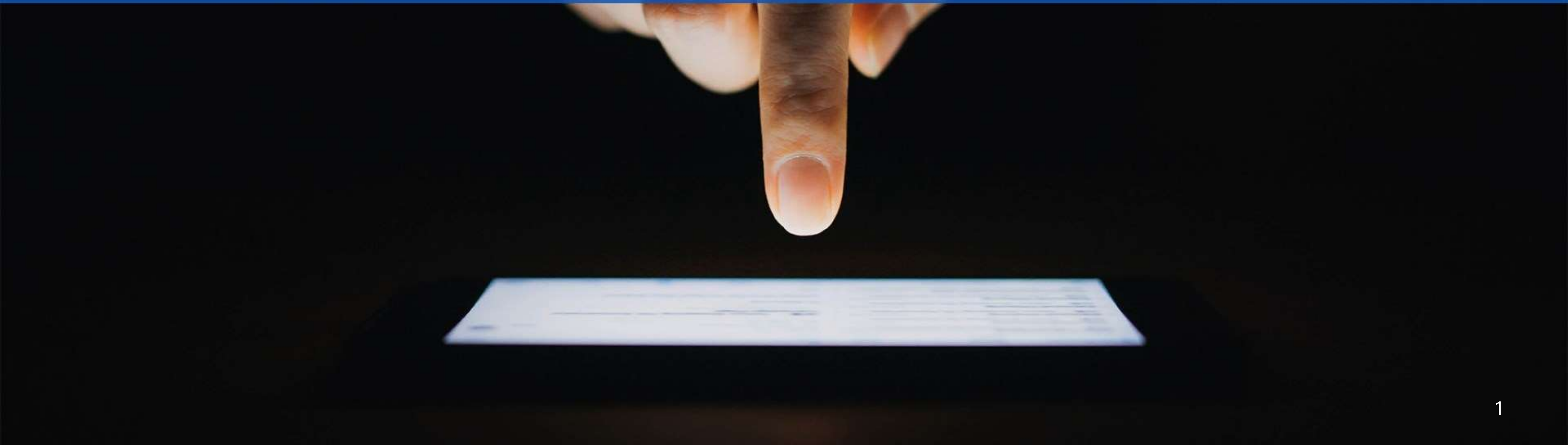
Responsible AI trade-offs in Insurance

Luca Baldassarre, Lead Data Scientist, Advanced Analytics CoE, Swiss Re



Luca Baldassarre

Luca Baldassarre - Lead Data Scientist
Advanced Analytics Centre of Expertise

What I do: *"I work with data scientists and business stakeholders to ensure AI is used and governed appropriately, and models are rigorously tested before deployed."*

Luca's Profile

16 years experience in data science, ML & AI; **PhD, 2 PostDocs, AgTech Startup, re/insurance**

Current focus on Analytics Governance: **establishing a company-wide AI governance framework that includes rigorous model validation and transparency and fairness assessments**

Key Learning Takeaway from Current Role: **Without proper governance model risks can cause serious harms**

