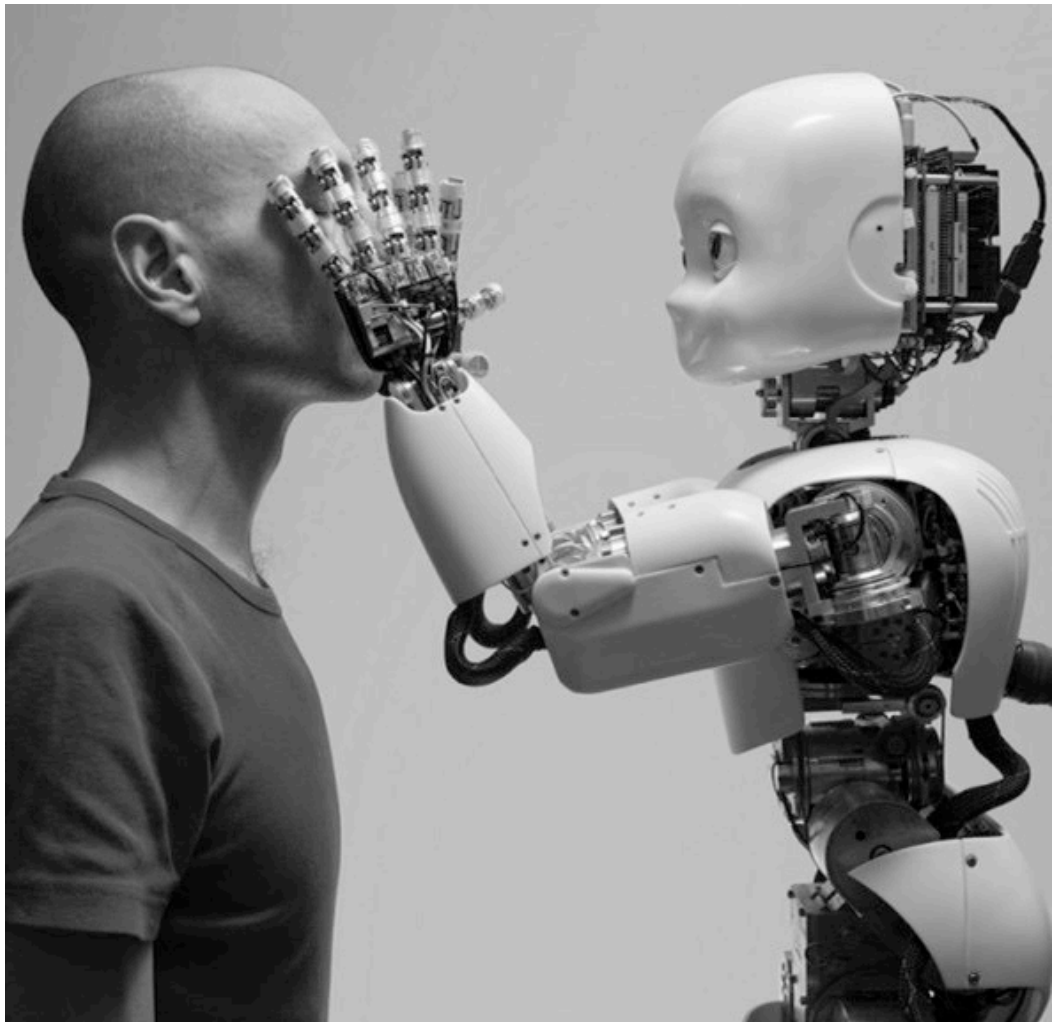
What 4.0 entails

- Robotics
- Voice and face recognition
- Quantum computing
- Collaborative learning
- Internet of Things
- Big data
- Blockchain

Our schools and universities are preparing students brilliantly – for the twentieth century

Benz factory in the 1890s

“AI is coming. To understand the stage we are with its arrival, we can draw an analogy from the car industry in 1886. Karl Benz had just invented the internal combustion engine. People had no idea how the invention would take off, or that it would transform human life across the planet. The comparison is wrong though in one respect. AI is far more wide-ranging than the car, and will carry humans much further.”



AI is infinitely seductive. It will know us better than our best friends, our parents, our partners. It probably already does. Under the guise of plausibility, is it opening our eyes, shielding our sight, or blinding us?

