ISO / IEC AI Workshop Series

Inaugural Workshop

Wael William Diab
ISO/IEC JTC 1/SC 42 Chair
ISO/IEC AI Workshop Chair

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

Welcome

Thank you for participating!

Participation Guidelines

Please use the Q&A function for questions

Please *do not use the chat function* for questions

This webinar is being recorded and will made available

Please keep your video off and audio muted unless speaking

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

Acknowledgments

Program Committee

- Heather Benko
- Norbert Bensalem
- Andrew Dryden
- Peter Deussen
- Rohit Israni
- Mike Mullane
- Catherine Nelson
- Wael William Diab

ISO

- Elisabeth Gasiorowski
- Catherine Infante
- Vivienne Rojas

IEC

- Stephen Dutnall
- Gabriella Ehrlich
- Ian Gardner
- Giulia Pizzi
- Laura Rocha

Program

SESSION 1

(24th May 13:00 – 16:00 UTC)

- Opening Remarks
- AI Applications
- Novel AI Standardization

SESSION 2

(25th May 05:00 – 08:00 UTC)

- Opening Remarks
- Emerging AI Requirements
- Emerging AI Technology Trends

SESSION 3

(25th May 22:00 – 26th 01:00 UTC)

- AI Applications
- Novel AI Standardization
- Closing Remarks

Session 1

Program	Talk Title	Track Chairs	Aprox Start Time (UTC)
Kickoff			13:00
Wael William Diab			
Opening Remarks			13:05
Gilles Thonet			
Phil Wennblom			
AI Applications		Rohit Israni Catherine Nelson	13:15
Mike Glickman	Health informatics and AI, the road to Interoperability		
Neil Frost	AI potential Applications in Transport		
Christophe Preube	AI finds awareness in standardization work – the White paper of ISO/TMB SMCC		
Novel AI Standardization Approaches		Peter Deussen Norbert Bensalem	14:25
Kimberly Lucy	Creating Trust in AI Through Standards: A Management System Approach		
Wo Chang	Data Quality for Analytics and Machine Learning		
Tim McGarr / Florian Ostmann	AI Standards Hub (UK)		
Viveka Bonde	Novel Standerdization Approach to AI Ethics		

Session 2

Program	Talk Title	Track Chairs	Aprox Start Time (UTC)
Kickoff		N/A	5:00
Wael William Diab			
Opening Remarks		N/A	5:05
Silvio Dulinsky			
Emerging AI Requirements		Catherine Nelson and Peter Deussen	5:10
Elham Tabassi	NIST AI Risk Management Framework		
Mariagrazia Squicciarini	The UNESCO Recommendation on the Ethics of AI - Setting the standards for a better and more inclusive future		
Liz Coll	Can standards deliver consumer trust and confidence in AI		
Daniel Loevenich	Evaluation Standards for Conformity Assessment of Trustworthy Cloud-based AI Applications		
Emerging AI Technology Trends		Norbert Bensalem and Rohit Israni	6:30
Shubhashis Sengupta	Emerging AI Trends – as seen by an Industry Practitioner		
Tilak Kasturi	How specialized AI drives value for Automotive Service Organizations		
William Uppington	AI Quality: The Next Big Challenge in AI		
Babak Hodjat	From Data to Decisions, and Back		

Session 3

Program	Talk Title	Track Chairs	Aprox Start Time (UTC)
Kickoff			22:00
Wael William Diab			
AI Applications		Rohit Israni and Catherine Nelson	22:05
Charalambos Freed	Safety considerations in autonomous products		
Hajime Yamada	Development of SRD 63416 “Ethical Considerations of AI when Applied in the AAL Context”		
Kenzo Nonami	AI Powered UAV and Future Prospects		
Fumihiro Maruyama	Use cases and AI application guidelines in international standardization		
Novel AI Standardization Approaches		Norbert Bensalem and Peter Deussen	23:30
Jochen Friedrich	AI Standardisation supporting regulatory needs		
Linzhing Meng / Mike Thieme	Developing a TS on assessment of machine learning classification performance		
Yonosuke Harada	Introduction of Governance Implications of AI Systems		
Paul Cotton / Milan Patel / Wei Wei	The foundational standards for AI – ISO/IEC 22989 and ISO/IEC 23053		
Closing Remarks		Wael William Diab	0:55
Wael William Diab	Overview and introduction of SC 42		
Program Committee	Insights from the workshop		

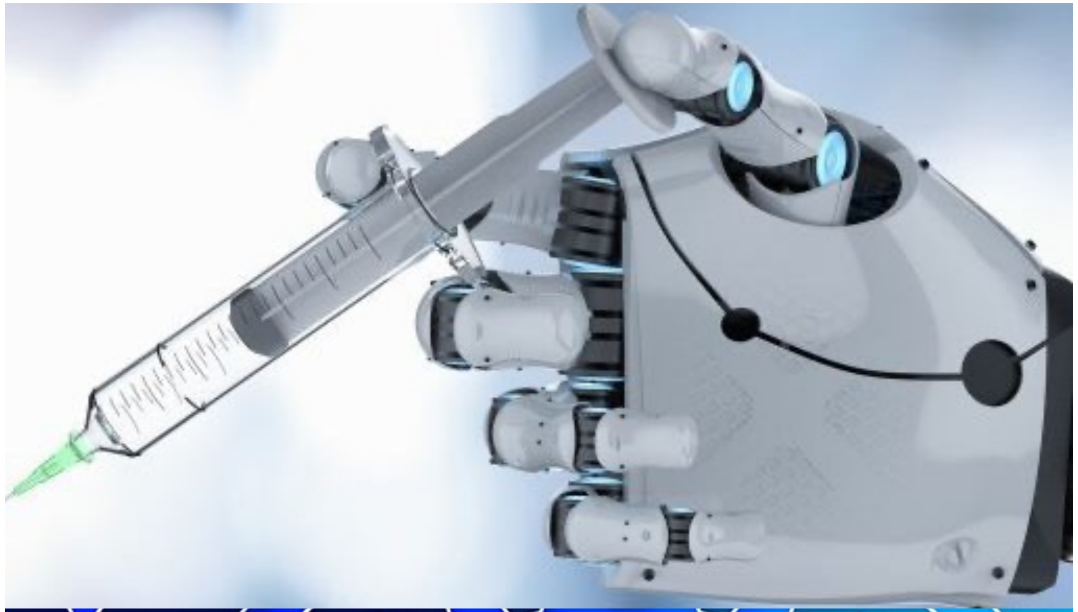
Thank you

Wael William Diab

wael.diab@gmail.com; Wael’s LinkedIn

Workshop Website

ISO/IEC AI LinkedIn Page



ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

NIST AI Risk Management Framework

NIST is actively engaging stakeholders to develop a risk management framework to map, measure, manage, and govern risks associated with AI technologies.

NIST AI Program

CONDUCT FOUNDATIONAL RESEARCH TO ADVANCE TRUSTWORTHY AI TECHNOLOGIES

ADVANCE AI RESEARCH AND INNOVATION ACROSS THE NIST LABORATORY PROGRAMS

ESTABLISH BENCHMARKS AND DEVELOP METRICS TO EVALUATE AI TECHNOLOGIES

PARTICIPATE AND LEAD IN THE DEVELOPMENT OF STANDARDS TO ADVANCE AI INNOVATION

CONTRIBUTE NIST'S TECHNICAL EXPERTISE TO DISCUSSIONS AND DEVELOPMENT OF POLICIES

ENSURE THAT NIST HAS RESOURCES AND EXPERTISE TO CARRY OUT ITS AI PROGRAMS

Open and public data

NIST Special Database 19 NIST Handprinted Forms and Characters Database

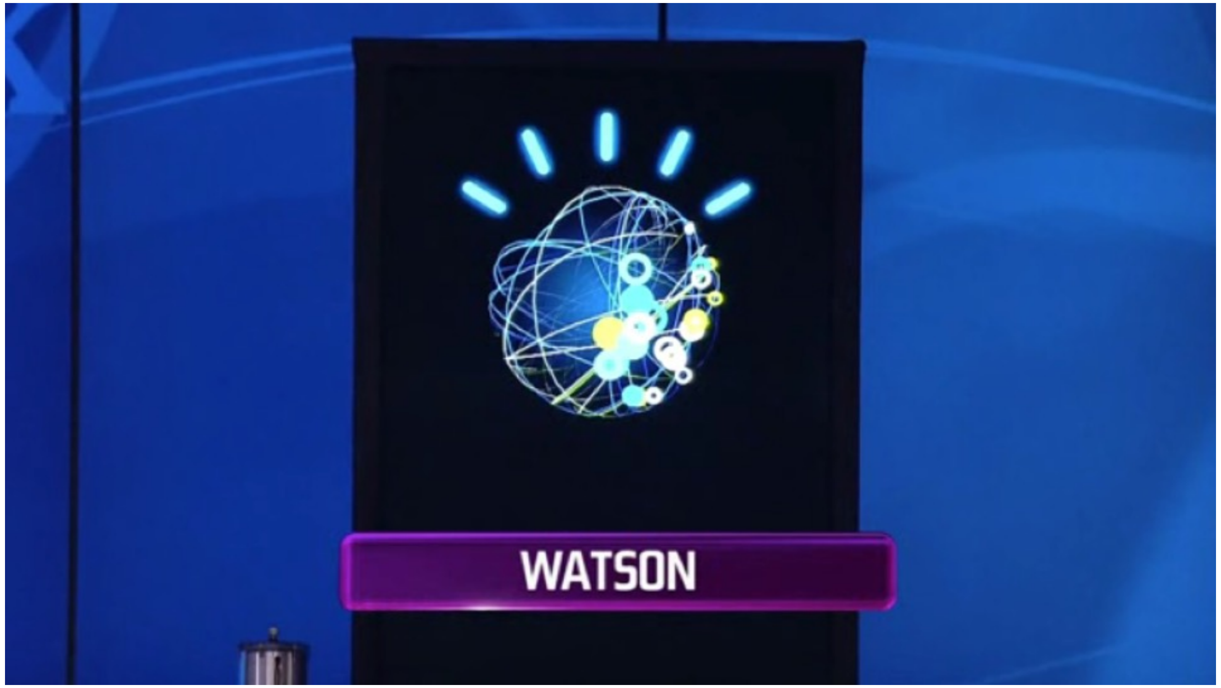
www.nist.gov/srd/nist-special-database-19

"Yet another advice: don't get fooled by people who claim to have a solution to Artificial General Intelligence. Ask them what error rate they get on MNIST or ImageNet."

Yann LeCun, 2014.
Director of AI Research, Facebook
and NYU professor

www.reddit.com/r/MachineLearning/comments/25lnbt/ama_yann_lecun/

Advance scientific disciplines via evaluations



"Our prior QA work took shape in the form of a QA system called PIQUANT. PIQUANT development started in 1999, predating our work in UIMA, and was funded by government research grants and tested against NIST evaluation data in the Text REtrieval Conference (TREC) QA track between 1999 and 2005."

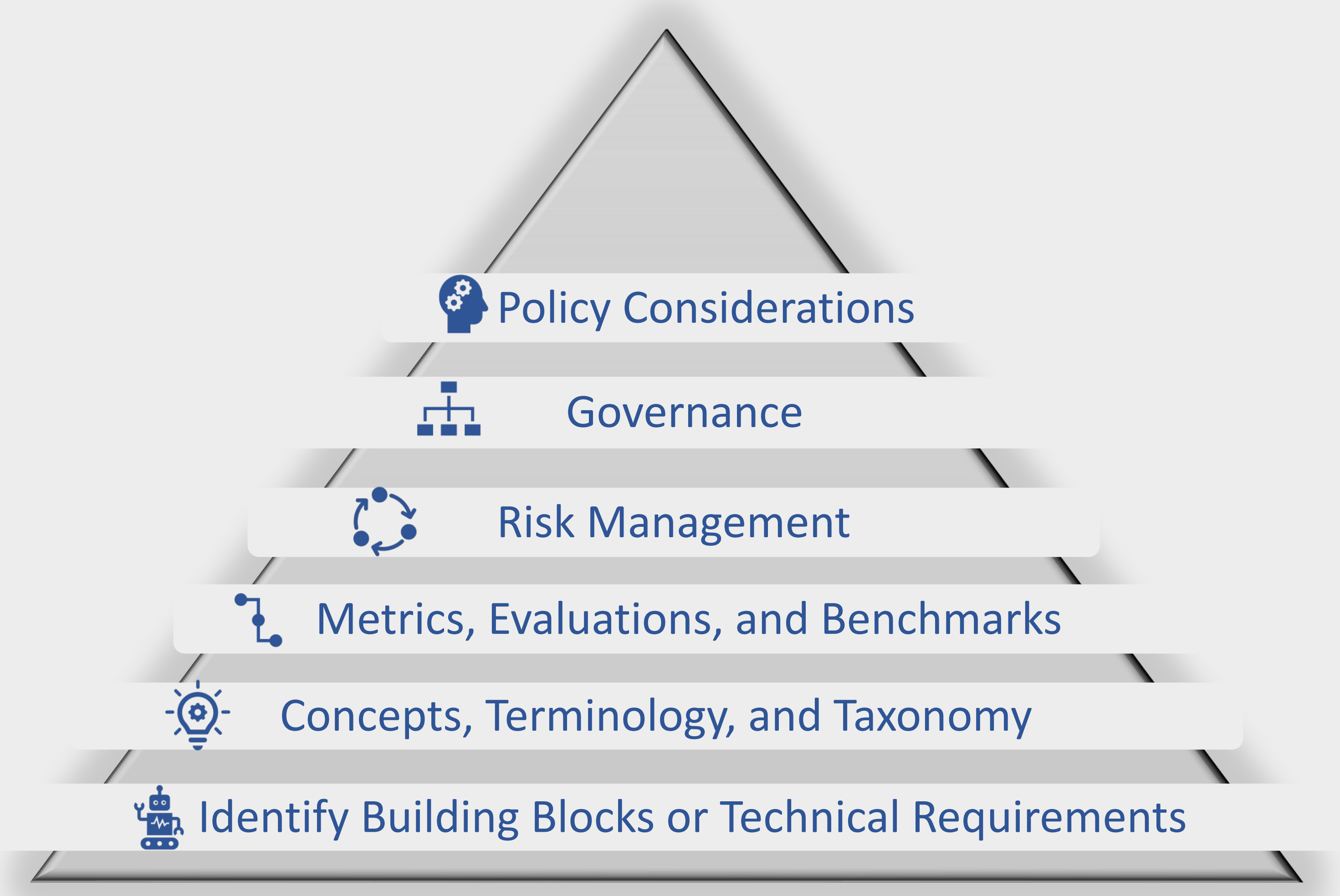
Introduction to "This is Watson". IBM Journal of Research and Development (Volume: 56, Issue: 3.4, May-June 2012)

Trustworthy and Responsible AI @NIST

Cultivate trust in the design, development, use and governance of artificial intelligence technologies and systems.

- Development of AI Risk Management
- Establishing National AI Advisory Committee
- AI Standards and Evaluation

Trustworthy AI's Foundation
From Technical Requirements to Policy Creation



	Technical Design Characteristics	Socio-Technical Characteristics	Guiding Principles Contributing to Trustworthiness
AI RMF Taxonomy	• Accuracy • Reliability • Robustness • Resilience or ML Security	• Explainability • Interpretability • Privacy • Safety • Managing Bias	• Fairness • Accountability • Transparency
OECD	• Robustness • Security	• Safety • Explainability	• Traceability to human values • Transparency and responsible disclosure • Accountability
EU AI Act	• Technical robustness	• Safety • Privacy • Non-discrimination	• Human agency and oversight • Data governance • Transparency • Diversity and fairness • Environmental & societal well-being • Accountability
EO 13960	• Purposeful and performance-driven • Accurate, reliable, and effective • Secure and resilient	• Safe • Understandable by subject matter experts, users, and others, as appropriate	• Lawful and respectful of our Nation's values • Responsible and traceable • Regularly monitored • Transparent • Accountable

Trustworthy AI: Risks & Characteristics

Technical

- Accuracy
- Reliability
- Robustness
- Resilience or Security

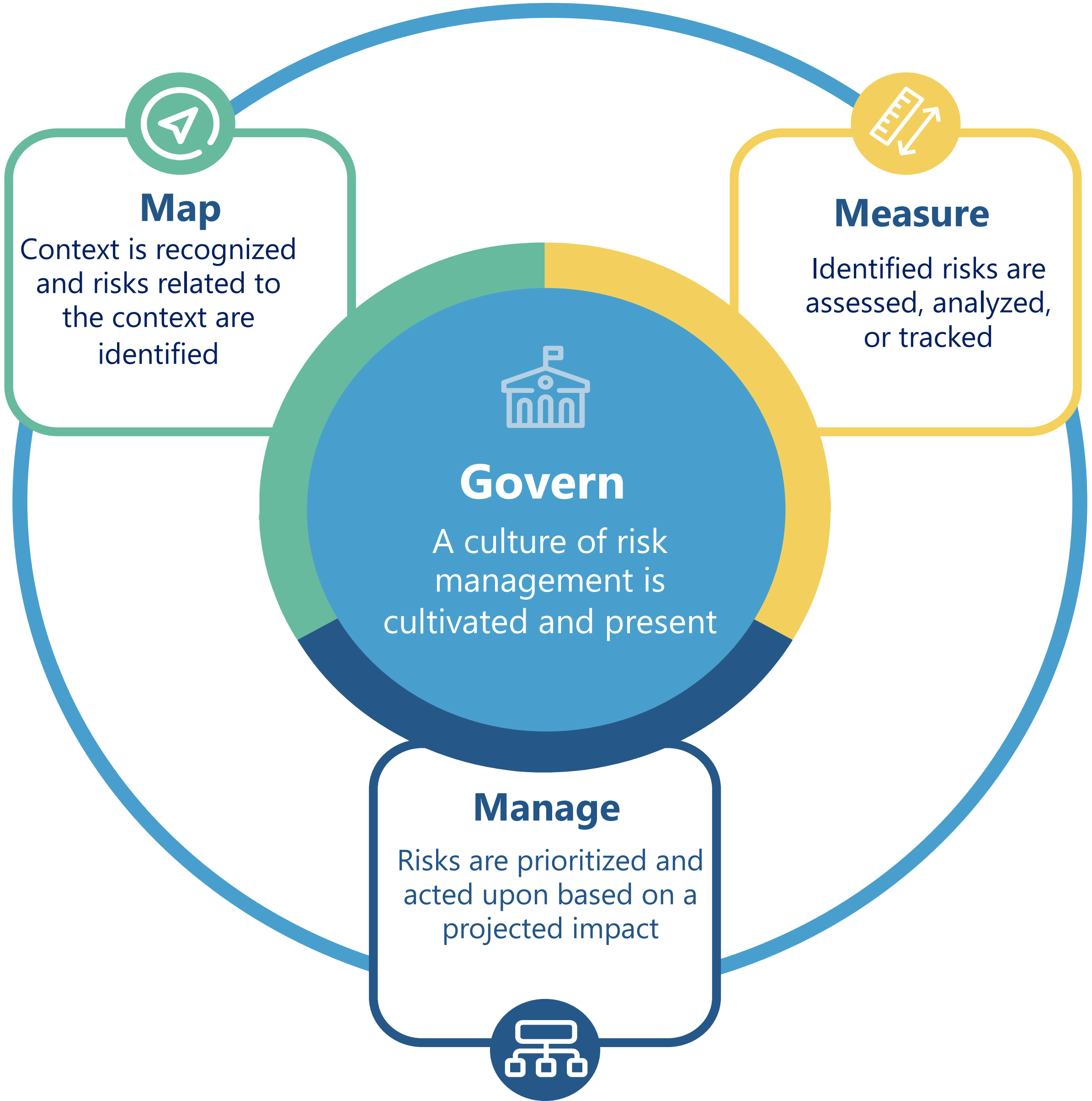
Socio Technical

- Explainability
- Interpretability
- Privacy
- Safety
- Managing Bias

Guiding Principles

- Fairness
- Accountability
- Transparency

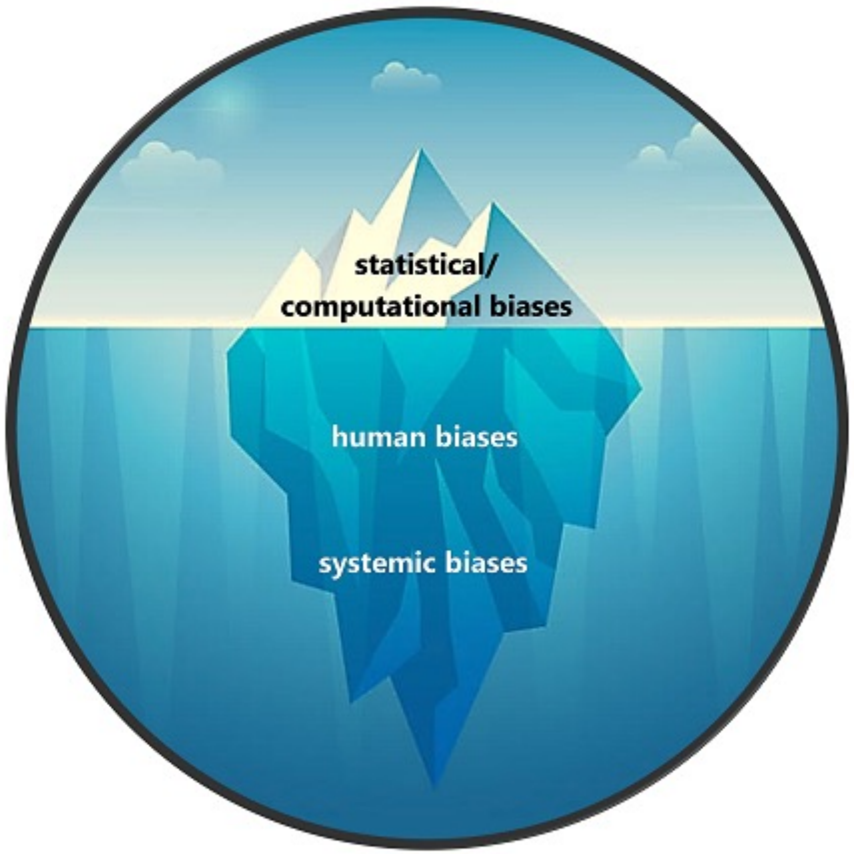
AI Risk Management Framework



AI RMF Stakeholder Engagement Timeline

Stakeholder Engagements

Initial RFI	Kickoff Workshop	Concept Paper	1st Draft Workshop 2	2nd Draft Workshop 3	AI RMF 1.0 Published
July-September 2021	**October 2021**	**December 2021**	**March 2022**	**Summer/Fall 2022**	**Early 2023**

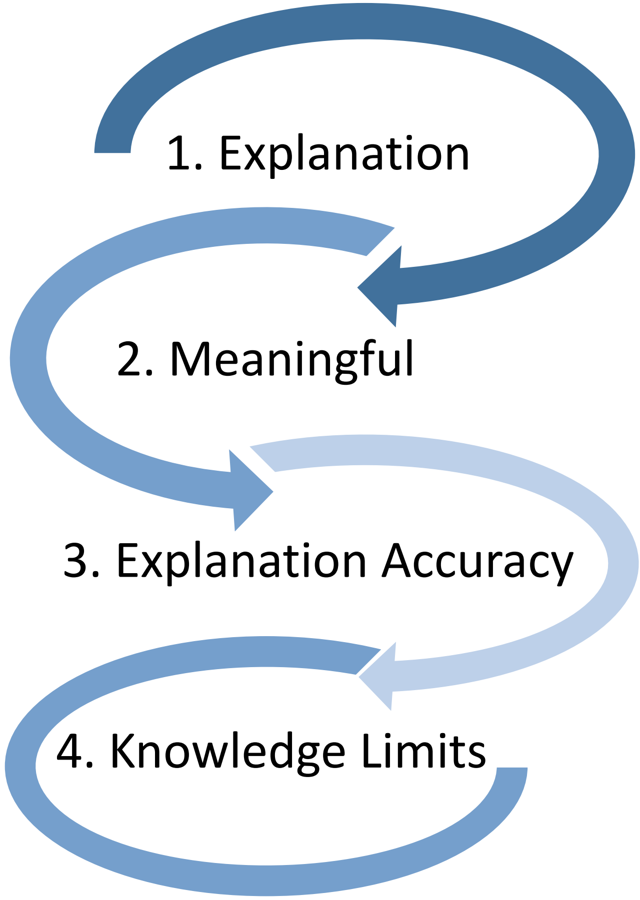


The Challenge of AI Bias

Special Publication 1270 - Towards a Standard for Identifying and Managing Bias in Artificial Intelligence

Interpretable and Explainable AI

Four Principles of Explainable AI

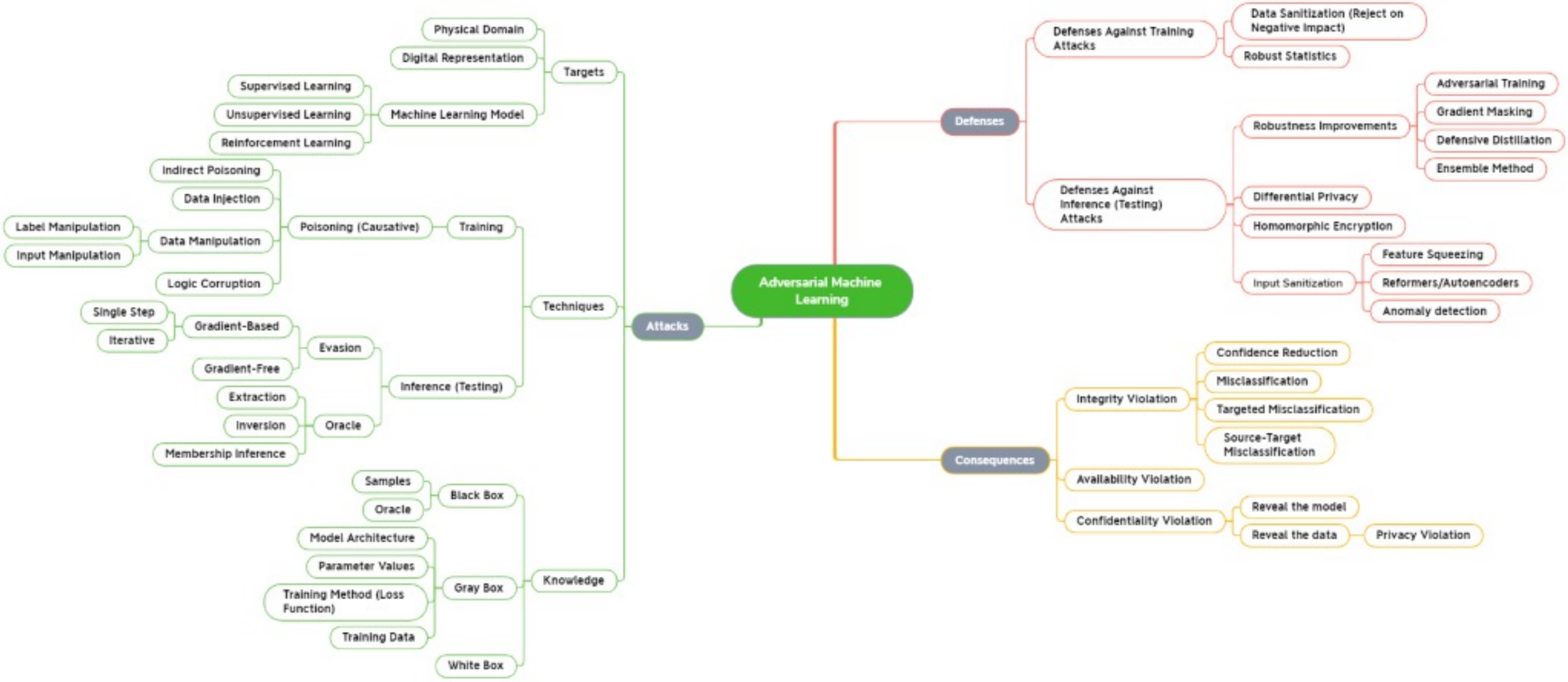


NISTIR 8312: Four Principles of Explainable Artificial Intelligence

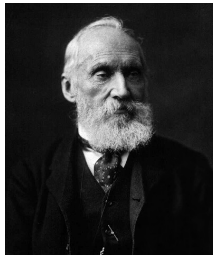
NISTIR 8367: Psychological Foundations of Explainability and Interpretability in AI

Secure AI

A Taxonomy and Terminology of Adversarial Attacks and Defenses on Machine Learning



AI Evaluations



“If you cannot measure it, you cannot improve it.”

Devise and revise metrics

Testbeds and test data

Assessing real world performance

Interoperable tests and test results

USG AI Standards Coordinator

Outreach to connect with all known federal efforts relating to AI standards development and use with the goal of community participants leveraging and learning from the successes of other participants.