How do I spend time outside of work: **with family and trail running**

2006 – Università di Genova

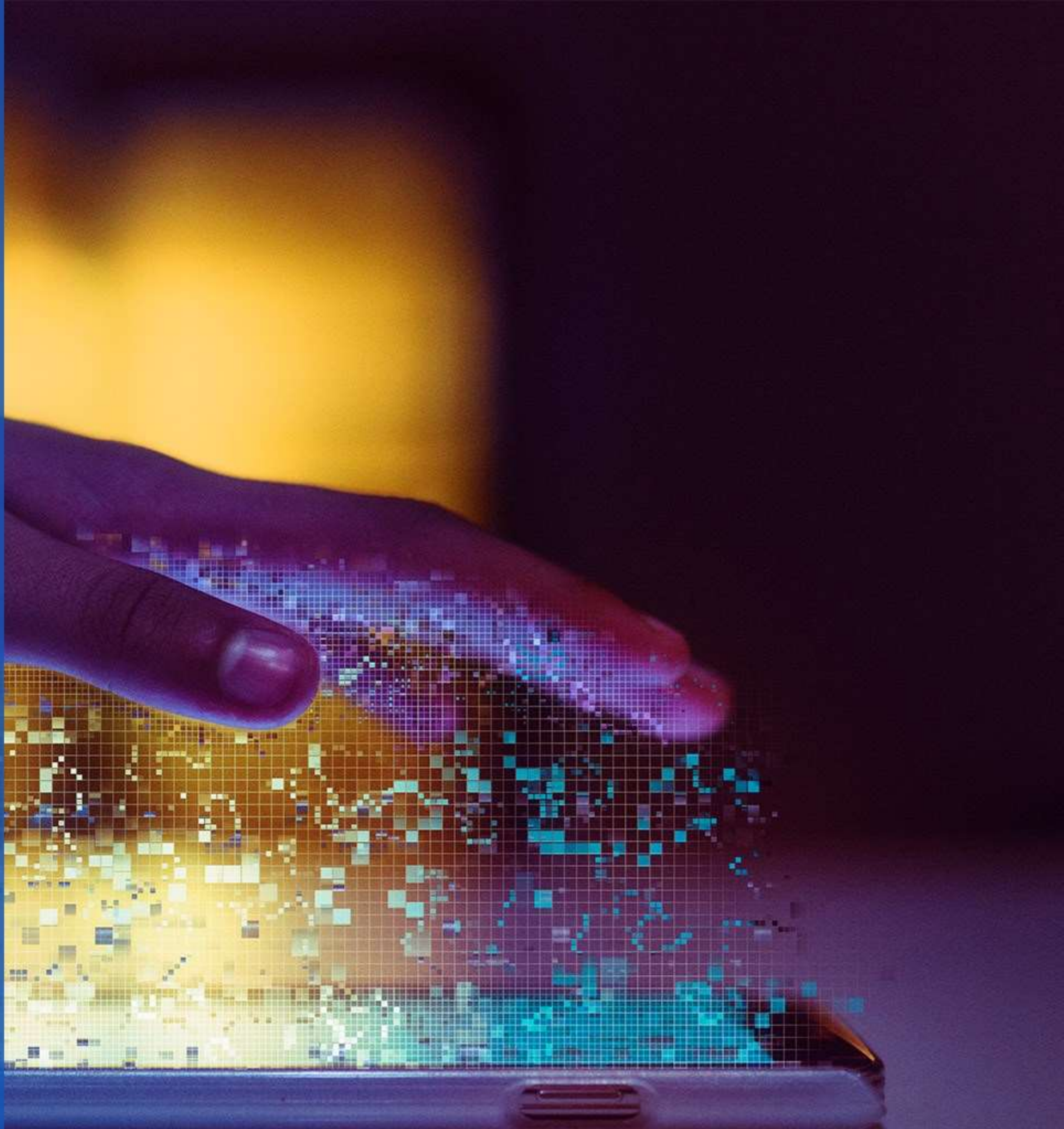
2010 – UCL

2012 – EPFL

2016 – GAMAYA

2019 – Swiss Re

What´s happening



AI is (finally) bringing value at scale across the insurance value chain

Product Development	Underwriting & Pricing	Sales and Distribution	Post Sales Services	Claims	Contracts
Analytics solutions to access and commodify new & existing risk pools	Streamlining underwriting with predictive modelling and personalized pricing	Data-driven lapse, retention and propensity to bind models	Post sales services include data visualizations, consulting services, portfolio & trend insights	Statistical methods & claims review approach to mitigate risk and save costs	Natural Language Processing to analyze contracts and automatically understand coverage

HOME > TECH

Amazon built an AI tool to hire people but had to shut it down because it was discriminating against women

Computer programs used in 46 states incorrectly label Black defendants as "high-risk" at twice the rate as white defendants

Natalia Mesa
Neuroscience
University of Washington

Over 50%

say that their organizations have adopted AI in at least one business function

Women less likely to be shown ads for high-paid jobs on Google, study shows

Automated testing and analysis of company's advertising system reveals male job seekers are shown far more adverts for high-paying executive jobs

The Machine
Making sense of AI

Bias persists in face detection systems from Amazon, Microsoft, and Google

Kyle Wiggers @Kyle_L_Wiggers September 3, 2021 10:40 AM

HOME > TECH

Amazon built an AI tool to hire people but had to shut it down because it was discriminating against women

Discriminatory pricing

Exploring the 'ethnicity penalty' in the insurance market

citizens advice

Tilly Cook
Aiden Greenall
Emer Sheehy

ates
s as
hite

hown ads f
study sho

lvertising system
erts for high-

n systems f
Amazon, Microsoft, and Google

Kyle Wiggers @Kyle_L_Wiggers September 3, 2021 10:40 AM


THE WALL STREET JOURNAL

What's News

Pelosi Visits Taiwan, Defying China

Robinhood Plans More Staff Cuts As Trading Plunges

Tactical Questions Follow Killing of al Qaeda Chief

Stephen King Takes The Stand

SoftBank Emerges As a Loser in Rout Of Tech, Again.

CEO urged staff to pour money into startup firms last year, at close to their market top

Equifax Sent Lenders Wrong Credit Scores

INSIDE

Over 50%

say that their organizations have adopted AI in at least one business function

AI governance – the regulatory landscape is quickly evolving (non exhaustive)

Principles | Guidance | Regulation proposal

Oct 2016 US NSTC Preparing for the future of AI

Jan2012 => May 2016 => May2018
GDPR EU released => adopted => inforced

Jan 2018 China AI standardisation white paper

Nov 2018 Singapore FEAT Principles

Jan 2019 Singapore IMDA Model AI Governance Framework

March 2019 Japan Social Principles of Human-Centric AI

Apr 2019 EU Ethics draft guidelines for trustworthy AI

May 2019 OECD Principles on AI

Jun 2019 China Governance Principles for the New Generation AI

Nov 2019 Australia AI Ethics Framework

Nov 2019 Hong Kong HKMA High-level AI Principles

Jan 2020 Berkman Klein Center review of existing sets of AI principles

May 2020 US federal data strategy released

Nov 2020 Singapore MAS Veritas Phase1 Fairness papers

Jan 2021 US National Artificial Intelligence Initiative Act promulgated

Mar 2021 US FTC guidelines on truth, fairness, and equity in AI

Mar 2021 Korea Financial Services Commission announces a new policy framework on insurance business

Apr 2021 European Commission Proposal for a Regulation laying down harmonized rules on AI - includes fines of up to 6% of a company's annual revenues for noncompliance

Apr 2021 Hong Kong The State of Ethical AI whitepaper

Sep2021 China Guidance ethical dev & use of AI

Feb 2022 EU Data Act

Feb 2022 Singapore MAS Veritas Phase2 FEAT whitepapers

Mar 2022 China guidelines on algorithmic recommender systems

Apr 2022 EU Digital Services Act package

(Some) Responsible AI principles

EIOPA

- Proportionality
- Fairness and non-discrimination
- Transparency and explainability
- Human oversight
- Data governance of record keeping
- Robustness and performance

Microsoft

- Fairness
- Reliability and Safety
- Privacy and Security
- Inclusiveness
- Transparency
- Accountability

Monetary Authority of Singapore (MAS)

- Fairness
- Ethics
- Accountability
- Transparency

European Union

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental well-being
- Accountability

Digital responsibility is critical to the success of an insurer's digital strategy and its risk assessment will be expected by regulators

- Customers are increasingly **unwilling to share data without transparent explanations**
- Governments and regulators are defining **digital ethics requirements**
- The **opportunities** that come with investments in data analytics and AI carry also important **risks**

Risks (not only ethical)

- **data risks** - issues associated with the collection of (large) datasets and their use (*privacy and confidentiality*)
- **algorithm risks** - issues arising from how algorithms recommend or make decisions (*bias and discrimination*)
- **compliance risks** – non-compliance with existing and upcoming laws and regulation

Costs of implementing an AI governance framework

- **Financial investment** in people and tools to develop and maintain framework
- **Upskilling** product owners, users, and developers to properly identify and address risks
- Increased **procedural overhead** to create accountability and enforce guidelines
- **Change management** friction

Trade-off #1

Get ready now or wait and see how regulation will be shaped?

