



OFFICIAL

Data Science Lifecycle

Procedural Instruction

Document ID (PPN)	TI-7143
TRIM Record Number	ADD2024/102285579
BCS Function	Technology and Information Management
Document Owner	Senior Director, Data Science & Applied AI
Approval Date	28 March 2025
Document Contact	Data Science & Applied AI Branch s. 22(1)(a)(ii) @homeaffairs.gov.au

Released by Department of Home Affairs
under the Freedom of Information Act 1982

OFFICIAL

OFFICIAL**Table of Contents**

1.	Purpose	3
2.	Scope	3
3.	Overview of the Data Science Lifecycle	3
4.	Accountabilities and Responsibilities	4
4.1.	Definition of roles	4
4.2.	RACI Matrix	5
5.	Project Types	6
6.	Requirements in Each Lifecycle Stage	6
6.1.	Identify	7
6.1.1	Roles & Responsibilities	7
6.1.2	Documentation & Artefacts	8
6.2.	Collect, Prepare and Approach	9
6.2.1	Roles and Responsibilities	9
6.2.2	Documentation & Artefacts	10
6.3.	Analyse, Build and Evaluate	11
6.3.1	Roles & Responsibilities	11
6.3.2	Documentation & Artefacts	12
6.4.	Deploy	12
6.4.1	Roles & Responsibilities	13
6.5.	Maintain, Monitor and Promote	14
6.5.1	Roles & Responsibilities	15
6.5.2	Documentation & Artefacts	15
6.6.	Decommissioning	16
7	Version Control	17
	Attachment A – Definitions	18
	Attachment B – Assurance and Control Matrix	19
1.1	Powers and Obligations	19
1.2	Controls and Assurance	19
	Attachment C – Consultation	20
1.1	Internal Consultation	20

 Released by Department of Home Affairs
under the Freedom of Information Act 1982

OFFICIAL

1. Purpose

This document provides a comprehensive and systematic roadmap for the delivery of data science and artificial intelligence (AI) projects within the Department of Home Affairs (the department) and the Australian Border Force (ABF). Data science and AI are already used in many areas of the department, helping to detect threats to our borders, facilitate trade and travel, automate routine tasks and drive policy monitoring and efficacy. With continued advances in technology, data science and AI can and should play a growing role in support of the portfolio's functions and strategic objectives.

Informed by industry best practice, this procedural instruction (PI) serves as a guide for data scientists, analysts, and product managers to follow, and defines the processes, responsibilities, milestones, and key interactions necessary for the successful delivery of data science and AI projects.

The delivery of a data science and/or AI product involves a continuous iterative process of model development, training, testing, and deployment – with a focus on data quality and model performance – whereas traditional software delivery emphasises the development, testing and deployment of code-based applications.

By establishing a common language and approach to data science and AI, this document aims to ensure alignment with departmental policies and accountabilities and foster collaboration and capability uplift across the department. Crucially, it seeks to encourage safe, accountable and ethical use of AI across all business areas in the portfolio.

2. Scope

Data science products are those that use scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data. This PI applies to all projects where algorithms are trained on the department's official administrative data. It covers projects undertaken by the Data and Economic Analysis Centre (DEAC), as well as other teams conducting data science and AI work in the department and the ABF.

In addition to the requirements detailed below, data science products should be delivered in line with other policies such as the department's Artificial Intelligence Ethics Policy – Policy Statement (SM-7141), the Privacy Impact Assessment Process – Procedural Instruction (TI-5413) and the Data Security and Access Management – Policy Statement (TI-3470).

Where third-party *generative* AI tools and services are utilised, an AI Risk Management Plan will be required. New products, particularly high risk AI systems, many also need to be assessed through the Commonwealth AI Assurance Framework.

3. Overview of the Data Science Lifecycle

The data science lifecycle is a systematic, iterative process that guides data science projects from conception to deployment and maintenance. Adopting the lifecycle helps to ensure that data science projects are executed efficiently and effectively, leading to sustainable and fit-for-purpose products.

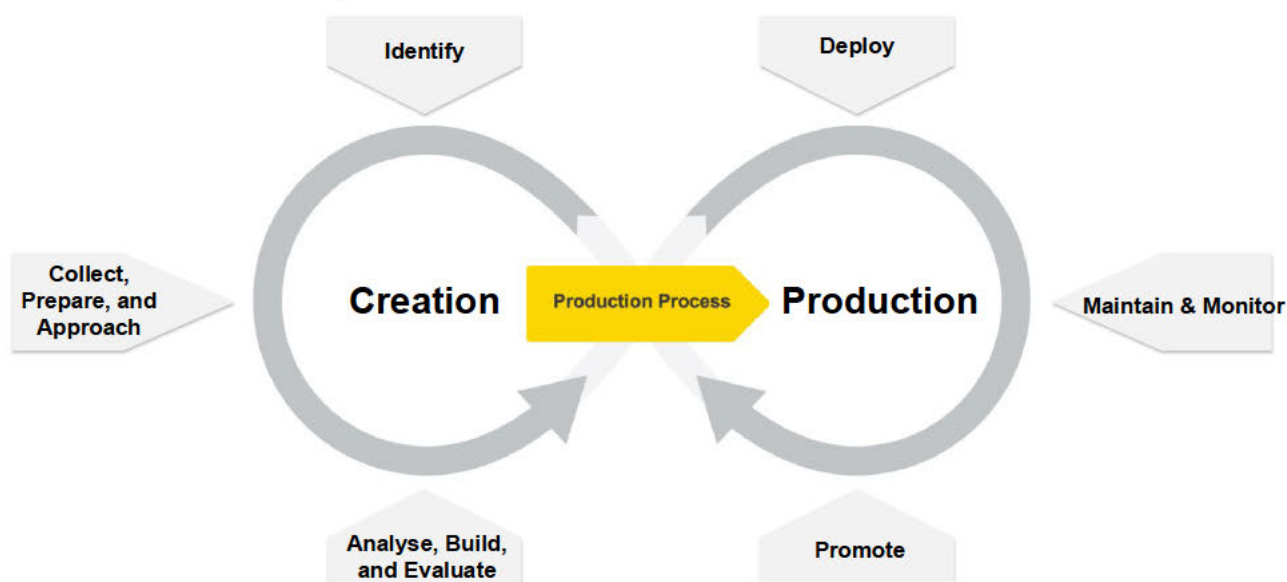
The stages of the data science lifecycle are set out in the diagram below.

1. Identify
2. Collect & prepare
3. Approach

Released by Department of Home Affairs
under the Freedom of Information Act 1982

OFFICIAL

4. Analyse
5. Build & evaluate
6. Deploy
7. Maintain, monitor & promote



Timeframes and requirements for each stage will vary depending on the specific project and business context. Stages of the lifecycle may occur simultaneously. Some elements may need to be repeated before embarking on the next stage in the lifecycle. Since data science is iterative in nature, this PI prioritises experimentation and feedback loops over strict adherence to linear steps.