Facilitate ongoing discussions between the U.S. private sector and federal agencies to strengthen private-public sector coordinator

From Innovation to Adoption

Fundamental Research

Applied Research

Standards + Best Practice Guides

Adoption

Image Credit: wsj.com

THANK YOU

Contact us via email at
aiframework@nist.gov

For more info on the NIST AI RMF, visit
https://www.nist.gov/itl/ai-risk-management-framework

17

The UNESCO Recommendation on the Ethics of AI

Setting the standards for a better and more inclusive future

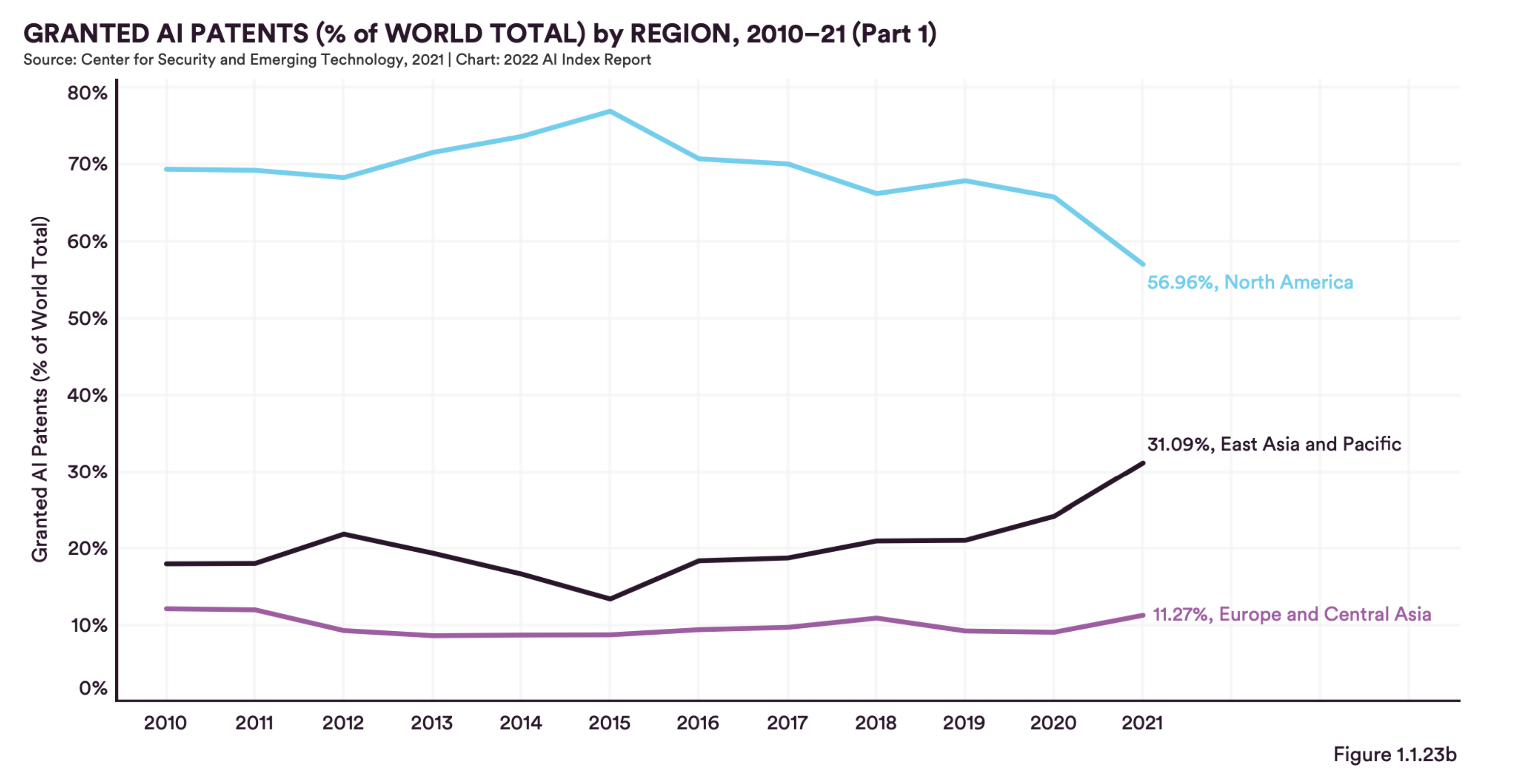
Mariagrazia Squicciarini, PhD

Chief of Executive Office
UNESCO Social and Human Sciences Sector

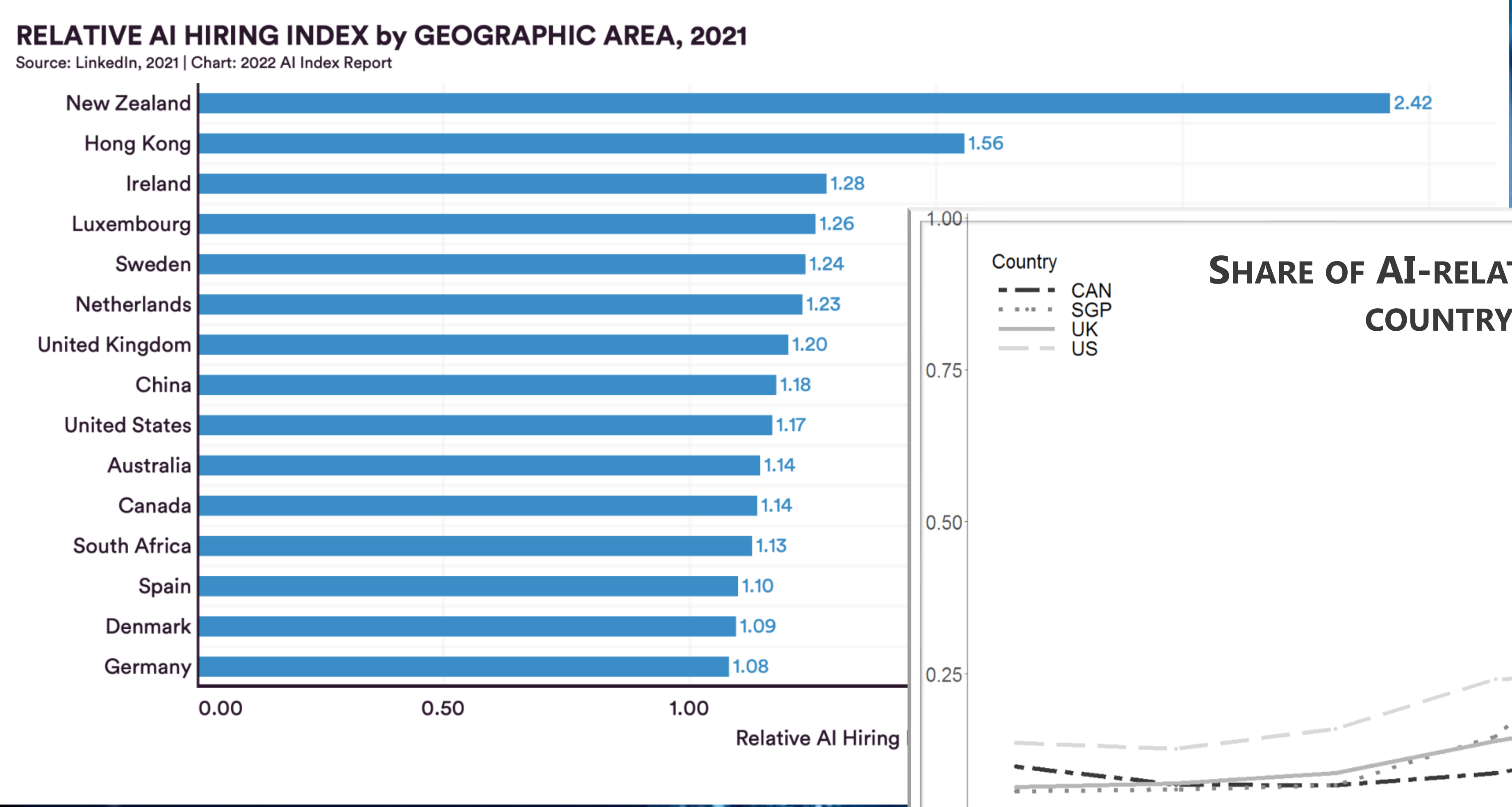
25 May 2022

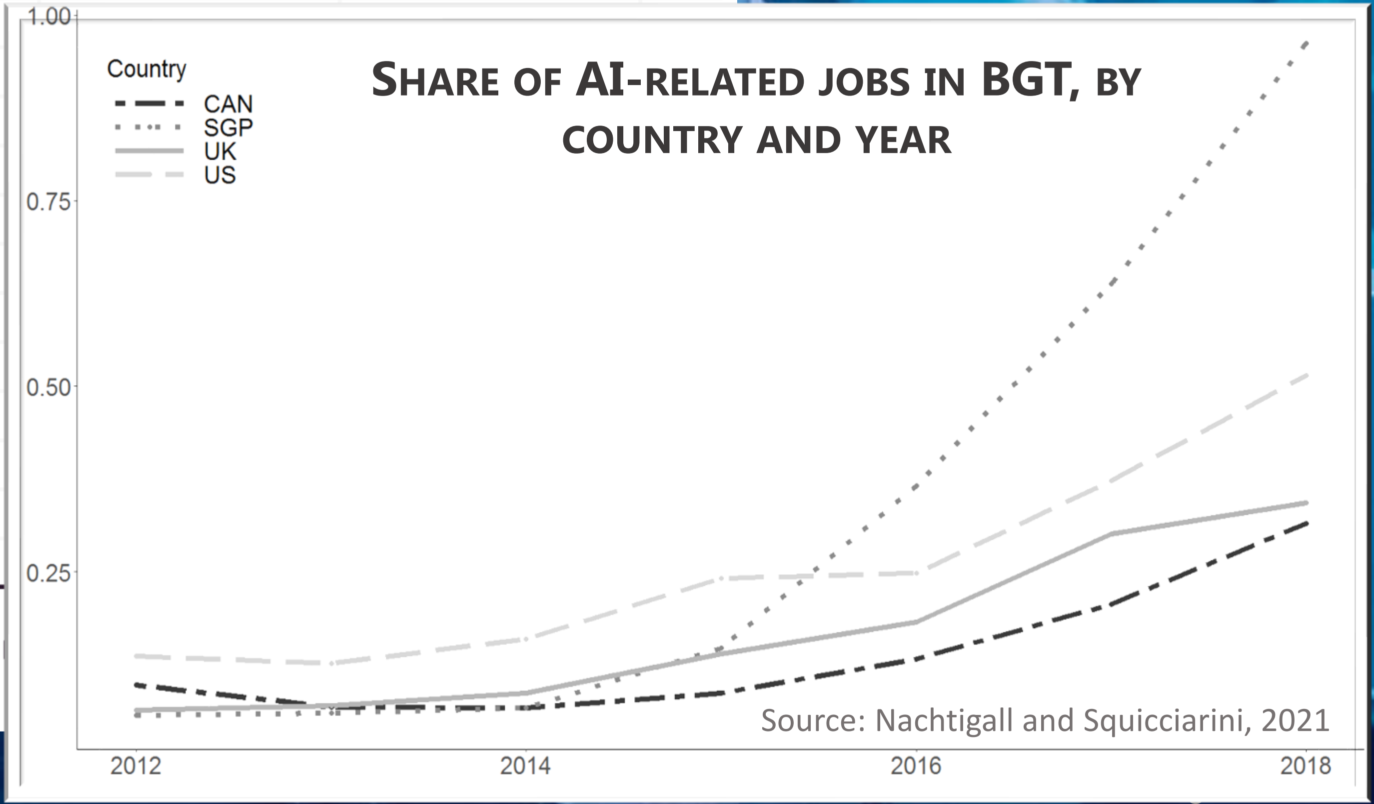
AI IS IN THE HANDS OF FEW

AI research and development is highly concentrated. The large majority AI patents are filed in and granted to groups in the U.S. and China.

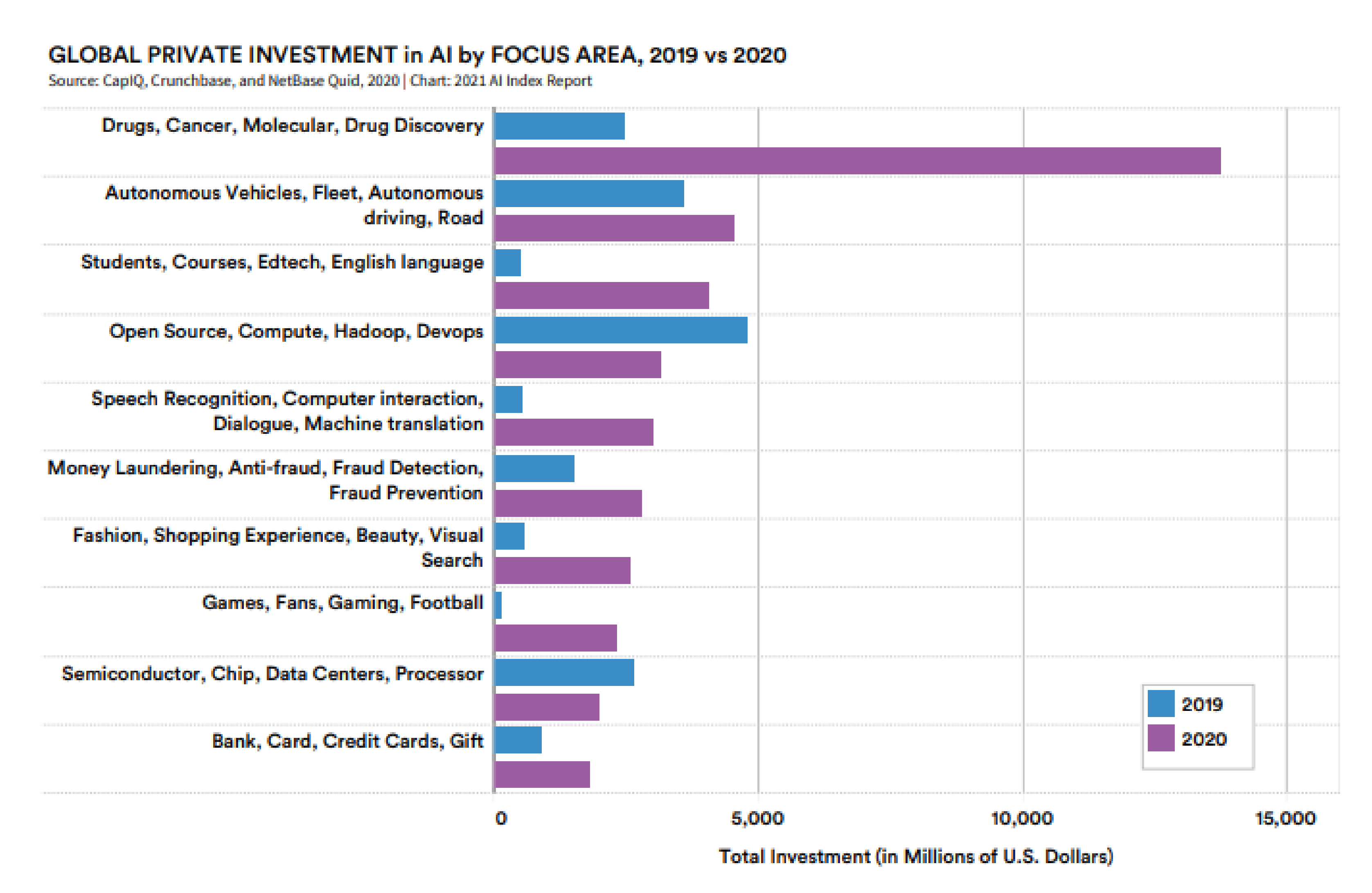


AI HIRING IS GROWING





… AS IS PRIVATE INVESTMENT



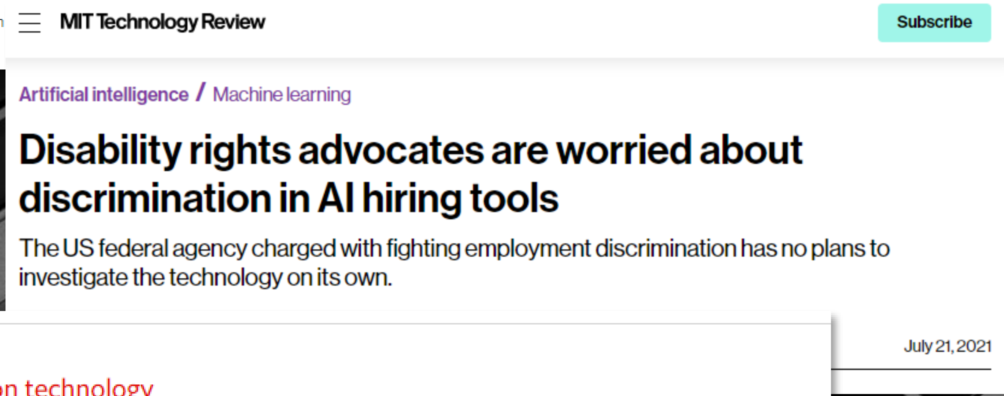
AI: FAST FORWARD TOGETHER

More and more public and private decisions are taken by artificial intelligence, and the outcomes are not free from biases or badly-defined frameworks.

Privacy, surveillance, and serious human rights concerns.


The rise of decision intelligence: AI that optimizes decision-making

By Sajan Kuttappa | 6 m


MIT Technology Review Subscribe

Artificial intelligence / Machine learning

Disability rights advocates are worried about discrimination in AI hiring tools

The US federal agency charged with fighting employment discrimination has no plans to investigate the technology on its own.

July 21, 2021


Face-recognition technology

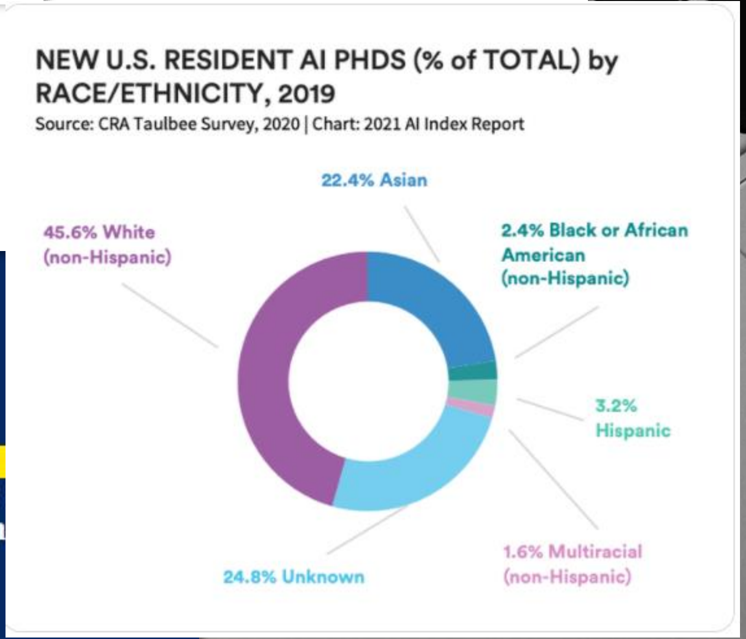
Computer programs recognise white men better than black women


euronews.

WORLD

Social scoring: Could that Facebook post stop you getting a loan or a mortgage? COMMENTS

By Euronews with Associated Press • 25/06/2020



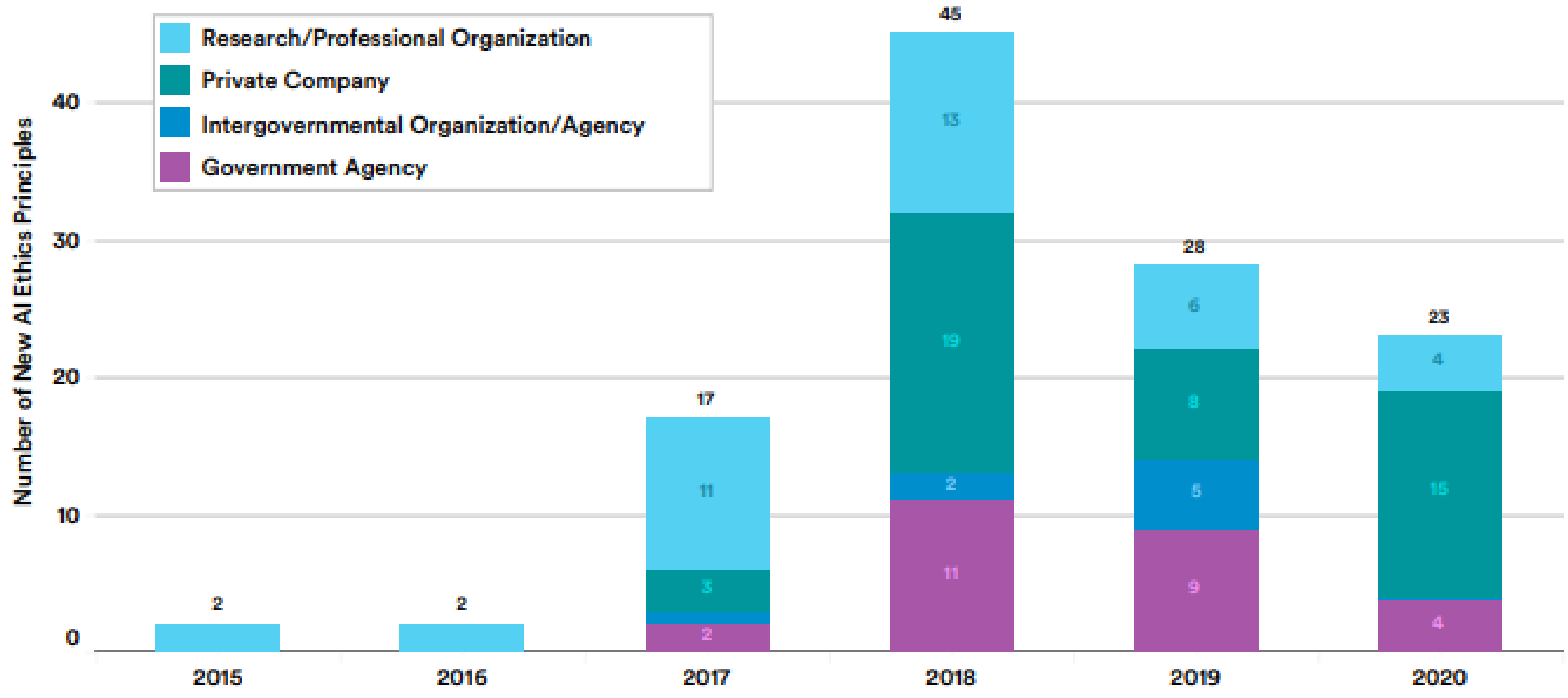
DOWNSIDES OF SELF-REGULATION

- Human Rights Violations
- Privacy Issues
- Diversity Inclusion
- Bias

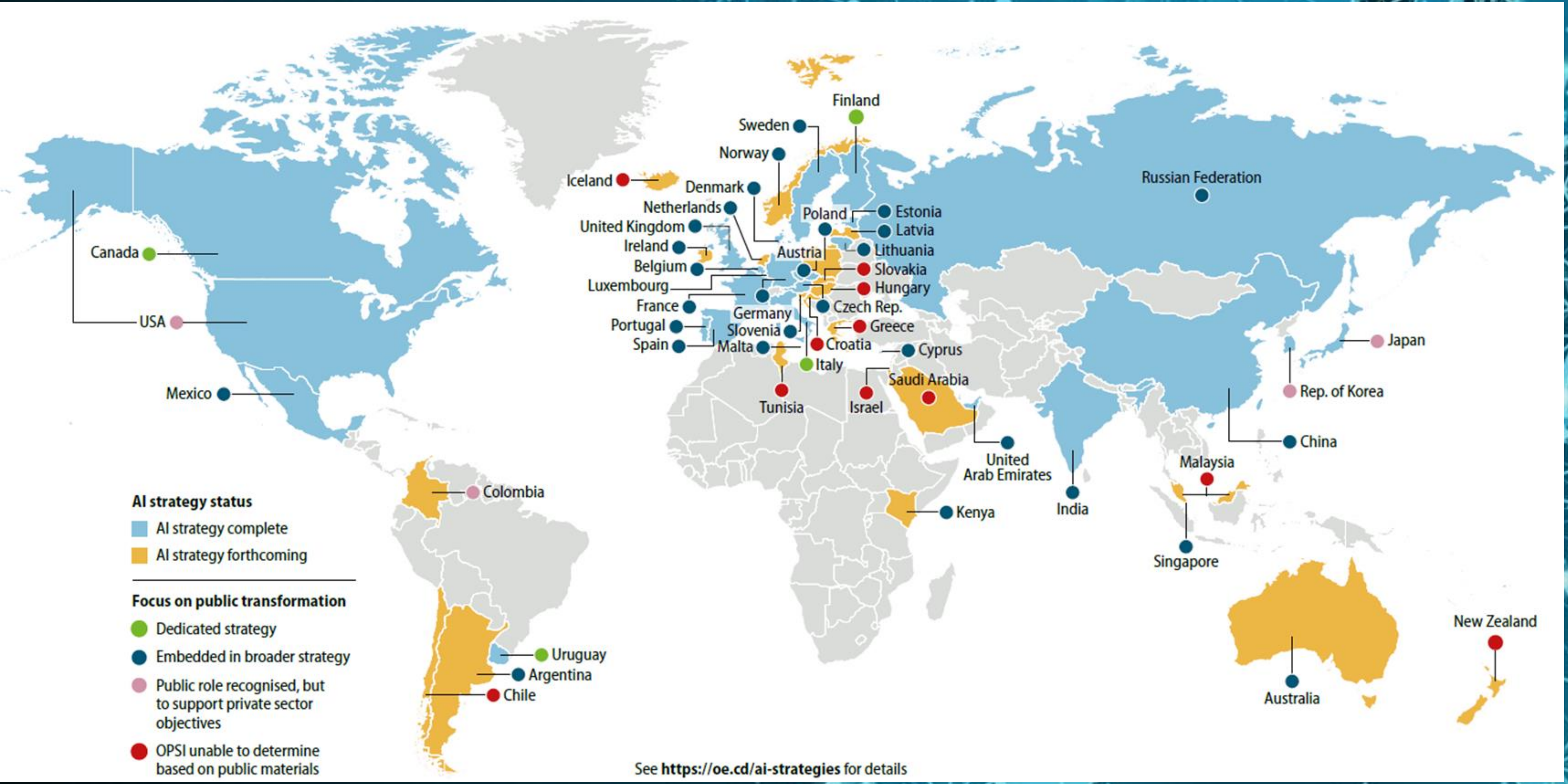
and more...

LACK OF A GLOBAL STANDARD AND BENCHMARKS

NUMBER of NEW AI ETHICS PRINCIPLES by ORGANIZATION TYPE, 2015-20
Source: AI Ethics Lab, 2020 | Chart: 2021 AI Index Report



LACK OF A GLOBAL STANDARD AND BENCHMARKS



Recommendation on the Ethics of AI adopted by acclamation
by 193 Member States of UNESCO, 23 November 2021

THE FRAMEWORK

The Recommendation is based on interconnected values and principles

Values

1. Respect, protection and promotion of human rights and fundamental freedoms and human dignity
2. Environment and ecosystem flourishing
3. Ensuring diversity and inclusiveness
4. Living in peaceful, just and interconnected societies

Principles

1. Proportionality and do no harm
2. Safety and security
3. Fairness and non-discrimination
4. Sustainability
5. Right to privacy, and data protection
6. Human oversight and determination
7. Transparency and explainability
8. Responsibility and accountability
9. Awareness and literacy
10. Multi-stakeholder and adaptive governance and collaboration

THE FRAMEWORK

Principles were there for many years, but the Recommendation goes into details of actions to ensure accountability, responsibility, transparency and necessary regulations to ensure the rule of law.

The Recommendation includes several action-oriented cross-sectoral policy chapters covering:

- **Data policy**
- **Gender**
- **Development & international cooperation**
- **Environment & ecosystems**
- **Health and social well-being**
- **Communication & information**
- **Education & research**
- **Economy & Labour**
- **Culture**

DATA

92% of businesses believe that high-quality data is the fuel for digital transformation

75% of businesses say that poor quality data has made it challenging to achieve their digital transformation plans

Only 15% of respondents believe that their existing systems are capable of producing clean data that can be trusted

Source: 2018 Study by the Supply Chain Resource Cooperative at North Carolina State University and IBM Watson

- DATA GOVERNANCE STRATEGIES
- RIGHT TO PRIVACY
- QUALITY DATASETS
- OPEN DATA

GENDER

300 millions fewer women than men access mobile internet	**25 %** of all tech jobs are held by women
Women are **TWICE** as likely as men to have lost their jobs due to the pandemic	**3 %** of female students in higher education choose ICT courses

GENDER

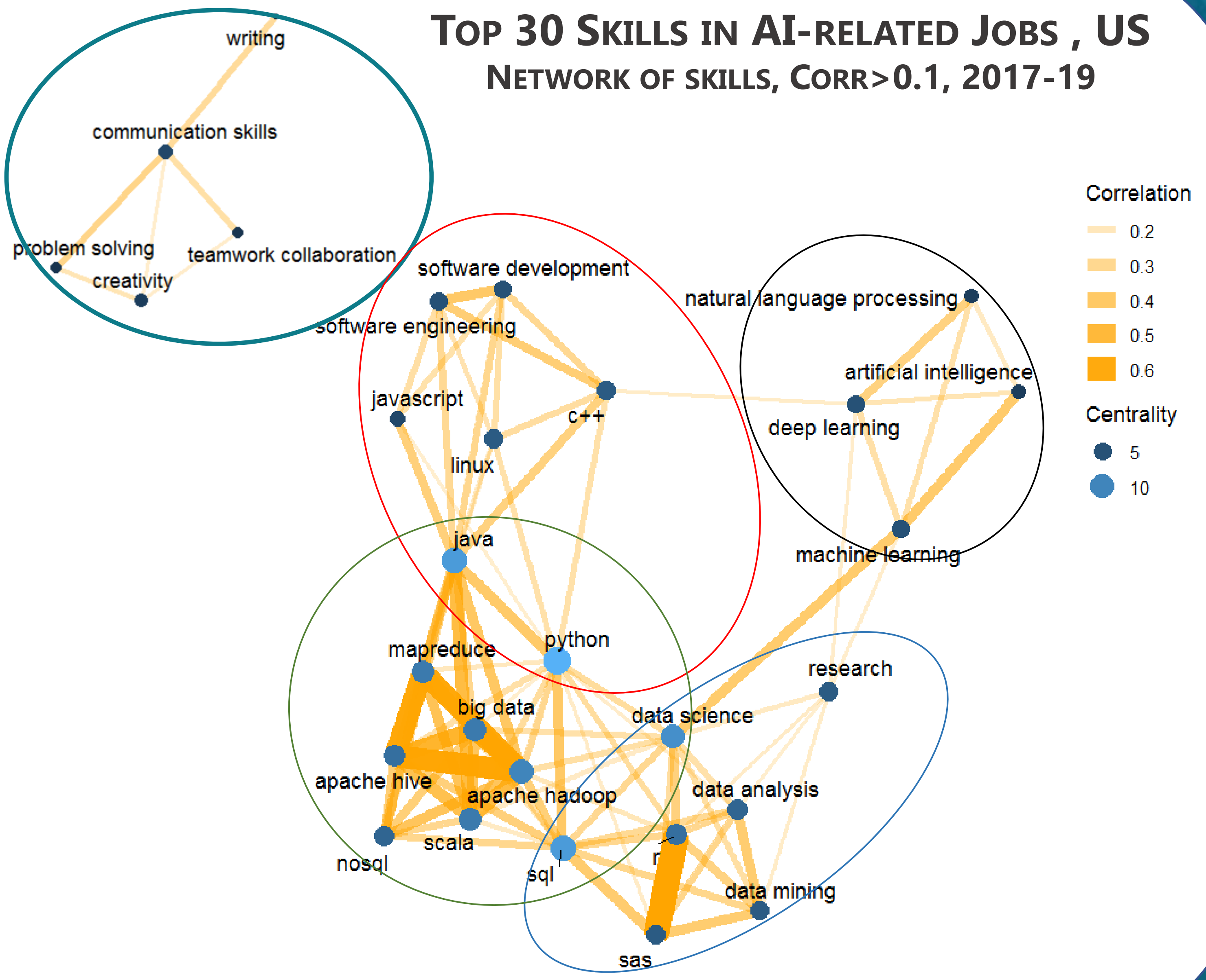
- FINANCING GENDER RELATED SCHEMES
- INCLUDING GENDER ACTION PLAN
- ADDRESSING WAGE AND EQUAL OPPORTUNITIES GAP
- ENCOURAGING FEMALE ENGAGEMENT IN AI
- INCREASING FEMALE PARTICIPATION IN STEM AND ICT
- ERADICATING GENDER STEREOTYPING AND BIAIS

EDUCATION AND SKILLS

- Top 30 skills appearing in AI-related jobs tend to be rather "general"
- Many of them relate to the "tools of the job", i.e. are software-related (e.g. MATLAB, R, SAS)
- Some technical skill bundles emerge
- A number of additional cognitive and socio-emotional skills are also required

Top 30 Skills in AI-related Jobs, US

Network of skills, Corr>0.1, 2017-19



Source: Samek, Squicciarini and Cammeraat (2021)

Assisting Member States to Implement the Recommendation

Ethical Impact Assessment (EIA)

Readiness Assessment Methodology

Building Institutional Capacity

Development of knowledge products

Assisting Member States to Implement the Recommendation

Observatory of Ethical AI

Global Forum on Ethics of AI

Women4EthicalAI

AI Experts without borders

Analytical works

Thank you!

m.squicciarini@unesco.org

More on the UNESCO Recommendation on the Ethics of AI:
https://en.unesco.org/artificial-intelligence/ethics

bsi.

…making excellence a habit.™

Emerging AI requirements

How can standards deliver consumer trust and confidence in AI?

Liz Coll, BSI Consumer and Public Interest Network Steering Group, and UK representative to the ISO COPOLCO AI task group.

Bsigroup.com/consumers

By Royal Charter

Outline

- Importance of standards for consumers
- Consumer stakeholder engagement in standards
- ISO COPOLCO

Importance of standards for consumers

Importance of standards for consumers

- Consumers **require protection** from: unsafe products, poor service, unfair treatment etc
- Standards are part of the '**consumer protection toolkit**' along with: legislation, regulation, industry codes, enforcement, information, advice
- **Minimize** deliberate or accidental detriment – could be physical, financial, psychological, wider economic/ environmental impact
- **Maximise** consumer rights: information, education, redress, fair choice, safety, healthy environment, privacy and security, access, inclusivity

Meeting consumer protection principles **builds trust and confidence** in products and providers

bsi.

Consumer stakeholder engagement in standards

bsi.

The consumer stakeholder

- Representative of an **independent** consumer organization, working in the interest of consumer and wider public interests
- Consumers International, ANEC, national consumer organisations, independent NSB committees.
- Consumer participation in standards adds value by:
  - Ensuring standards address real problems and experiences faced by real people – crucial for technology standards
  - Technical content is based on relevant data from sources outside of usual scope
  - Products and services are fit for purpose, safer, fairer, better quality, more sustainable, more accessible and inclusive

bsi.

Minimizes **risk of harm** and **improves confidence**

COPOLCO

- COPOLCO is an **ISO policy committees** advising/reporting to the ISO/Council.
- COPOLCO coordinates and encourages **consumer input** into technical and policy work in ISO to help consumers benefit from standardization
- Provide a **consumers' network** to exchange information
- **Advise ISO on policies** and actions to respond to consumers' needs
- **Make recommendations** on current and potential standardization work – Privacy by Design, Vulnerable Consumers, Mobile Banking, Cross-border trade of second hand goods etc
- Strategic priorities 2022: Sustainability, Digitalization and Vulnerability

Consumers and Artificial Intelligence

bsi.

Consumer experience of AI

- Types of tasks AI-enabled technologies perform (in combination or separately)
    - sorting, classifying, pattern recognition image recognition, voice recognition and natural language processing, predicting events, assessing significance, deciding options, enacting options etc

- Much is labelled as AI when simply an algorithm or when use AI technologies for a portion of their functionality.
- Public understanding is mixed due to jargon and hype
- General purpose vs narrow purpose

bsi.

Current use cases

- Use cases in action - common in some consumer areas, infancy in others:
    - Voice recognition services (fraud detection/identify verification, verbal chatbots, home assistants)
    - Facial recognition via door entry system (Amazon Ring) or in predictive policing and crowd control (USA)
    - Healthcare or other social care entitlements distribution (USA), sentencing guidelines
    - Recommendations for content, search etc
    - Products/Appliances – energy efficient washing machines etc
    - Recruitment – identification of people from data on their skills, experience, job titles, pay grades etc.
    - Running services – public transit, security systems, traffic management systems, health research etc etc

Is AI different to other technologies?

- AI systems and tools continually learn and adapt their outputs - there is not a static entity to easily observe or test.
- AI technologies are **socio-technical** - operate in human contexts, to fulfil human-defined goals. They reflect the values and choices of those who build and use them (CofE AI, human rights, democracy, and the rule of law, 2021)
- Performance and outcome of an AI product or system does not only depend on its technical components, but on decisions about who uses the technology for what purpose, in what context. (ANEC Position Paper on EU AIA, 2021)
- These decisions have the potential to involve and impact on fundamental rights, cause harm and lower confidence and trust (ANEC)

Is AI markedly different to other technologies?

- While AI is different from other sets of technologies, many of the same issues are present– data, security, bias, lack of information, safety, discrimination etc.
- And same consumer rights apply:
    - Right to Transparency, Explanation and Objection
    - Right to Accountability and Control
    - Right to Fairness
    - Right to Non-discrimination
    - Right to Safety and Security
    - Right to Access to Justice
    - Right to Reliability and Robustness

BEUC, AI rights for consumers, 2019

bsi.

Developing responsible AI

- Developing quickly and being rolled out in markets and public services, without clear governance or regulation.
- Predictions that the use of AI will become more **ubiquitous**. As intro to this track says, it also must be **responsible**.
- What do we need from standards to make sure we develop responsible AI that consumers have trust and confidence in?

bsi.

Global Risk Perception of AI Decision Making

Will the development of machines or robots that can think and make decisions in the next 20 years mostly cause harm or mostly help?

Region	Mostly Harm	Mostly Help	Neither	Do not Know
Latin America & Caribbean	49	26	19	6
North America	47	41	12	0
Europe	43	38	15	5
Central Asia	34	36	17	13
Middle East	33	38	19	10
South Asia	33	31	17	19
Africa	31	41	16	12
Southeast Asia	25	37	21	17
East Asia	11	59	12	18

Based on sample of 154,195 respondents in 142 countries
Global Attitudes towards Artificial Intelligence (AI), University of Oxford 2021. Analysis of survey data from the 2019 World Risk Poll

bsi.