Which model risks can you keep ignoring?

Focus on Fairness

Swiss Re

MAS's (re)definition of the Fairness Principles

Justifiability

F1 **Individuals or groups of individuals** are not **systematically disadvantaged** through AIDA*-driven decisions, unless these decisions can be **justified**

F2 Use of **personal attributes** as input factors for AIDA-driven decisions is **justified**

Accuracy and Bias

F3 Data and models used for AIDA-driven decisions are **regularly reviewed** and validated for **accuracy and relevance**, and to minimise **unintentional bias**

F4 AIDA-driven decisions are regularly reviewed so that **models behave as designed and intended**

*AIDA : Artificial Intelligence and Data Analytics (AIDA) systems

Veritas Fairness Assessment Methodology – Key Concepts

Personal Attributes

Attributes about individuals considered sensitive enough to require justification for use as the basis of decisions.

Example: person's ethnicity or gender

(Unintentional) bias

Systematic disadvantage FSI is not aware of, coming from data, model or how they are used.

Example:, New product designed to be delivered only online could exclude customers with less digital literacy

Fairness Objective

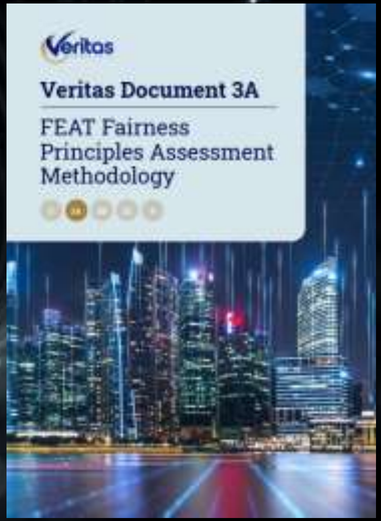
What the FSI must achieve to meet its fairness principles for an AIDA System.

Example: Males and females have same opportunity to get a loan.

Fairness Metric

Mathematical definition of fairness objective.

Example: Equal Opportunity or False Negative Rate Balance



(Unintentional) Bias - Types & Mitigation: Examples

	1	2	3	4
AIDA System Development Lifecycle	**Define System Context & Design**	**Prepare Input data**	**Build & Validate**	**Deploy & Monitor**
Where Measurable and Cognitive Bias can enter **Human judgement is involved at every step of algorithm development!**	Exclusion Bias Confirmation Bias Availability Bias	Representation Bias Measurement Bias Proxy Bias Pre-processing Bias Historical Bias Decision fatigue	Aggregation Bias Evaluation Bias Social norms	Deployment Bias Drift Bias
Example of Measurable Bias	Exclusion Bias: New product designed to be delivered only online could exclude customers with less digital literacy	Representation Bias: Women under-represented in the dataset of customers with car loans to be used as model development dataset	Aggregation Bias: New to bank customers and long-time customers included in same model to predict fraud risk	
Example Mitigation	Include alternative channel to online	Upweight women in the development dataset	Build separate models or post-processing calibration mitigation	

Fairness Objectives and Fairness Metrics

Group Fairness

V.S.

Individual Fairness

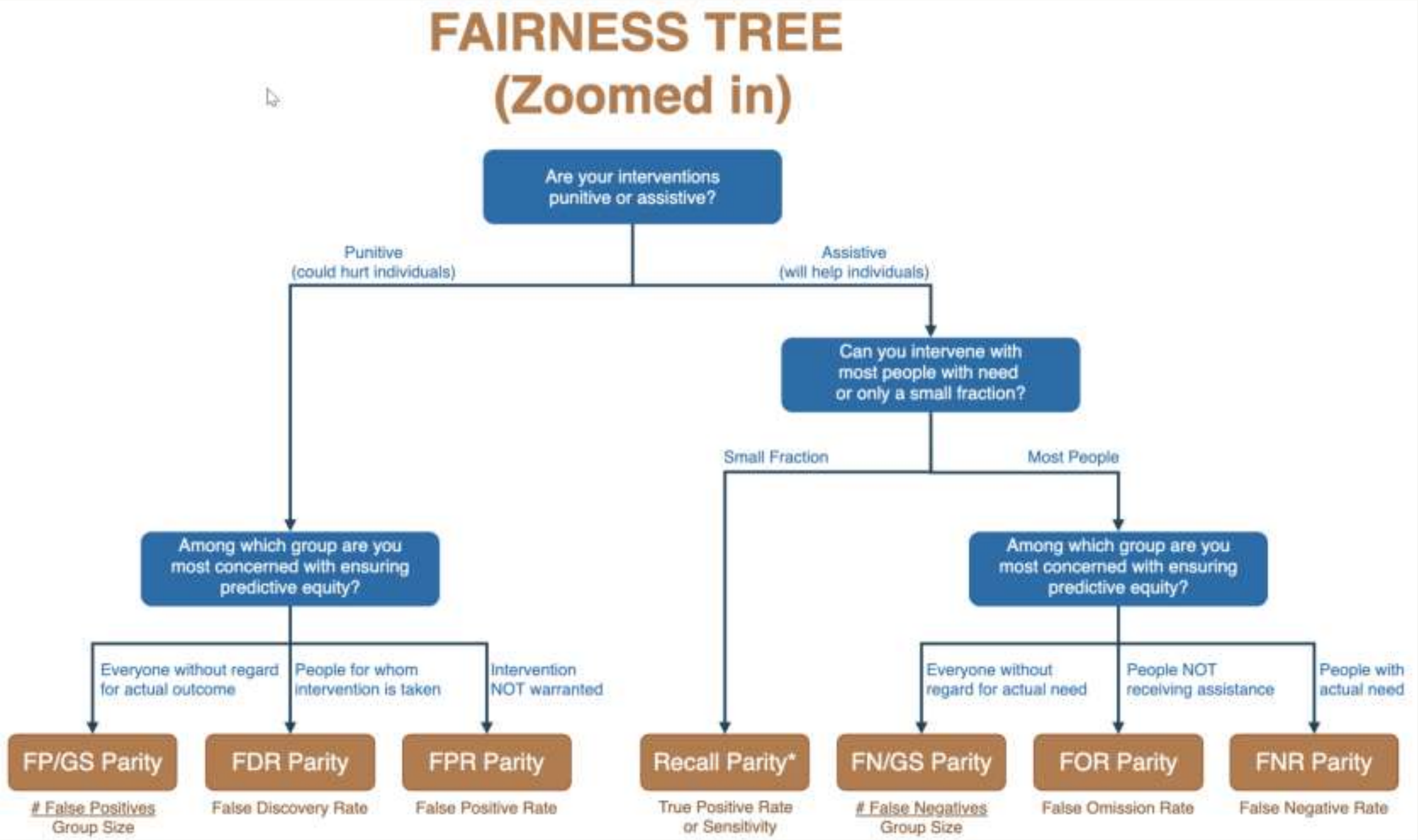
Population groups are defined, and the outcomes between groups compared.