4. Definition of Roles

Setting clear roles and responsibilities for each data science project is critical. Delineating tasks based on expertise ensures the right individuals handle the appropriate tasks, preventing overlaps and ensuring no aspect of the project is overlooked. This structure not only streamlines execution, but ensures that efforts aren't duplicated and all essential tasks are covered.

Data science projects typically involve the following role types. The specific project scope will dictate how roles are allocated across the team, and may require individuals to work flexibly and undertake multiple roles. On larger projects with multiple business users or complex interlocking delivery components, additional roles and governance arrangements may need to be created.

Role	Description
Business Sponsor	A representative of the business area that champions the development, deployment, and outcomes of the project. They are responsible for providing access to business data and processes, providing regular feedback to ensure the product delivers business value and ensuring the product is used as intended, by authorised officers, once deployed. The business sponsor is either an SES officer or has SES-level support for the project, including SES endorsement of any funding and resourcing commitments.

Released by the Department of Home Affairs under the Freedom of Information Act 1982

OFFICIAL

Role	Description
Product Manager	Responsible for overseeing and guiding the project from initiation to completion, ensuring it stays on schedule, within budget, within scope and delivers on project objectives. They facilitate communication among team members and stakeholders, and manage risks, changes and budget. Once a project is delivered, the product manager remains accountable for promoting and improving the product to maximise its value and return on investment.
Technical Lead	Responsible for providing guidance on the technical aspects of the project and leading the technical delivery of the product. They make critical decisions to ensure best practices and applicable ICT processes are followed.
Data Scientist	Applies statistical, computational, and machine learning techniques to solve business problems and contribute to the delivery of the product. They design, build, and validate models to derive insights and address the project's objectives.
Data Engineer	Designs data storage and pipelines, and ensures data is readily available, reliable, and optimised for use by data scientists.
Product Support	Provides ongoing support, maintenance and monitoring for the product post-deployment to ensure it remains usable. Responsible for promoting the product to end users.

4.1. RACI Matrix

The Responsible, Accountable, Consulted, Informed (RACI) process is a responsibility assignment matrix used to clarify the roles and responsibilities in projects and processes. The roles are as follows:

- **Responsible (R):** Role that actually completes the task or process
- **Accountable (A):** Role that is ultimately answerable for the completion of the task. Often this person is the key decision maker.
- **Consulted (C):** Roles who provide information or expertise necessary for the task and with whom there is two-way communication
- **Informed (I):** Roles that need to be kept updated on progress or decisions. This is typically a one-way communication.

Tasks/Phases	Product Manager	Tech Lead	Data Scientist	Data Engineer	Product Support	Business Sponsor	ICT
Identify	AR	R	C	C	-	C	C
Collect & Prepare	A	C	C	R	-	C	I
Approach	A	R	C	C	-	C	C
Analyse	C	A	R	R	-	I	I

Released by Department of Home Affairs under the Freedom of Information Act 1982

OFFICIAL

Build & Evaluate	C	AR	R	C	-	C	C
Deploy	C	AR	C	C	C	I	R
Maintain, Monitor & Promote	R	C	C	C	AR	C	C

5. Project Types

Under this PI, the documentation and artefacts required at each stage of the Data Science Lifecycle depend on whether the project is delivering a proof of concept, prototype or product. To ensure consistency of language and approach, these three concepts are defined below.

	Proof of Concept (PoC)	Prototype	Minimum Viable Product (MVP) / Product
Goal	Prove technical feasibility	Demonstrate potential value to business	Validate that the product works as intended
Purpose	Identify feasibility and technical hurdles	Identify feasibility of use case in a controlled or simulated setting	Identify feasibility of use case in real setting
Development time	< 4 weeks	< 6 weeks	Months
Target users	Users are internal to the department	Internal and external testers, champions or early adopters	End users may be internal or external
Data	Typically uses non-production data	Typically uses a sample of production data	Production data
Duration	30 day max life	60 days max life (with a one-time extension to 90 days in limited circumstances)	Long-term

6. Requirements in Each Lifecycle Stage

Creation Phase

The creation phase is dedicated to exploration, development, and refinement. It revolves around gathering data, experimenting with models, and generating insights. The outcome of the creation phase is a set of ready-to-deploy models or analytical solutions, along with documentation, benchmarks, and reproducible code.

Released by Department of Home Affairs under the Freedom of Information Act 1982

OFFICIAL

6.1. Identify

The identify stage defines the project's objectives, scope, stakeholders, and the resources required. This stage aims to ensure everyone involved has a clear understanding of the project's purpose, goals, timeframes and constraints.

The key questions to answer here are:

- What problem does this project aim to solve?
- What are the boundaries of the project? What is in and out of scope?
- Who has a vested interest in this project? Who are the primary beneficiaries or potential users?
- What resources (human, technical, financial) are needed, and are they readily available?
- What is the expected duration of the project and the requested completion timeline? Do these timelines align?
- What potential risks can we foresee and how can we mitigate them?
- How will we measure the success of this project?

The end of the identify stage should result in a project plan for the business area to agree to and a decision to progress before a product manager is committed to the project. It's better to 'fail fast' in the identify stage rather than embark on a project with an unclear value proposition.

6.1.1 Roles & Responsibilities

Role	Responsibilities & Accountabilities
Business Sponsor	- Define the problem and identify data sources available. - Work with the Product Manager and Tech Lead to refine scope, and define success criteria and metrics. - Commit resourcing (financial and human) to support all stages of the lifecycle.
Product Manager	- Gather requirements with a focus on identifying how data science and AI can improve business processes and systems. - Identify resourcing requirements for business sponsor to agree to provide. - Complete project initiation document and project plan if required. - Consult with Architecture Design Authority on the appropriate solution design. - Engage with Data Governance and Strategic Policy section (DGSP) on data privacy and AI ethics requirements, including the Artificial Intelligence Ethics Policy Statement and the potential need for Data Security and Assess Management (DSAM). - Identify other key stakeholders and assemble the project team.
Tech Lead	- Identify technical requirements or constraints. - Advise on the selection of tools, platforms, and technologies, working with ICT. - Work with Product Manager to estimate timelines related to technical tasks.
Data Scientist	- Provide initial insights into possible approaches and methodologies. - Identify potential data sources.
Data Engineer	- Assess the infrastructure needs. - Evaluate data sources for integration and transformation requirements. - Provide initial feedback on data storage, retrieval, and processing capabilities.

Released by Department of Home Affairs under the Freedom of Information Act 1982

OFFICIAL

Role	Responsibilities & Accountabilities
ICT	- Provide advice on supported platforms for build and deployment and any relevant security or technical requirements. - Provide advice and support on architectural and governance processes to support delivery timelines.

6.1.2 Documentation & Artefacts

The following documentation is required during the identify stage.