“AI will be 'either best or worst thing' for humanity”

“Every aspect of our lives will be transformed. In short, success in creating AI could be the biggest event in the history of our civilisation”

Steven Hawking

“Artificial intelligence is the biggest risk we face as a civilisation and needs to be checked as soon as possible”

Elon Musk

The First Revolution - The *Dawn* of Learning some five million years ago

The Second Revolution – *Organised* Learning
i.e. 5000 years ago, cities sprung up on four rivers

The Second Revolution:
The first schools and the first universities

The first University, Bologna, 1088

The Third 3.0 Revolution – The Printing Press

Mass Learning: 500 years ago

and mass education at the time of the Industrial Revolution

We are still living in the 3.0 education revolution model

Our schools are fundamentally the same as in 1600 – teachers at the front, children in rows, teacher exposition, writing on board at the front of the class, homework, exams, marking, grades, rankings, reports, teachers complaining about workload, parents complaining about heads…

Until…

The 4.0 Revolution – AI etc.

4.0 is already transforming

- Healthcare
- Transport
- Shopping
- Law firms
- Accountancy
- Agriculture
- Banking

Five enduring problems with the 3.0 factory/third education revolution model

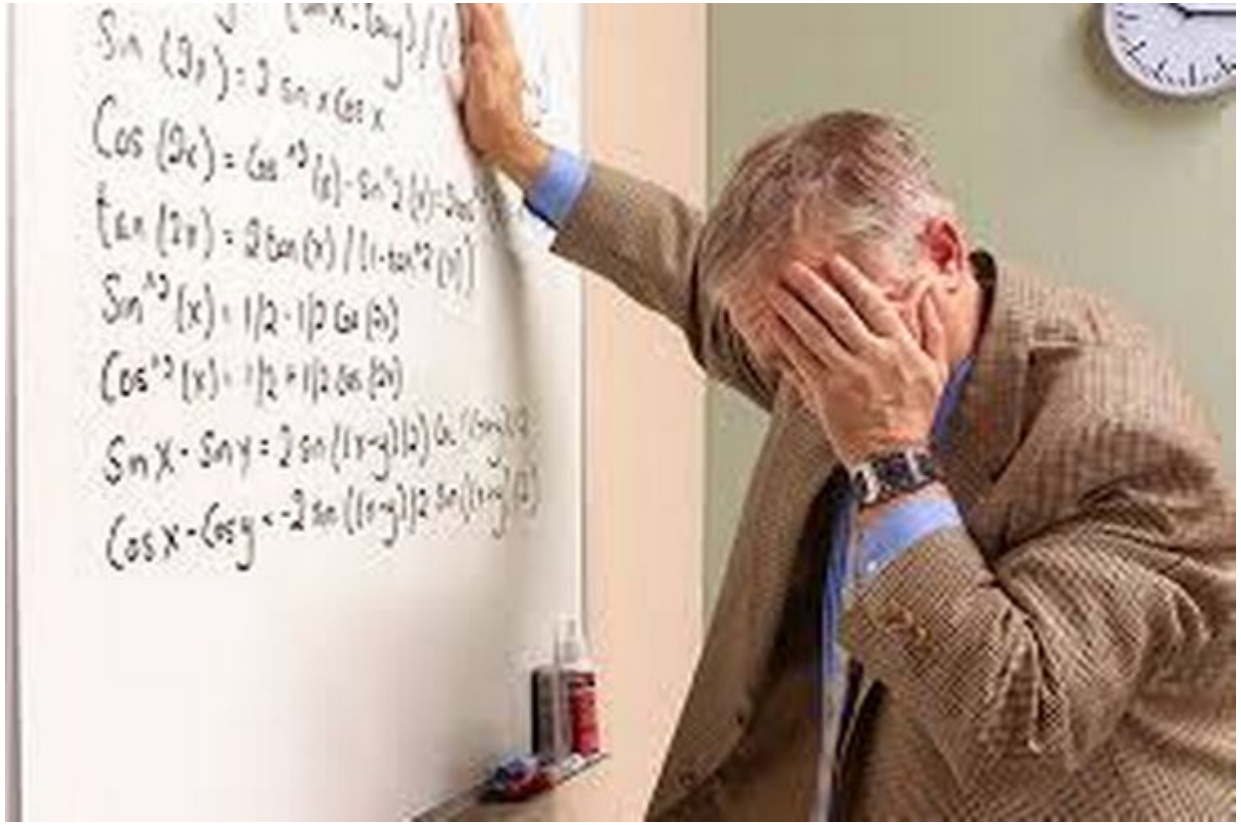
1. Failure to achieve social mobility

Solved by: An ‘Eton’ Teacher For All

2. One size doesn’t fit all – age not stage

Solved by: personalised learning – each child moves at own pace in each subject

3. Teaching drowned out by administration



Solved by: personalised marking, feedback, administration etc. by machines

4. Narrow focus on just cognitive ability

Solved by: learning across all *8* intelligences

5. Some advances in education already: Homogenisation not individuation

Solved by: learning individuated to each student, boosting their mental health and sense of self

4.0 will impact on all five stages of learning

1. *Selection* of appropriate material for teaching.
2. *Organisation* of the learning space and time available.
3. *Presentation* of the learning material to the student in a form to produce optimal learning/understanding.
4. Regular *assessment*, tests and exams, with presentation of real time, personalised, *feedback.*
5. Report writing and overall assessment of student performance and *suitability* for *next levels* of study or employment.

We are not even preparing our school students for the world of work.

- Oxford Martin School 2013
- David Deming - Harvard working paper 2015
- Richard and Daniel Susskind – *The future of the professions 2015*
- McKinsey Global Institute January 2017
- IPPR, Carys Roberts, 2017
- PriceWaterhouseCoopers March 2017
- Oxford Martin/Pearson/Nesta 2018
- Future of Work Centre 2019 – half life of a skill in 1995 was 26 years: now 4.5 years McKinsey – by 2030 workforce will spend 55% more time using technology. AI is advancing 10X faster and at 300X the scale of the Industrial Revolution.