What might erode consumer trust and confidence?

Consumers are positive about potential benefits of AI in some applications but concerns about individual/societal impact. They express fears that:

- it will lead to an increase in abuses related to the use of their personal data
- companies are using AI to manipulate their decisions
- governments are using AI to control their citizens
- AI is dangerous because machines can fail
- AI could lead to unfair discrimination against particular individuals or social categories
- it is not clear who is accountable in case AI is not secure or causes harm

BEUC nine country survey, 2019

Role for standards

- Not all consumer concerns and risks can be dealt with via standards
- Legislation is very important in setting out parameters and expectations (as EU AI Act is poised to do)
- Looking at concerns and risks, there are obvious areas where standards can help eg, data quality, privacy and security, environmental efficiencies, Transparency, redress, design of consumer interfaces for maximum accessibility and understanding

Potential areas for standardization

Design:

Risk assessment
Impact assessment
Data procurement, extraction and analysis: reviewing for bias /blanks
Ethics in design
Stakeholder consultation

Development:

Data quality
Data training
Data collection: clarity on consent
Model testing: reviewing for bias and negative impact
Safety
Autonomy
Cybersecurity

Deployment:

Misleading claims
Explainability: visuals, language comprehension etc
Audit: technical documentation and record keeping
Supervision
Transparency
Quality management
Dispute resolution
Accessibility
Privacy and data protection

Transversal:

- Governance
- Ethics
- Environment and sustainability
- Conformity Assessment
- Terminology
- Classifications

Consumer requirements within AI standards

- **Risk assessment** standards content to address needs and expectations:
  - Likelihood of the harm occurring
  - Immediacy of the harm
  - Foreseeable use of the AI system
  - Measures and metrics for short and longer term impacts of the AI system on:
    - health, safety, fundamental rights, consumer rights, societal effects, impact on democracy, rule of law or environmental impact, as well as the potential for economic harm.

Consumer requirements within AI standards

- **Inclusion and access** standards content to address needs and expectations:
  - Accessible and usable devices and services for users of AI systems and services
  - Design for all/access to older people and people with disabilities
  - Multi-cultural and multi-linguistic aspects
  - Consumers expect to be able to use products, services and systems enabled by AI whatever their age or ability..

bsi.

Consumer requirements within AI standards

- **Transparency and information** standards content to address needs and expectations
- AI systems entail very complex technical processes which are hard to explain and understand for consumer users of products and services.
  - Explanations about how decisions are made and how the system works are tailored to the context and provided in useful and comprehensible format suitable for user and include future potential risks of AI.
  - Information on the nature of data use and potential privacy implications for users of AI products and services is clear and readily available.
  - Information is also clearly and readily available on how to object to decisions or seek human recourse.

bsi.

Consumer requirements within AI standards

- **Cybersecurity** standards content to address needs and expectations:
- The personal information and personally identifiable information in use in AI systems increases the cybersecurity risks.
  - Technical requirements such as secure authentication and security updates, incident management and abuse alerts, encryption etc.
  - Organisational requirements such as information on product security status and support, expected product lifetime etc

Consumer requirements within AI standards

- **Quality management** standards content to address needs and expectations. A quality management system which includes:
  - Robust consumer relationship management
  - Systems for dispute resolution, complaints
  - Customer redress
  - Effective objection systems for dubious decisions
  - Recall of harmful products
  - Post-marketing monitoring system

bsi.

Thank you

Liz Coll, BSI Consumer and Public Interest Network Steering Group, UK representative to the ISO COPOLCO AI task group.

Bundesamt für Sicherheit in der Informationstechnik

Deutschland Digital•Sicher•BSI•

Evaluation Standards for Conformity Assessment of Trustworthy Cloud-based AI Applications

Daniel Loevenich, 16. Mai 2022

Agenda

Understanding AI Applications

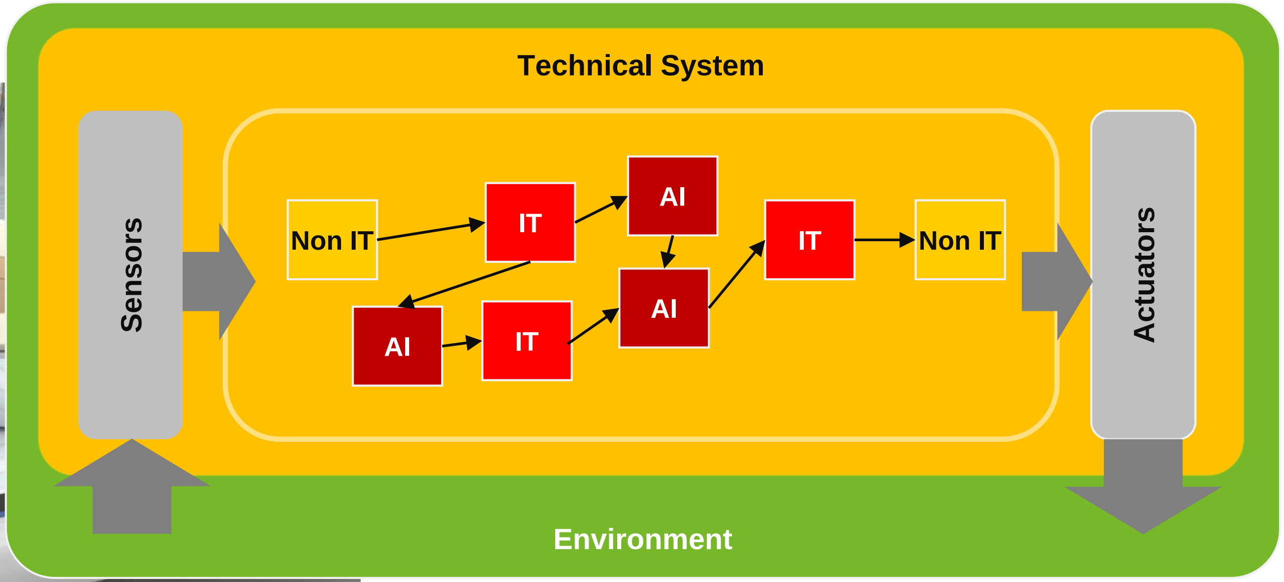
Understanding ISO/IEC AI Standards

Horizontal Trustworthy AI Conformity Assessment

Project Success Factors

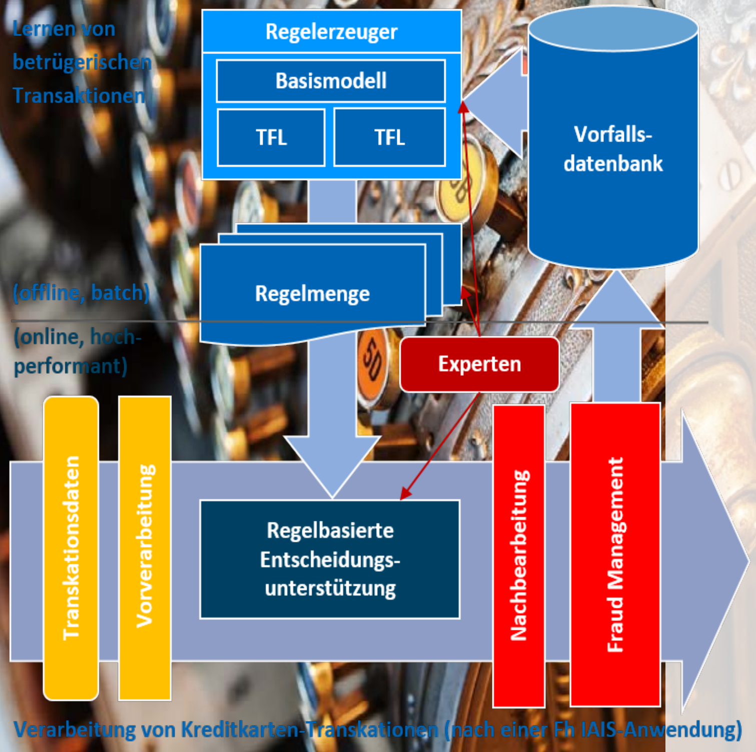
A uniform Framework for Horizontal AI Conformity Assessment

Understanding AI Applications



Understanding AI Applications

- **Conformity Assessment** considers the complex technical system (e.g. a vehicle) as a whole entity in the application context
- **Risks and requirements for the AI-components** have to be derived from this
- Such **Operationalizations** are essentially **mappings** into AI requirements
- This results in evaluation requirements for each component of usually hybid AI solutions (**application-agnostic** perspective)



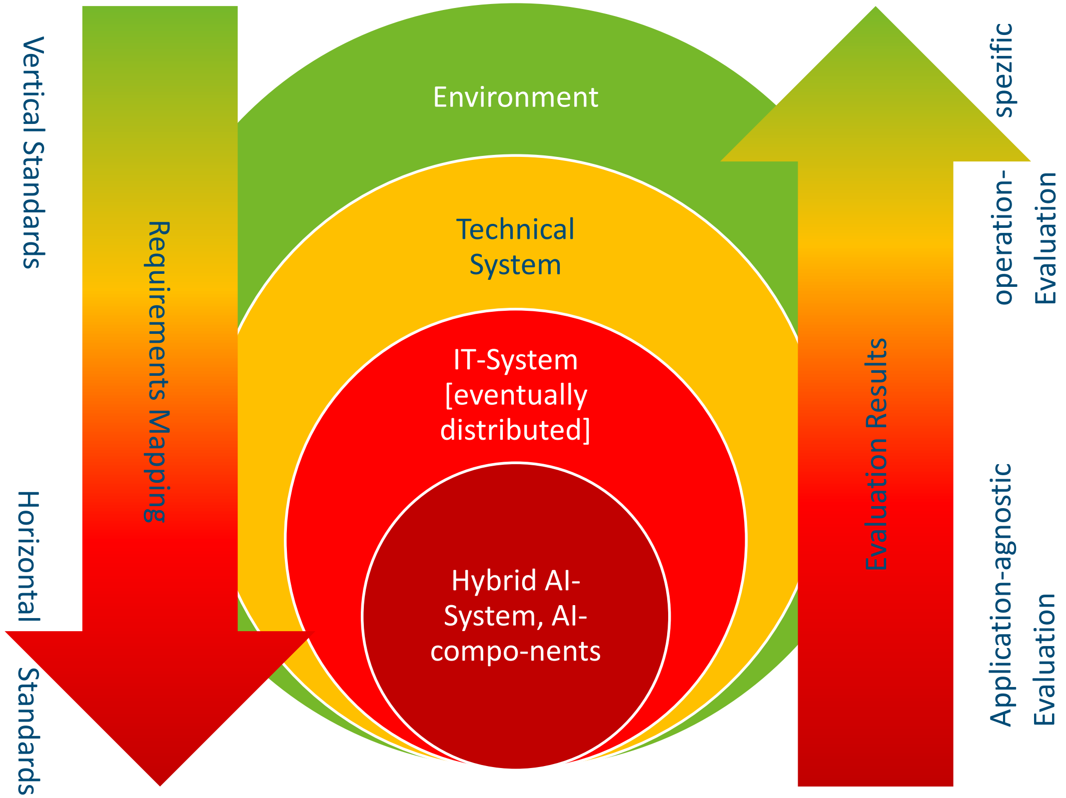
Understanding AI Applications

- A technical system usually follows **existing test regulations for supply chains** with multiple actors, e.g. OEMs
- Industry expects **cascadable evaluations for (distributed) AI-systems**
- AI test results must be **composable** (for instance in case of individual component checks of hybrid systems)
- A horizontal standard for trustworthy AI must be embeddable into AI management systems (**AIMS**)

Understanding ISO/IEC AI Standards: Testing and Evaluation

Horizontal and Vertical Standards Concept:

- **Downgrading Risks** to AI Components
- **Conformity Testing and Evaluation** of AI Components (all CA types)
- Upgrading and **Composition** of Evaluation Results
- **Uniform Conformity Assessment Framework**
- Application of schemes and **introduction to markets worldwide**



Horizontal Trustworthy AI Conformity Assessment: Strategic Objectives

Trustworthiness of the **entire Supply Chain** becomes transparent with **Conformity Assessment**

The necessary **evaluation criteria and test procedures** are to be developed

This evaluation bases shall be **applicable to hybrid solutions** as well as **embedded components**

EU-AIA: The evaluation bases serve as foundation for a **horizontal AI standard**

An **easy market access for SMEs** with acceptable costs will be facilitated

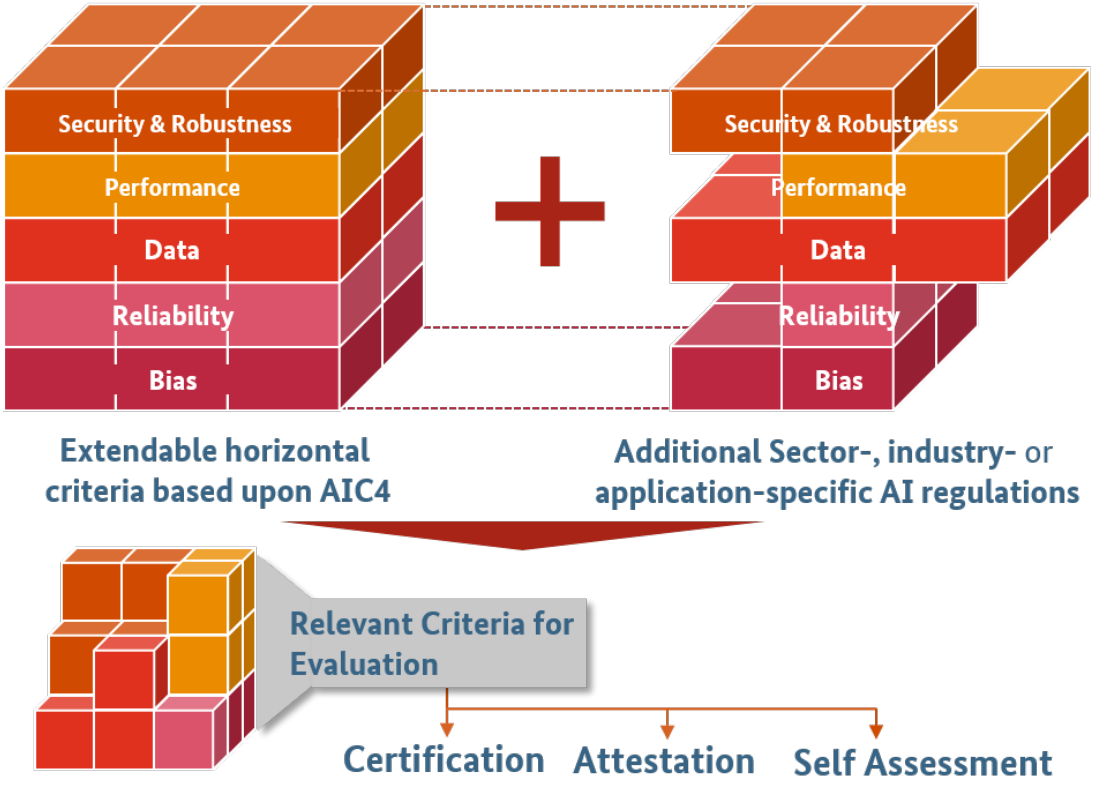
Vertical and sectoral standards should be based a horizontal standard for trustworthy AI

Uniform Evaluation Framework for Trustworthy AI Applications

Horizontal Trustworthy AI Conformity Assessment: Deliverables

An extendable set of **horizontal, application agnostic criteria** (“**Trustworthy AI Services Evaluation Criteria**”),

An extendable set of valid **AI evaluation procedures** (“**Trustworthy AI Services Evaluation Methodology**”), applicable to all three types of conformity assessment (self assessment, attestation and certification),



Horizontal Trustworthy AI Conformity Assessment: Deliverables

A proposal for an **application procedure** to extend the methodology and to implement it within the ongoing standardization process,

A proposal for an **application procedure** to extend this criteria and to implement the procedure within the ongoing standardization process,

A **procedure for mappings** of vertical application specific requirements into horizontal criteria requirements,

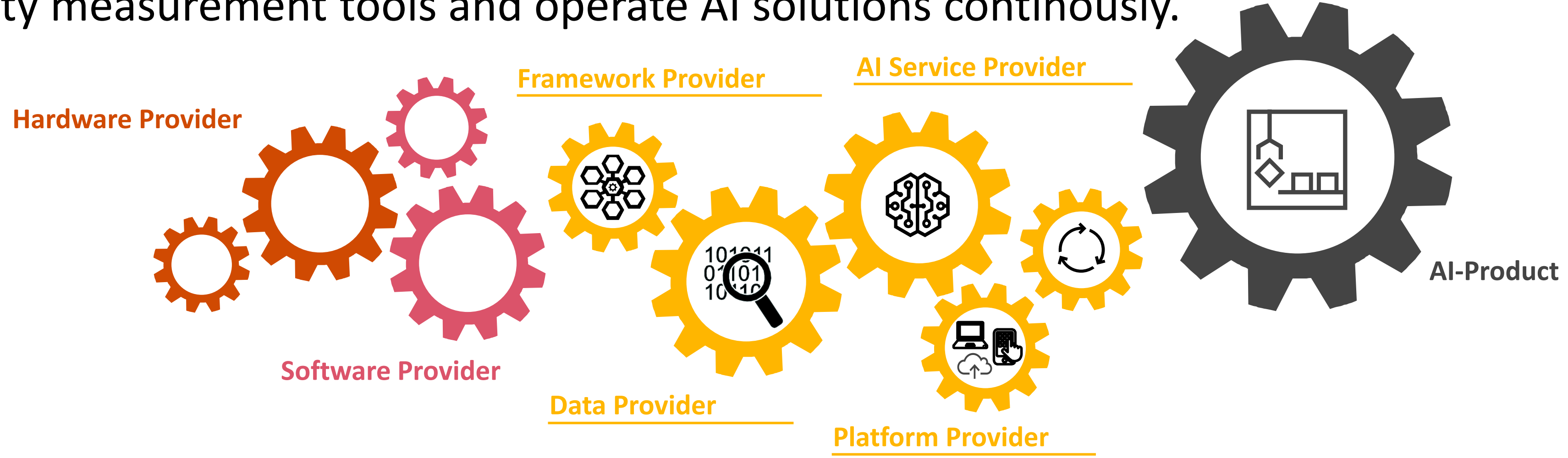
The **Guidance documents for production, application, and support** for all parties involved in the corresponding AI Eco-Systems. In the context, the framework establishes additional guidelines on how to integrate evaluation activities into an AIMS.

Horizontal Trustworthy AI Conformity Assessment: Project Roadmap

Definition of evaluation bases	**Applicability and Market Penetration**	**Publication Phase**	**International Standardization**
Q3 - 2022	Q1 - 2023	Q4 22/Q3 23	2025
Development of **criteria, methodology, scheme** with use cases	**Validation** and Acceptance on basis of relevant Use Cases	Publication within **three Workshops -** Europe, USA, Asia	Transfer and Harmonization in **hEN/ISO Standard**

Project Success Factors: Drivers for Technology Transfer

Cloud Providers drive AI technologies within the compete AI supply chain. They offer full service enterprise customer support with AI experts for development, IDEs, Frameworks and quality measurement tools and operate AI solutions continously.

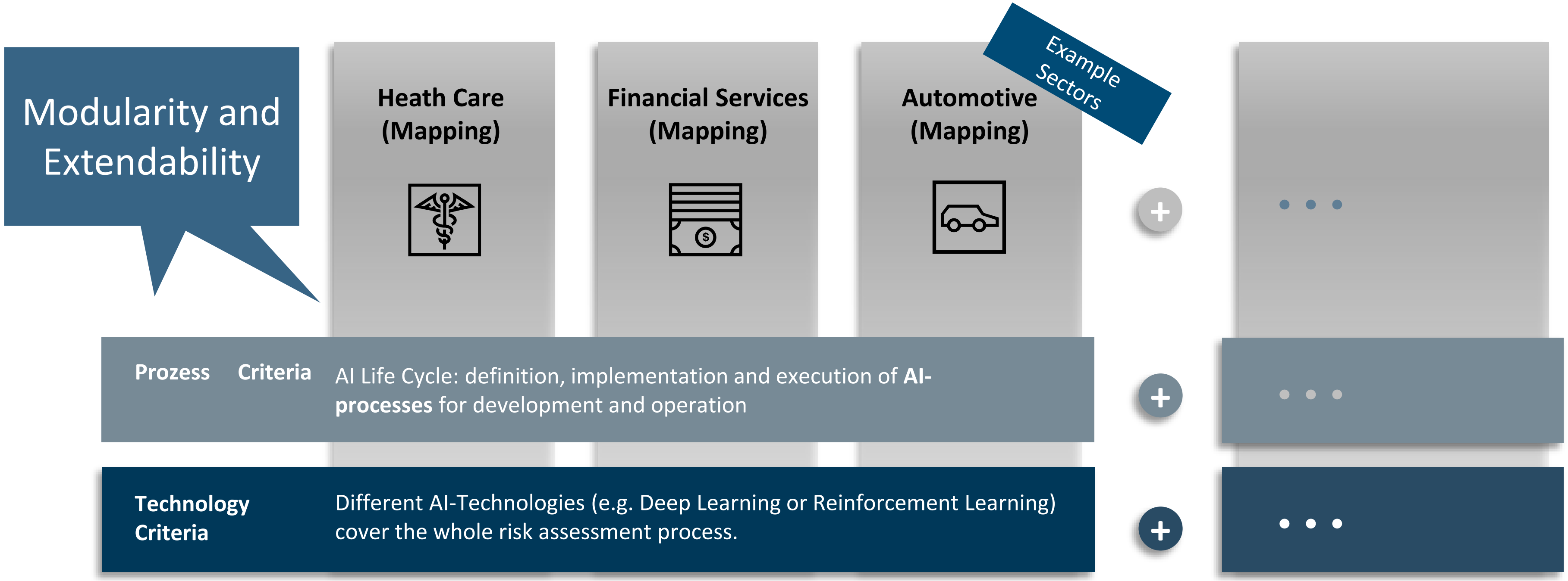


A project for an horizontal standard for trustworthy AI shall use these resources.

Project Success Factors:
Use Case Projects Outline

	Phase 1:	Phase 2:	Phase 3:	Phase 4:
Focus:	Scoping, Risk Analysis & TOE	Evaluation Procedure Development	Procedure Execution	Tool and Framework Validation
Outcome (Dokuments):	Mapping System Description Minutes	Procedure Description Protocol	Evaluation Documents (dependable on the conformity assessmant type) Evaluation Report	Evaluation Documents Evaluation Report
Format:	Workshops	Coordination Conference with unanimous vote	Depends on the type of conformity assessment	Certification Report
	Mandatory: Generalization of Requirements up to Criteria and Evaluation Process Definition		**Optional: Pilot Evaluation, Procedure Validation, Assurance Methods Assessment (Levels)**	

Project Success Factors : Sub-Projects and Use Cases

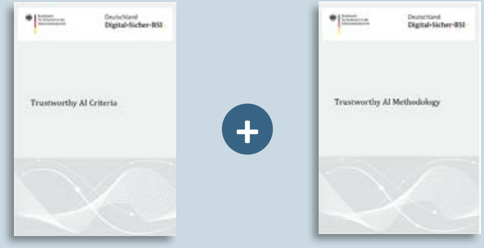


A uniform Framework for Horizontal AI Conformity Assessment: End User Acceptance

Open questions for dealing with AI in a trusting manner …

- How can responsible use of AI be demonstrated accross the entire supply chain of an AI system?
- Which requirements are relevant for my AI systems and how do I fulfill them according to my area of responsibility?
- How can I efficiently implement requirements with regard to my use case - specific systems and processes?
- How do I efficiently demonstrate compliance with increasing AI regulations?

… are addressed by applying the criteria and method documents



Provision of extendable and horizontal criteria for trustwothy AI applications

Provision of a Validation, Testing and Certification attempt

AI Service Providers as important multiplicators

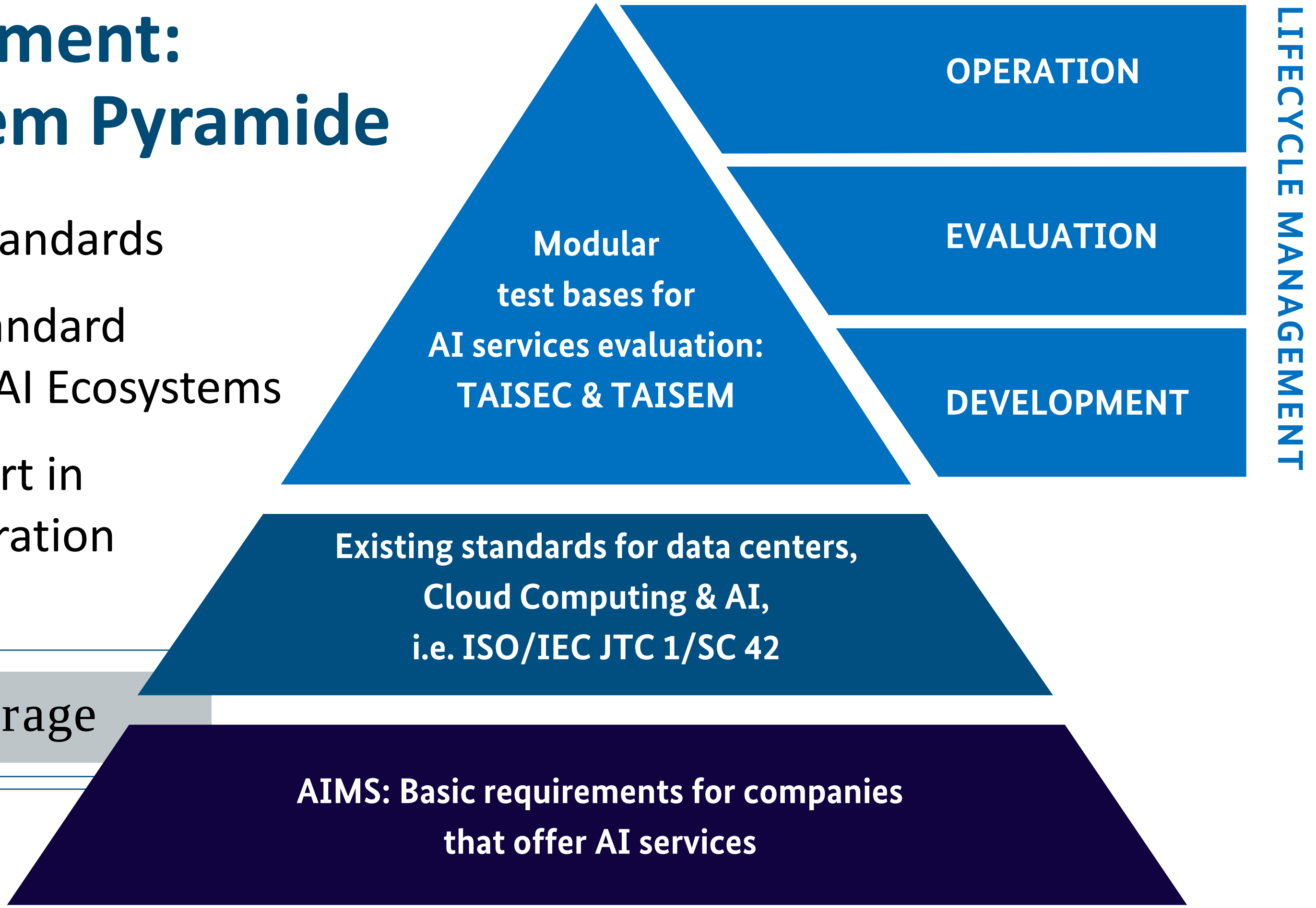
A uniform Framework for Horizontal AI Conformity Assessment: Standardization System Pyramide

Transfer of AI Trustworthiness Standards

Conformity Assessment of AI-Standard Solutions lead to Acceptance of AI Ecosystems

Worldwide End Customer Support in regulated sectors: Market Penetration

Complete AI Ecosystem Coverage



Project Partners

Development, Validation, Market Penetration

Hyperscaler:

- AWS
- Alibaba
- Google
- Huawei
- Microsoft
- Oracle
- SAP
- Dt. Telekom

Research Projects:

- DFKI
- Fh IAIS (Program „Certified AI“)
- With Use Cases: GAIA-X

Standardization

Evaluation, Certification and Accreditation:

- DIN/VDE
- CEN/CENELEC AI JTC21
- **ISO/IEC JTC 1/SC42**

- AI Quality und Testing Hub
- DFKI Osnabrück
- Fh IAIS
- PricewaterhouseCoopers
- Verband der TÜV

Thank you for your attention!

Deutschland
Digital•Sicher•BSI•

Contact:

Daniel Loevenich

daniel.loevenich@bsi.bund.de
Tel. +49 (0) 228 9582 5395
Fax +49 (0) 228 10 9582 5395
Mobil +49 (0)171 30 29 824

Bundesamt für Sicherheit in der Informationstechnik (BSI)
Referat TK 23: Grundsatz, Strategie und Nachweise in der Künstlichen Intelligenz

Godesberger Allee 185-189
53175 Bonn
www.bsi.bund.de

KI-KOMPETENZZENTRUM DES BSI

Das BSI als die Cyber-Sicherheitsbehörde des Bundes gestaltet Informationssicherheit in der Digitalisierung durch Prävention, Detektion und Reaktion für Staat, Wirtschaft und Gesellschaft.

Emerging AI Technology Trends

Shubhashis Sengupta

accenture

ISO JTC 1 IEC

INFORMATION TECHNOLOGY STANDARDS

SC 42 – Artificial Intelligence

Introduction

5 Key Technology Innovation Trends are shaping the future of AI engineering and Adoption

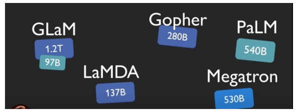
- **Large Language and Image Learning Models**
- **Gaining Efficiency and Sustainability in AI**
- **Synthetic Data comes to the Rescue**
- **AI and Intelligence**
- **AI Automation at Scale**

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

accenture

TREND1: Large-Scale “Meta-Learning” Models

- **Transformer models – learning from Terabytes of data and Billions of Parameters**
- **3 key Innovations –**
  - Attention
  - Self Attention
  - Positional Embedding
- **MIT calls these Foundation Models**



NVIDIA/**Megatron-LM**

- **Usage and adaptation are changing dramatically –**
  - Few Shot and Zero-Shot learning through “Smart Prompts”
  - Fine Tuning of models through LLRD, SWA, Adaptive Tokenization, LoRA
  - Model access, open sourcing and BERTOLOGY
- **Meta Learning as models learn representations for multiple tasks** (summarization, translation, Q&A, classification, lang generation, sentiment analysis etc.)

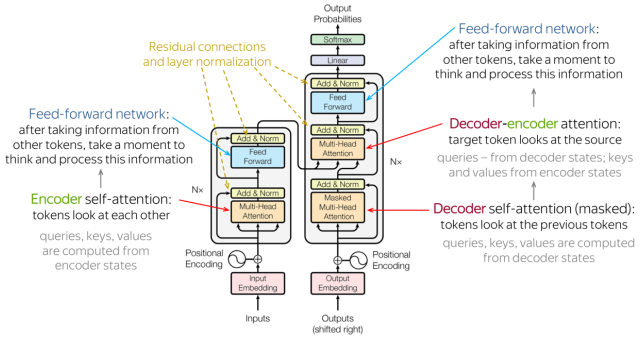
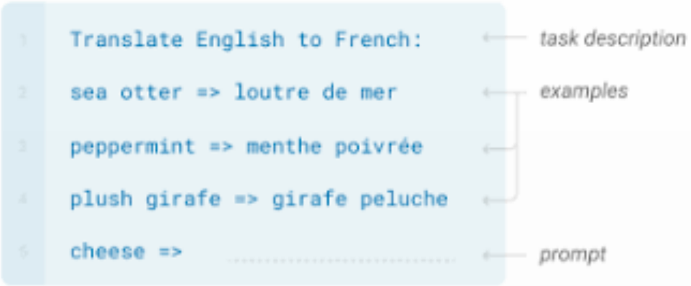


Image courtesy: Cross Validated, Stack Exchange



Source: https://arxiv.org/abs/2005.14165

- **Similar advancement in Image Language Modeling and Generation –**
  - DALL-E, DALL-E2
  - Multi-modal Transformers

Image courtesy: https://twitter.com/OpenAI/

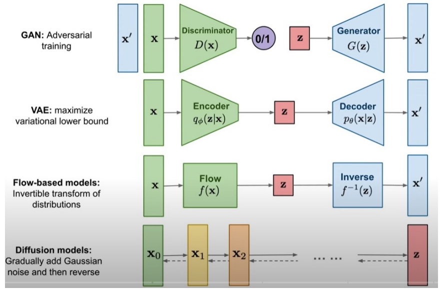


Image courtesy: lilianweng.github.io

INFORMATION TECHNOLOGY STANDARDS

SC 42 – Artificial Intelligence

TREND2: Efficient AI (Faster, Cheaper, More Sustainable)

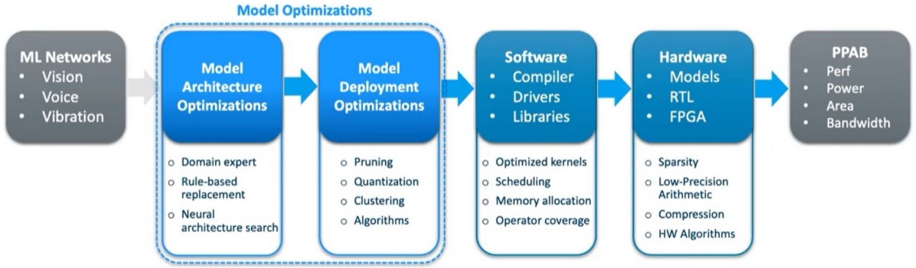
- **Large Models are Costly and not Environment friendly**
  - Google Switch (1.6 T params) takes 1 MUSD to train and emits 284T of Co2e
  - Training GPT-3 once may take up to 800K USD

Model	Hardware	Power (W)	Hours	kWh·PUE	CO_2e	Cloud compute cost
$\text{Transformer}_{base}$	P100x8	1415.78	12	27	26	\$41–\$140
Transformer_{big}	P100x8	1515.43	84	201	192	\$289–\$981
ELMo	P100x3	517.66	336	275	262	\$433–\$1472
BERT_{base}	V100x64	12,041.51	79	1507	1438	\$3751–\$12,571
BERT_{base}	TPUv2x16	—	96	—	—	\$2074–\$6912
NAS	P100x8	1515.43	274,120	656,347	626,155	\$942,973–\$3,201,722
NAS	TPUv2x1	—	32,623	—	—	\$44,055–\$146,848
GPT-2	TPUv3x32	—	168	—	—	\$12,902–\$43,008

Table 3: Estimated cost of training a model in terms of CO_2 emissions (lbs) and cloud compute cost (USD).[7] Power and carbon footprint are omitted for TPUs due to lack of public information on power draw for this hardware.