Document	Purpose	PoC & prototype	MVP / product
Project Proposal Template	Required for prototype and deployed products. Compiles key product information such as the project scope, goals, success criteria, business value, delivery timelines, security requirements (including data security) and potential project risks.	Not required	Required <i>Template at ADD2024/3792606</i>
Plan on a page	Required for smaller scale projects such as POCs and experimentation. At a high level, captures key information like goal/objective, delivery details, timelines, resources, funding and business value/impact measurements.	Required	Optional <i>Template at ADD2024/3792705</i>

Released by Department of Home Affairs
under the Freedom of Information Act 1982

OFFICIAL

6.2. Collect, Prepare and Approach

The collect and prepare stages involve gathering the required data from various sources, processing and cleaning this data to make it suitable for analysis. The approach stage involves deciding on the most appropriate data science methodologies to use to achieve the project's objective. These can be quite iterative and may require significant data preparation, experimentation and consultation to determine the right approach. The key questions to answer are:

- Where will the data come from? Is it internal, external, or a combination of both?
- Is the data complete, accurate and relevant? What kind of cleaning and pre-processing is required?
- How will the data from different sources be integrated? Are there any inconsistencies across sources and if yes, how will this be reconciled?
- Are there any privacy concerns with the data? Does it include Personally Identifiable Information? Are there regulations or guidelines that need to be followed?
- How much data is expected? Where and how will it be stored?
- Who needs access to the data and how will it be provided?
- What tools or infrastructure are required for data collection and preparation?
- What data science methodology is the most appropriate to achieve the project's objectives and solve the business problem?
- What systems and tools does this methodology need?
- How can the user access and interact with the product? What is required for the user interface?

6.2.1 Roles and Responsibilities

Role	Responsibilities & Accountabilities
Business Sponsor	- Provide access to business data as required. - Advise on data classification and security.
Product Manager	- Work with the Data Engineer and technical teams to ensure the right data is being collected. - Discuss privacy and security risks with Business Sponsor and initiate the Data Security and Access Management process and/or Privacy Impact Assessment as required.
Tech Lead	- Evaluate and recommend methodology for data collection and preparation, consulting with Data Scientist and Data Engineer. - Design approach for technical delivery.
Data Scientist	- Define data requirements based on the project's objectives. - Explore initial data to identify quality issues or insights. - Collaborate with Data Engineer on data pre-processing requirements. - Provide advice on the right data methodology to use based on their analysis of the data.
Data Engineer	- Identify appropriate data sources. - Design and execute data collection and pre-processing.
ICT	- Consulted on the solution design and platforming of the product.

OFFICIAL

Role	Responsibilities & Accountabilities
	- Provide advice and support on access to systems, environments and data, including on any security requirements that may apply.
Independent peer reviewer (recommended)	- Reviews approach to validate if it is appropriate from an ethical, technical and business perspective.

6.2.2 Documentation & Artefacts

The following documentation is required during the collect, prepare and approach stages.

Document	Purpose	PoC & prototype	MVP / product
High-level design document	Provides an overview of an entire system, identifying the main components that would be developed for the product and their interfaces.	Required (Light touch)	Required
Datasheet	Provide structured summaries of datasets with explanations of processes and rationale that shape the data and describe how the data may be used to train or evaluate models.	Required	Required <i>Template at ADD2024/1889326</i>
Data Risk Minute <i>Only required if data is stored in the Connected Information Environment</i>	Allows decision makers to understand and accept the data risks, recommended mitigations and residual risks before the project progresses to the Build Phase. The minute is approved by the Chief Data Officer as the CIE Senior Responsible Officer.	Not required	Required <i>Template at ADD2020/6230871</i>
Threshold Privacy Impact Assessment	A preliminary assessment to determine a project or asset's potential privacy impacts and whether it is a 'high privacy risk project' that requires a full PIA under the Privacy (Australian Government Agencies – Governance) APP Code 2017.	Not required	Required <i>Template at ADD2020/3143122</i>
Privacy Impact Assessment	A systematic assessment of a project that identifies the impact that the project might have on the privacy of individuals, and sets out recommendations for managing, minimising or eliminating that impact.	Not required	Required if deemed necessary by the Threshold Privacy Impact Assessment <i>See Privacy Impact Assessment Process – Procedural Instruction (TI-5413)</i>

Released by Department of Home Affairs under the Freedom of Information Act 1982

OFFICIAL

6.3. Analyse, Build and Evaluate

The analyse, build and evaluate stages make sense of the prepared data and develop predictive or descriptive models or other data science products to best deliver the project's objectives. This stage also tests and validates the performance of these products, while ensuring the product is accessible and user-friendly. The key questions to answer here are:

- Which patterns, trends, relationships, or irregularities emerge from the data?
- What algorithms or techniques are best suited for the project's goals?
- How well does the model perform with regard to the agreed performance metrics e.g. accuracy, precision, recall?
- Are the results and insights aligned with business objectives and actionable?
- Is the model exhibiting any unwanted biases? Is it fair across various demographic or segment groups in accordance with the AI ethics policy?
- How transparent or explainable are the model's outputs and recommendations, particularly if the model will be making substantially automated decisions with a legal or similarly significant effect?
- Based on evaluation, what changes or improvements are required or worth exploring further?

6.3.1 Roles & Responsibilities

Role	Responsibilities & Accountabilities
Business Sponsor	- Review and provide feedback on functionality and usability during the build and evaluate stage.
Product Manager	- Ensure technical delivery is in line with project scope, purpose, timeframes and budget. - Coordinate collaboration between different roles. - Maintain communication plan with all relevant stakeholders.
Tech Lead	- Supervise technical aspects. - Provide guidance in methodology, technology and tool selection.
Data Scientist	- Develop, train and refine models and applications. - Appraise model performance.
Data Engineer	- Undertake data analysis, feature engineering and cleansing to support model build.
ICT	- Consulted on the development environment where the model and applications will be built. - Provide support where ICT platforms are impacted.
Independent peer reviewer (recommended)	- Review the model code and/or model card to validate that it is robust and trustworthy.

Released by Department of Home Affairs under the Freedom of Information Act 1982

OFFICIAL

6.3.2 Documentation & Artefacts

The following documentation is required during the analyse, build and evaluate approach stages.

Document	Purpose	PoC & prototype	MVP / product
Model card	Explains how the model has been built and trained, as well as the context in which the model is intended to be used, details on the performance evaluation procedures and other relevant information. Model cards increase transparency by communication information about trained models to broad audiences.	Required <i>Template at ADD2024/3874439</i>	Required <i>Template at ADD2024/3874439</i>
Code repository	Storage location for code and code documentation.	Required <i>Managed in Azure DevOps</i>	Required <i>Managed in Azure DevOps</i>
Model artefacts	The model itself, any components required to re-build the model – e.g. lineage, links to training data, hyper-parameters.	Required	Required

Production Phase

Once the creation phase has been completed, the business sponsor will need to determine whether the project has demonstrated enough value to proceed to the production phase. The production phase ensures that models built during the creation phase are properly validated and operationalised, then deploys the model in a real world environment and makes it available for business use.

There may be a number of approval processes between the creation and production phases and this will likely differ from project to project. However, as an overarching principle, approvals should be clearly provided and documented before a model moves to production. Models that do not move to production, for example a quick proof of concept, are expected to be decommissioned at the end of their 30 day lifespan. Often the move from creation to production also involves decisions around the appropriate shelf-life of the product and the sustainment funding and product support that may be required going forward.