- ✓ More common practice today to assess fairness at group level
- ❖ need to be aware that **compound effects** may exist at the intersection of the attributes of certain groups (e.g., gender, ethnicity, disabilities, etc.)
- ❖ a **prioritized selection** of at-risk groups is recommended **in consultation with stakeholders**

The notion of similarity in outcomes for similar individuals is defined, and outcomes between similar pairs of individuals compared.

- × Assessing fairness at too granular a level may be too complex to execute and can be limited by the collected features, privacy aspects or consent provided by individuals
- × Currently no standard definition of 'similarity of outcomes' – thus subjective.

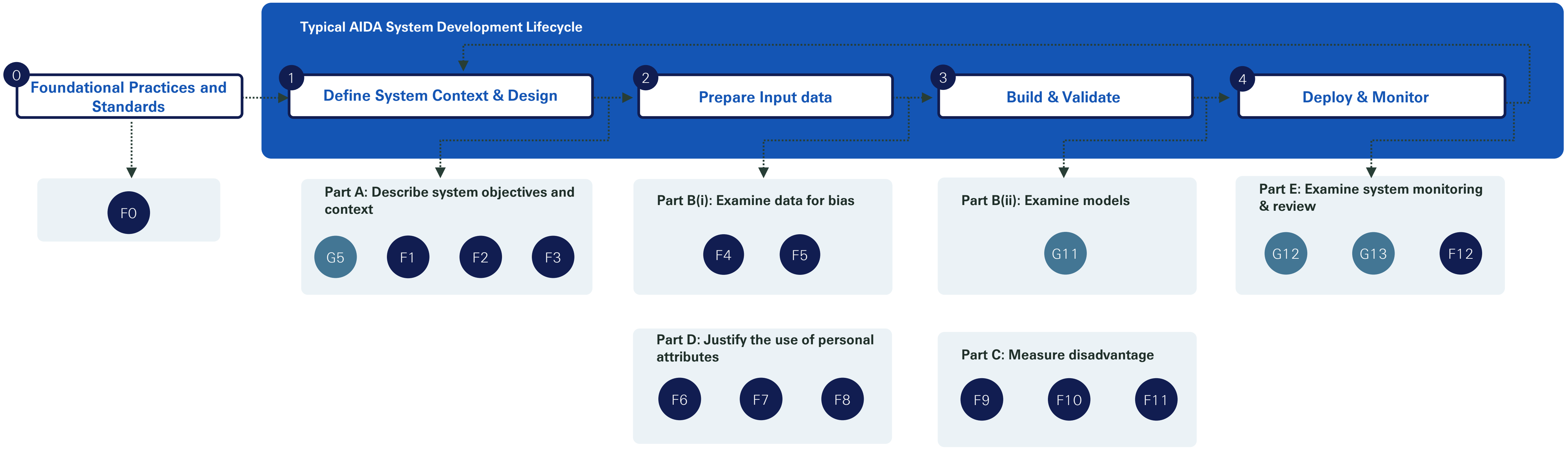
Fairness Objectives and Fairness Metrics



Source: http://www.datasciencepublicpolicy.org/our-work/tools-guides/aequitas/

Fairness Methodology – Checklist mapped to AIDA development lifecycle

13 Fairness specific questions* (F0-F12) embedded into a typical AIDA Development lifecycle + 4 general questions** (G5, G7-9) to be applied to each use cases per proportionality