4.0 Harbingers of the future

AltSchools

Summit public schools

School of One

Khan Lab School

Riverbend School, Chennai

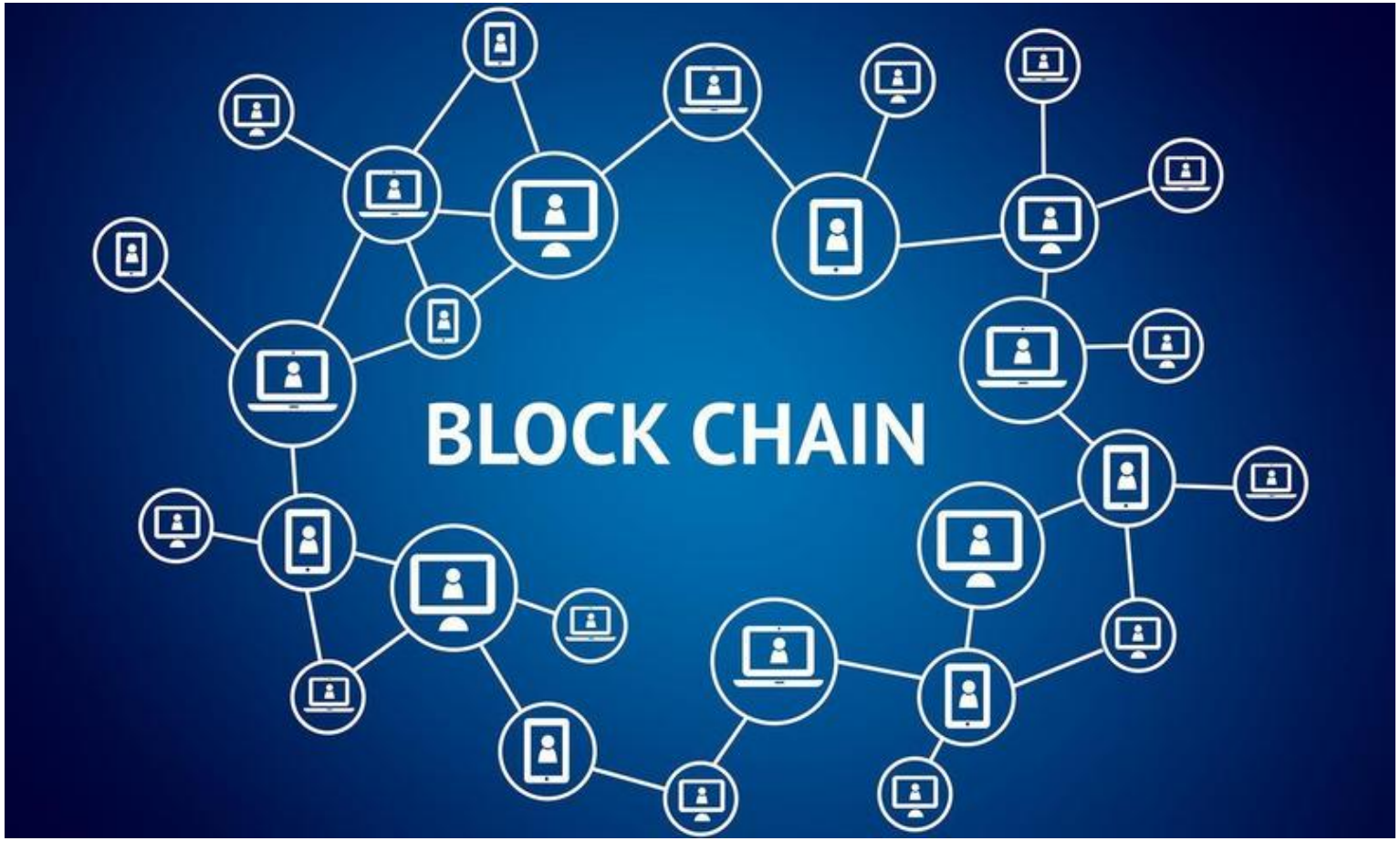
The five examples of 4.0 universities

1. The end of the lecture hall at the University of Northampton, UK

2. The ‘C-Campus’ or bilateral/trilateral degree, Australia/China

3. Virtual degrees or ‘nanodegrees’ from Udacity, USA

4. The Blockchain, the University of One and Woolf University, UK



5. No universities at all – Holland and everywhere

Risks of 4.0

1. Enhance social *immobility*

2. Education becomes *narrower*

3. Teachers lose *control* over students

4. Teachers become *deskilled/deprofessionalised*

5. Students become *infantilised* (like taxi drivers)

Risks of 4.0 (continued…)

6. Students become bored and *demotivated*

7. Students become *anti social* and live more egocentric lives

8. Mental health *damaged* – sense of reality and efficacy lost

9. Erosion of human/good *values*

10. Students *preyed on* by evil people/tech companies/governments.

Risks of 4.0 (continued…)

11. *Impersonation* by teachers/ youth leaders/ politicians;

12. *Incomprehension* gaps between young and old;

13. AI/4.0 develops *a life of its own*;

14. The real world appears *dull/stale* compared to the virtual world;

15. Erosion of *parenthood*: 4.0 machines will ‘know’ children better than any parent could.

AI/4.0 poses profound ethical questions for young people

Children used to be safe in their home and in their bedrooms

The 3.0 revolution made them much more vulnerable- their bedroom now became a hut in a wasteland

The 4.0 revolution strips away the hut. Now the young person is totally exposed

We lost the race against 3.0
We cannot lose against 4.0