6.4. Deploy

After development and testing, the focus shifts to deploying the product in a real world environment and making it available for testing and use by business areas.

Questions to ask here include:

- How can the model be best integrated into existing systems or platforms, maximising reusability and minimising technical debt?
- What system-specific cyber security risks must be considered and remediated in order to attain an authority to operate?
- How efficiently will the model perform in real-time or high-demand scenarios?

Released by Department of Home Affairs
under the Freedom of Information Act 1982

OFFICIAL

- Who will assume ongoing responsibility for the product once it is fully deployed? Who will be the business and technical system owners? What support arrangements are needed and how will they be resourced?

It is important to align with ICT architectural and technical governance processes in the stage, particularly:

- Roles and Responsibilities of Business and Technical System Owners – Policy Statement (TI-6262).
- ADD2019/4384796 - Protection Rating Target: a structured approach to identifying the level of security required in a system and the typically required artefacts associated with the introduction of a new or changing system/product.
- ADD2019/5217575 - ICT Environment Release Criteria: provides the expected prerequisites before ICT environment, infrastructure, or platform components can be provisioned.

6.4.1 Roles & Responsibilities

Role	Responsibilities & Accountabilities
Business Sponsor	- Determines product ownership, including identification and resourcing of business and technical owners and Product Support, once the model is successfully deployed.
Product Manager	- Oversees the deployment and coordinates lines of technical effort. - Provides advice on the appropriate lifespan of the product and ongoing support needs once deployed.
Tech Lead	- Work with ICT and cyber security teams to conduct the security impact assessment and seek an authority to operate. - Deploy the model to the appropriate environments. - Conduct testing to ensure performance and stability.
Data Scientist	- Kept informed on deployment progress and testing. - Assist with troubleshooting model errors or output issues as required.
Data Engineer	- Kept informed on deployment progress.
Product Support	- Consulted on the support implications of the deployed product and the ongoing level of support and maintenance that is appropriate.
ICT	- Provide advice on technical pre-requisites and standards that must be met to enable model deployment. - Provision and enable the ICT products and/or platforms supporting model deployment, including at lower level environments (development, test, staging etc). - Assist and support testing activities across environments to ensure performance and stability.

Document	Purpose	PoC & prototype	MVP / product
Security Impact Assessment	A product of the cyber security impact management process that identifies and quantifies system-specific security risks and provides recommendation to mitigate	Not required	Required if using production data

Released by Department of Home Affairs under the Freedom of Information Act 1982

OFFICIAL

Document	Purpose	PoC & prototype	MVP / product
	unacceptable risks. A security impact assessment is required before an Authority to Operate or Interim Authority to Operate is issued.		<i>Guidance at ADD2019/4384796</i>
Authority to Operate	Provides permission for a system/product to operate with other existing systems and networks.	Not required	Required Exceptions may be granted where the technical platform already has an authority to operate. Requires approval from the Technical Director, Data Science Branch.

6.5. Maintain, Monitor and Promote

Once deployed, the model must be continually monitored – both from a technical perspective and to ensure it performs as intended and does not lead to unintended adverse outcomes. The model should be monitored for drift, latency, failure rate, and infrastructure metrics and steps taken to remediate if these metrics go beyond acceptable thresholds. Robust mechanisms should also be put in place to gather feedback from users and business areas.

Questions to ask here include:

- Once deployed, is the model performing efficiently in real-time scenarios?
- How does the model handle increased data volume or changes in data structure?
- How and how frequently will the model's performance be monitored? How will poor performance be remedied and what are the acceptable thresholds for performance?
- To what extent does the model solve the users' business problem?
- Do end users require further workshops or training guides to increase utilisation?
- What are the touchpoints for regular feedback? How will feedback from stakeholders or systems be incorporated into model refinement?
- How will updates or refinements to the model be managed and re-deployed?
- If the model is part of a decision-making process, is that process documented, auditable and able to be contested or reviewed?

Once models are deployed and being actively monitored, the next step is to increase visibility, adoption, and utilisation through active engagement with stakeholders and end users. Rapid business adoption provides greater opportunities to stress test the product and receive user feedback. Engaging with stakeholders to understand how they use the model allows for continuous alignment between technical solutions and business needs. The promote stage is also about identifying opportunities for growth, using or re-using the product in other business domains, and ensuring that the model remains impactful and relevant.

The Data Science Lifecycle encourages a pattern of continuous integration and continuous deployment. As such, the initial deployment is only the beginning, and models need refinement and retraining to adapt to the

Released by Department of Home Affairs
under the Freedom of Information Act 1982

OFFICIAL

changing environment. Poorly performing or deteriorated models will likely move back through to the creation phase for rebuilding, or for a decision to decommission it.

6.5.1 Roles & Responsibilities

Role	Responsibilities & Accountabilities
Business Sponsor	- Identify and seek sustainment funding for Product Support if required. - Embed safe and ethical use of the product into business processes and workflows. - Provide regular feedback to the Product Manager on performance and outcomes from a business perspective.
Product Manager	- Track and report on model performance against benchmarks and success criteria. - Lead the development of SOPs, user guides and model maintenance and monitoring plan. - Calculate the value impact of the product. - Incorporate feedback or issues identified into plans for product refreshes, upgrades or enhancements.
Tech Lead	- Monitor model performance at the technical level. - Inform the development of SOPs and model maintenance and monitoring plan. - Hand over ongoing technical maintenance to the appropriate areas of Product Support.
Data Scientist	- Refine model based on feedback or issues identified. - Work with Product Manager to develop SOPs and model maintenance plans and hand over to Product Support.
Product Support	- Assume responsibility for ongoing product support and maintenance of relevant documentation and user guides. - Manage minor issues regarding use, access and availability. - Escalate major issues to other stakeholders for remediation as required.
ICT	- Monitoring the model in production. - Provide performance and usage logs. - Provide technical support as required under any agreed support arrangement.

6.5.2 Documentation & Artefacts

The following documentation is required at the maintain, monitor and promote stages.

Document	Purpose	PoC & prototype	MVP / product
Standard Operating Procedure (SOP)	Step-by-step instructions for updating the product, refreshing or retraining the model and fixing common issues.	Not required	Required

OFFICIAL

Document	Purpose	PoC & prototype	MVP / product
User guides	Provides detailed instructions to the end user on how to use the product.	Not required	Required
Maintenance & Monitoring Plan	Sets out how the product will be maintained including considerations of resourcing, business continuity procedures and any relevant service level agreements or support agreements. The plan should include requirements and procedures to monitor for model drift and bias.	Not required	Required

6.6. Decommissioning

All models and applications will eventually be retired or decommissioned. A thorough decommissioning process is important for maintaining data security, preventing misuse of a product and reducing the proliferation of 'shadow IT' or technical debt across the department.

Questions to ask here include:

- Which business users, if any, will be impacted by the decommissioning of the product?
- What performance criteria indicate that a model is no longer fit for purpose?
- How can learnings from the product be captured, including any performance evaluations, intellectual property or code?