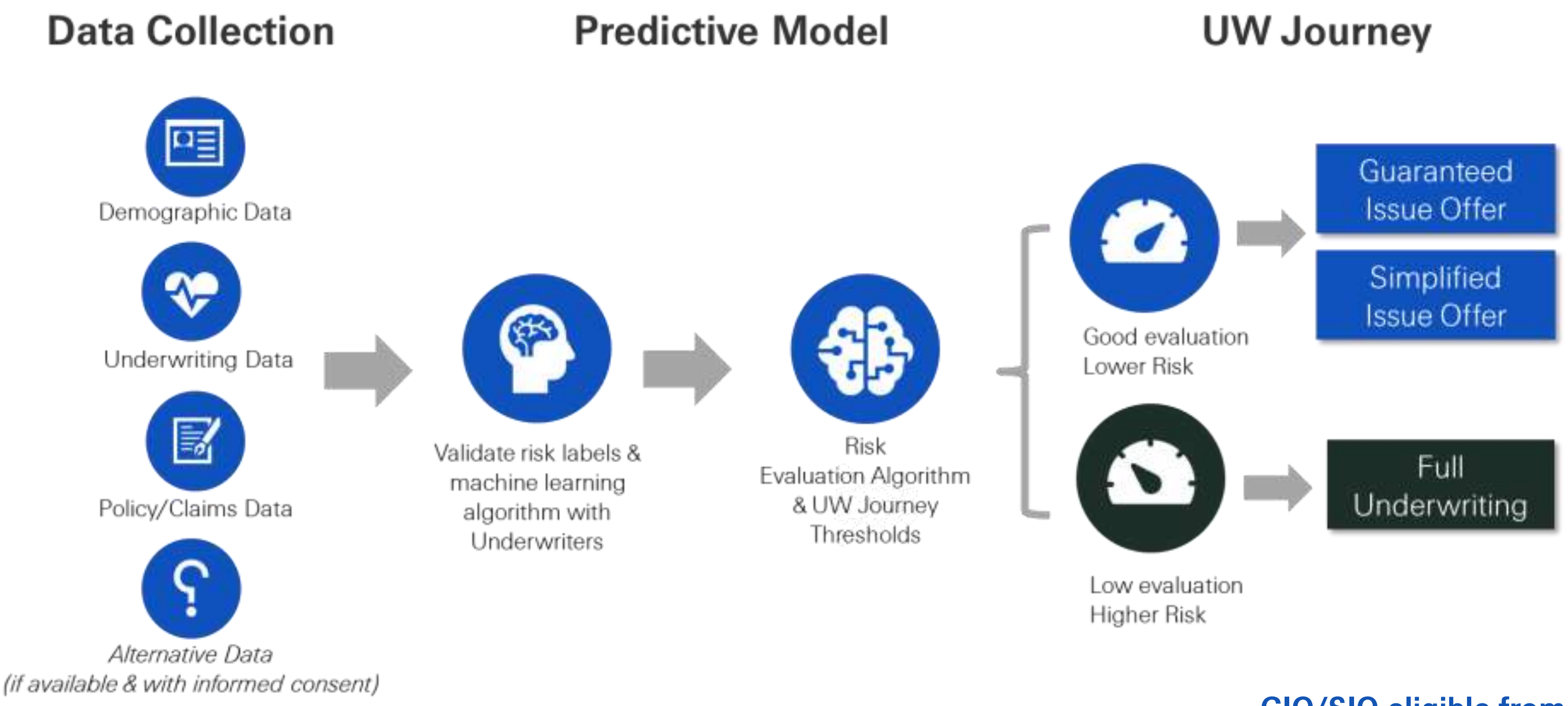
*Based on fairness risk of the AIDA System, the FSI can decide all or a subset

**General Questions means relevant for all FEAT principles and should be included in assessments of each principle

Trade-off #2

How to balance the assessment´s completeness and its adoption?

Fairness assessment methodology applied to predictive underwriting use case



63% **GIO/SIO eligible from existing customers with no price loading**

Fairness assessment methodology applied to predictive underwriting use case

- **False Negative Rate Ratio** was chosen as the most appropriate fairness metric:
  - For the eligible population (those that should get simplified underwriting) the rate of false negatives (those that aren't offered simplified underwriting) does not differ by over 20% among subgroups
- **Gender & Ethnicity** were selected for this particular use case for fairness assessment
- For **Ethnicity** – the fairness metric is within the defined threshold
- For **Gender** - **post-processing mitigation** of split gender thresholds bring the fairness metric within the acceptable threshold, while meeting the primary commercial objectives, i.e., when optimizing gender fairness, the model's balanced accuracy drops slightly, but is still above the minimum required.

	Single Threshold	Threshold Male	Threshold Female	Commercial Objective: Balanced Accuracy	Commercial Objective: Precision	Fairness Objective: FNR Ratio
Required Constraints	NA	NA	NA	>82%	>96.00%	0.8 < < 1.2
× Single threshold for max balanced accuracy	0.54	NA	NA	83.04%	96.34%	1.56
♦ Split threshold for max balanced accuracy	NA	0.54	0.59	83.27%	96.45%	1.42
★ Split threshold for max balanced accuracy constraining for fairness	NA	0.48	0.59	83.14%	96.00%	1.18

Within Threshold Range
Outside Threshold Range

Trade-off #3

How to balance the model's financial performance and its fairness?

Focus on Transparency

Did you Know ?

41%
of consumers believe AI will improve their lives in some way

(Source: Strategy Analytics)

33%
of consumers think they're using tech that features AI

(Source: Pega)

In reality,
77%
already uses an AI-powered service

(Source: Pega)

A disturbing, viral Twitter thread reveals how AI-powered insurance can go wrong

Lemonade tweeted about what it means to be an AI-first insurance company. It left a sour taste in many customers' mouths.

By Sara Morrison | May 27, 2021, 1:30pm EDT

I like growth stocks
@GrowthLike

Replying to @EigenegiE @LincoIns_Finger and @Lemonade_Inc

Like These?

@Lemonade_Inc

Lemonade is built on a digital substrate – we use bots and machine learning to make insurance instant, seamless, and delightful.

This puts us at a data advantage—in

Lemonade @Lemonade_Inc · 1T

A typical homeowners policy form has 20–40 fields (name, address, bday...), so traditional insurers collect 20-40 data points per user.

AI Maya asks just 13 Q's but collects over 1,600 data points, producing nuanced profiles of our users and remarkably predictive insights. (2/7)

Data

150 87 54

@Lemonade_Inc

This ultimately helps us lower our loss ratios (aka, how much we pay out in claims vs. how much we take in), and our overall operating costs.