What will the Institute for Ethical AI in Education do?

1. *Identify* the existing forms of governance, ethical principles, guidelines, standards and regulations relevant ethical AI in education;
2. Produce a *framework* for Ethical governance for AI in education for the UK;
3. Produce a *roadmap* for the development of inclusive, responsible, explainable, interpretable, verifiable and agile ethical governance for AI in Education that will protect people from disadvantage, ill, harm;
4. Build *public knowledge* and appropriately critical trust in AI in education through public engagement;
5. *Demand more* from our large technology companies in terms of ethical practice and ethical education and training for educators, trainers, parents and students;

What will the institute for Ethical AI in Education do? (continued…)

6. Demand support for our *Start-up and SME* technology community to ensure their ethical practice;

7. Demand *ethics training* for everyone involved in education or training directly or indirectly;

8. Ensure that ethical AI in education *minimises extra burdens* to educators, learners, and parents beyond their needing to understand what is required for them to protect themselves, their students, their employees, or their family from;

9. Publish an *ethical code of conduct* for those working to develop and use AI for educational and training purposes;

10. Provide *ethical training and approval protocols* for anyone developing or using AI in education and training to encourage ethical transparency and publicly ethical practice;

Conclusion:

We need AI machines to teach our students to become more fully *human* - the education system currently deploys humans to teach our young to become more like *machines*.

DEBATE

Moderator: Sven Mastbooms