The Business Sponsor or Business System Owner for the product will usually make the formal decision to retire or decommission it. The Product Manager will then work with the other stakeholders to ensure that all code bases and documentation created throughout the lifecycle are properly archived, system accesses are removed and data is appropriately secured or deleted.

7. Accountabilities and Responsibilities

Role	Description
Data and Economic Analysis Centre (DEAC)	Responsible for Home Affairs' data and AI strategy, AI governance and compliance, specialised AI products, data science and analytics services and governance of third-party AI tools.
Chief Data Officer	Home Affairs' senior responsible officer for department usage of AI.
Product Managers within Data Science and Applied AI Branch (DSAI), DEAC	Leads the delivery of data science and AI products and services within the department and ABF. Provides advice and guidance on use of data science and AI methodologies. As owners and practitioners of the Data Science Lifecycle – Procedural Instruction, DSAI sets the processes and procedures to be followed for the delivery of data science and AI projects.

Released by Department of Home Affairs
under the Freedom of Information Act 1982

OFFICIAL

Role	Description
Information and Communications Technology (ICT)	Sets technology direction and standards for the department and is responsible for ICT security. Provides advice and guidance on the appropriate deployment solution to meet a project's objectives. If the deployment platform is ICT-owned, ICT maintains and supports the technical infrastructure that products are deployed to.

8. Version Control

Version number	Date of issue	Author(s)	Brief description of change
1	01 November 2023	s. 22(1)(a)(ii)	Initial Draft
2	28 March 2024	s. 22(1)(a)(ii)	Reviewed and updated
3	23 April 2024	s. 22(1)(a)(ii)	Reviewed following discussion among DSAI EL2s and Senior Director
4	8 May 2024	s. 22(1)(a)(ii)	Reviewed following CDO feedback
5	17 June 2024	s. 22(1)(a)(ii)	Reviewed following staff feedback
6	28 June 2024	s. 22(1)(a)(ii)	Final review before departmental consultation
7	12 July 2024	s. 22(1)(a)(ii)	Released for two weeks of all staff consultation to 26/7/2024
8	14 October 2024	s. 22(1)(a)(ii)	Revised following workshop with key stakeholders on 17/9/2024
9	28 March 2025	s. 22(1)(a)(ii)	Final revision based on PPCF comments and approved for publication

Released by Department of Home Affairs
under the Freedom of Information Act 1982

OFFICIAL

Attachment A – Definitions

Term	Acronym (if applicable)	Definition
Data Science		Data science is an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data.
Artificial Intelligence	AI	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. ¹
Responsible, Accountable, Consulted, Informed	RACI	A model that describes the participation by various roles in completing tasks or deliverables for a project or business process.

Released by Department of Home Affairs
under the Freedom of Information Act 1982

¹ OECD definition and adopted by the Commonwealth AI Assurance Framework

OFFICIAL**Attachment B – Assurance and Control Matrix****1.1 Powers and Obligations**

Implementation of this PI does not require any powers, delegations or authorisations.

1.2 Controls and Assurance

Related Policy	Data Security and Access Management Policy Statement and Procedural Instruction (TI-3470).
Procedures/Supporting Materials	Project proposal template, s. 47E(d) Plan on a page template, s. 47E(d) Datasheet template, s. 47E(d) Data risk minute template, s. 47E(d) Threshold privacy assessment template, s. 47E(d) Privacy Impact Assessment Process – Procedural Instruction (TI-5413) Model card template, s. 47E(d)
Training/Certification or Accreditation	Nil
Other required job role requirements	Nil
Other support mechanisms	Data Science & Applied AI Branch s. 22(1)(a)(ii) @homeaffairs.gov.au
Escalation arrangements	For additional support or critical matters, contact Chief Data Officer, Data and Economic Analysis Centre, at s. 22(1)(a)(ii) @homeaffairs.gov.au.
Recordkeeping	Content Manager (TRIM)
Program or Framework	Information Governance
Job Family Framework Role	All roles under the Data and Research job family and ICT and Digital Solutions job family

Released by Department of Home Affairs
under the Freedom of Information Act 1982

OFFICIAL

Attachment C – Consultation

1.1 Internal Consultation

- Data Governance and Strategic Policy Section
- Data Services Branch
- Chief Data Officer
- IT Service Management Branch
- Group Manager, Technology and Major Capability Group
- Threat Risk and Intervention Systems Section
- Risk, Data and Insight Systems Branch
- Customs and Border Modernisation Branch
- Organisational Compliance & Assurance Section

Released by Department of Home Affairs
under the *Freedom of Information Act 1982*

OFFICIAL



Australian Government

Department of Home Affairs

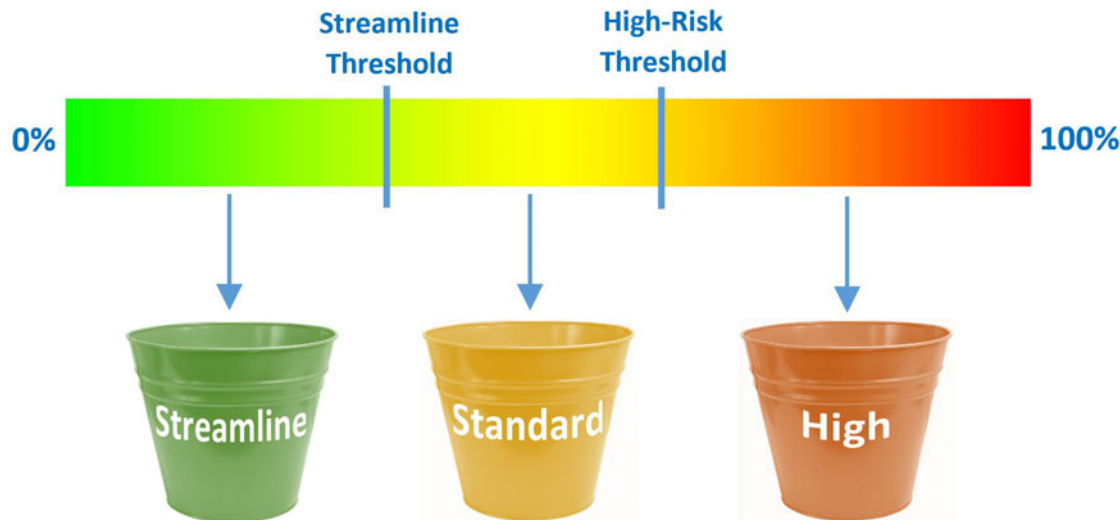
s. 47E(d)

Immigration and Integrity Systems Section (IISS)

2025

s. 47E(d)