In Q1 2017, our loss ratio was 368%

Lemonade @Lemonade_Inc · 1T

Antwort an @Lemonade_Inc

Something else that's friggin terrible: our mistake! That 71% reflects the removal of the impact of catastrophes in Q1 21, as noted in our Q1 shareholder letter. This

5:03 PM · May 26, 2021

101 5 Copy link to Tweet

Lemonade
@Lemonade_Inc

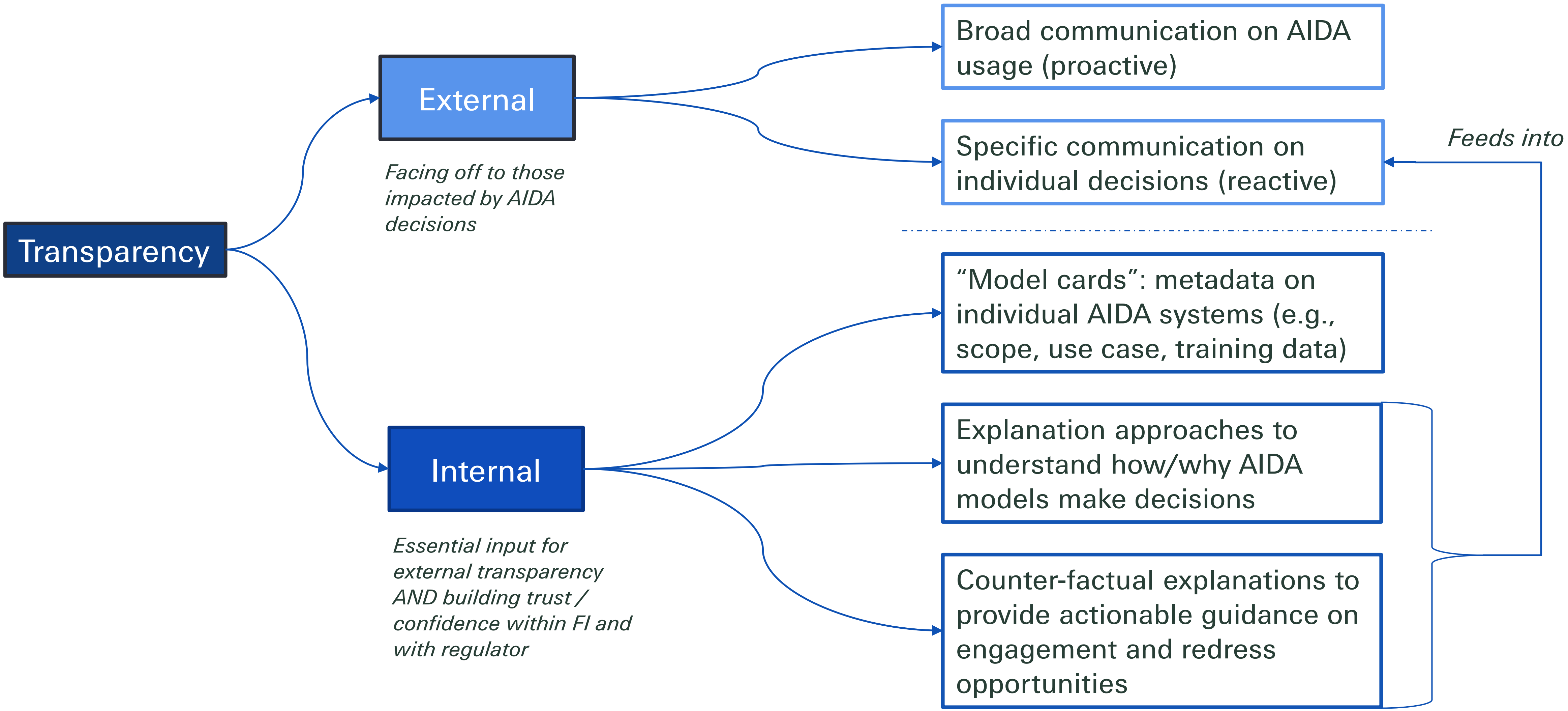
So, we deleted this awful thread which caused more confusion than anything else.

TL;DR: We do not use, and we're not trying to build AI that uses physical or personal features to deny claims (phrenology/physiognomy) (1/4)

