AI Governance for Actuaries

Samuel Chu

Aug 2024

Agenda

1. Landscape of regulations
2. FEAT principles and how to apply them
3. Case study: Predictive Underwriting
4. Challenges & Considerations:

State of AI Regulations in Singapore

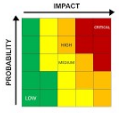
	Legislation				Guidelines		
Harms	**Personal Data Protection**	**Consumer Protection / Online Safety**	**Fair Employment Practices**	**Intellectual Property Office of Singapore**	**MAS: FEAT Guidelines**	**IMDA: Model AI Governance Framework**	**MOH: AI in Healthcare Guidelines**
Misuses data or personal information	✓			✓	✓	✓	✓
Produces an incorrect output	✓				✓	✓	✓
Provides unfair or unreasonably harsh treatment		✓			✓	✓	✓
Provide misleading advice or info		✓			✓	✓	✓
Discrimination based on protected attribute			✓		✓	✓	✓
Causes physical, economic or psychological harm		✓			✓	✓	✓

Definition of AI

MAS Definition:
*"AIDA" refers to artificial intelligence or data analytics, which are defined as **technologies** that **assist** or **replace human decision-making**.*

OECD Definition of "AI System"
*An AI system is a **machine-based system** that, for explicit or implicit objectives, infers, from the **input** it receives, how to generate **outputs** such as predictions, content, recommendations, or decisions that can **influence physical or virtual environments**. Different AI systems vary in their levels of **autonomy and adaptiveness** after deployment.*

Firms can calibrate actions and requirements under their internal governance framework based on the materiality of the AIDA-driven decisions.

- Extent to which AIDA is used in decision-making
- Complexity of AIDA model
- Extent of automation of process AIDA-driven decision-making
- 

Severity and probability of impact on different stakeholders, including individuals
- Monetary and financial impact
- Regulatory impact
- Options for recourse available

FEAT Principles

Fairness

Justifiability

1. Individuals or groups of individuals are not **systematically disadvantaged**
2. Use of **personal attributes** as input factors for AIDA decisions is **justified**

Accuracy and Bias

1. Data and models used are **regularly reviewed** and validated for **accuracy** and **relevance** to minimize unintentional bias.
2. AIDA-driven decisions are **regularly reviewed** so that models behave as designed.

Ethics

1. Use of AIDA is aligned with **firm's ethical standards**
2. AIDA-driven decisions are held to at least the **same ethical standards as human-driven decisions**

Accountability

Internal Accountability

1. Use of AIDA is approved by an **appropriate internal authority**
2. Firms using AIDA are accountable for both **internally** and **externally** sourced models

External Accountability

1. Data subjects are provided with **channels to enquire about, submit appeals for and request reviews** of AIDA-driven decisions.
2. **Verified and relevant supplementary data** provided by data subjects are taken into account when performing a review of AIDA-driven decisions.

Transparency

1. Use of AIDA is **proactively disclosed** to data subjects as part of general communication
2. Data subjects are provided, upon request, **clear explanations on what data is used** to make AIDA-driven decisions and **how the data affects the decision**
3. Data subjects are provided, upon request, **clear explanations on the consequences that the AIDA-driven decisions may have on them**

FEAT Principles

Fairness

Code of conduct, Guidance Notes?

Justifiability

1. Individuals or groups of individuals are not **systematically disadvantaged**
2. Use of **personal attributes** as input factors for AIDA decisions is **justified**

Data Quality, Control Cycle

Accuracy and Bias

1. Data and models used are **regularly reviewed** and validated for **accuracy** and **relevance** to minimize unintentional bias.
2. AIDA-driven decisions are **regularly reviewed** so that models behave as designed.

Ethics

Professionalism and Ethics

1. Use of AIDA is aligned with **firm's ethical standards**
2. AIDA-driven decisions are held to at least the **same ethical standards as human-driven decisions**

Accountability

Appointed/Certifying Actuary

Internal Accountability

1. Use of AIDA is approved by an **appropriate internal authority**
2. Firms using AIDA are accountable for both **internally** and **externally** sourced models

Underwriting? Pricing?

External Accountability

1. Data subjects are provided with **channels to enquire about, submit appeals for and request reviews** of AIDA-driven decisions.
2. **Verified and relevant supplementary data** provided by data subjects are taken into account when performing a review of AIDA-driven decisions.