Source: https://blog.learningtree.com/carbon-footprint-ai-deep-learning/

- **Making AI training and Optimization efficient at Data Centers**
  - Algorithmic Optimization – Pruning and Clustering
  - Deployment Optimization – Equalization, Fold-Batch norms, Fused layers, Quantization
- **Training speedups: FFT models, Pathway Dataflow for parallelism**
- **How to induce Pruning / Sparsity without Model Degradation**

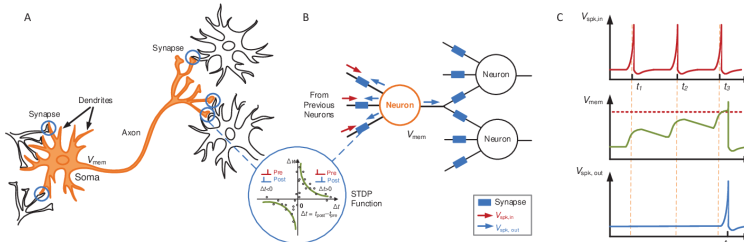


Trained Models -> Candidate Models -> NetOpt Models -> DepOpt Models

Image Courtesy: Qualcomm

- **Efficient Edge Computing Architecture**
  - TinyML models for low AI footprint
  - Optimized models like MobileNet V1 (DS-CNN),V2
  - Neuromorphic Architecture (SNN) – extremely low power architecture

TINY ML



https://www.tinyml.org/

Spiking Neural Network: https://en.wikipedia.org/wiki/Spiking_neural_network

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

accenture

TREND3: Synthetic Data for AI

- **Right Data is the most Difficult Ask**
  - Getting right domain data for Model Training is getting increasingly difficult
    - GDPR Regulation, Sensitive (PHI, PII) data, Localization and Fiduciary implications, Data Security
- **Open Source and Linked Data exploitation has hit limits**

LINKINGOPENDATA
W3C SWEO Community Project

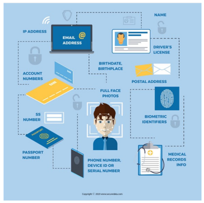
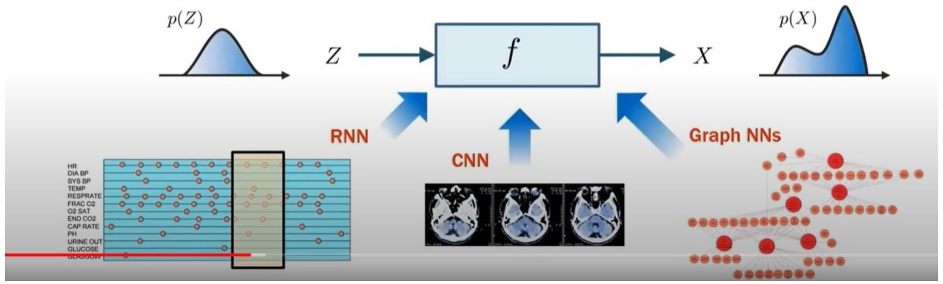


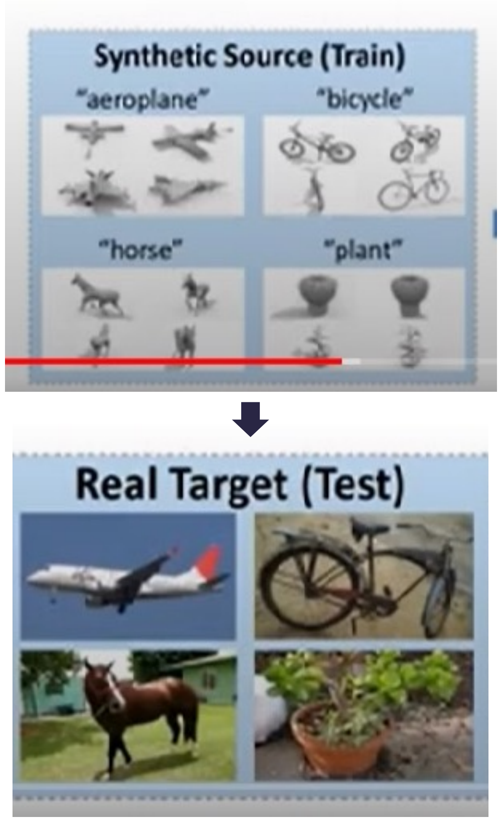
Image Courtesy: iri.com
Image Courtesy: SecureDrive

- **Data Generation for AI**
  - Data De-identification
  - Data Augmentation
  - Fully Synthetic Generative Modeling
    - Models with Implicit likelihood (GAN -> generates by comparison)
    - Models with Explicit likelihood (VAE, Fully Observable Models)
    - Diffusion Network (very realistic)



Source: ICML Tutorial on Synthetic Data Generation (https://www.youtube.com/watch?v=_EEH9HU2EE0)

- **Quality and Trustworthiness**
  - Should not copy but preserve relationship/correlation
  - Statistically meaningful
  - Predictivity, Diversity, Realism, Privacy Preservation



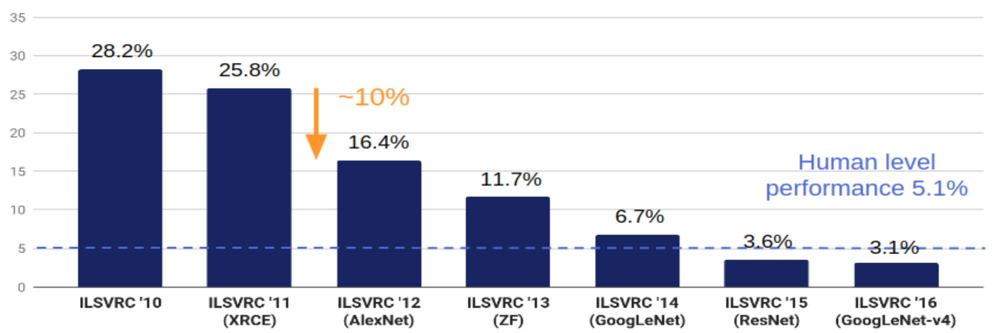
ISO JTC1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

accenture

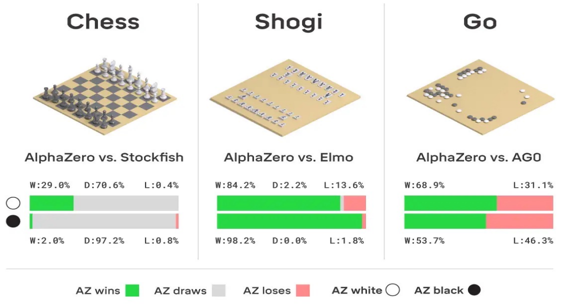
TREND4: Cognitive AI (Can AI augment / partner with Humans?)

- **Narrow AI supremacy (already achieved)**
  - AI is already performing at “superhuman” levels for many specific jobs (Q&A, Reading Comprehension, Image Segmentation)
  - Generalization to multi-task level is the key challenge

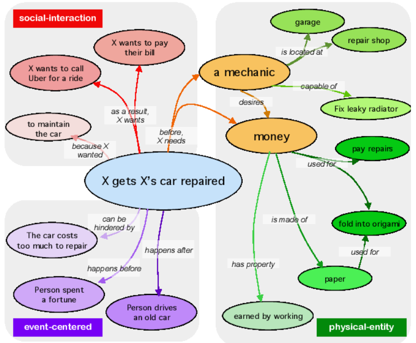
ImageNET competition classification top-5 error (%)



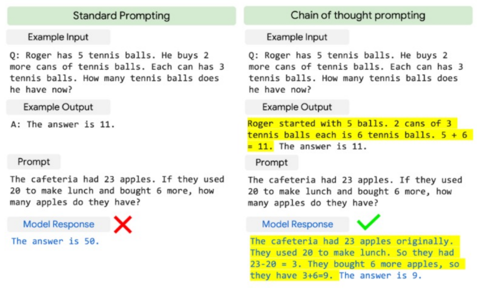
- **Towards Complex and Creative Tasks (In Progress)**
  - Planning and Strategy (Games) – Deep Reinforcement Learning (AlphaZero) vs Rules
  - Common Sense Reasoning – Atomic
  - Analogy based Reasoning, Forward Chaining
  - NLI / NLE
  - Multi-step Reasoning with Explanation (logically solving Math Words Problem) – Google PaLM
  - Code generation - CoPilot
  - KG based reasoning (link prediction, message passing)
  - Image Incongruity Detection
  - Winograd Schema



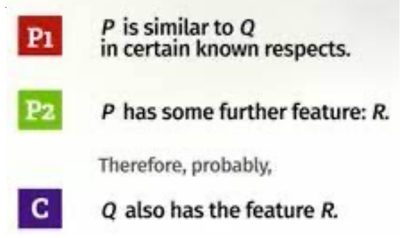
https://www.deepmind.com/blog/alphazero-shedding-new-light-on-chess-shogi-and-go



Atomic: An Atlas for Machine Common Sense (https://arxiv.org/abs/1811.00146)



Standard prompting versus chain-of-thought prompting for an example grade-school math problem. Chain-of-thought prompting decomposes the prompt for a multi-step reasoning problem into intermediate steps (highlighted in yellow), similar to how a person would approach it.



https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html

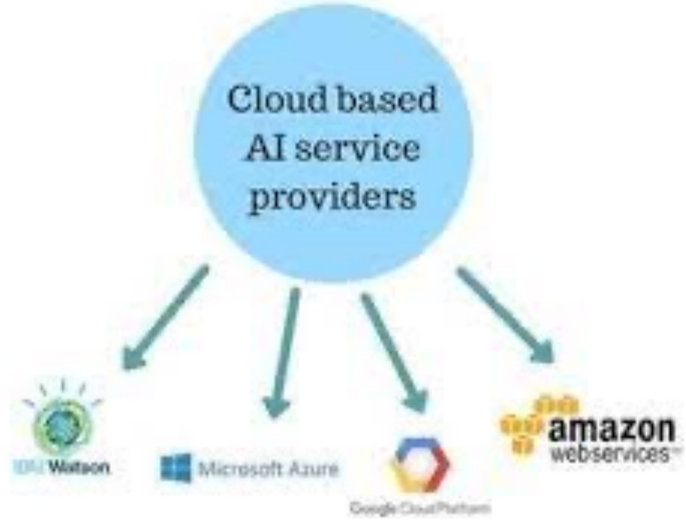
Achieving AI Singularity? Not Yet

INFORMATION TECHNOLOGY STANDARDS

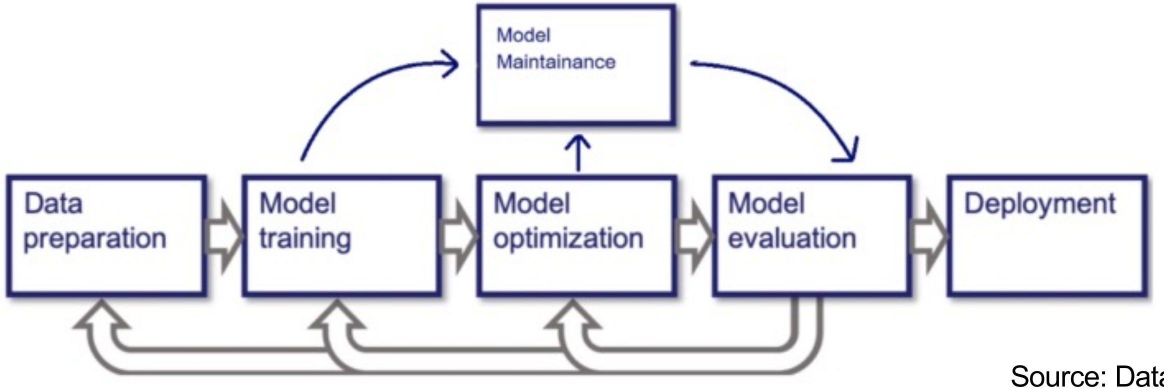
SC 42 – Artificial Intelligence

TREND5: AI Automation at scale; AI as a Service (AAAS)

- **Cloud-based pre-trained and (slightly) trainable models**
  - Easy to set up, operate and consume
  - SaaS - Mainly driven by AutoML, templates, large pre-trained models.
  - PaaS / Workbench facilities – SageMaker, Azure, Vortex, Einstein, Watson
  - Industry / Domain AI Solution (e.g., Accenture Solutions.AI)



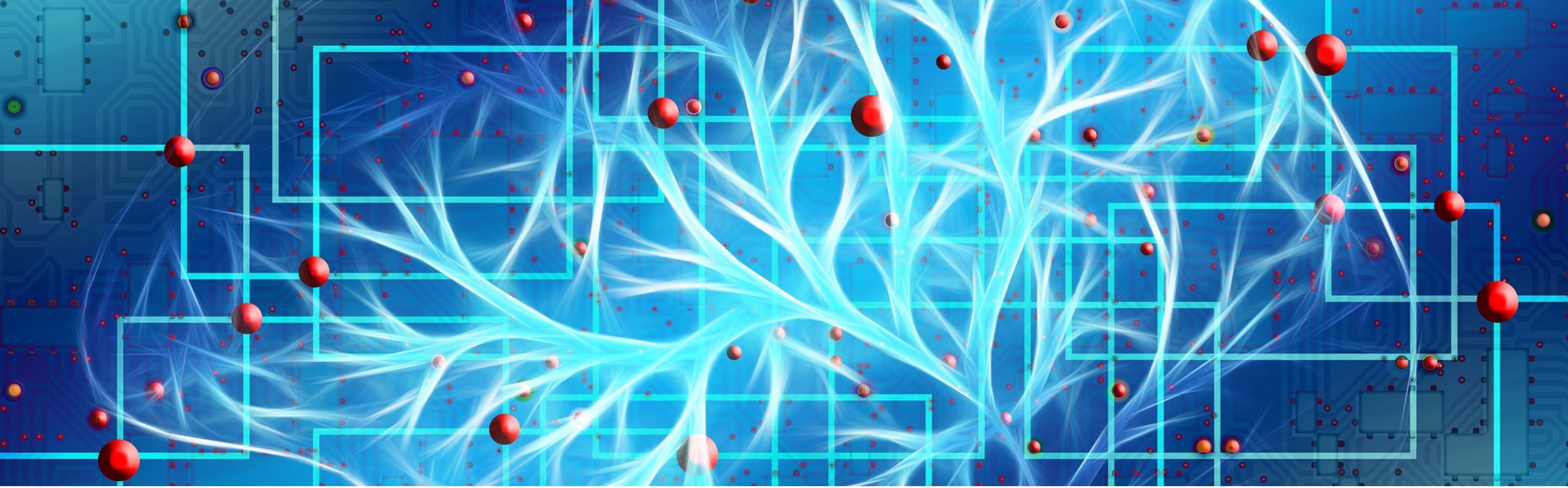
- **Key technical issues**
  - Input Data Quality - Sample, Shape, Coverage, Bias, Class Imbalance, Anonymity and Privacy, Noise, Drift. EDA and Feature Engineering.
  - Model Engineering – Layers/ Stacks, Weights, Residuals, Connections, Density / Sparsity, Loss Function, Back Propagation
  - Model Quality –
    - Traditional Metrics: Precision, Recall, f-measure, AUC
    - Benchmarks
      - Single Task (BLUE, ROUGE), Multi Task (GLUE, Super GLUE), Complex Task (BIG_Bench)
  - Inference – Representation, Realism, Trust and Explainability (local – LIME, SHAP, Counterfactuals; Global – Gradient-based), Understanding Impact



Source: Datatron

INFORMATION TECHNOLOGY STANDARDS

SC 42 – Artificial Intelligence



Summary

In this presentation, we discussed 5 key technology innovation trends for enterprise AI. As AI has got mainstream, it is now time to assess and analyze the medium and long-term implications of these trends, especially from the point of view of standardization, interoperability, and trustworthy adoption.

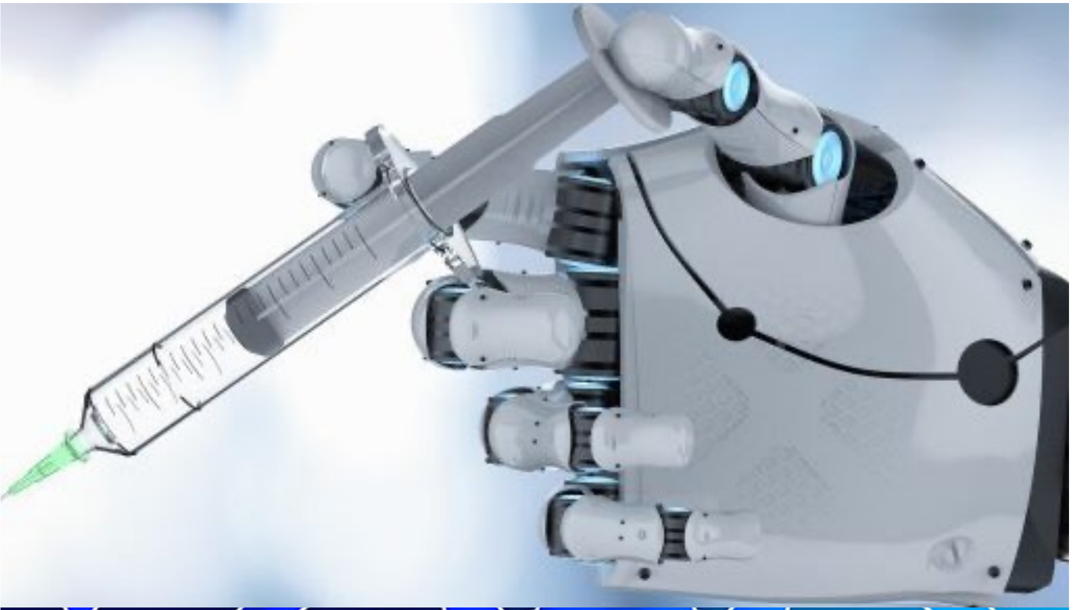
Thank you

Shubhashis Sengupta

Shubhashis.Sengupta@accenture.com

https://www.accenture.com/us-en/about/accenture-labs-index

https://www.linkedin.com/in/shubhashis/



INFORMATION TECHNOLOGY STANDARDS

SC 42 – Artificial Intelligence

AI Quality: The Next Big Challenge in AI

Will Uppington, CEO &
Co-Founder, TruEra

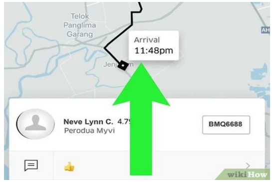
ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

Introduction

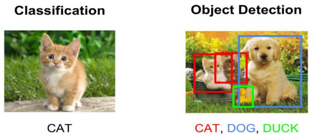
Data Scientists have always focused on the performance of Artificial Intelligence/Machine Learning (AI/ML) systems. Recently, however, as the use of ML has moved from consumer to Enterprise and Government applications, increasing attention is being paid to the potential risks related to non-transparent ML systems. Enterprise organizations have struggled to move ML pilots into mainstream deployment. Regulators throughout the US and the world have issued warnings around the potential misuse of artificial intelligence in lending, employment, medical and other applications. Lawmakers, such as in the EU and New York City, have proposed or passed new laws regulating the risks of AI. To build and operate successful ML systems, developers need to account for both the performance and risks of AI/ML. This session will cover how a systematic approach to evaluating, testing and monitoring AI Quality leveraging end-to-end AI explainability technology can address the dual challenge around ML performance and risk.

Use of AI Moving from Consumer to Enterprise

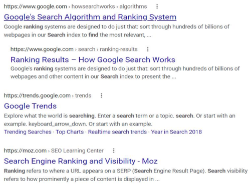
Machine Learning First Used For



ETA



Cat detection

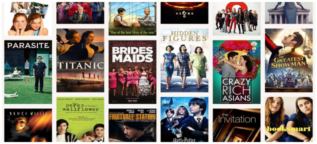

https://www.google.com › howsearchworks › algorithms
Google's Search Algorithm and Ranking System
Google ranking systems are designed to do just that: sort through hundreds of billions of webpages in our Search index to find the most relevant, ...

https://www.google.com › search › ranking-results
Ranking Results – How Google Search Works
Google's ranking systems are designed to do just that: sort through hundreds of billions of webpages and other content in our Search index to present the ...

https://trends.google.com › trends
Google Trends
Explore what the world is searching. Enter a search term or a topic. search. Or start with an example. keyboard_arrow_down. Or start with an example.
Trending Searches · Top Charts · Realtime search trends · Year in Search 2018

https://moz.com › SEO Learning Center
Search Engine Ranking and Visibility - Moz
Ranking refers to where a URL appears on a SERP (Search Engine Result Page). Search visibility refers to how prominently a piece of content is displayed in ...

Search ranking



Movie recommendation

Google Uber NETFLIX Microsoft

It's now being used for

Zillow

Homebuying

standard chartered

Lending

Pricing

amazon

Hiring

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

This Change Is Hurting Some Companies

Zillow®

Zillow to stop flipping homes for good as it stands to lose more than $550 million, will lay off a quarter of staff

Last Updated: Nov. 3, 2021 at 12:36 p.m. ET
First Published: Nov. 2, 2021 at 4:21 p.m. ET

—

By Jon Swartz

41

'We've determined the unpredictability in forecasting home prices far exceeds what we anticipated and continuing to scale Zillow Offers would result in too much earnings and balance-sheet volatility,' CEO tells investors

Hard to Build High Performing ML Models

This Change Is Hurting Some Companies

Apple Card algorithm sparks gender bias allegations against Goldman Sachs

Entrepreneur David Heinemeier Hansson says his credit limit was 20 times that of his wife, even though she has the higher credit score

Enterprise ML Models Bring New Risks

This Change Is Hurting Some Companies

ARTIFICIAL INTELLIGENCE

Our weird behavior during the pandemic is messing with AI models

Machine-learning models trained on normal behavior are showing cracks — forcing humans to step in to set them straight.

By Will Douglas Heaven

May 11, 2020

Hard to Maintain Performance & Manage Risks Over Time

Regulators Scrutinizing AI Risks

Nov 2018	Monetary Authority of Singapore – Fairness, Ethics, Accountability & Transparency
Nov 2019	Hong Kong Monetary Authority – High Level Principles on AI
Jun 2020	Banque De France – Governance of AI in Finance
Aug 2020	National Association of Insurance Commissioners - AI guiding principles
Sep 2020	Office of the Supdt. of FIs – Tech. Risk Consultation including AI
Mar 2021	CFPB, OCC, Fed Reserve, FDIC, NCUS - Use of AI/ML
Apr 2021	*European Commission – Draft AI law (Credit = high risk AI)*
Jun 2021	European Insurance & Occupational Pensions Authority - AI governance
Jul 2021	Bundesbank and BaFin – ML in Risk Models
Dec 2021	NYC Law on Automated Employment Decision Tools
Feb 2022	Bank of England/ FCA – AI Public Private Forum report

"at the CFPB, we will also be closely watching for digital redlining, disguised through so-called neutral algorithms, that may reinforce the biases that have long existed."

- Rohit Chopra, CFPB

High Failure Rates in Enterprise AI

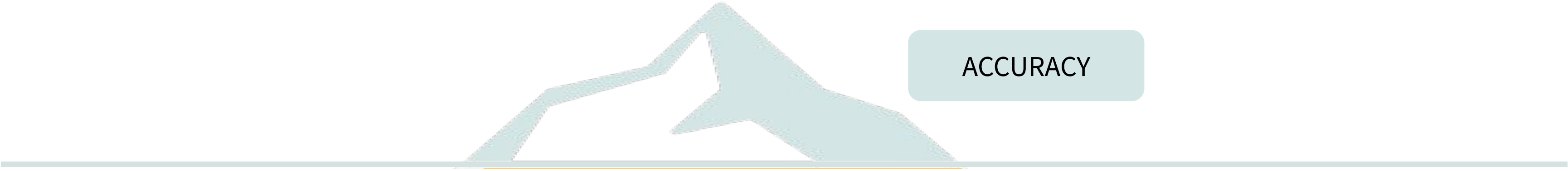
Failure rates for analytics, AI, and big data projects = 85% – yikes!

July 23, 2019 by Brian T. O'Neill

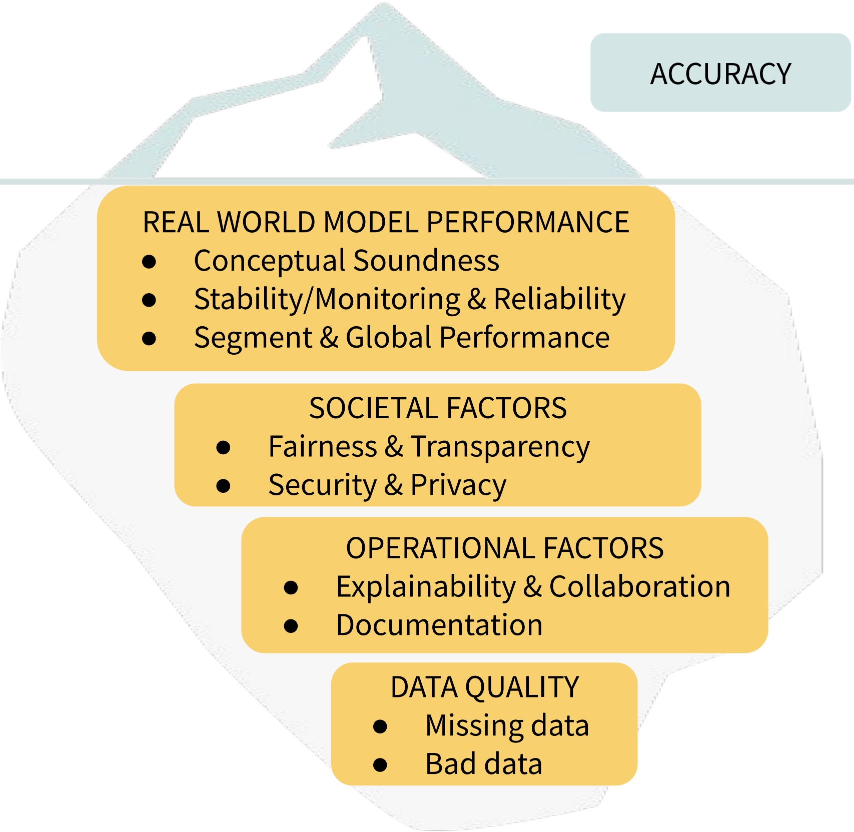
Enterprise AI Demands a New Approach: AI Quality

Enterprise AI requires AI Quality: balance between performance & risk mitigation

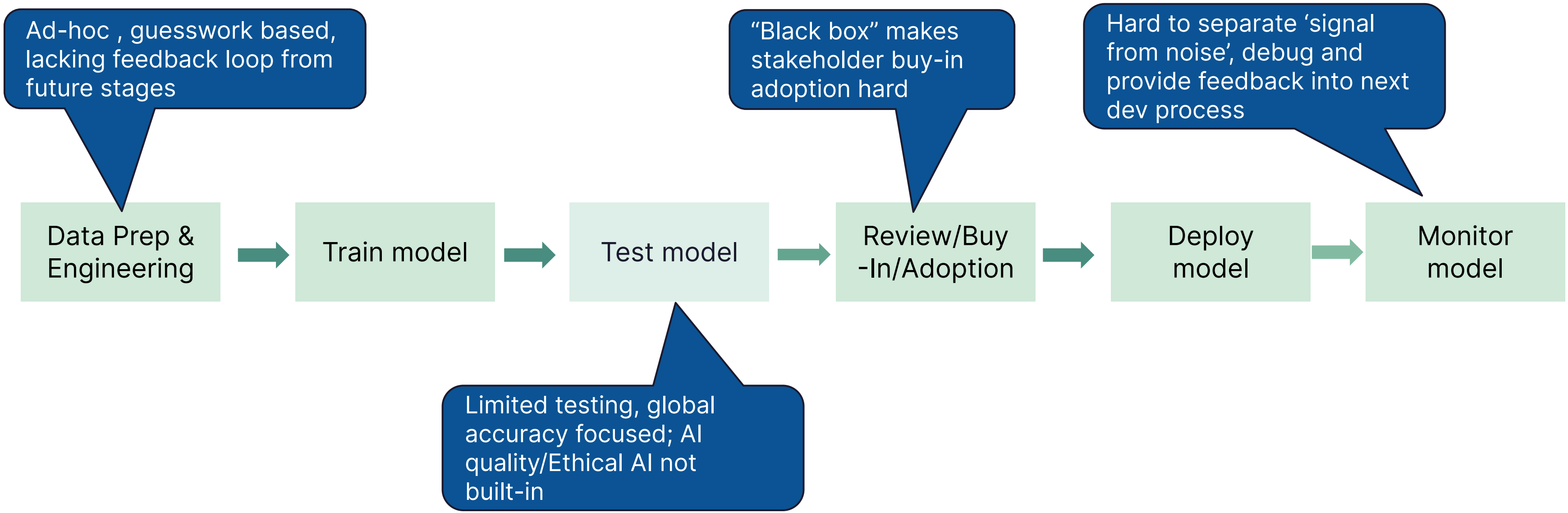
What is AI Quality?



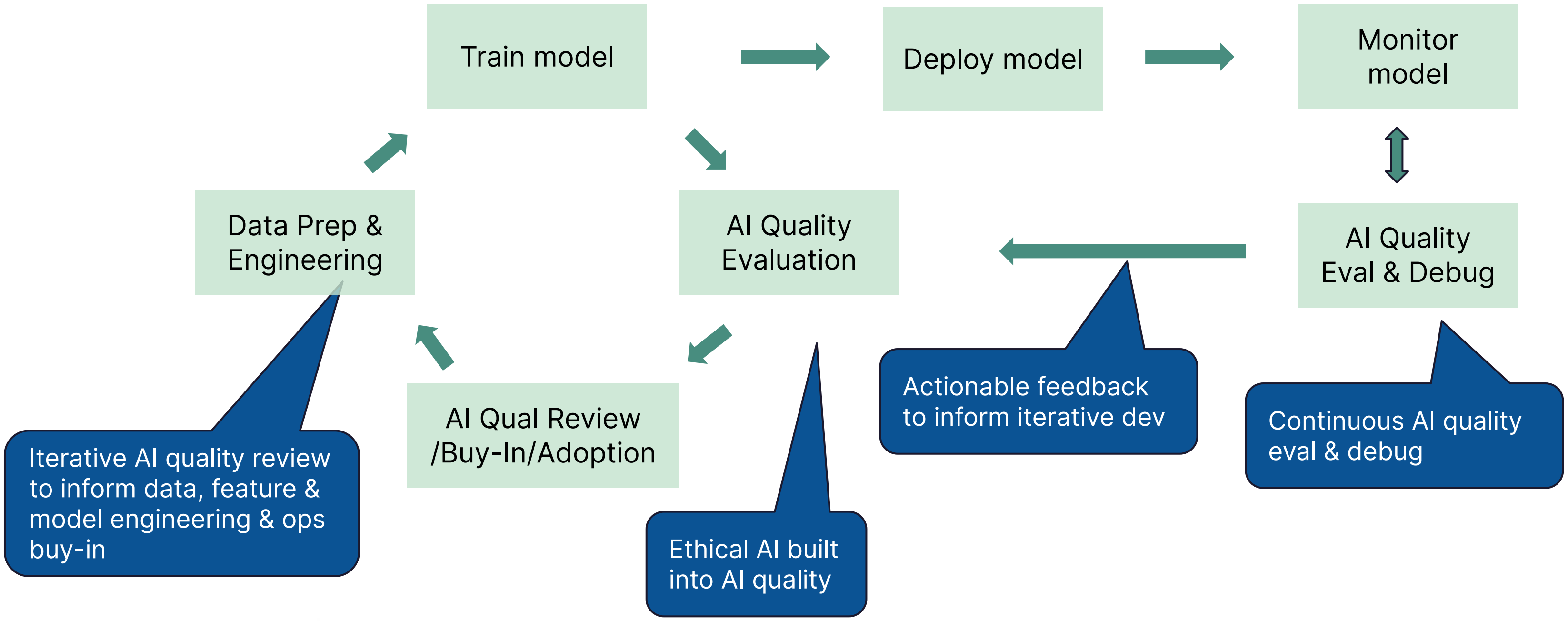
What is AI Quality?