4:45 PM · May 26, 2021

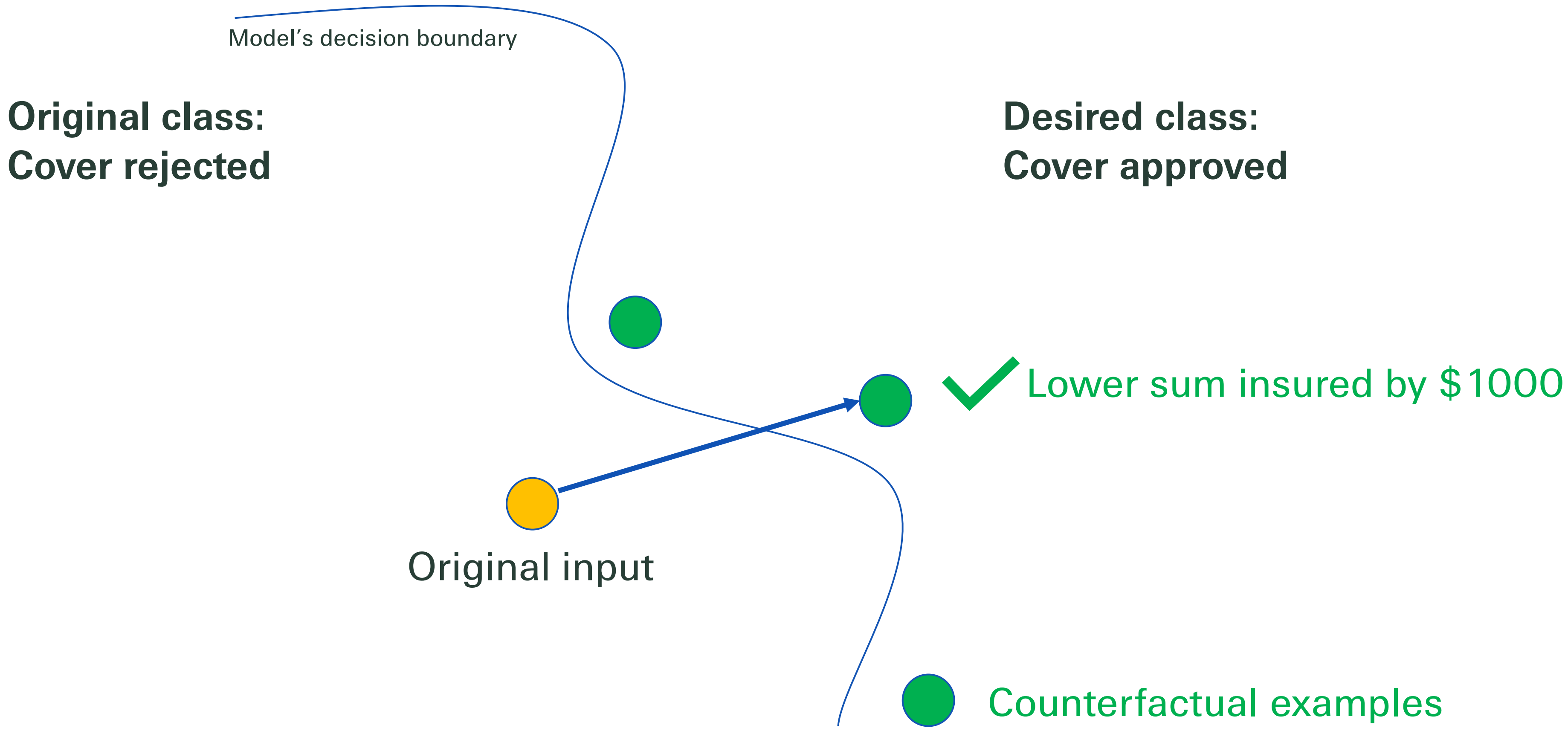
307 372 Copy link to Tweet

Different flavours of transparency to address specific concerns



Counterfactual examples and explanations

If Andrea's application is rejected, how can she improve her outcomes for the future?
Need to provide **actionable information** on what she can change.

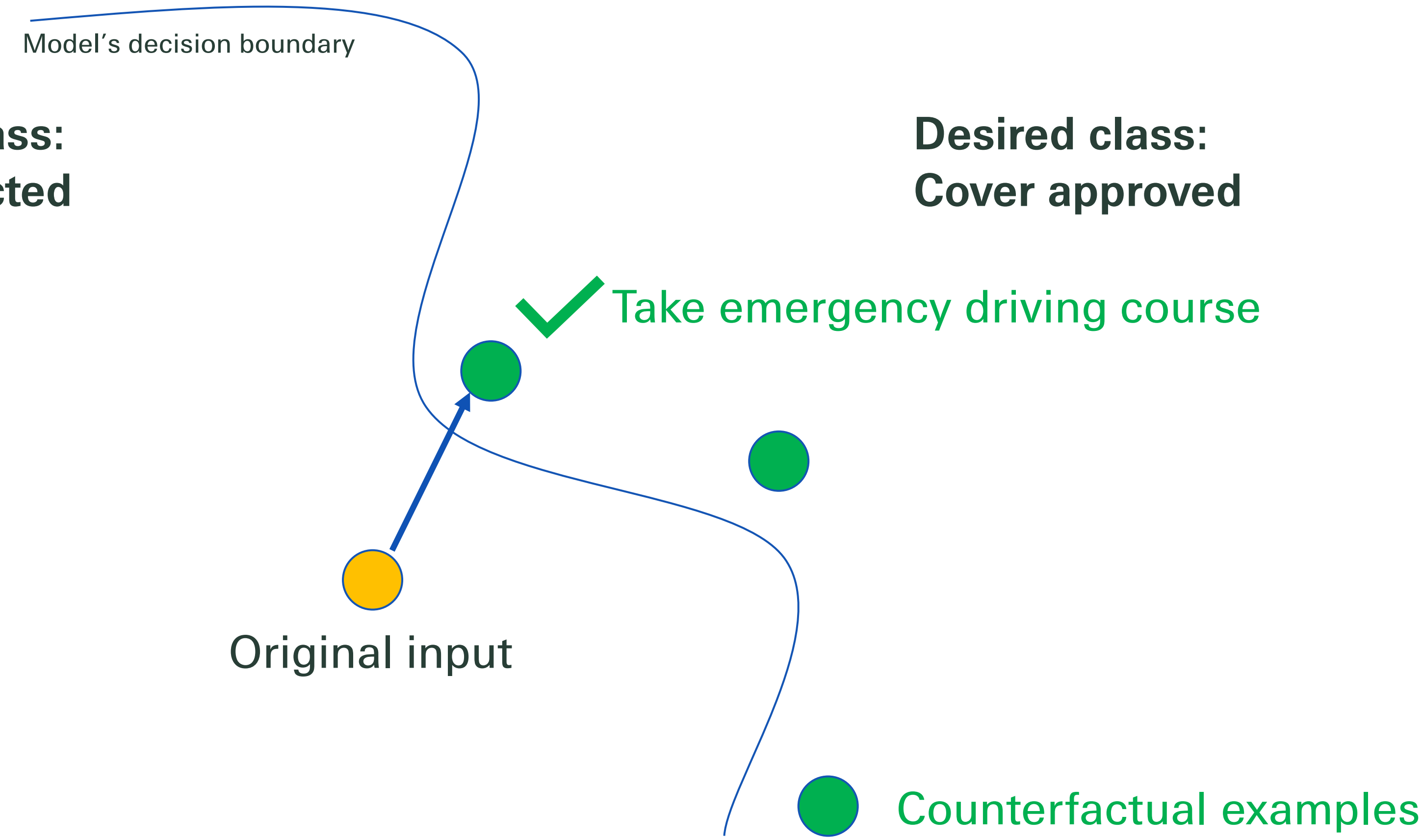


Counterfactual examples and explanations

If Andrea's application is rejected, how can she improve her outcomes for the future?
Need to provide **actionable information** on what she can change.

Original class:
Cover rejected

Desired class:
Cover approved



Recourse frameworks must satisfy a set of principles

Validity

Will this suggestion change the outcome? → reduce your sum insured by 50$!

Actionability

Can the user follow up on the suggestion? → Increase your age by 5 years!

Proximity

Is this the smallest change possible? → reduce miles travelled per year by 10k!

Sparsity

Is the user forced to change many things? → Change car, reduce sum insured, change driving behaviour, take emergency driving training

Privacy

Will this suggestion leak customer data? → Follow Andrea's profile

Trade-off #4

How to provide external explanations to build trust while avoiding leaking intellectual property or making it easier to game the system?