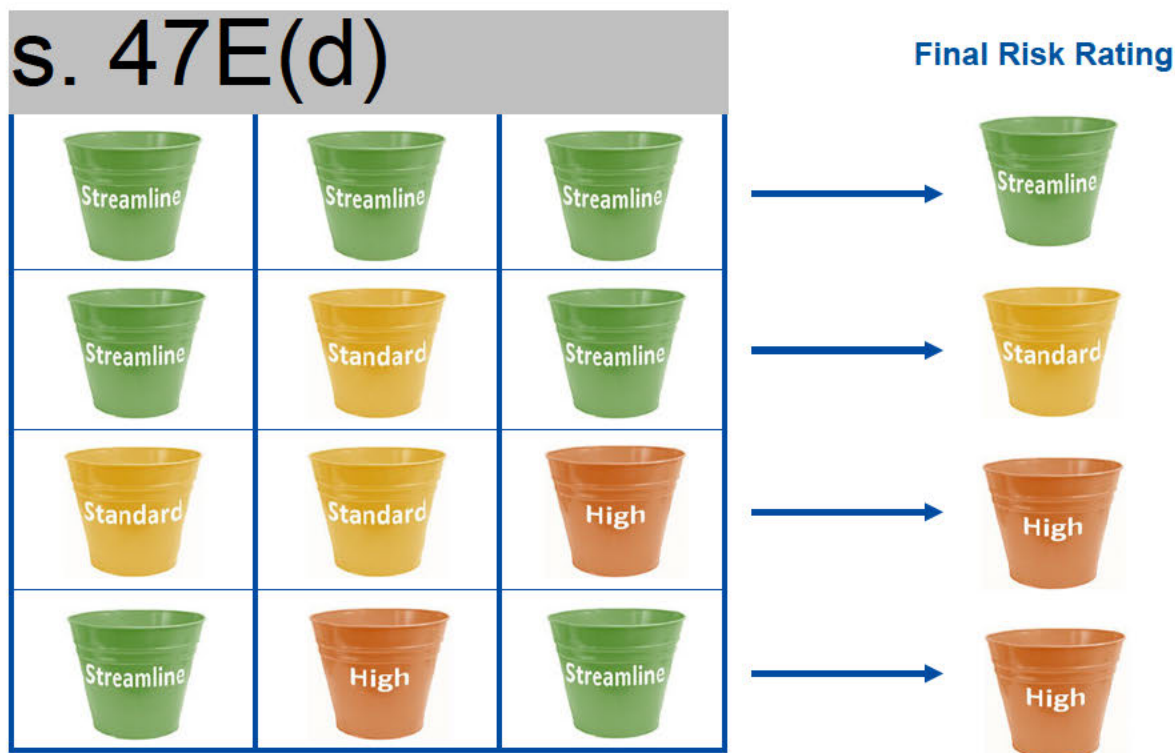
- The model initially makes a probabilistic prediction between 0 and 100% on the risk of *each spoke*
- If that risk is more than a business-determined high-risk threshold, it is considered *high*
- If that risk is less than the business-determined low-risk threshold, it is considered *streamline*
- Otherwise, that spoke is considered to be of *standard* risk



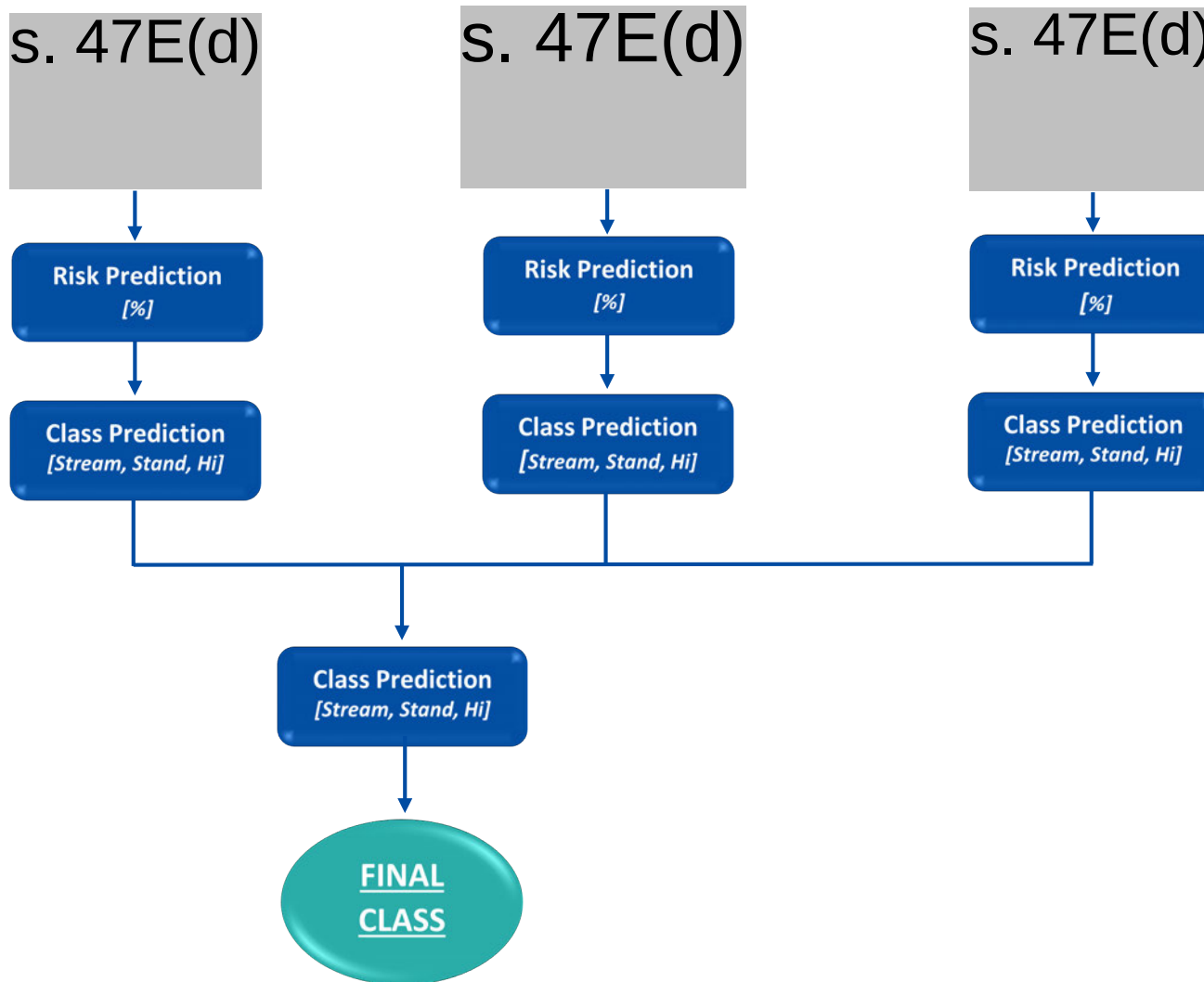
s. 47E(d)

- From this, we generate a final risk classification as the highest risk classification of all spokes
- This is the final risk rating provided to a Visa Processing Officer