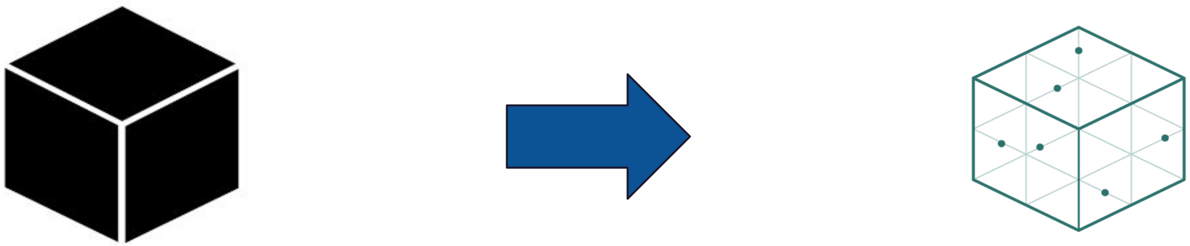
Current AI Methodologies & Tools Not Sufficient



Need Iterative AI Quality Tools & Mindset



AI Explainability is Key to AI Quality



Black Box AI | **AI Explainability** | **Transparent, Managed AI**

AI Explainability Enables AI Quality Tooling

Explainability

Error Analysis

Model Monitoring

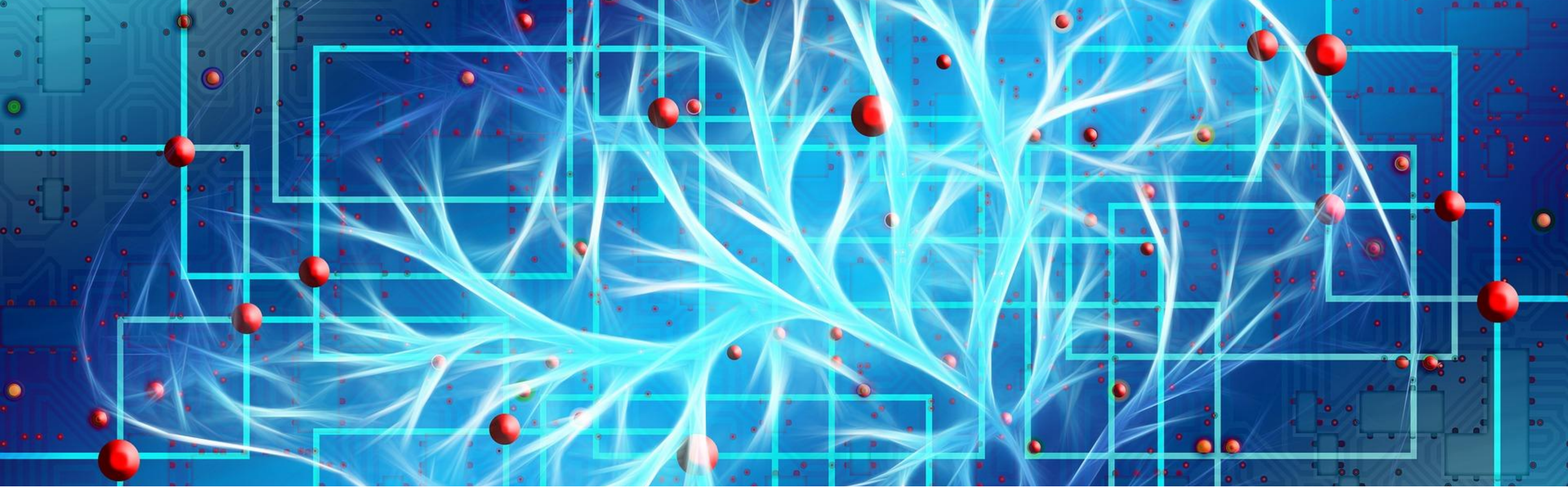
Bias & Fairness

Conceptual Soundness

Segment Analysis

Root Cause Analysis

Reporting



Summary

Succeeding in Enterprise AI requires balancing performance and risk and using AI Quality methodologies and tooling throughout the life cycle

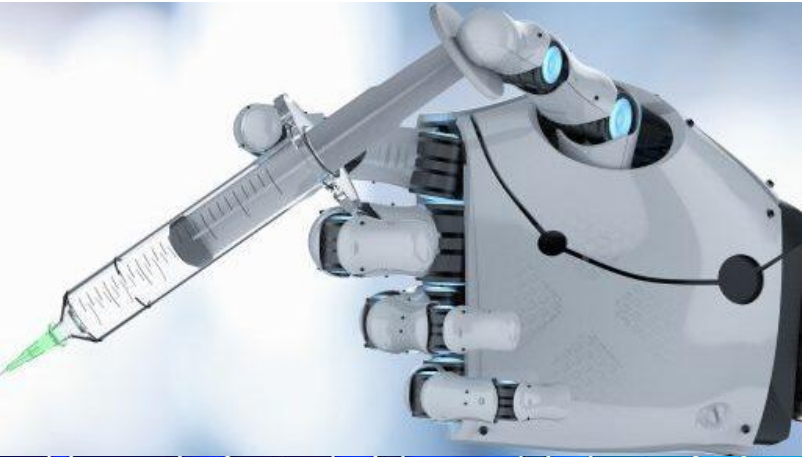
AI Explainability, error analysis, root cause, responsible AI and monitoring technologies can enable iterative AI quality

Thank you

Will Uppington

will@truera.com

www.truera.com

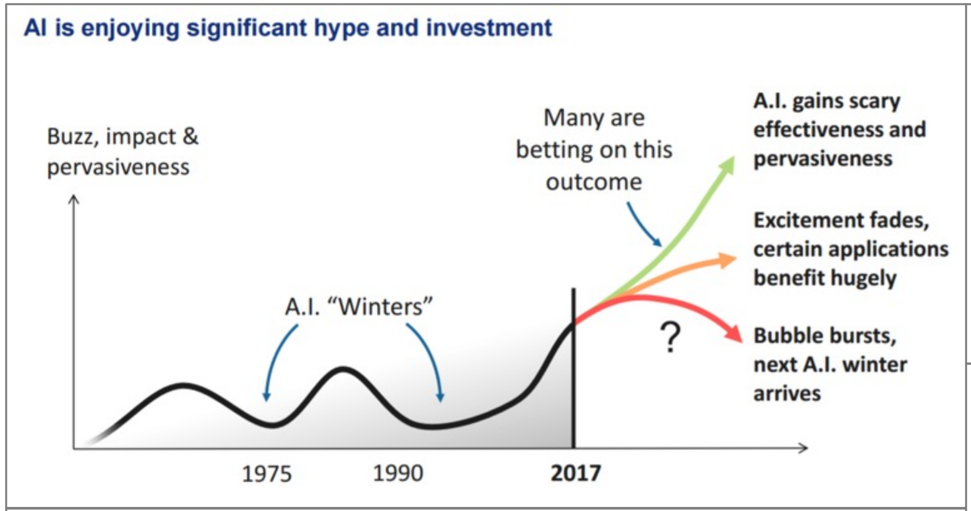


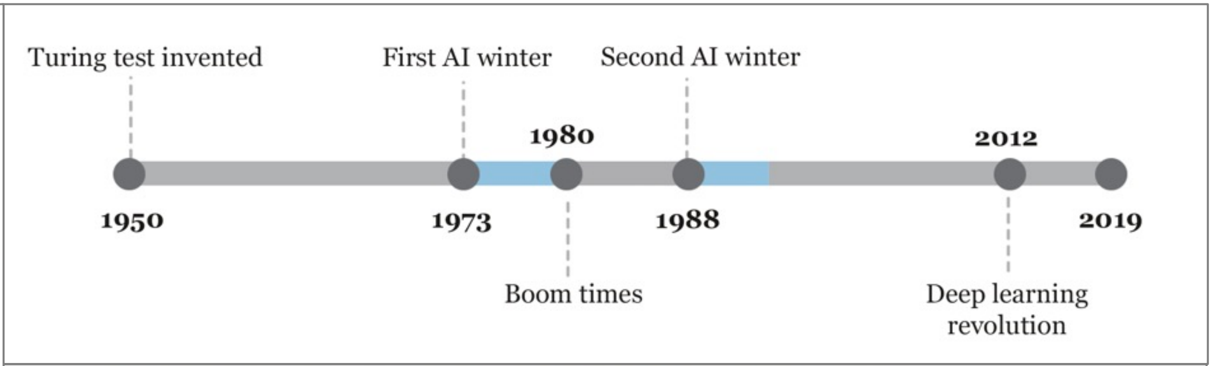
From Data to Decisions and Back

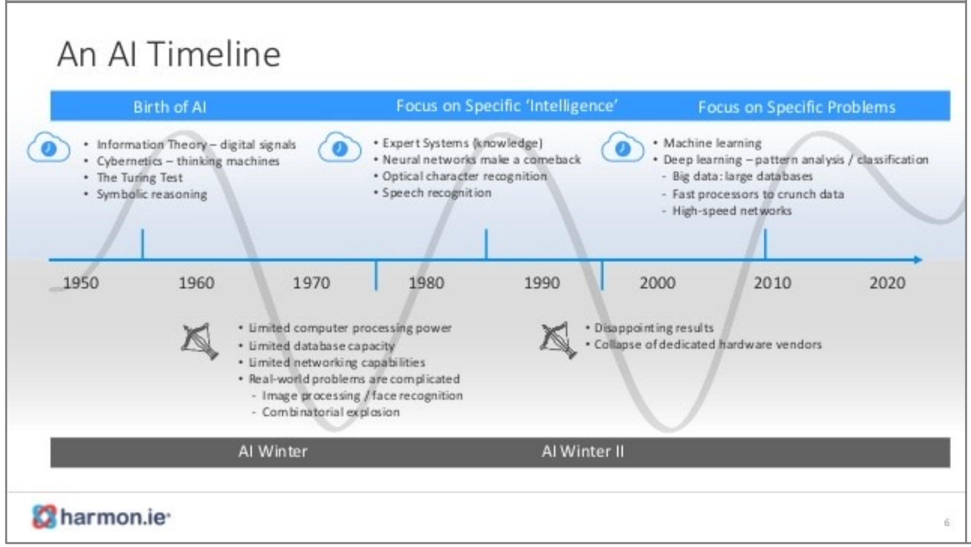
Babak Hodjat
CTO-AI

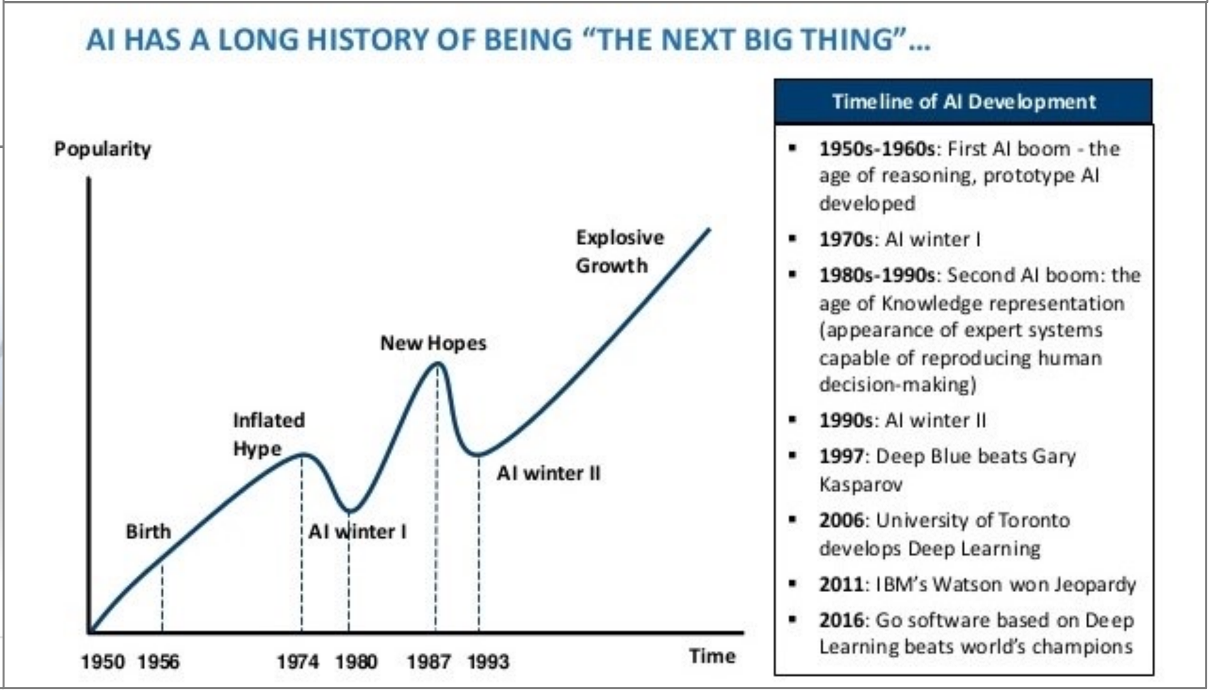
May 24th, 2022

Another AI Winter?









cognizant

There is Always Room For AI

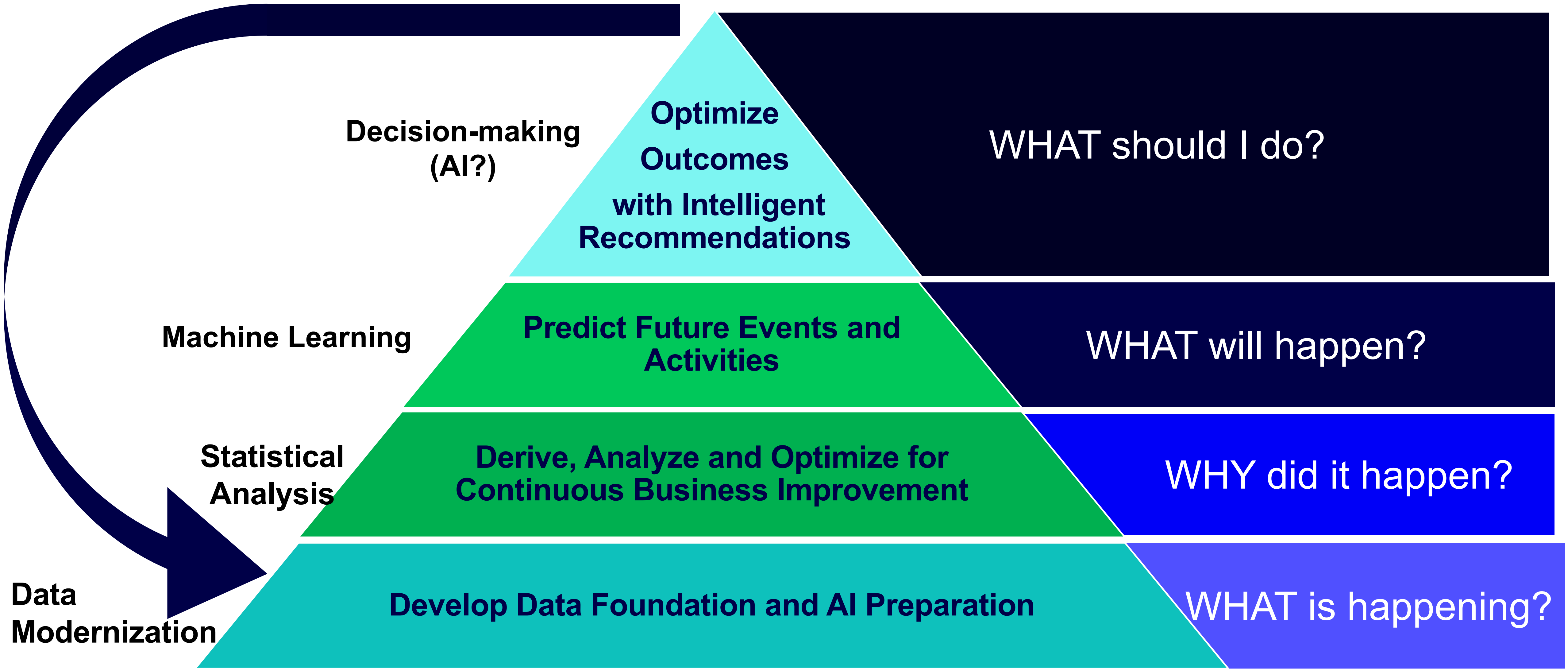
Science	Ideology
Known	Unknown

Software Engineering	AI
Programable	ML + Search + Heuristics

What Should we Expect from an AI System?
AI Should Tell Us:

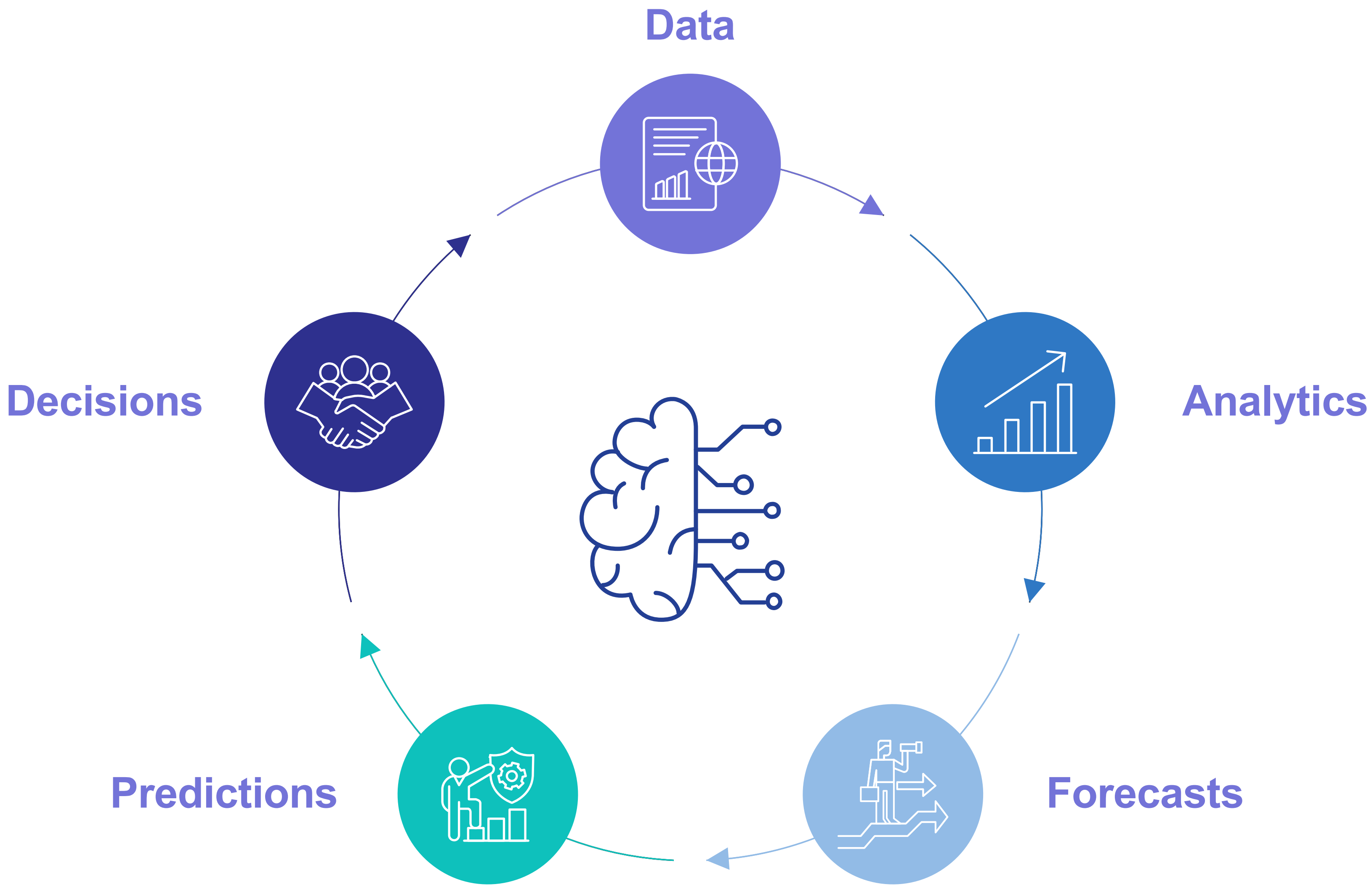
- What to do
- When to do it
- What would happen if we do it
- How much should we trust it
- What would happen if we do something else…

The AI Journey Should come Full Circle



cognizant

The Scope of AI Enablement: Step-by-step vs Wholistic



What is wrong with this picture?

Discovery + Planning

Design the solution

1. Assess the data available within XXX
2. Mapping of data elements
3. Evaluate the metadata availability and fitment
4. Finalize the technical architecture

Data Preparation + Setup

Prepare the Data

1. Setup technology Infrastructure
2. Gather metadata and other data
3. Engineer the metadata into required format
4. Ingest the data and training dataset

Asset Evaluation

MVP for aiding asset evaluation

For a job input by the agency, predict:

1. How likely it is to pass each dimension of evaluation along with
2. An estimation of certainty in this prediction
3. Identify and mark the probable areas of concern on the asset

Campaign Effectiveness

MVP for aiding campaign effectiveness

Predict whether:

1. The job is likely to receive regulatory complaints after it has been approved
2. The campaign will be effective
3. Identify and mark areas of probable concern on the asset

Suggestion

MVP for aiding asset design

Suggest how:

1. Elements of the job could be improved
2. Allow users to edit the suggestions
3. Predict how likely is it for the edits t
   a. Pass evaluations
   b. Receive complaints, and
   c. Perform well in the marketplace

Data

Decisions

cognizant

AI Hesitancy

Our data is a mess/not ready

DIFFERENT SPELLING

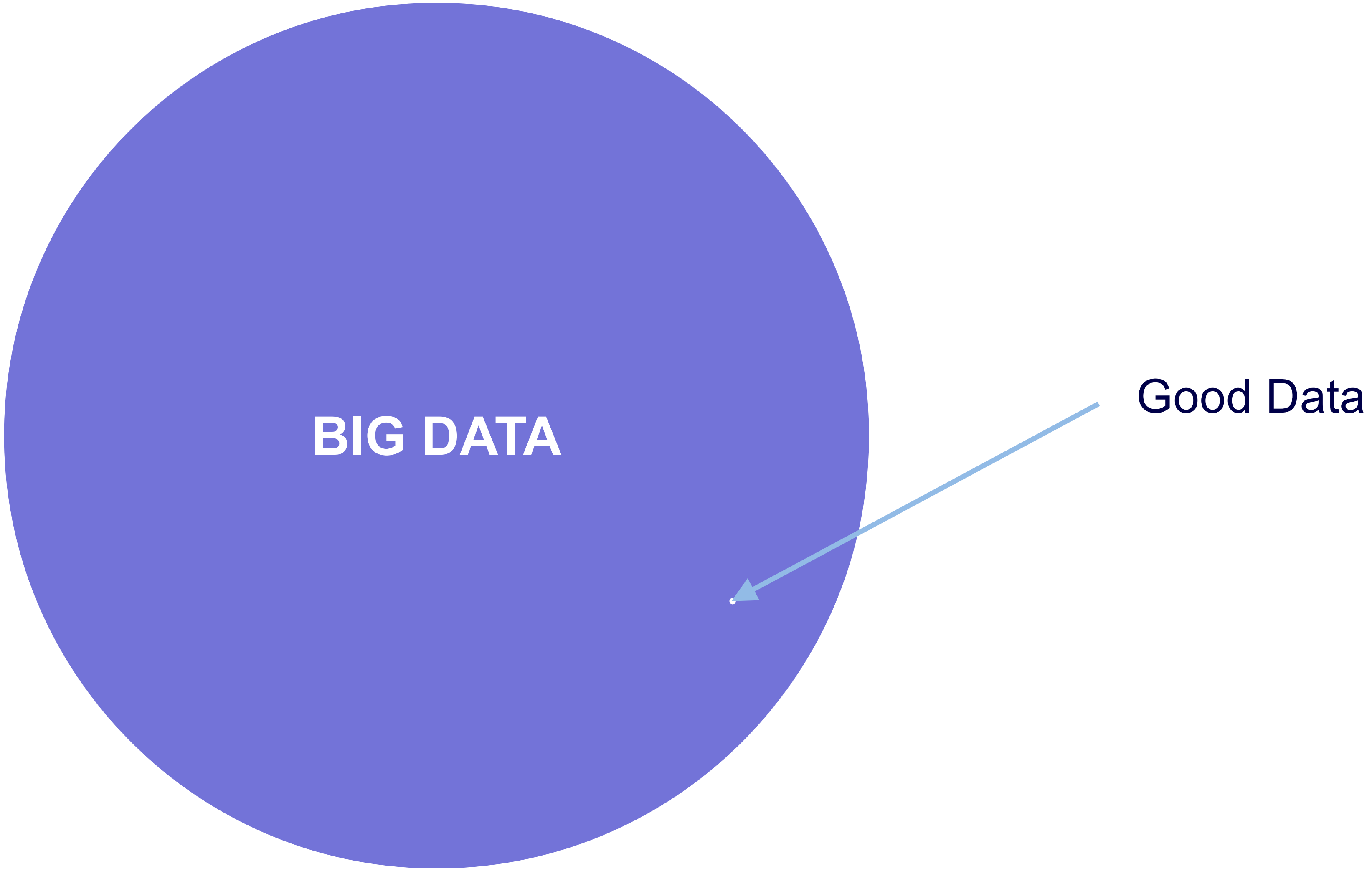
EMPTY ROWS

ONE PERSON, MULTIPLE ENTRIES

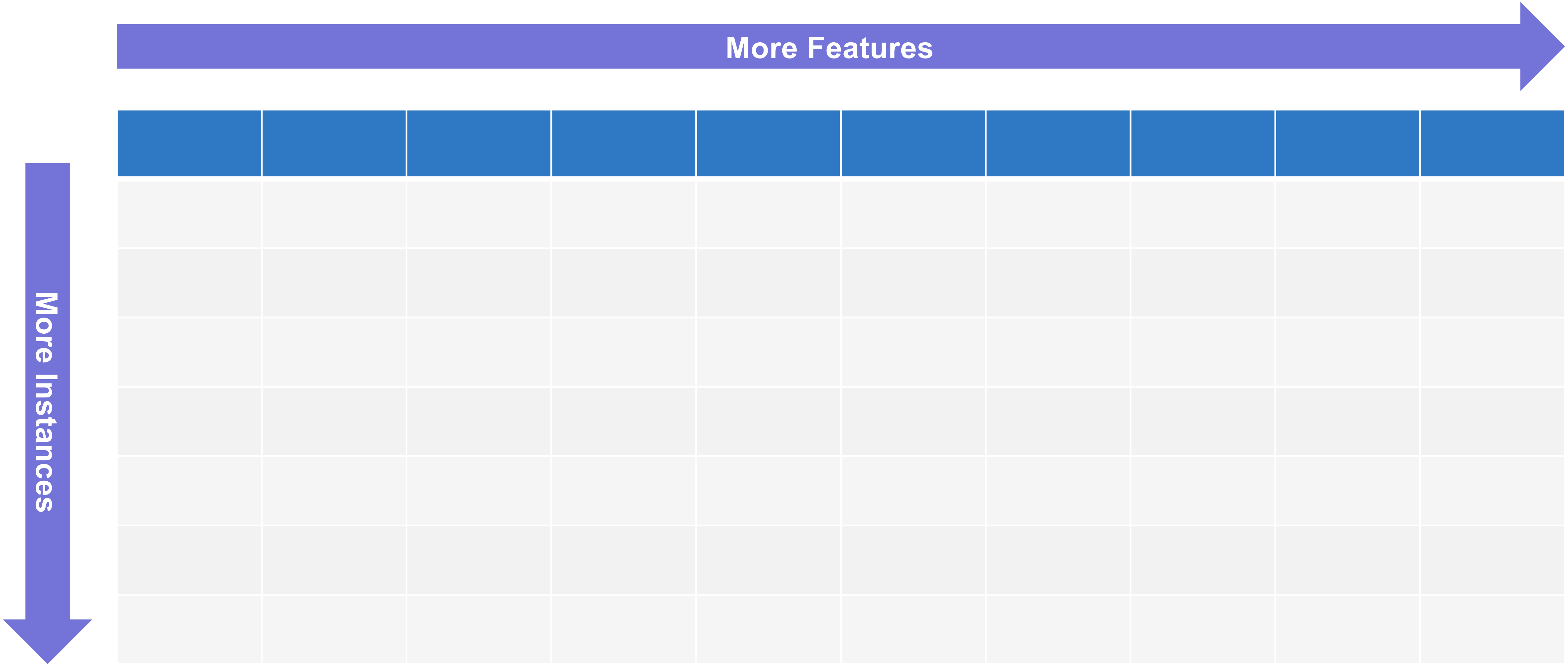
Timestamp	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
12/5/2016 13:18:46	Chris Staples, Golden Age	Kaytranada, 99.9%	Michael Kiwanuka, Love & Ha	Arkells, Morning Report	Hannah Georgas, For Evelyn
12/5/2016 13:19:50	Radiohead, A Moon Shaped	James Blake, The Colour in A	Bon Iver, 22, A Million	Chance the Rapper, Coloring	Anderson .Paak, Malibu
12/5/2016 13:19:51	Frightened Rabbit, Painting o	Wilco, Schmilco	Weezer, White Album	Drive-By Truckers, American	Band of Horses, Why Are You
12/5/2016 13:19:55	Childish Gambino, "Awaken,	Basia Bulat, Good Advice	Kanye West, The Life of Pabl	Frank Ocean, Blonde	Tacocat, Lost Time
12/5/2016 13:20:09	Nada Surf, You Know Who Yo	The Jayhawks, Paging Mr. Pr	The Posies, Solid States	Purusa, Soundtrack	
12/5/2016 13:20:10	William Tyler, Modern Countr	Hiss Golden Messenger, Hea	BJ Barnham, Rockingham	Mandolin Orange, Blindfaller	The Avett Brothers, True Sad
12/5/2016 13:20:15	22 A Million, Bon Iver	Coloring Book, Chance the R	We've Got It From Here, Thx	Moon Shaped Pool, Radiohea	The Colour in Anything, Jame
12/5/2016 13:20:19	The Posies, Solid States				
12/5/2016 13:20:28					
12/5/2016 13:20:31	ZAYN, Mind of Mine	Ariana Grande, Dangerous W	Solange, A Seat at the Table	Chance the Rapper, Coloring	dvsn, SEPT. 5TH
12/5/2016 13:21:01					
12/5/2016 13:21:07	BTS, Wings	Agust D, Agust D	Lindsey Stirling, Brave Enoug	Babymetal, Metal Resistance	Red Velvet, Russian Roulette
12/5/2016 13:21:11	Hope Sandoval, Until the Hur	Madness, Can't Touch Us No	Suzanne Vega, Lover, Belove	Poliça, United Crushers	Underworld, Barbara Barbara
12/5/2016 13:21:40	Beyonce, Lemonade	Radiohead, A Moon Shaped	Parquet Courts, Human Perfo	Cate Le Bon, Crab Day	Nels Cline, Lovers
12/5/2016 13:21:43	Zayn, Mind of Mine	Banks, The Altar	Beyonce, Formation	Rihanna, Anti	Jojo, Mad Love.
12/5/2016 13:21:59	Josh Ritter, Sermon on the R	Lori McKenna, The Bird & the	Sara Watkins, Young in All the Wrong Ways		
12/5/2016 13:23:06	Tancred, Out of the Garden	Jimmy Eat World, Integrity Bl	Sam Beam & Jesca Hoop, Lo	The Lumineers, Cleopatra	Sara Watkins, Young In All Th
12/5/2016 13:23:21	Anderson .Paak, Malibu	Toro y moi, Live from Trona	Kendrick Lamar, untitled unm	Frank Ocean, Blonde	Chance the Rapper, Coloring
12/5/2016 13:23:59	Weezer, Weezer	Death Grips, Bottomless Pit	Jon Bellion, The Human Cond	David Bowie, Blackstar	Explosions in the Sky, The W
12/5/2016 13:24:09	RVRB, FADED	Views, Drake	Latigo, Wilderado	TLOP, Kanye West	22 a million, Bon Iver
12/5/2016 13:24:16	Teenage Fanclub, Here	The Posies, Solid States	Phantogram, Three	Wilco, Shmilco	Monkees, good times
12/5/2016 13:25:10	RVRB, FADED	RVRB, FADED	RVRB, FADED	RVRB, FADED	RVRB, FADED
12/5/2016 13:25:11	AJJ, The Bible 2	Jeff Rosenstock, WORRY	Eric Bachmann, S/T	Joyce Manor, Cody	Kanye West, Life of Pablo
12/5/2016 13:25:26	Care Le Bon, Crab Day	David Bowie, Blackstar	Weaves, Weaves	Trentmøller, Fixion	Solange, a Seat at te Table
12/5/2016 13:25:26	RVRB, FADED	RVRB, FADED	RVRB, FADED	RVRB, FADED	RVRB, FADED
12/5/2016 13:25:28	Carseat Headrest, Teens of D	Beyonce, Lemonade	Radiohead, Moon Shaped Po	Margaret Glaspy, Emotions a	Dawes, We're All Gonna Die

cognizant

We don’t have enough Data

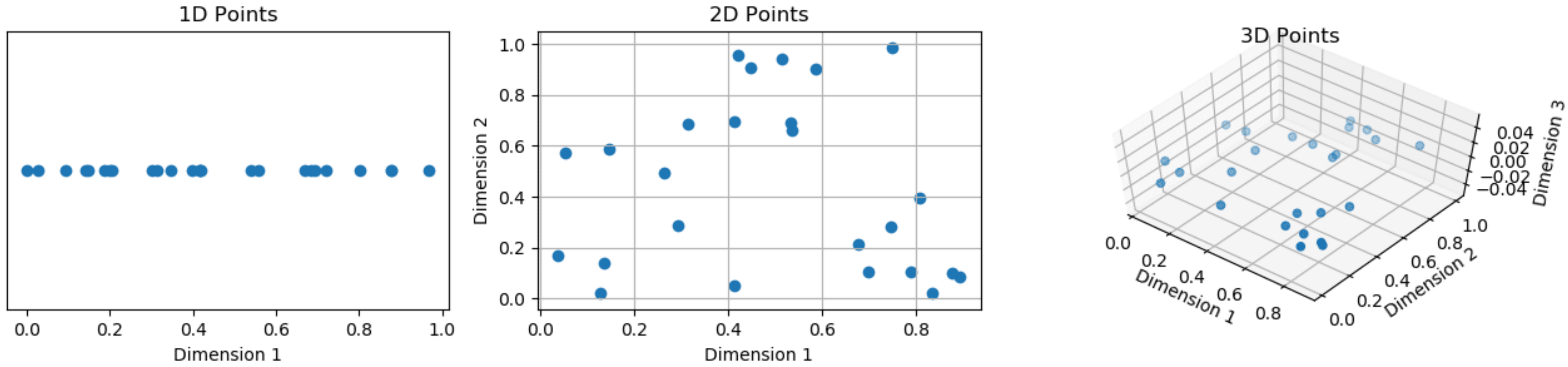


More data does not mean, more features!