AI Governance

Advanced Analytics come with new challenges to be addressed across the entire analytics lifecycle from scoping, development, deployment, to monitoring

The proliferation of advanced analytics activities creates additional risks to be mitigated...

CHALLENGES

Regulatory pressure

Internal requirements

- Model explainability
- Data reliability, value & principles
- Model governance roles and responsibilities
- Model reproducibility and scalability
- Ethics, Non-Discrimination & Bias
- Consistent monitoring of models

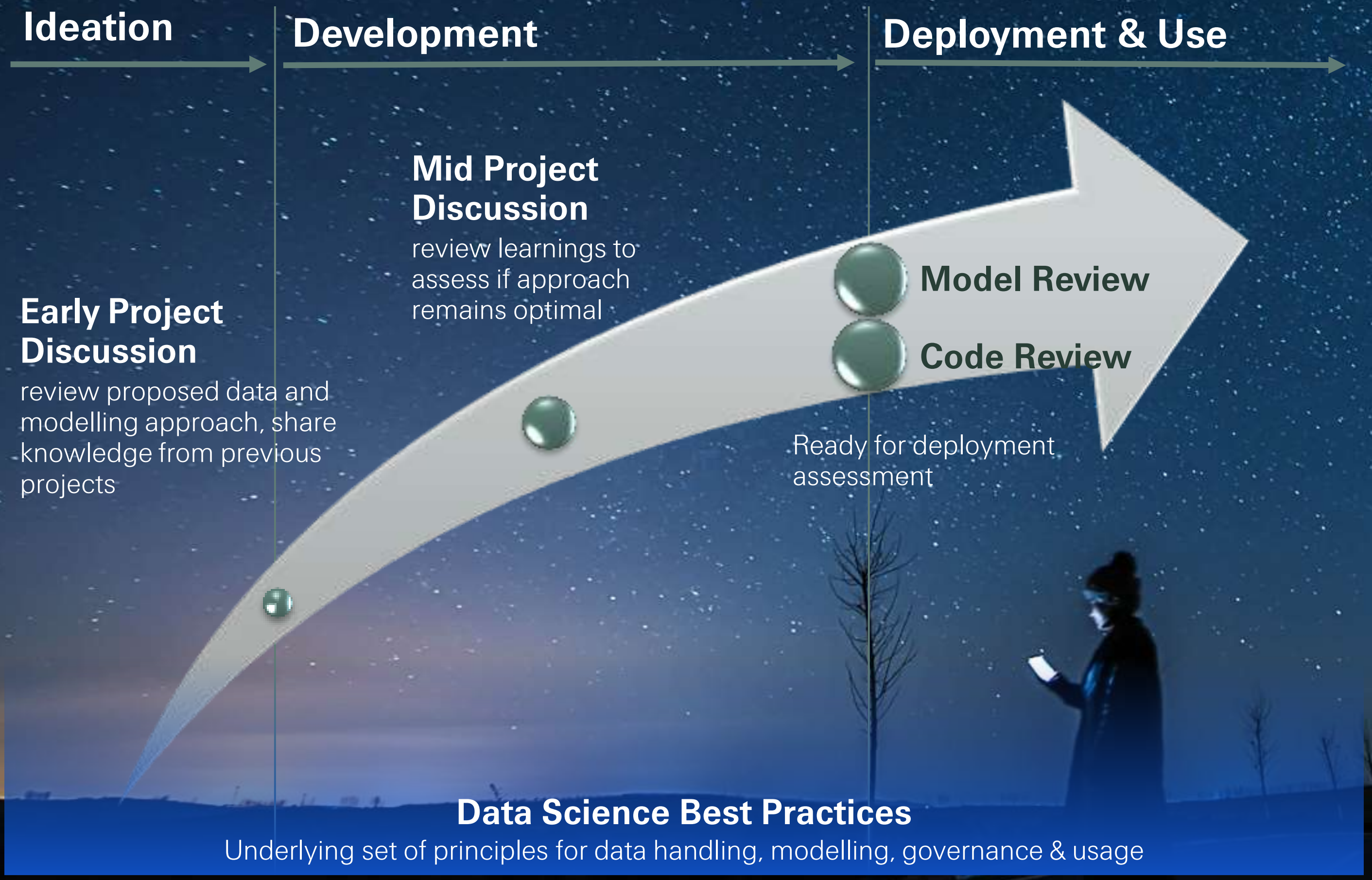
... hence the need for a comprehensive governance embedded in our current standards & policies

- Develop guidelines to facilitate and monitor **responsible use** of data & models
- Create **guidance** on **model design** and **development** to improve **scalability** of use cases
- **Systematize independent reviews** of model components to support **auditability**
- Assign clear **responsibilities** and **accountabilities**

Non exhaustive – illustrate our current focus

Analytics Governance run by our experts

Ensuring Analytics Quality Standards for the Group through the Advanced Analytics **Peer Review Framework**



1. Approach & solution are **fit-for-purpose**
2. **Performance metrics** are well-target
3. **Data quality** is sufficient
4. Model **assumptions, parameters & limitations** are appropriate
5. **Code** is of high-quality and well documented
6. Model has been properly **tested & validated**
7. **Responsible** use of data & models

Sharing our best practices since 2019

Trade-off #5

How to balance governance, speed-to-market and costs?

Some recommendations

01
Foster awareness as the topic keeps increasing in importance among regulators and civil society

02
Adopt a risk-based approach and pilot governance, e.g., starting from a CoE

03
Create transparency on AI/ML use cases and models, e.g., via a model catalogue

04
Strengthen best practices via peer reviews, onboarding and community building activities

05
Listen to all of your stakeholders and iterate to refine the governance approach to fit your organization

"Much has been written about AI's potential to reflect both the best and the worst of humanity.

As leaders, it is incumbent on all of us to make sure we are building a world in which every individual has an opportunity to thrive."

Andrew Ng

Thank you!

Any questions?

Legal notice