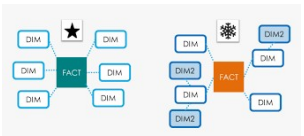
Transparency

Actuarial Advice / Report

1. Use of AIDA is **proactively disclosed** to data subjects as part of general communication
2. Data subjects are provided, upon request, **clear explanations on what data is used** to make AIDA-driven decisions and **how the data affects the decision**
3. Data subjects are provided, upon request, **clear explanations on the consequences that the AIDA-driven decisions may have on them**

FEAT in MLOps

Business Problem



Gather data

Model Training

Diagnostics

Commercial / Technical Review

Deployment framework

Ethics
Set up values / principles / commitments

Fairness
Justify use of personal attributes

Fairness
Test for individuals or groups being systematically disadvantaged

Transparency
Clear explanations on what data is used and how it affects decision

Fairness
Justification for unfairness

Ethics
Same ethical standards as human decisions

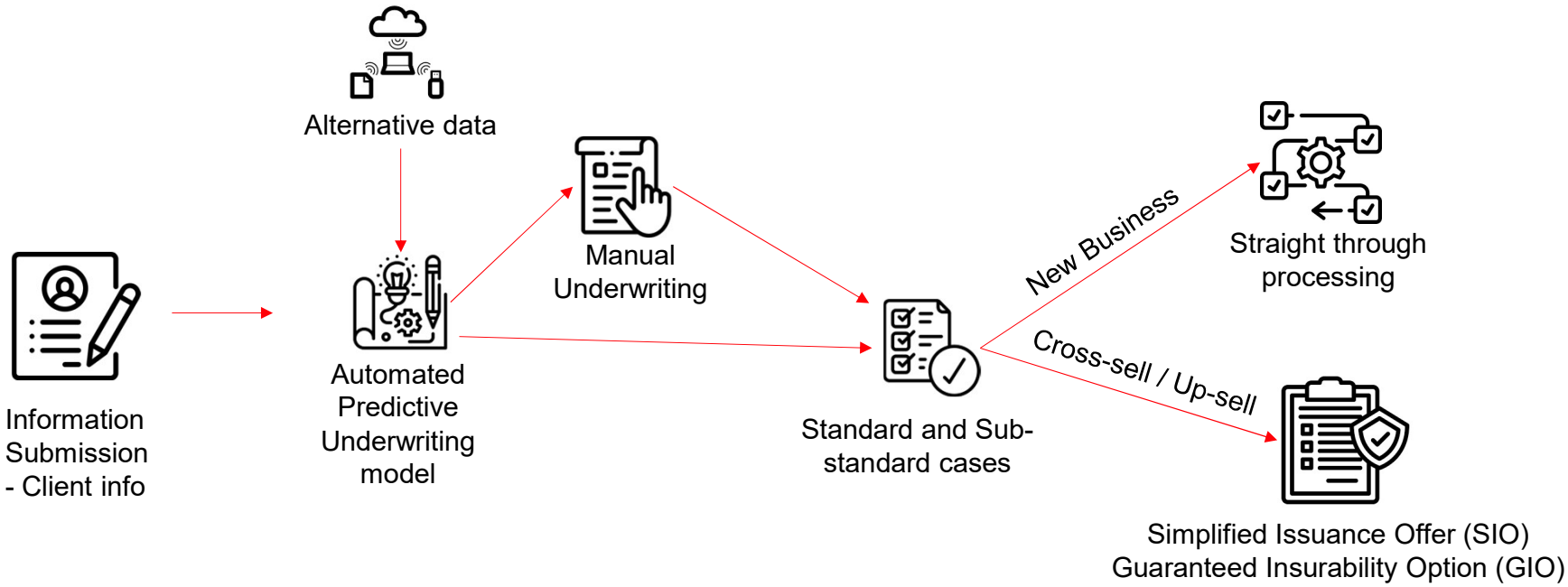
Fairness
Regular review for accuracy and relevance