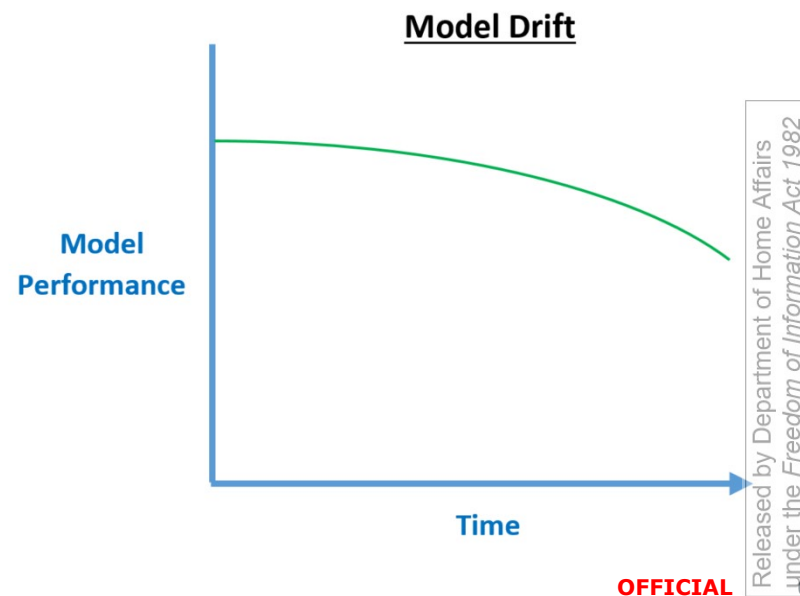
Example:



s. 47E(d)



- Models are created and tuned to perform the best against a particular dataset provided at the time
- Overtime, the characteristics and patterns of incoming data can experience unforeseen changes that models are not prepared for
- Predictive models tend to decrease in performance with time – referred to as *Model Drift*
- Using inaccurate risk systems in visa processing could lead to misinformed immigration decisions at great financial, societal, reputational and social costs for the department, immigrants and Australian society at large



“The department must continually monitor and review both internally developed and third-party AI systems to ensure they are operating reliably and in accordance with their intended purpose and government frameworks.

...

*The System Business Owner is responsible for ensuring these steps are undertaken at every stage across the AI system’s lifecycle from model selection, data preparation and model training **through to model performance monitoring and maintenance.**”*

Excerpt from Artificial Intelligence Ethics Policy (2024)

Section 4.8: Reliability and Safety

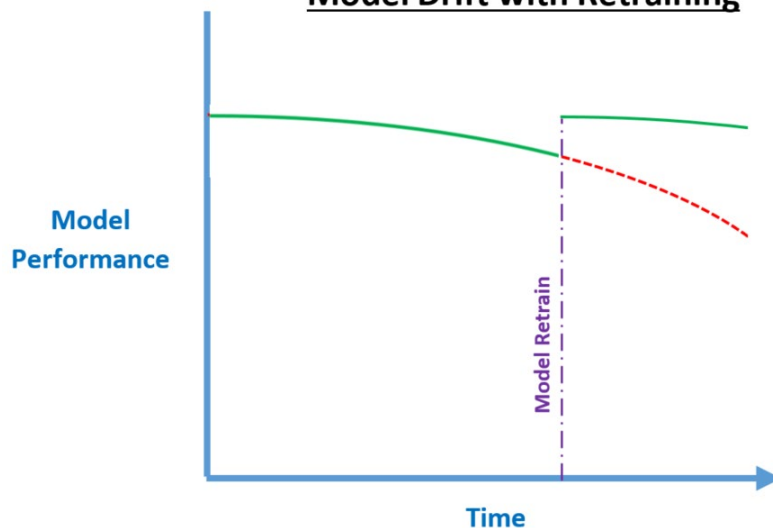
Department of Home Affairs

s. 47E(d)

Types of Model Drift include:

- **Seasonal Drift** – reoccurring cyclical patterns of change in data – *e.g. recurring annual patterns in tourism to Australia influencing immigration*
- **Gradual Drift** – slow, long-term changes in input data – *e.g. growing political unrest leads to a slow increase in emigration from a country*
- **Sudden Drift** – a sudden, unexpected event affecting data distribution – *e.g. the COVID-19 pandemic suddenly limiting immigration to Australia*

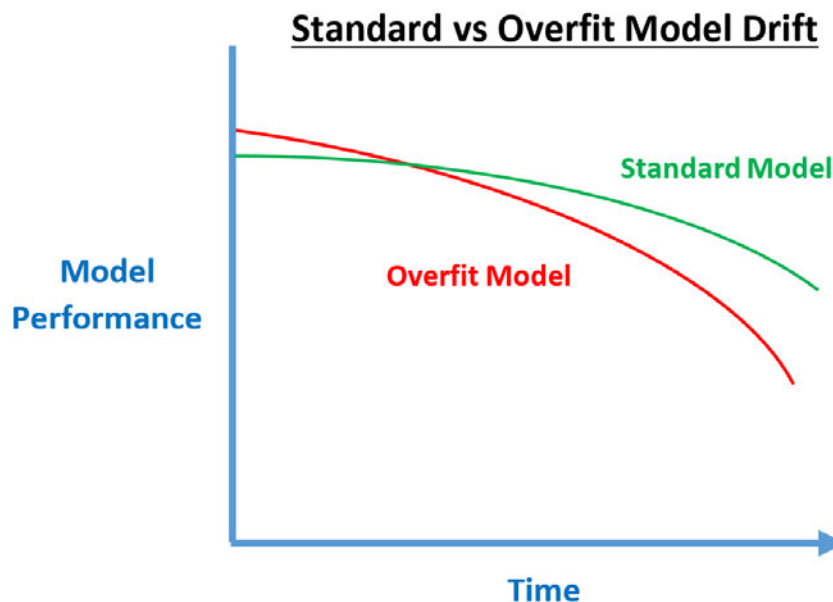
Model Drift with Retraining



Decline in performance that comes about from drift should be addressed by retraining models, i.e. rebuilding them with updated parameters that respond better to the new characteristics of incoming data