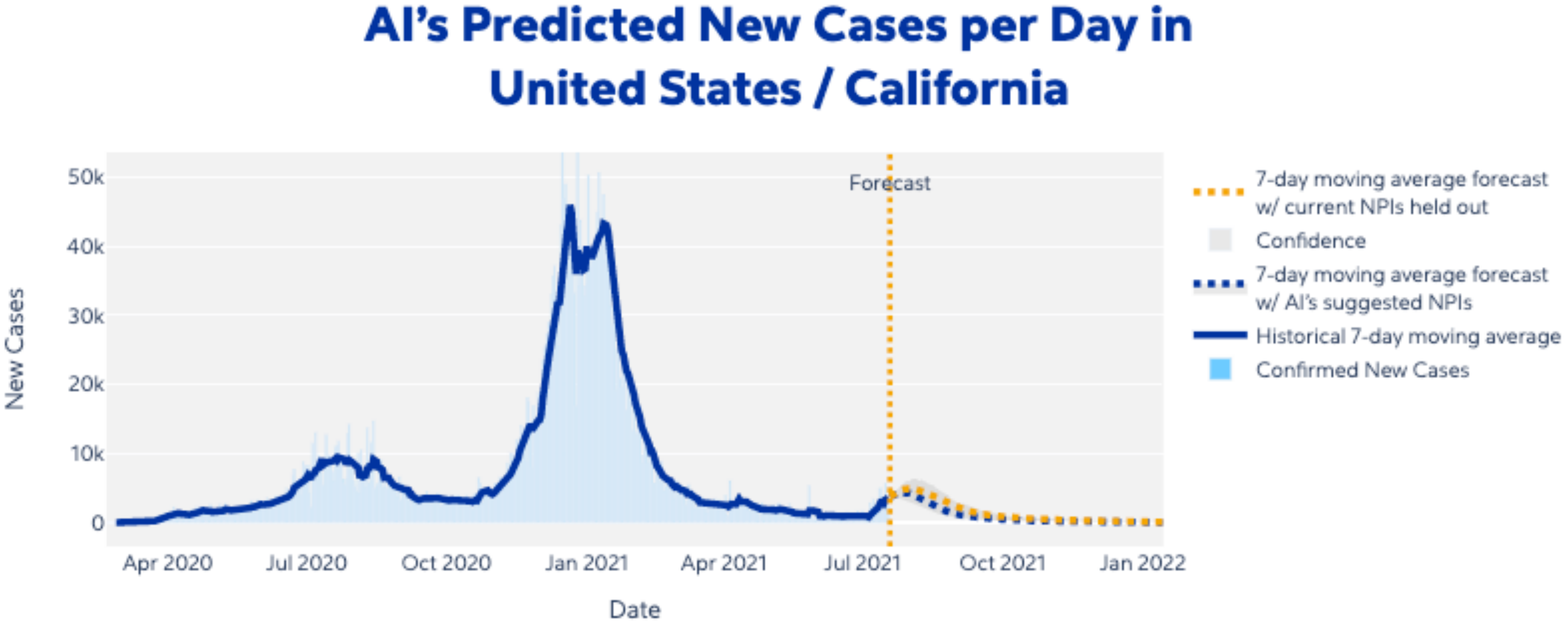
The Curse of Dimensionality

More features means exponentially more observations to avoid overfitting



cognizant

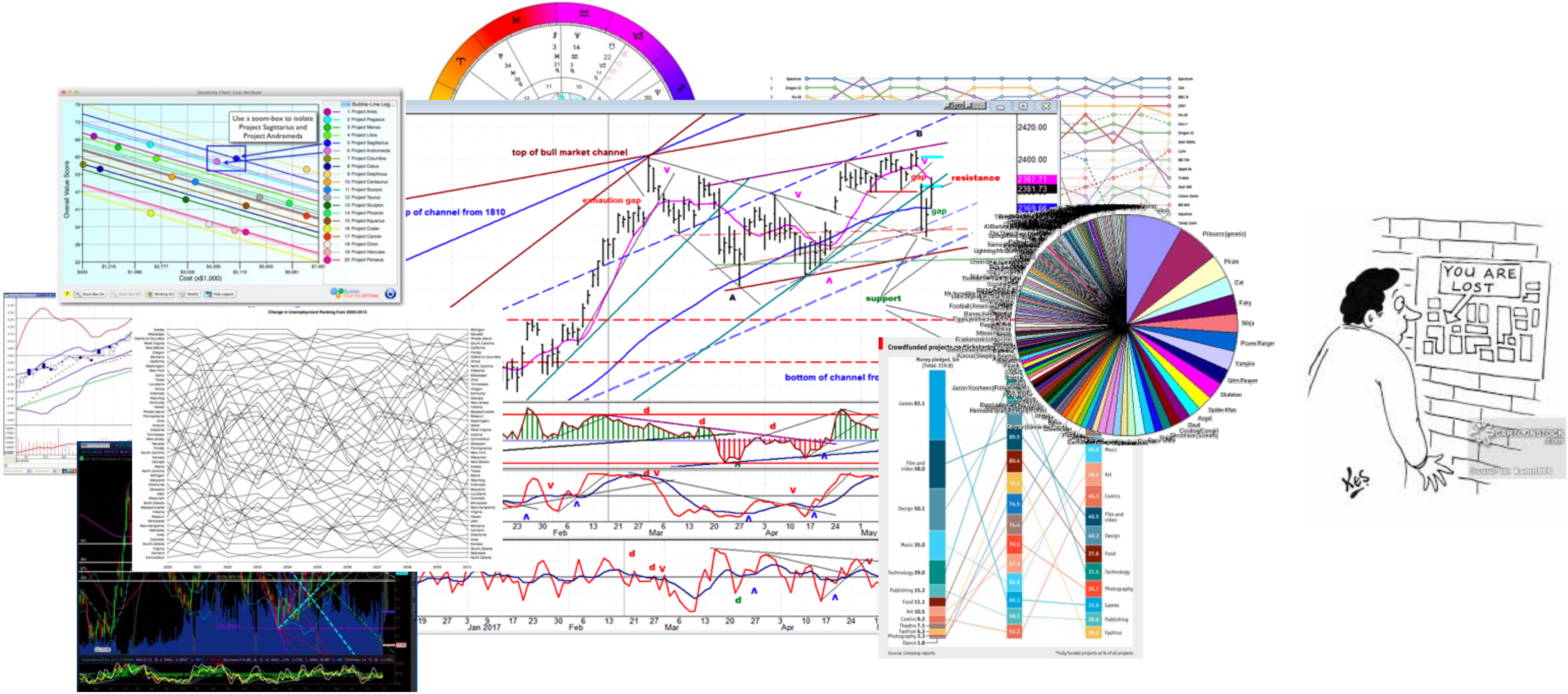
Outcomes Timeseries Captures Effect of Missing Features



No One-Size-Fits-All AI

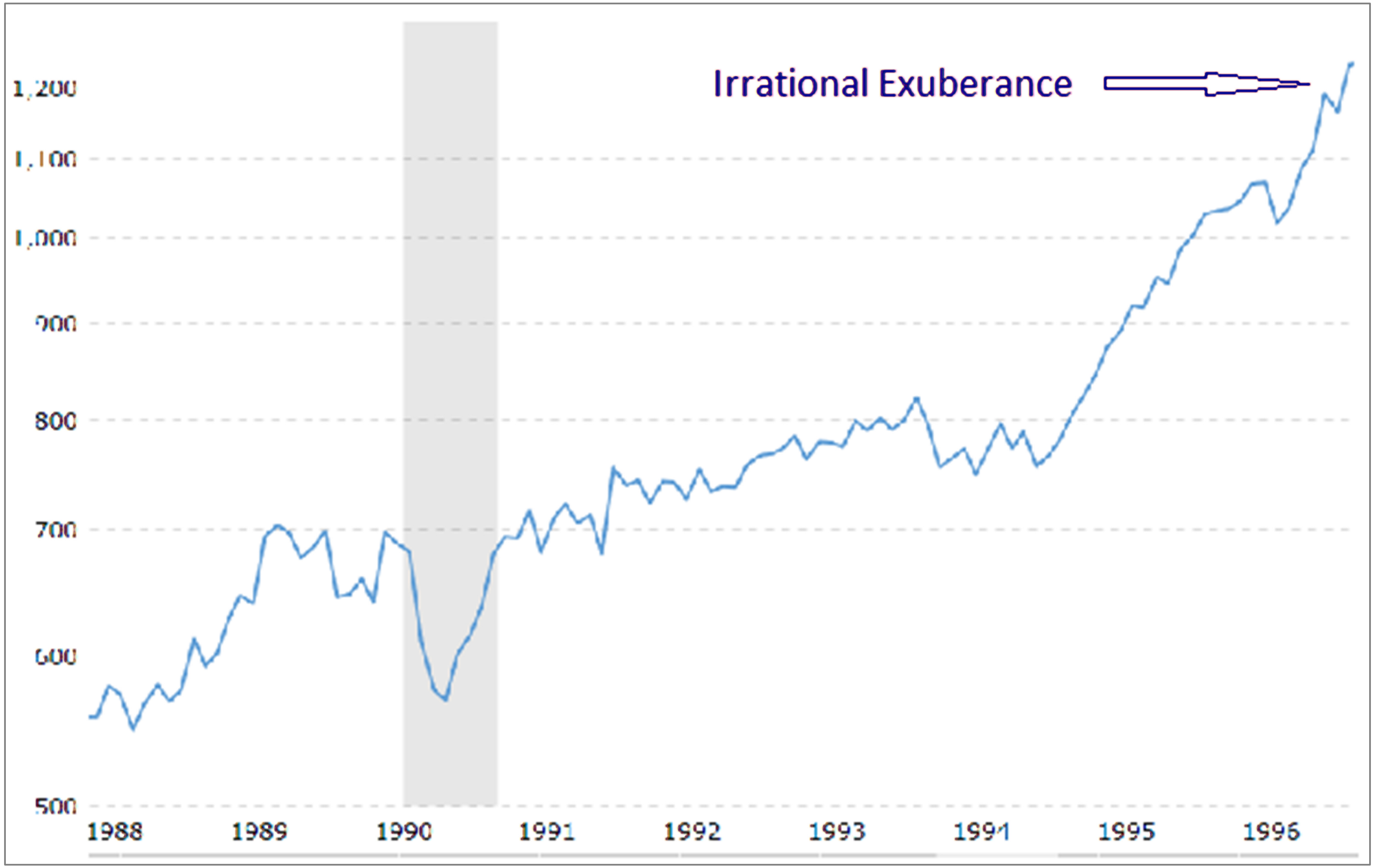
Domain Specific Models are the new IP

Need to get Insights First

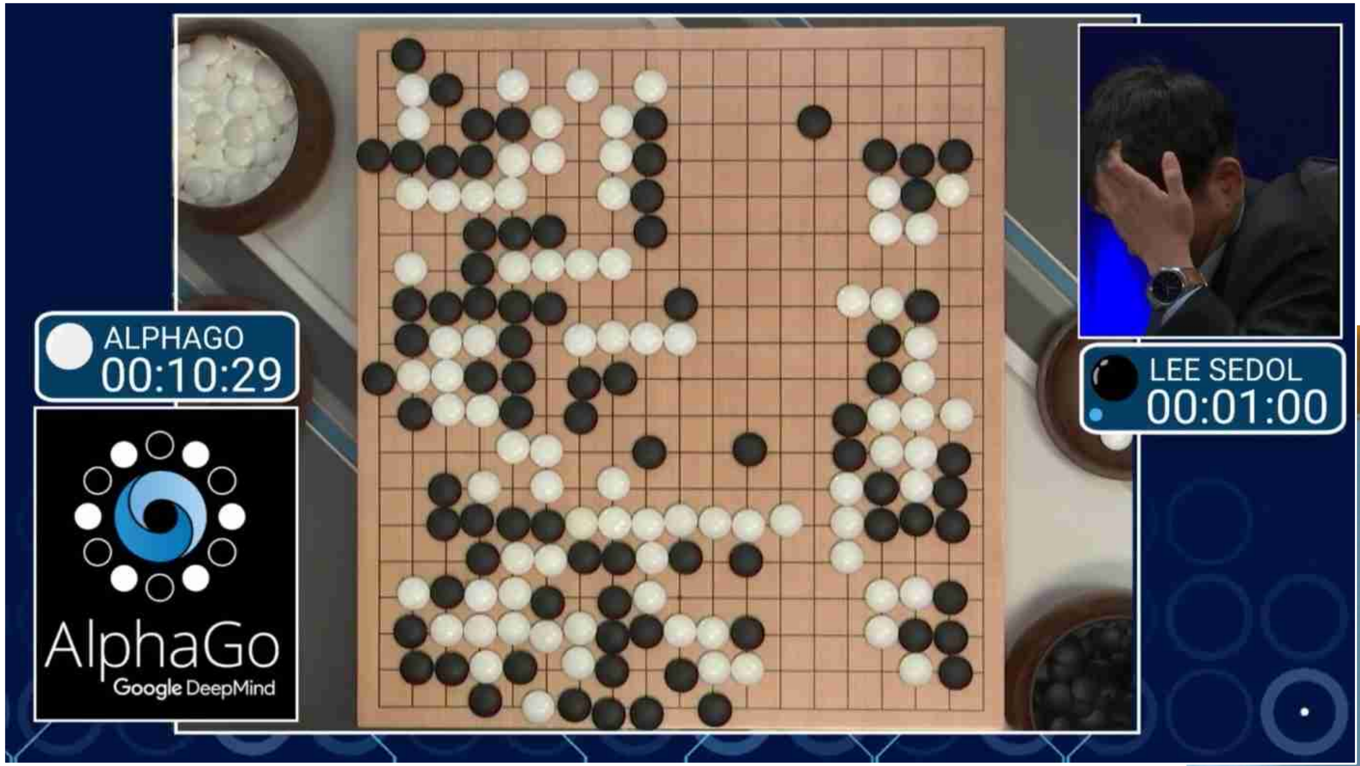


cognizant

What if Insights Show we are Good Already?



Human Decision-Making is Better



All we need is better insights - we can make the decisions

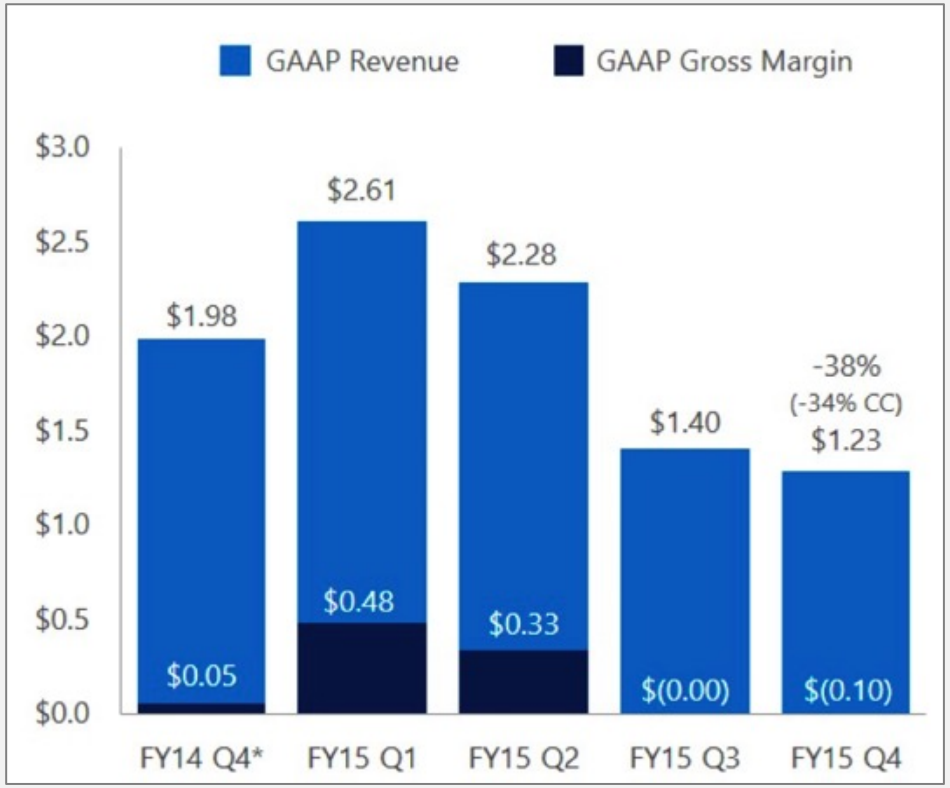
Projections Show we're Missing Revenue Targets!

Should we cut costs?

Sell at a discount?

Retarget sales?

Invest in growth?



We care about sales – Cost is for a different department

There is never a single objective!

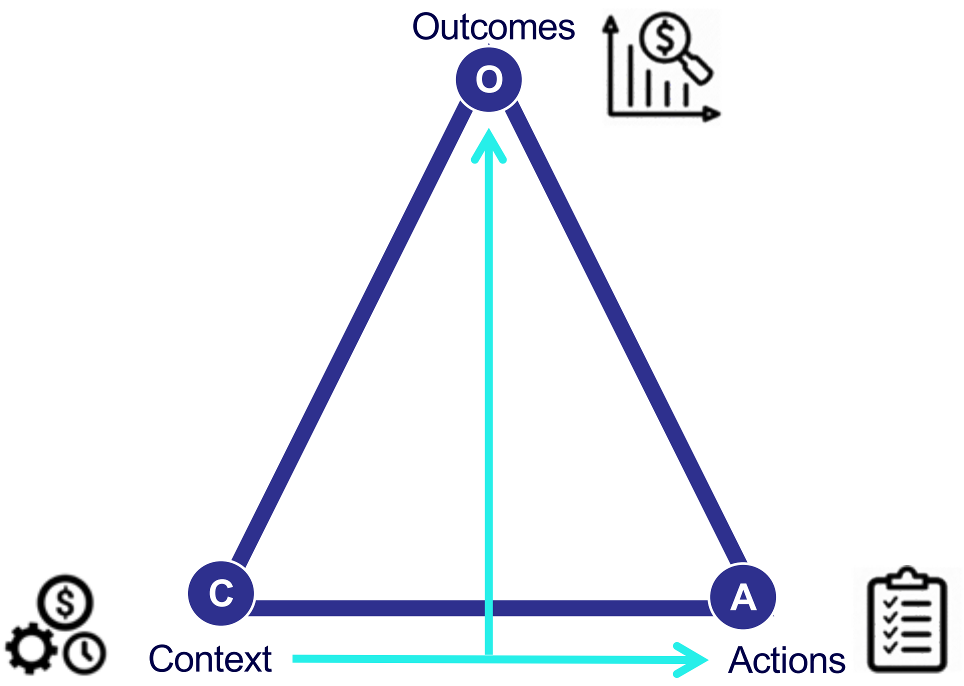
We often want to:

- Maximize revenue, and
- Maximize sales, while
- Minimizing cost, and,
- Minimizing risk

Worried about bias and ethics of using AI

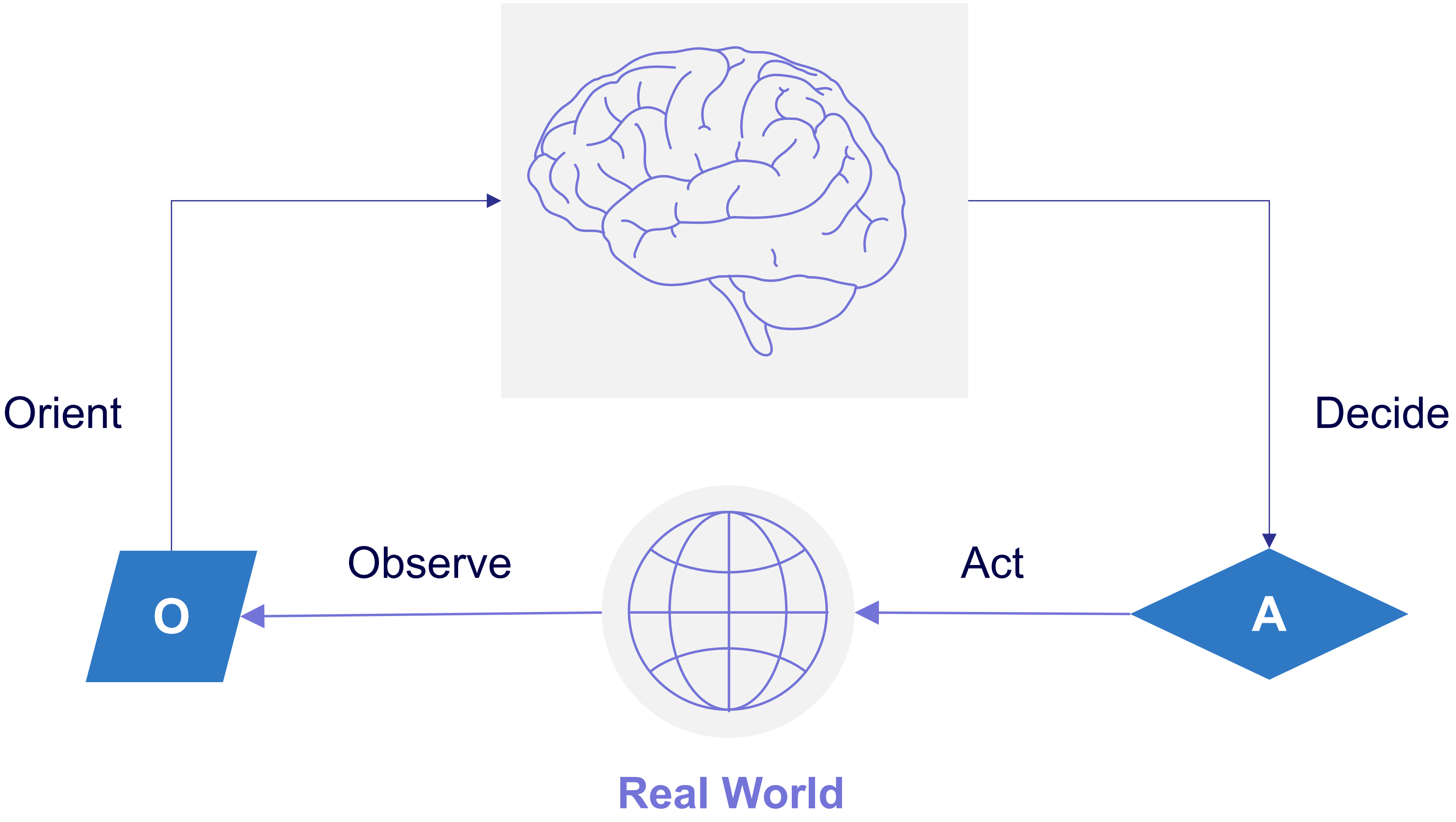
"And finally this is Phil Goodwin. He's part of the problem."

We make Decisions all the time

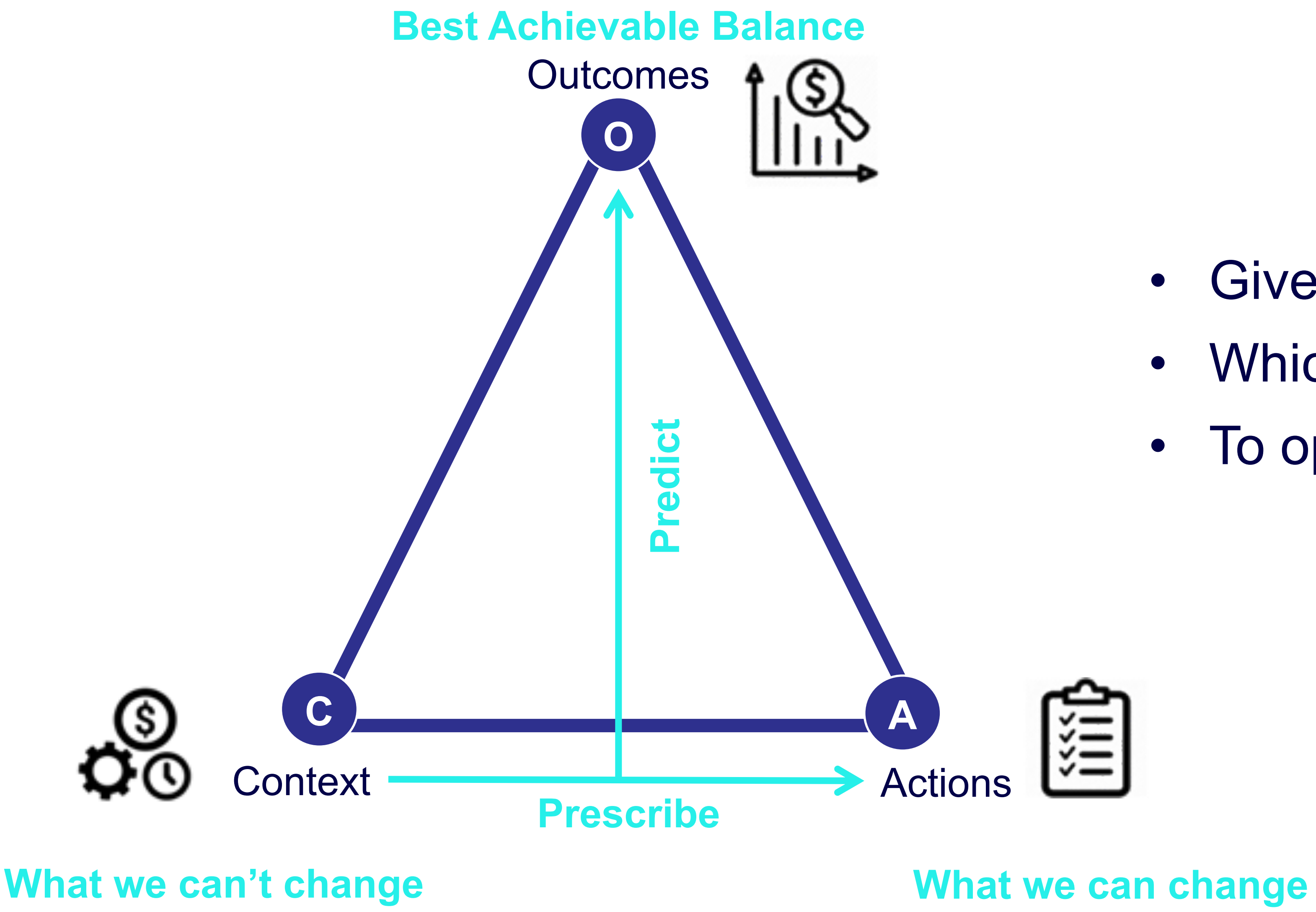


- Given a **context** (C)
- Which **actions** should I take (A)
- To optimize some **outcomes** (O)?

Adaptive Decision-Making: The OODA Loop

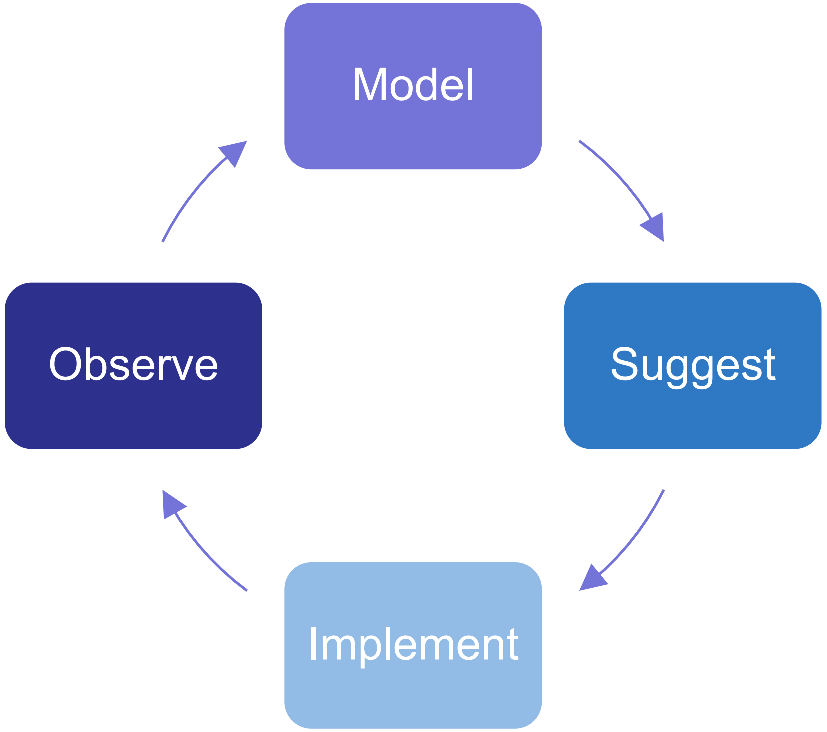


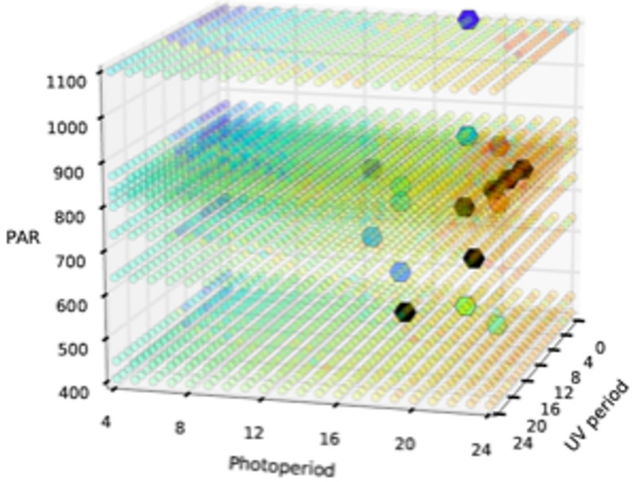
How do we make decisions?



- Given a **context** (C)
- Which **actions** should I take (A)
- To optimize some **outcomes** (O)?

AI-Enabled Decisioning (and Data Collection)





Important Questions AI can Answer

Predictor:

What will happen?

Prescriptor:

What should we do?

Uncertainty:

How much can we trust it?

Scratchpad:

What-if we do something else?

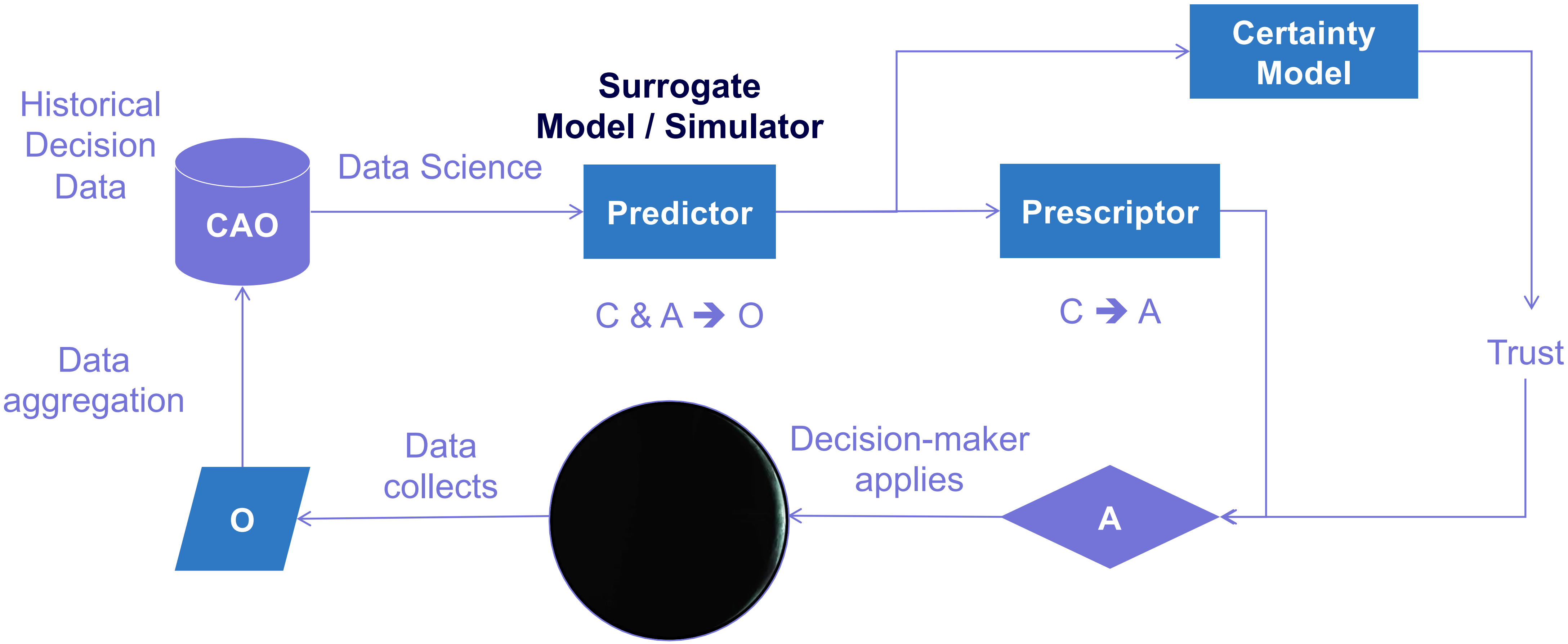
Impact Estimation (Retrospective):

What if we had done something else?