Accountability
Channel to enquire / appeal about AIDA decisions

Transparency
Proactive disclosure to data subjects

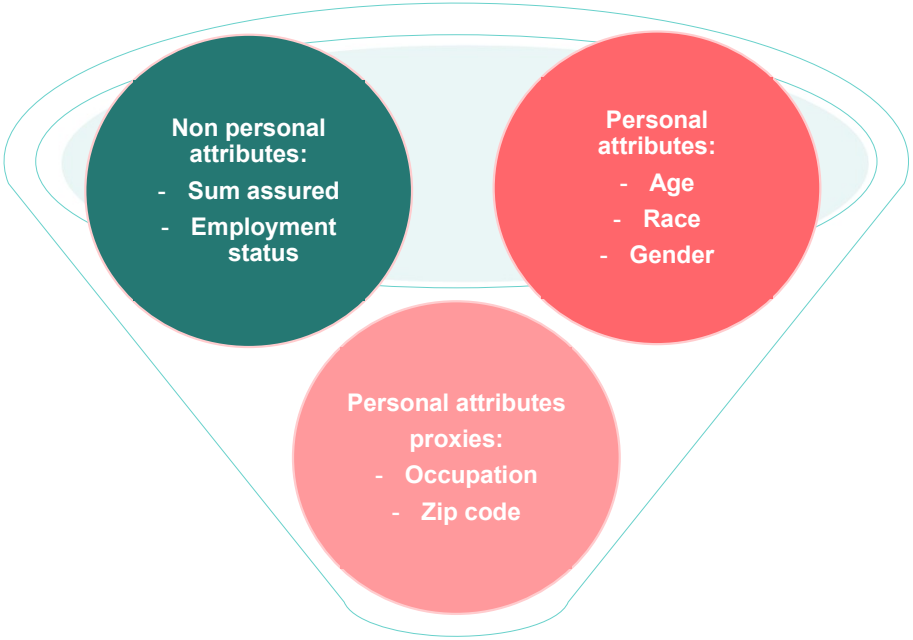
Accountability: Approval by appropriate internal authority

Predictive Underwriting



Predictive Underwriting: Fairness, Use of Personal Attributes

1. Identify the personal attributes

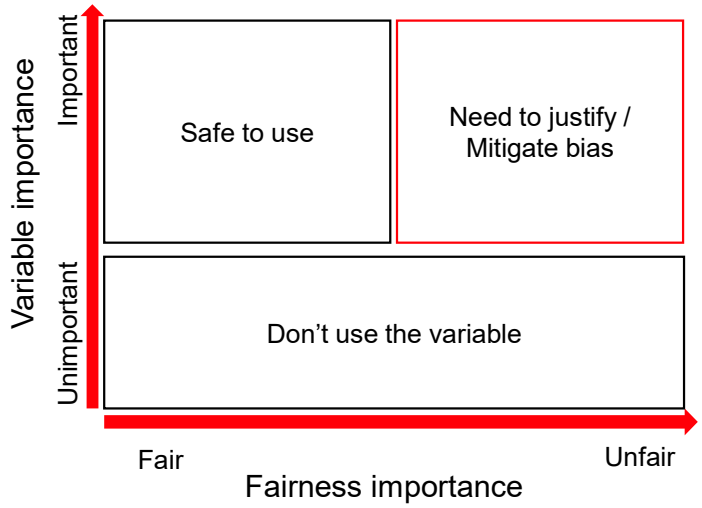


How to identify personal attributes?

- Regulation
- Market practice
- Socio-cultural acceptance
- Organizational values

2. Justify inclusion of every personal / proxy attribute

- Regulatory or essential to product
- Material for fairness objective / historical bias in training data
- Modifiable nature of attributes
- Material for commercial objective, benefit of inclusion outweighs the drawbacks.



Predictive Underwriting: Fairness (Cont.)

https://github.com/mas-veritas2/veritastool

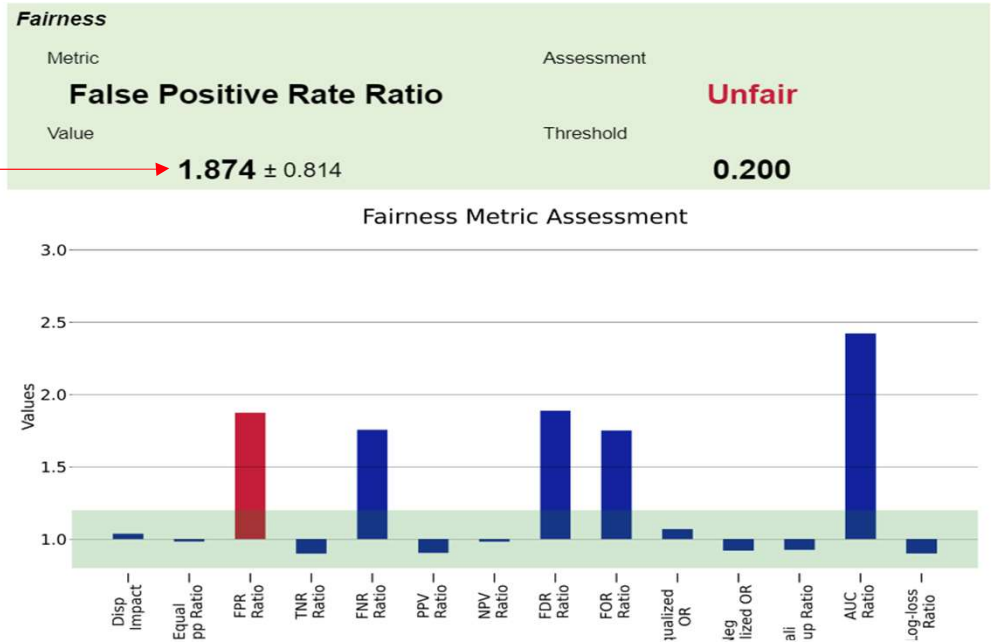
	Predicted Non-standard	Predicted Standard
Actual Non-standard	118	2
Actual Standard	27	415

	Predicted Non-standard	Predicted Standard
Actual Non-standard	139	6
Actual Standard	20	520

$FPR = \frac{27}{27+415} = 0.061$ ➡ 6.1% of male customers are predicted Non-standard although they are actually standard

$FPR = \frac{20}{20+520} = 0.037$ ➡ 3.7% of female customers are predicted Non-standard although they are actually standard

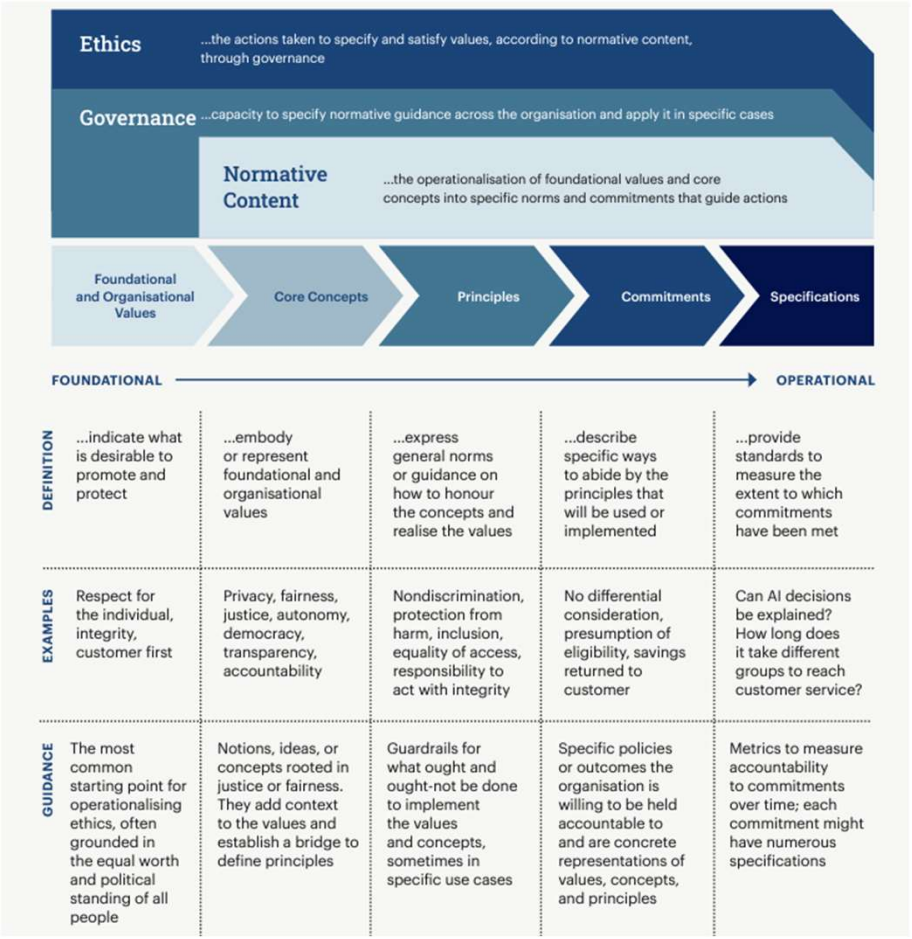
$FPRratio = \frac{0.061}{0.037} \cong 1.874$



$$FPR^1 = \frac{FP^2}{FP^2 + TN^3} \qquad FPR^1ratio = \frac{FPR^1 for\ male}{FPR^1 for\ female}$$