Multiple-Objectives:

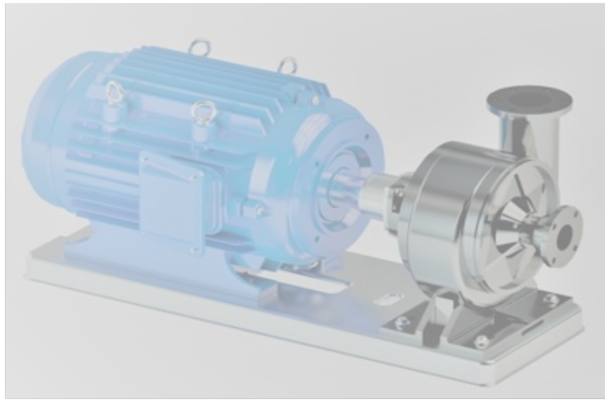
What if we want a different balance of outcomes?

How it all comes together



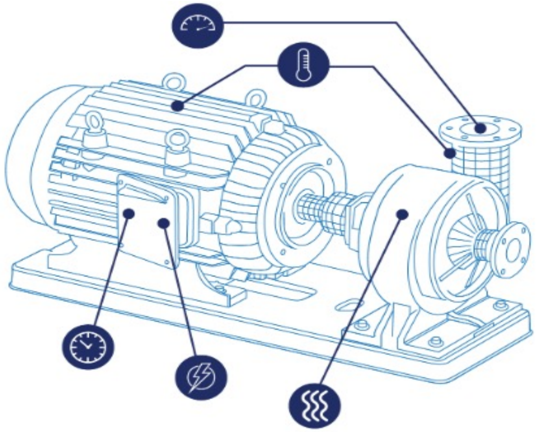
Digital Twins

Physical environment



- Design Systems
- Enterprise Data
- Sensor Data
- Historical data
- Image data
- Acoustic data
- ....

Digital Twin Instance



- Vibration
- Heating Temperature
- Flow Speed
- Viscosity
- Vapor pressure
- Power Consumption monitor
- Chamber humidity
- Current and voltage
- ....

Prediction and Optimization

- Simulation
- What-if scenarios
- Prediction and prescription
- Optimization
- Predictive maintenance
- Remaining useful life

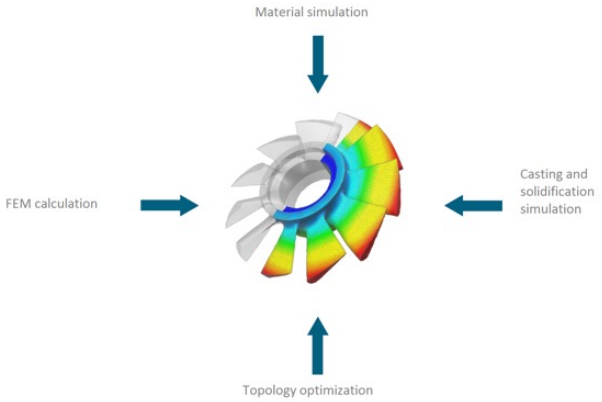
Rigid Simulations

Make Assumptions
- May be unavailable
- May be wrong
- May be inaccurate
- May be insufficient

Create Model
- Requires knowledge of mechanisms and dynamics

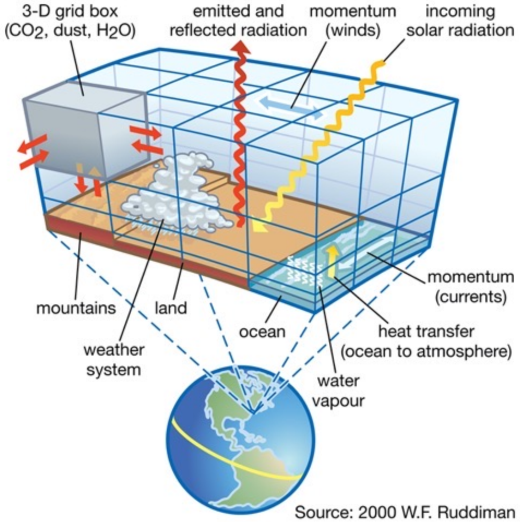
Run Model
- Often Expensive at scale
- Often inaccurate in miniature

Interrogate Model
- Rigid



Digital Twin

Concept diagram of climate modeling



Source: 2000 W.F. Ruddiman

SEIR: Susceptible Exposed Infected Recovered Compartment Model

Refer: Julie C. Blackwood & Lauren M. Childs (2018) An introduction to compartmental modeling for the budding infectious disease modeler, Letters in Biomathematics, 5:1, 195-221, DOI: 10.1080/23737867.2018.1509026

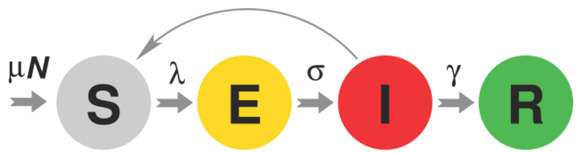
$$\frac{dS}{dt} = \mu N - \lambda(I)S - \mu S,$$

$$\frac{dE}{dt} = \lambda(I)S - (\sigma + \mu)E,$$

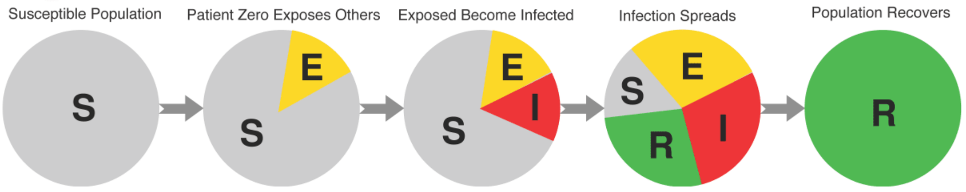
$$\frac{dI}{dt} = \sigma E - (\gamma + \mu)I,$$

$$\frac{dR}{dt} = \gamma I - \mu R,$$

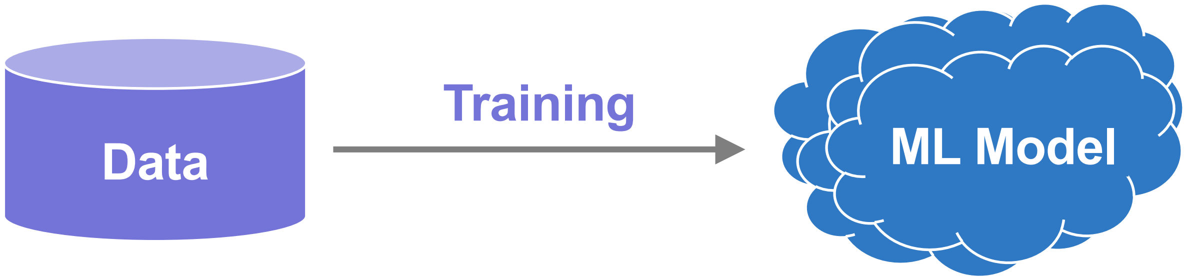
Infected Transfer Contagion



S - Susceptible, Fraction of Population Susceptible to Infection
E - Exposed, Fraction of Population Exposed / Not Yet Contagious
I - Infected, Fraction of Population Infected and Contagious
R - Recovered, Fraction of Population Recovered



ML Models

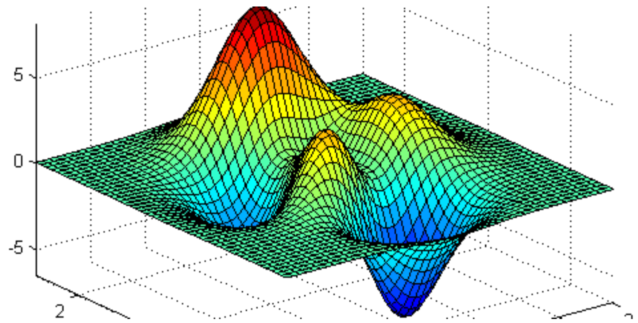


Robust

Data Driven

Does not require (as much) knowledge of the domain

Can be applied at scale



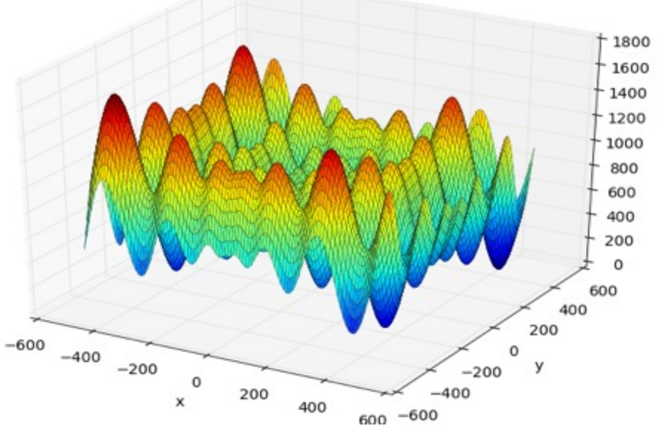
<u>BUT</u> **not enough**

Data-hungry

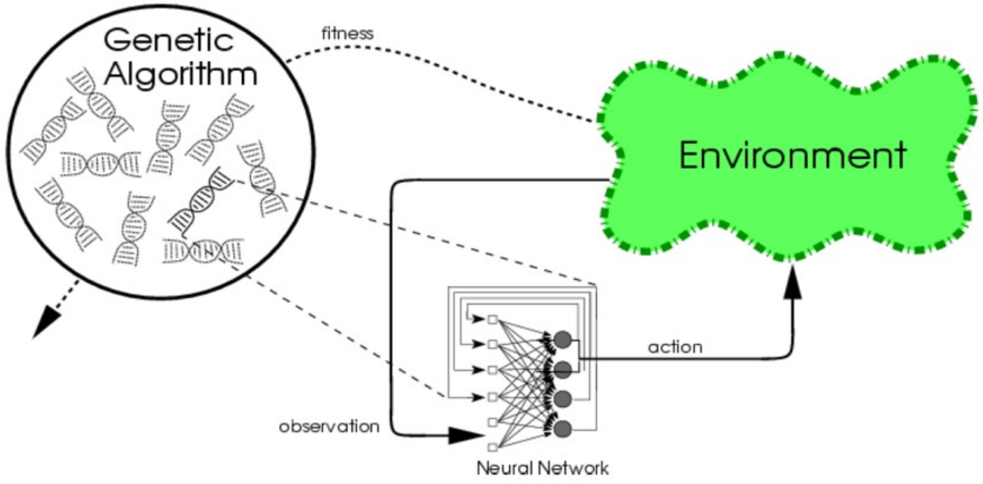
Training could be expensive/time consuming

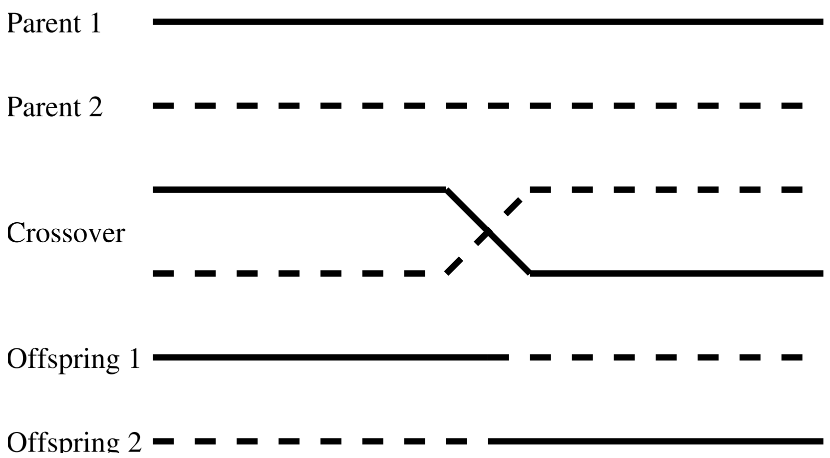
Not adaptive

Gradient Descent is not good at large, deceptive, multi-objective search spaces



Why use Evolutionary AI?



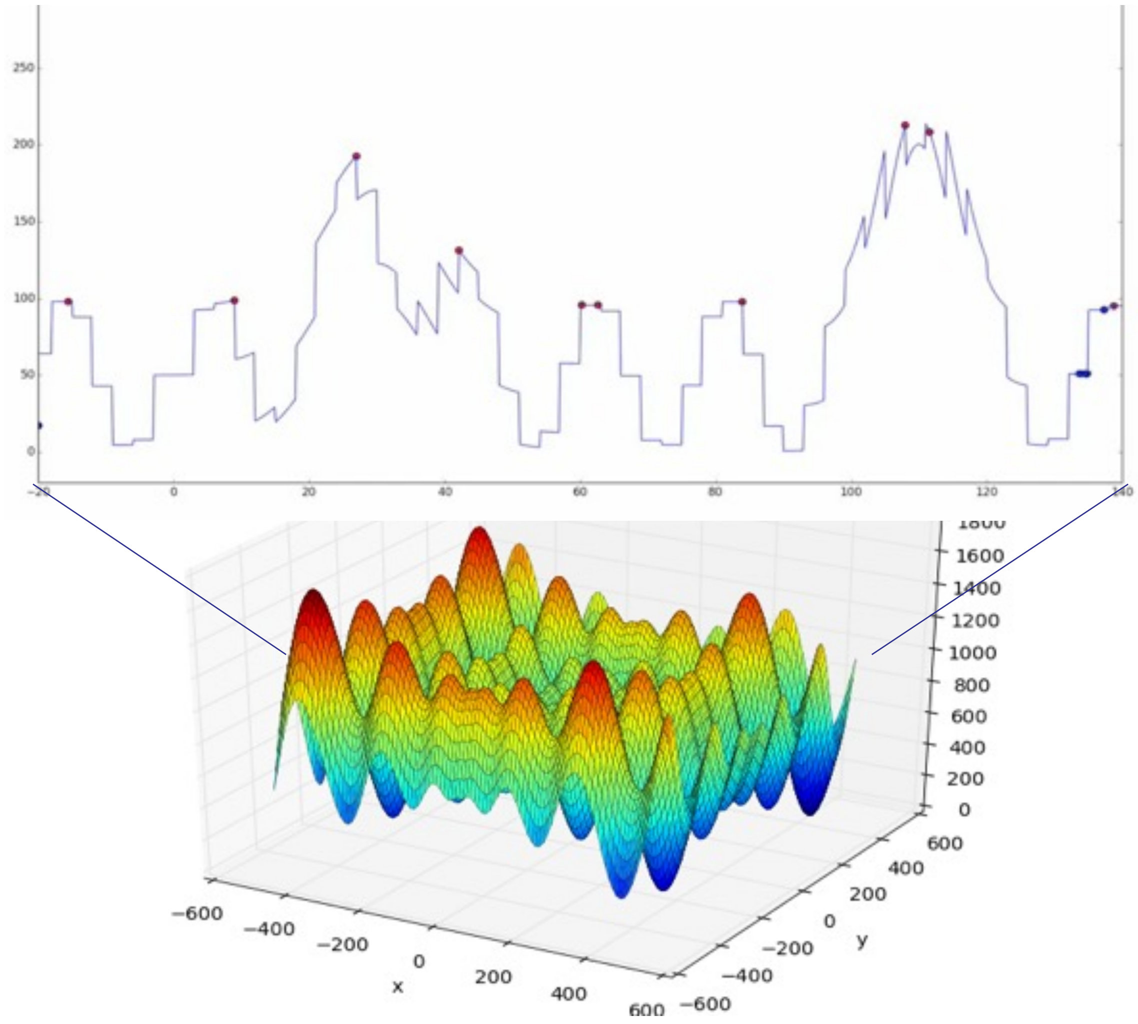


- A population of encodings (e.g., lists or trees)
- Decoded into individuals that are tested
- Good individuals retained, bad thrown away

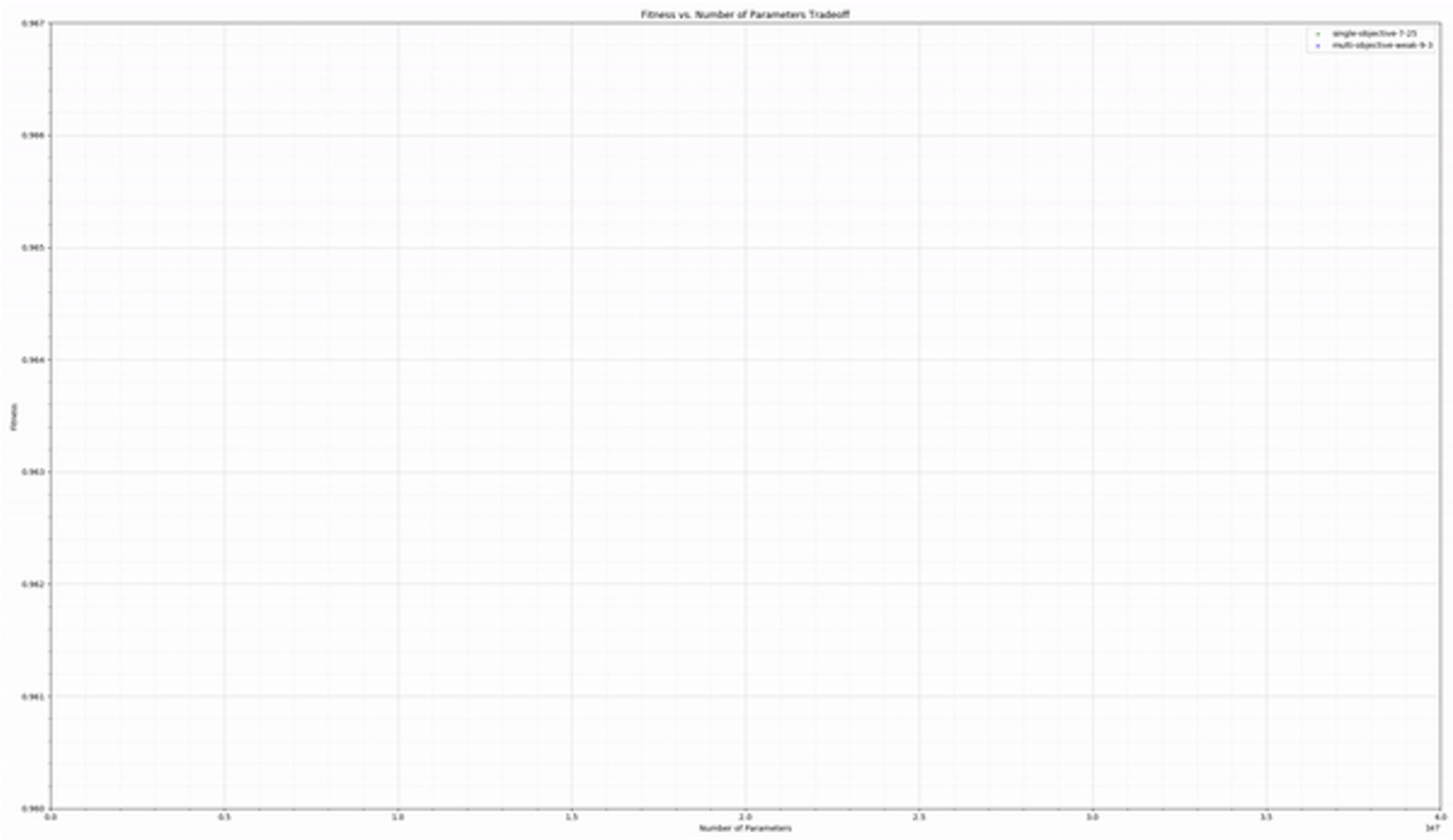
- New individuals generated from the parent encodings
  - Crossover: Combine building blocks from two parents
  - Mutation: Create new building blocks

Evolution is Efficient

- Works in large spaces (e.g., 2^2^70)
- Multiple variables optimized at once (e.g., up to 1B)
- Multiple objectives and novelty get around deception

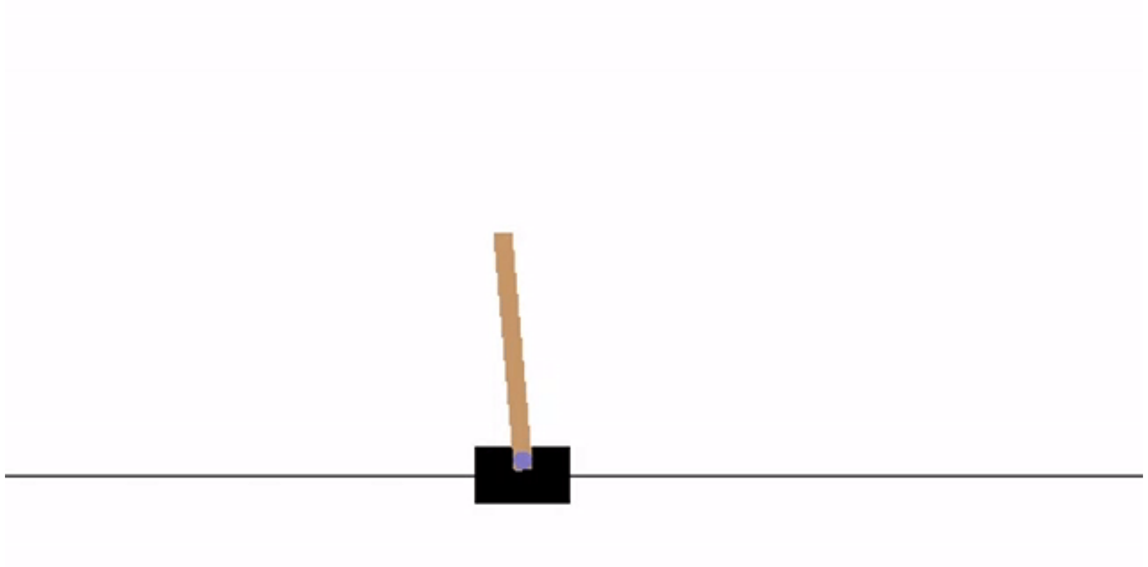


Evolution can handle Multiple Objectives



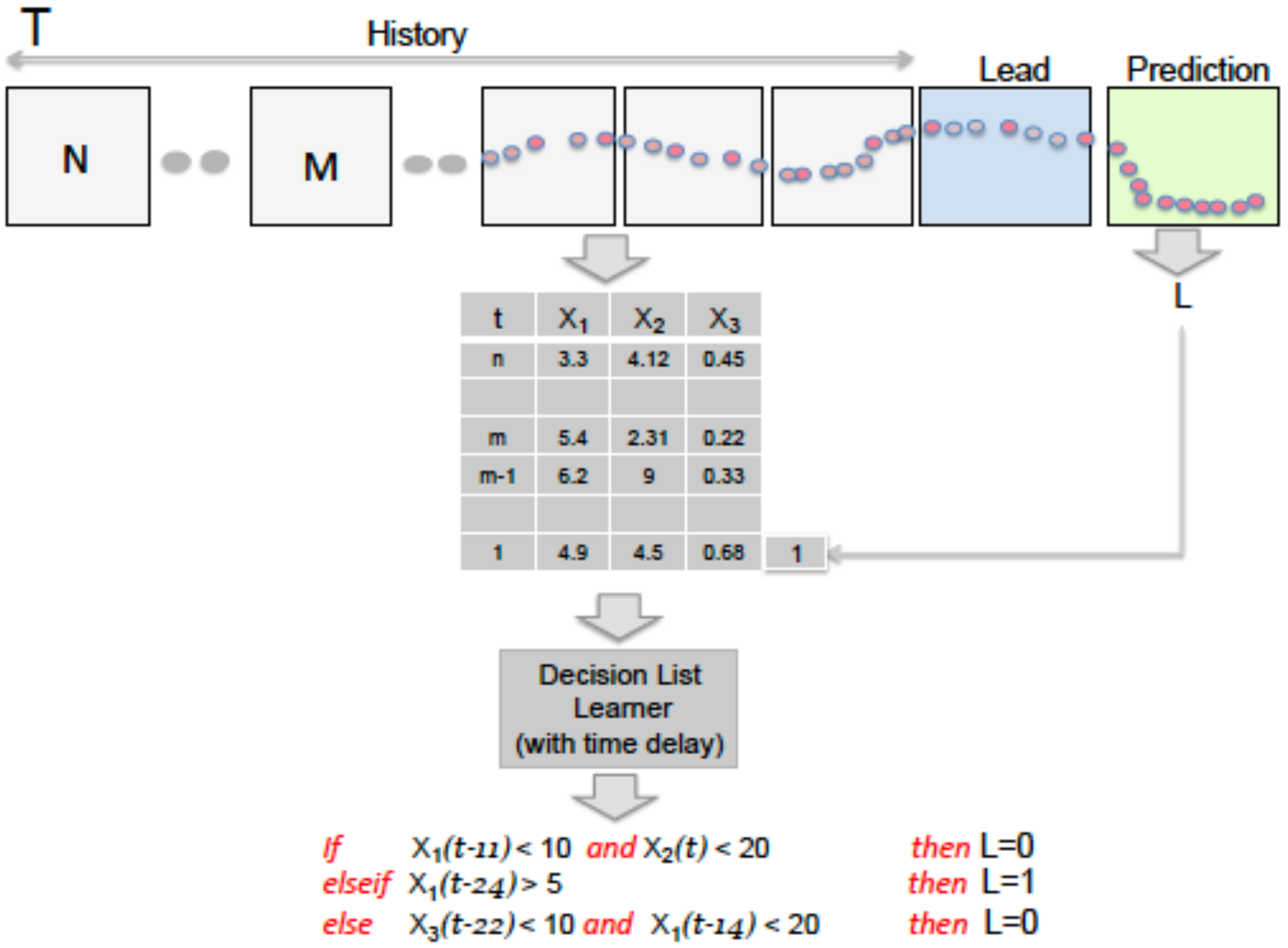
- Animation: Pareto front by generation for single-objective (green) vs. multi-objective (blue)
- Single-objective focuses on improving **largest networks**
- Multi-objective focuses on improving the **entire curve**
- Result: Multi-objective finds much smaller models for the majority of performance values.

Evolution Can Create Explainable Solutions



- 3_53 Rules agent:
- <22978> (0.11*S_1^3 < 0.87*S_2) ---> **LEFT**
- <23022> Default Action: **RIGHT**

cognizant



Each rule is a variable length conjunction of conditions with an associated class prediction. In the evaluation each condition compares a lagged value or the current value of the time series to a threshold (decision boundary). Decision lists in LEAF have a variable number of rules and conjunctive clauses in each rule.

This representation is different from many other classifiers e.g. Decision Trees, simple Decision Lists, Support Vector Machines and Logistic Regression, which requires every time lagged value or an aggregate to be explicitly presented as a different feature.

Diabetes Rules

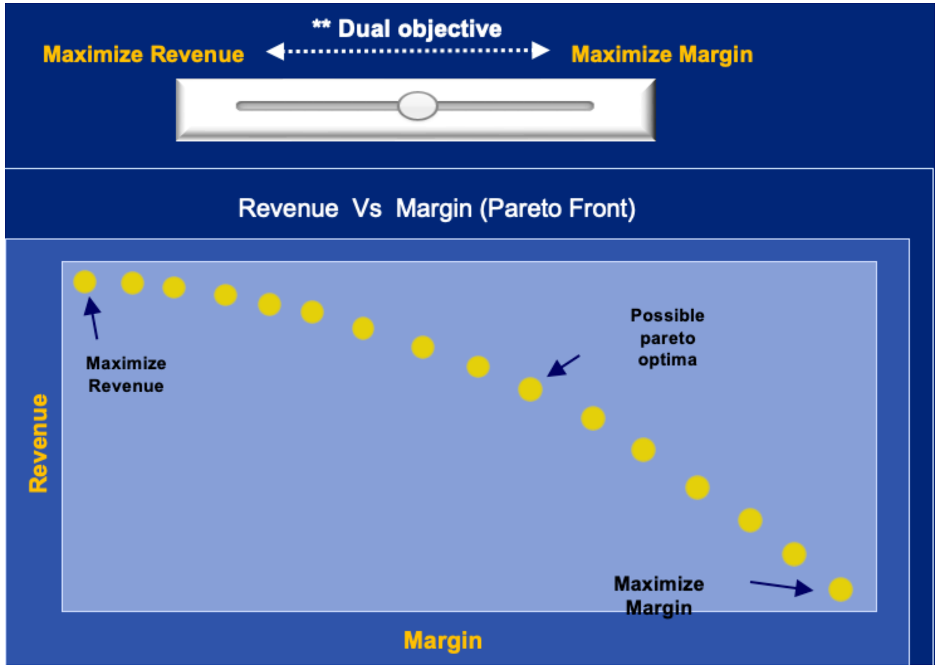
1. (diag_1_injury <= 2.2*diag_1_respiratory) and (diag_2_neoplasms > 33) —> 18
2. (age_[60-70) >= 1.2*age_[10-20))   --> 7
3. (diag_2_diabetes >= 0.82*diag_1_digestive) --> 17
4. (admission_source_id_Court/Law Enforcemenct >= 1.045*diag_3_circulatory) --> 9
5. (age_[80-90) >= 2.29*admission_type_id_Newborn) --> 14
6. (age_[30-40) <= 0.47*age_[50-60)) and
   (admission_type_id_Newborn <= 0.47*age_[50-60)) and
   (diag_2_respiratory >= 1.86*diag_3_musculoskeletal) and
   (diag_1_diabetes >= 1.42*age_[30-40)) and
   (0.18*gender >= 0.38*race_Asian) and
   (0.12*race_AfricanAmerican >= 0.15*admission_source_id_Court/Law Enforcement) and
   (0.03*race_Hispanic <= 0.76*admission_source_id_Emergency Room) --> 15
7. Default 8

Diabetes Rules Interpretation

1. If Patient has a respiratory problem but not a neoplasms problem prescribe **metformin-pioglitazone**
2. If the person any age except 60-70 prescribe **glyburide**
3. If the person has Diabetic Diagnostic result as positive prescribe **glipizide-metformin**
4. If the person has a positive circulatory diagnosis prescribe **pioglitazone**
5. If the admission type is Newborn prescribe **tolazamide**
6. If the person is between 30 and 60 years old and has a positive musculoskeletal diagnosis and is Asian and was sent by court/law enforcement and was admitted in emergency, prescribe **insulin**
7. If none of the above, prescribe **glipizide**

Some Examples

Pricing Optimization



	LEAF Recommendation	Scenario Planning
Price	$608 **	$397
Quantity	3942 **	8990
Revenue	$1,600,000 **	$2,300,000
Margin	$440,000 **	-$300,000

Improving pricing strategy for a leading home appliance manufacturer

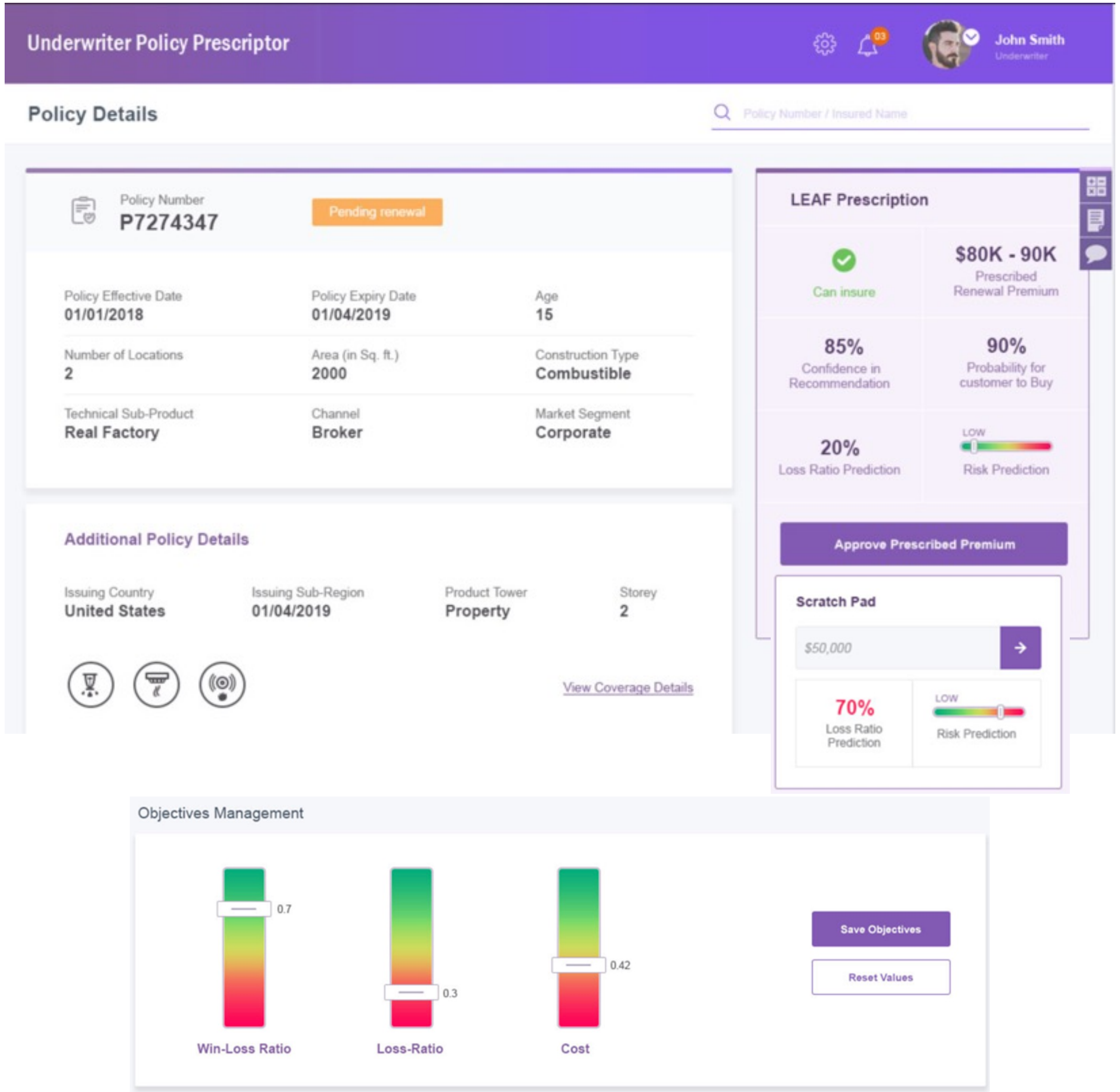
- Using historical data, build an adaptive digital twin of the market
- With evolutionary optimization, maximize conflicting objectives of revenue and margin