1. *FPR: False Positive Rate*
2. *FP: False positives*
3. *TN: True Negatives*

Ethics & Accountability



Veritas Document 3B – FEAT Ethics and Accountability Principles Assessment Methodology

Ethics

Societal Norms
- Non-discrimination
- Inclusion
- Fairness
- …

Values / Concepts
- Agility
- Trust
- Empathy
- …

➡ **Principles**
- Autonomy and respect of persons
- Meet and exceed privacy and security expectations
- Human involvement in AI decision-making

➡ **Commitments / Specifications**
- No discrimination based on personal attributes (percentage of personal attributes that have fairness value above threshold)

Accountability

Internal Governance Structures
- AI system and tech owner
- Line 1,2,3 controls
- Approval authority

Risk Identification and Controls
- Operational risk framework

Stakeholder Interaction
- Management and board awareness
- Internally and externally sourced model accountability

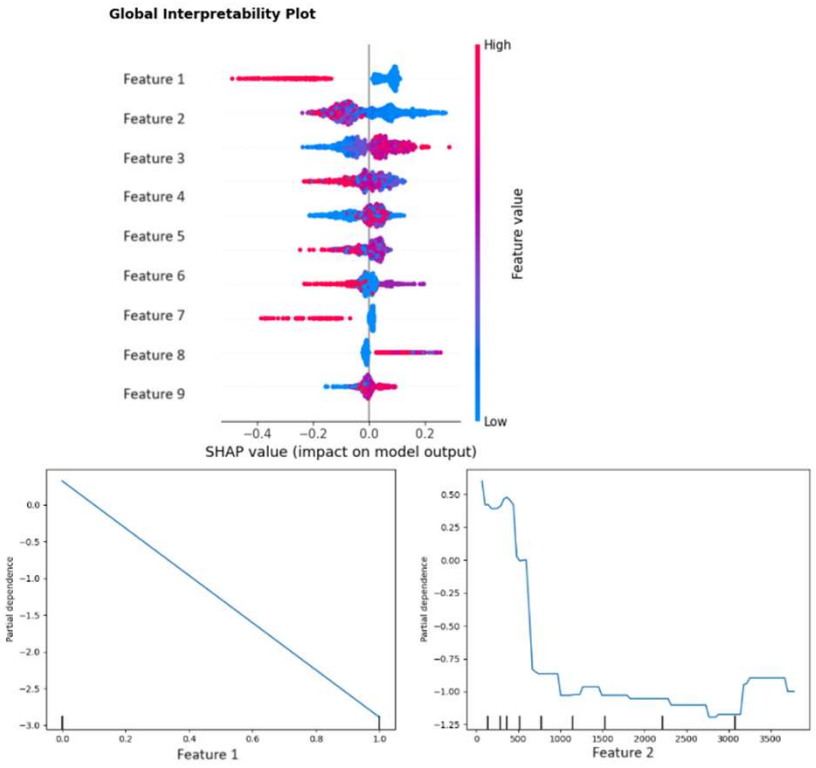
Data Subjects
- Avenue for feedback and recourse for customers
- Supplementary data

Predictive Underwriting: Transparency

https://github.com/mas-veritas2/veritastool

Internal Transparency: Explainability

pred_underwriting_obj.explain()



External Transparency : Data Subjects

Disclosure to customers that their decision is driven by AI

- Proactive vs Reactive
- General vs Specific
- Informational vs Action Orientated

Channel for enquiry

Clear Explanations:

- How was decision made?
- Top reasons behind decision?
- What actions that could have enabled more favorable outcome?
- Future impact?
- Redress options

Lessons Learned

Principles, Governance Risk Control

3 Lines of Defence

- Model owner
- Model risk management
- Internal Audit

Risk-based Governance

Risk framework commensurate with the level of potential harm

Adapting Global Policies

Governance framework that spans multiple geographies

Processes & Technology

Processes to keep pace with evolving tech

- MLOps vs AIOps

Responsible AI Awareness

- Cross-functional teams
- Senior stakeholder education

Culture & Training

AI literacy

- Mandatory certifications

AI Talent

- Qualitative vs Quantitative concepts
- POC vs Production
- End-user education

Use-case Specific

Low materialiaty use cases

- Full fledged framework too onerous for majority of productivity use cases.
- Fairness assessment can be gamified

New tools and techniques

- Unstructured data
- Unsupervised / deep learning
- Non-Python based models

Veritas Document 5 - From Methodologies to Integration