Result: A Pareto front of the best possible tradeoffs
- The user can select a desired tradeoff from it
- RIO indicates uncertainty in the predictions

Scratchpad allows adjusting the price by hand
- The digital twin predicts the resulting revenue and margin

cognizant

Insurance Underwriting



Improving pricing strategy for a leading property insurance company

- Using historical data, build an adaptive digital twin of the properties
- With evolutionary optimization, maximize conflicting objectives of win/loss, risk, and cost

Result: A Pareto front of the best possible tradeoffs

- The user can select a desired tradeoff from it
- RIO indicates uncertainty in the predictions

Scratchpad allows adjusting the premiums by hand

- The digital twin predicts the resulting risk and win/loss

Ship Trim Optimization

Improving ship speed/fuel efficiency for a shipping company

- Using historical data, build an adaptive digital twin of the ship
- With evolutionary optimization, maximize conflicting objectives of speed and fuel consumption

Result: A Pareto front of the best possible tradeoffs

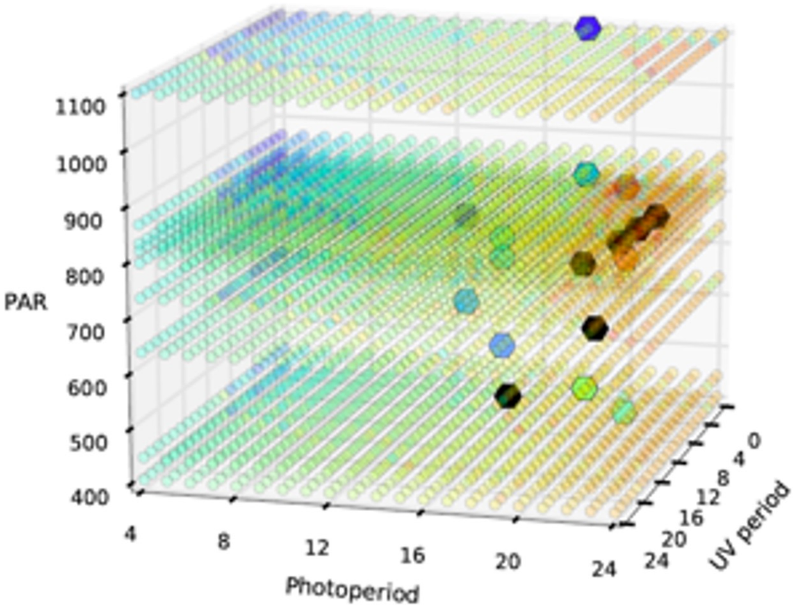
- The user can select a desired tradeoff from it
- RIO indicates uncertainty in the predictions

Scratchpad allows adjusting the premiums by hand

- The digital twin predicts the resulting speed and fuel consumption

Optimizing Growth Recipes for Agriculture

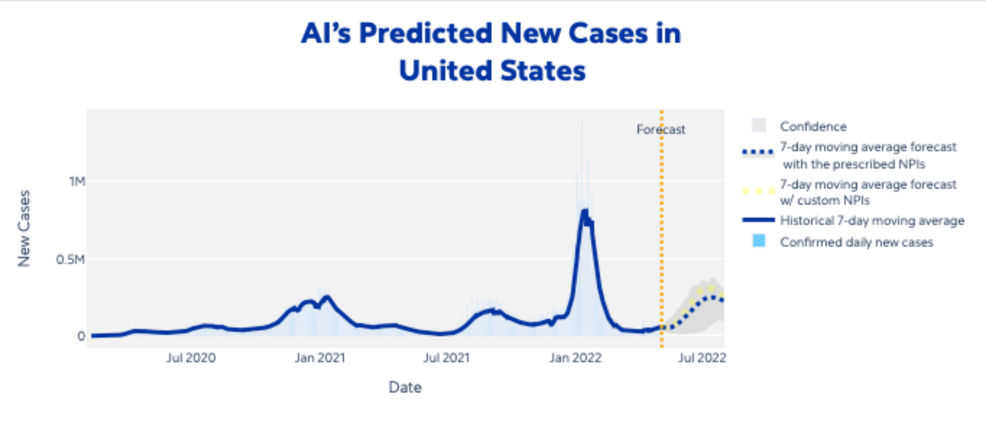
Environments at the OpenAg Lab of MIT Media Lab

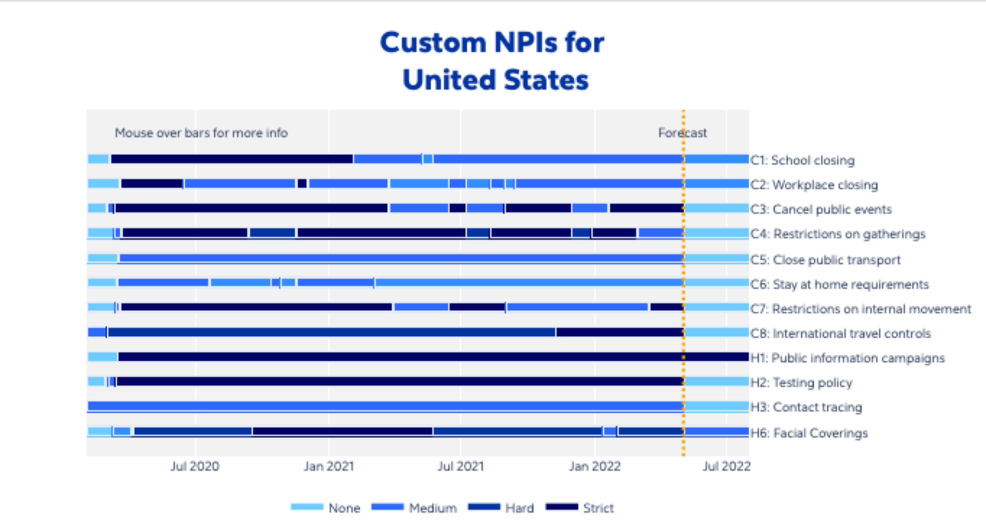


Recipes discovered for light (darker is better)

- Build an adaptive digital twin of how plants grow in hydroponic growth environments
  - Light, temperature, water, nutrients etc. computer controlled
- Evolve growth recipes against the twin
  - Millions new recipes created and evaluated; only a few hundred actually planted
- Discovered that basil does not need to sleep!
  - Grows best when lights are on 24hrs/day
  - An insight that surprised domain experts

Optimizing COVID-19 Response





Interactive demo: https://evolution.ml/demos/npidashboard

Build an adaptive digital twin of COVID-19 cases over time
- Predict number of cases in different countries
- In response to non-pharmaceutical interventions (NPIs)

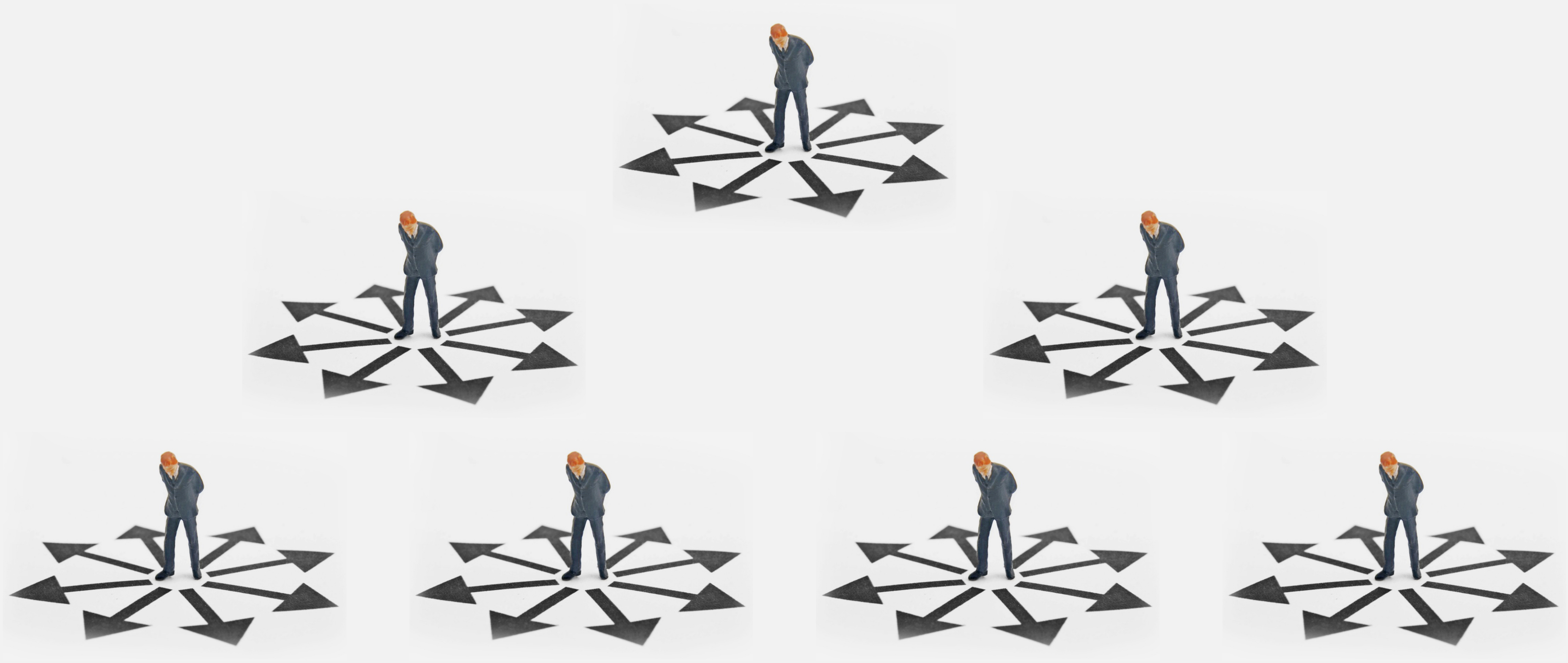
Discover best interventions against the twin
- Resulting in smallest number of cases
- With minimal economic cost

Not just what will happen, but what we should do about it!
- Scratchpad, uncertainty to help explore alternatives

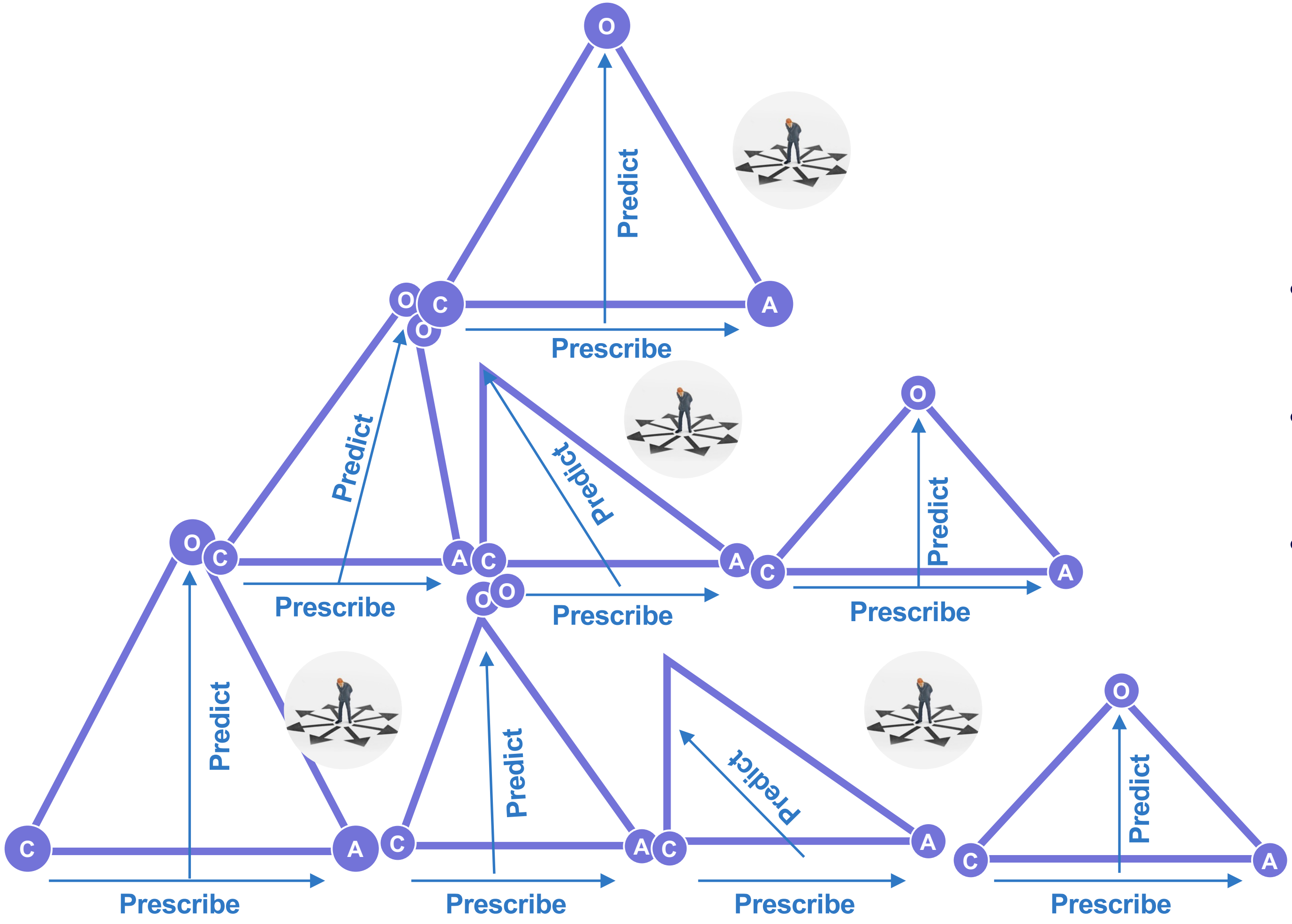
Retrained continuously since May 2020
- Adapting to the different stages of the pandemic
- Generalizing from experiences across the world

Recommendations about two weeks in advance, e.g.
- May 2020: Focus on schools and workplaces (i.e. indoors)
- Sept 2020: Focus on gatherings, travel restrictions
- March 2021: India lockdown
- July 2021: Delta surge on countries with low rates so far
- March 2022: Masking to avoid a second Omicron surge

An Organization Consists of Many Decision Makers



Scale in Entangled Decision-Loops



- More Accurate Decision
- More Efficient Decision-making
- Better Outcomes

Thank You

How specialized AI drives value for Automotive Service Organizations

Tilak Kasturi

INFORMATION TECHNOLOGY STANDARDS

SC 42 – Artificial Intelligence

Automotive Ecosystem Use Cases

Fleets | Insurers | OEMs | Dealers | Aftermarket Service Providers | Government Agencies | Parts Suppliers | Consumers

Vehicle Sales Forecast
Repair Diagnostics
Predictive Maintenance
Road Capacity Planning
Vehicle Design
Residual Value
Parts Production Forecast
Repair Estimation
Usage Based Insurance

Manufacture | Sell | Operate | Repair | Remarket | Salvage

- New / Used Vehicle Sale Prices
- Production Costs
- OEM Repair Procedures
- Parts Pricing (Listed / Paid)

- Labor Rates Incurred
- Replacement Part SKUs
- Vehicle Design Specifications
- Repair Cycle Time

- Parts Replacement Occurrence
- Driver Behavior
- Collision Severity
- Parts Performance and Wear

- Diagnostic Trouble Codes
- Emissions Levels
- Shop Equipment Utilized

Source: Jefferies

Automotive Data is unstructured, complex.

$2	Trillion total global market	~1.4	Billion vehicles in operation globally
2%	of US nominal GDP	12.2	Years average age of light vehicles (US cars & light trucks)
4.7	Million jobs in the US	100K	make, model, engine, transmission combinations

Automotive service focus is to minimize breakdowns

Predict and Prescribe

Right Part, right place, right time

Challenges with Automotive service data for AI Applications

- Availability of data for research community
- Proprietary and heterogeneous data
- Large NLP models are not trained on “automotive data”
- Data is too noisy
  - Textual data has industry and local acronyms, incomplete sentence structures, etc.
  - Connected car data is not continuous, lacks coverage

State-of-the-art NLP Trends

- Transformers changed the landscape and is here to stay
- Thrust on low-resource languages; multi-lingual NLP going mainstream
- Breaking barriers: Supervised --> Unsupervised --> Self-supervised learning (hybridization)
- Low-code NLP on the rise
- Multimodal systems gaining momentum

Advancements in developing large NLP Models

- BERT started the trend towards dynamic models
- Race towards larger NLP models
- Megatron-Turing NLG model at 530B parameters!
- Powering downstream tasks
  - Semantic search
  - Automated dialogue generation
  - Summarization
  - machine translation
  - Commonsense reasoning

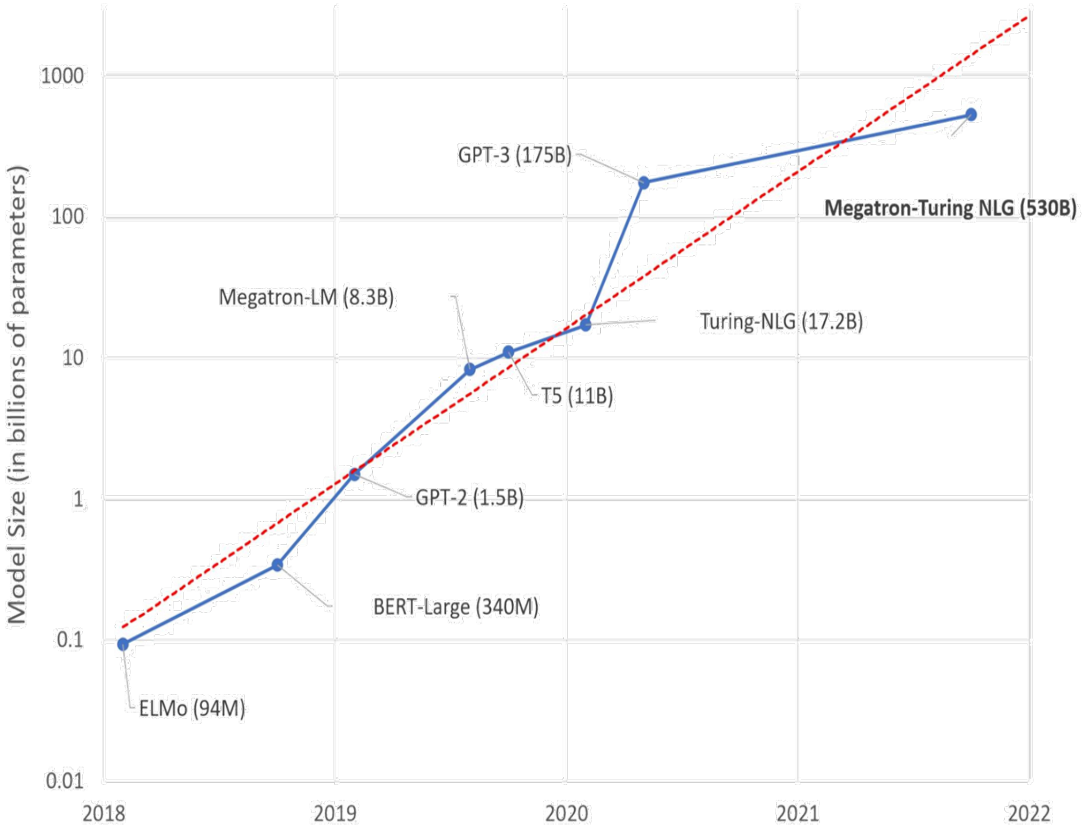


Figure 1. Trend of sizes of state-of-the-art NLP models over time

Source: Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, the World's Largest and Most Powerful Generative Language Model - Microsoft Research

Large Language Models are not meant for industry specific use-cases

Models should drive specific decisions

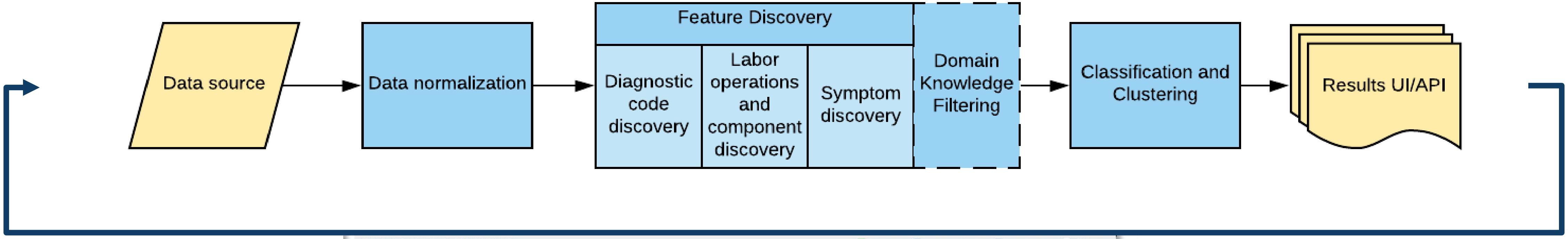
When will breakdowns happen and why?
Which repair will resolve the issue?
Which issues are a safety concern?

Models need to be explainable

Models should have high confidence

Models need 'deep contextual understanding' in **specialized domain**

MLOps & Cloud native AI Pipeline drive actionable decisions



Odometer 90,062 / 2019 / BMW X3

RO Lines

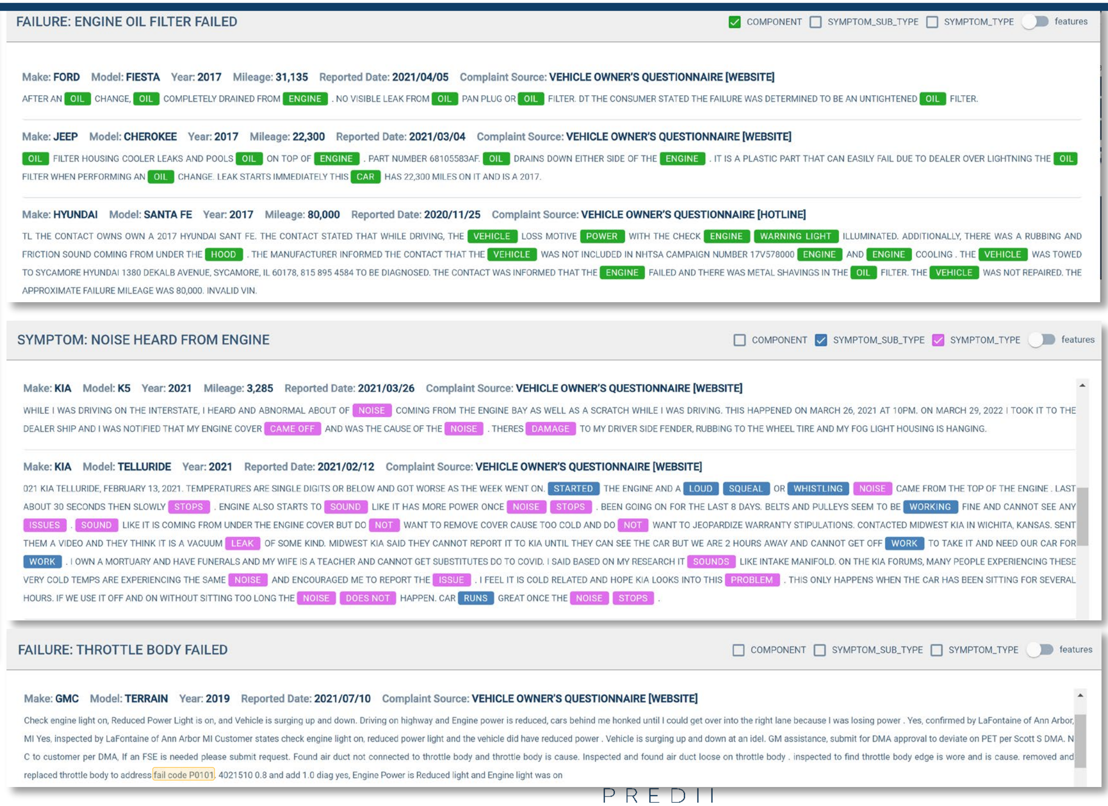
18221 : Part - Throttle Body

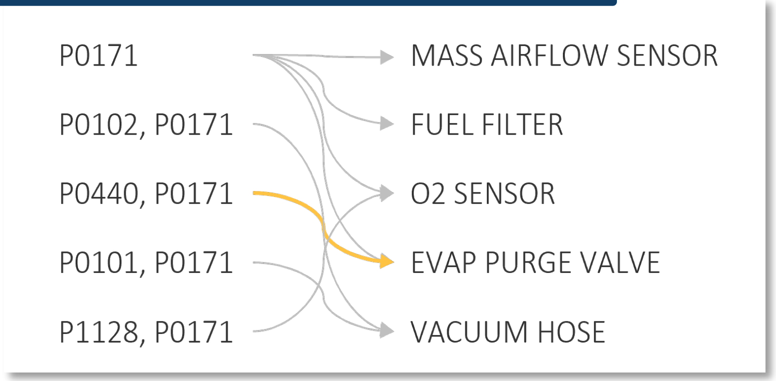
18219 : Part - Oil

18220 : Part - Oil Filter

16272 : P115E Labor - Noise in engine Removed oil pan and balance shafts to inspect crank. Found high amount of meta debris in the pan. Found the #3 rod bearing had spun and chewed up the crankshaft and rod

18213 : Labor - LONG BLOCK - Remove & Replace - LX - Consists of a Cylinder Block fitted with Pistons, Rings, Connecting Rods, Crankshaft and all Bearings, Cylinder Head(s), Camshaft(s), Timing Chain or Belt and Sprockets or Gears.<\n>Includes (where applicable): Clean





Two examples of Future trends in Automotive Service

"John, Your 2019 BMW X3 needs service"

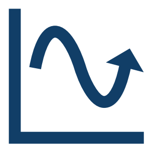
"What are my vehicle's top Safety Concerns?"

“John, Your 2019 BMW X3 needs service.”

1. Vehicle IOT Trigger

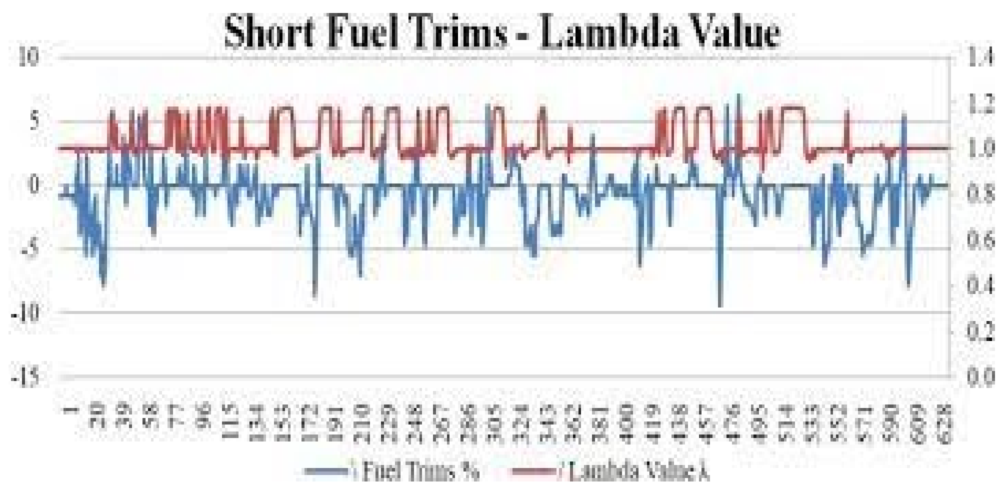
DTC, PID via Telematics

2. Upstream Analytics



Min, Max, Anomalies *isolated*

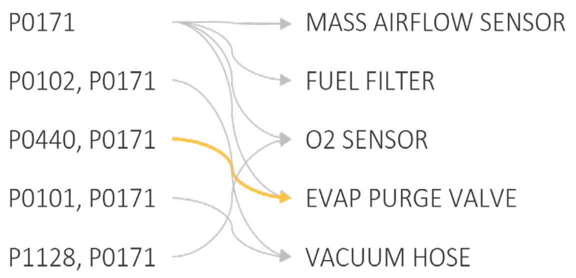
DTC(P0171,P0440),Short Fuel Trims(%),O2 Sensor(v)



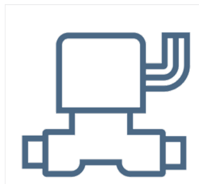
event=startup&ssnId=06f4c1505469fb0eff1b&timestamp=10:55%20AM&system=Electrical=&frameNumber=141&Door%20Lock=true&Power=on&Temp=98.8&Roller=active&Rate=1.2&Pressure=67……….

3. Overlay & Correlate

Overlaying with insights from repair orders (textual)

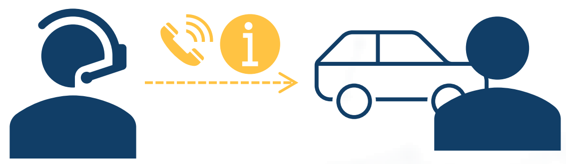


4. Predict Parts Replacement



Vehicle YMME & mileage specific

5. Pro-active CRM



6. Parts Ordering

BMW Vapor Canister Purge Solenoid - Genuine BMW 13907618647

$97.99

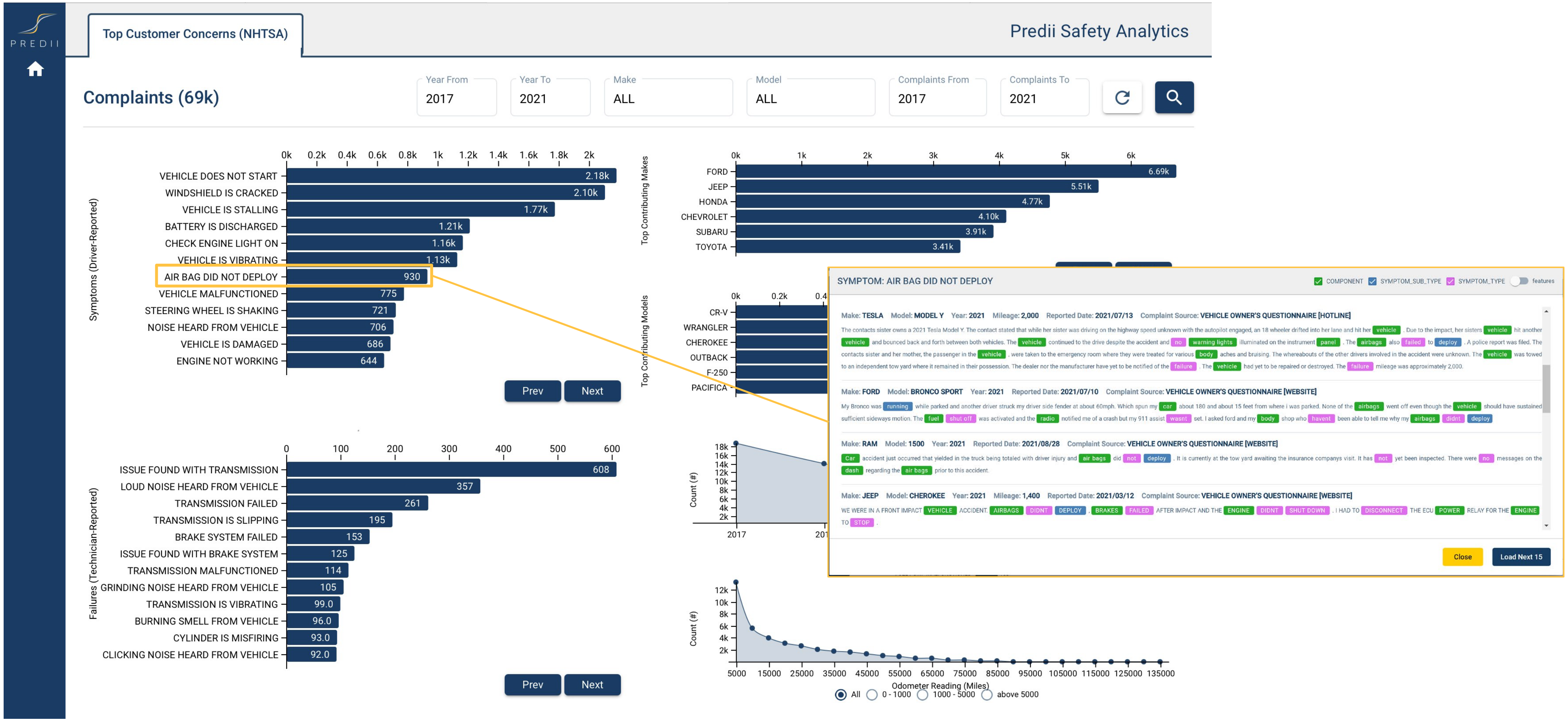
FCP Euro

Free shipping

7. Vehicle Repair

PREDII

“What are my vehicle’s top Safety Concerns?”



Important Conferences and Journals

- ACL
- EMNLP
- COLING
- NAACL
- EACL
- AACL-IJCNLP
- TACL
- CL
- AAAI, IJCAI, NeurIPS, CVPR, ICML, ICLR, ICCV, ECCV
- + many relevant workshops

Twitter: #NLProc

Thank you

Tilak Kasturi
tilak@predii.com
www.predii.com