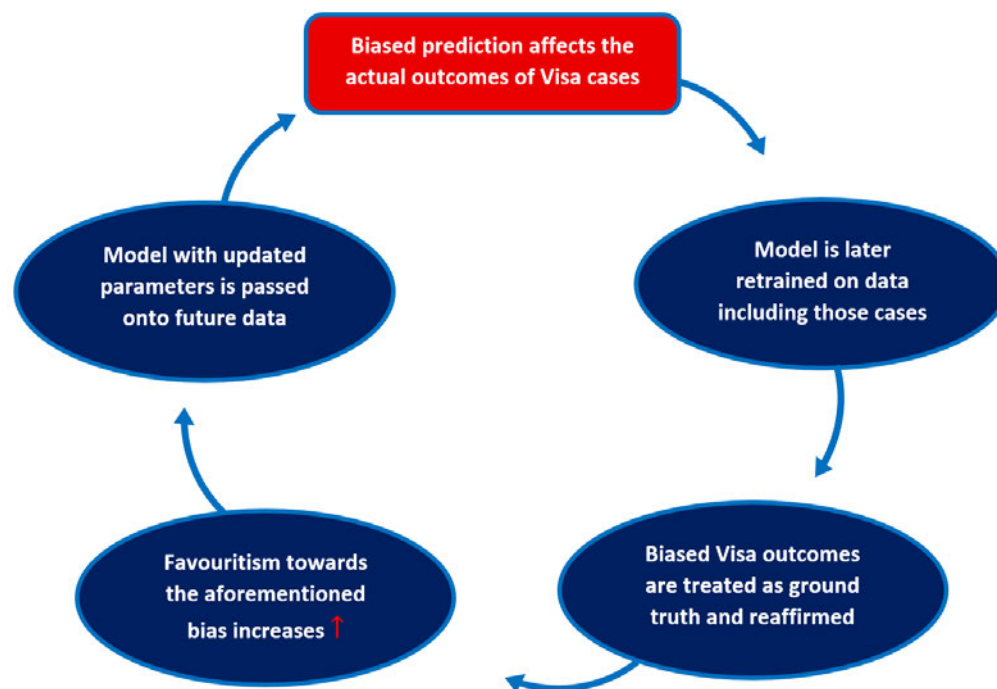
Overfitting

- Models are refined using the input characteristics and their outcomes for a training dataset
- Models that are fit too strictly to the data they are fit on may not perform well on unseen data
- These models tend to be more sensitive to changes in data and experience more severe and rapid model drift



- **Feedback Loop**

- Some adverse outcomes are influenced by the predictions of the model, s. 47E(d)
- Predictions of the model and their biases are feed into later training and taken as truth
- Creates self-fulfilling prophecy amplifies biases with retraining and time
- This can not be addressed by our testing process and should be cautiously considered



- Part of responsible machine learning use and management is ensuring that models are up to date and performing well
- Even if nothing seems wrong, we ourselves should still get occasional check-ups and screenings to make sure we are in good health; it's the same for models!
- We need a test to be able to identify when a model is *sick*, and develop informative evaluation procedures to help *diagnose* what is causing it
- This is vital to ensuring effective risk assessment and, ultimately, fair Visa outcomes in immigration processing



- Predictions are classified using a *Confusion Matrix*
- We group the classes so that we have a 2-class basis to compare to the actual outcome, *i.e.* *low risk* → *negative*, *standard or high risk* → *positive*
- In the context of our data, and s. 47E(d) specifically:
 - **True Positive:** non-compliant visa applicants correctly flagged as standard or high-risk, with a *match* between prediction and actual class
 - **False positive:** applicants flagged for risk that are actually compliant with visa conditions
 - **False Negative:** applicants predicted as low-risk that actually did not comply with visa conditions
 - **True Negative:** compliant visa applicants correctly predicted as low-risk, with a *match* between prediction and actual class

	Actual Positive	Actual Negative
Predicted Positive	True Positive	False Positive
Predicted Negative	False Negative	True Negative

- **Class imbalance** – many visa programs have a very small number of adverse outcomes compared to compliant applicants
- Targeting the minority class in testing when the classes are imbalanced like this can be like *searching for a needle in a haystack*
- Evaluation metrics are quantitative measures of the performance of a model in making predictions
- Some evaluation metrics, such as accuracy, respond poorly to class imbalance



$$\text{Accuracy} = \frac{\text{correct predictions}}{\text{correct predictions} + \text{incorrect predictions}}$$

- **Example:** consider a cohort where only 1% of applicants will result in adverse outcomes
 - Then test a model that naively ALWAYS predicts risk class as negative/low-risk
 - This model guesses class correctly 99% of the time and thus has 99% accuracy
 - However, it NEVER correctly assigns risk predictions for the positive minority
 - Accuracy does not fairly capture the poor predictive performance of this model

Misclassification cost –

- *False negative* misclassifications are generally more harmful than *false positive* misclassifications. **Comparative example:**
 - A *false negative* leading to **S. 47E(d)** can incur hefty financial and time costs through administration, legal fees, and compliance monitoring
 - A **False Positive** may lead to 20 minutes of unnecessary time spent in over-processing an application and a slightly increased workload
 - There may be incentive to accept more false positives if it reduces false negative misclassifications
 - It's about finding a balance to reduce overall cost
- Furthermore, some negative misclassification outcomes may be more expensive than others **S. 47E(d)** etc. all come with different costs)

True positive prioritisation –

- We have a targeting challenge where we are more focussed on identifying true positives (high-risk applications) than true negatives
- Our selected metric should target true positive identification as the primary success factor

Thus, our testing process needs to balance and account for all three of these unique challenges compared to a standard performance evaluation problem:

- ***Varying misclassification costs***
- ***Class imbalance***
- ***Targeting of the Positive class***



β Score

- Need a metric that can handle class imbalance and dynamic misclassification costs
- Key terms:
 - **Precision** – the proportion of cases predicted as positive risk out of all cases that resulted in adverse outcomes;

$$precision = \frac{True\ Positive}{All\ positive\ predictions}$$

- **Recall** – the proportion of cases resulting in adverse outcomes out of all cases predicted as positive that resulted in adverse outcomes;

$$recall = \frac{True\ Positive}{All\ positive\ cases}$$

β Score

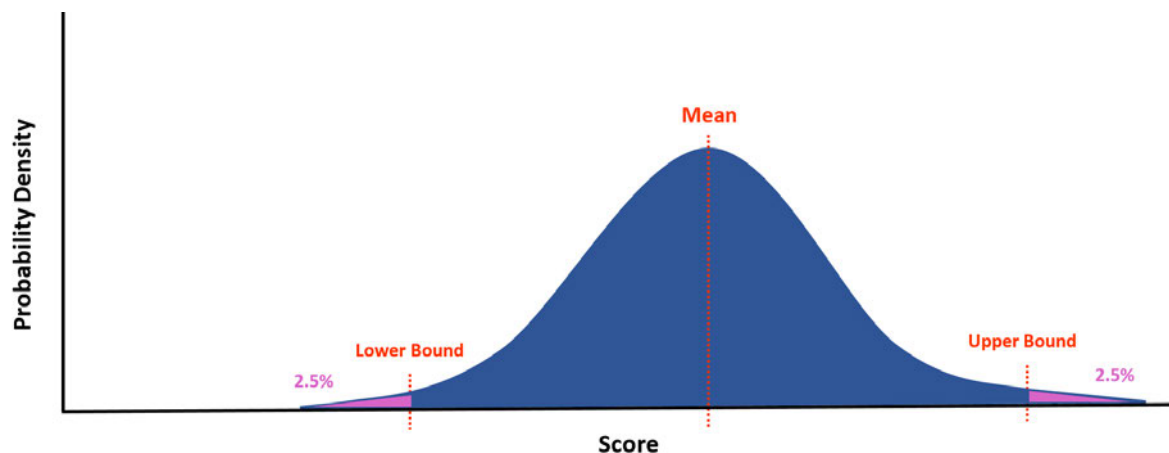
- F_β Score *balances precision and recall based on the risk tolerance for a visa program*
 - **i.e. it balances assessment of the accuracy of all positive predictions, and correctly categorising all adverse outcomes**
- The result is a performance measurement between 0 and 1
- We can then flag a model for retraining and updating when measured performance falls below this level

F-Score Calculation:

s. 47E(d)

β Score: Confidence

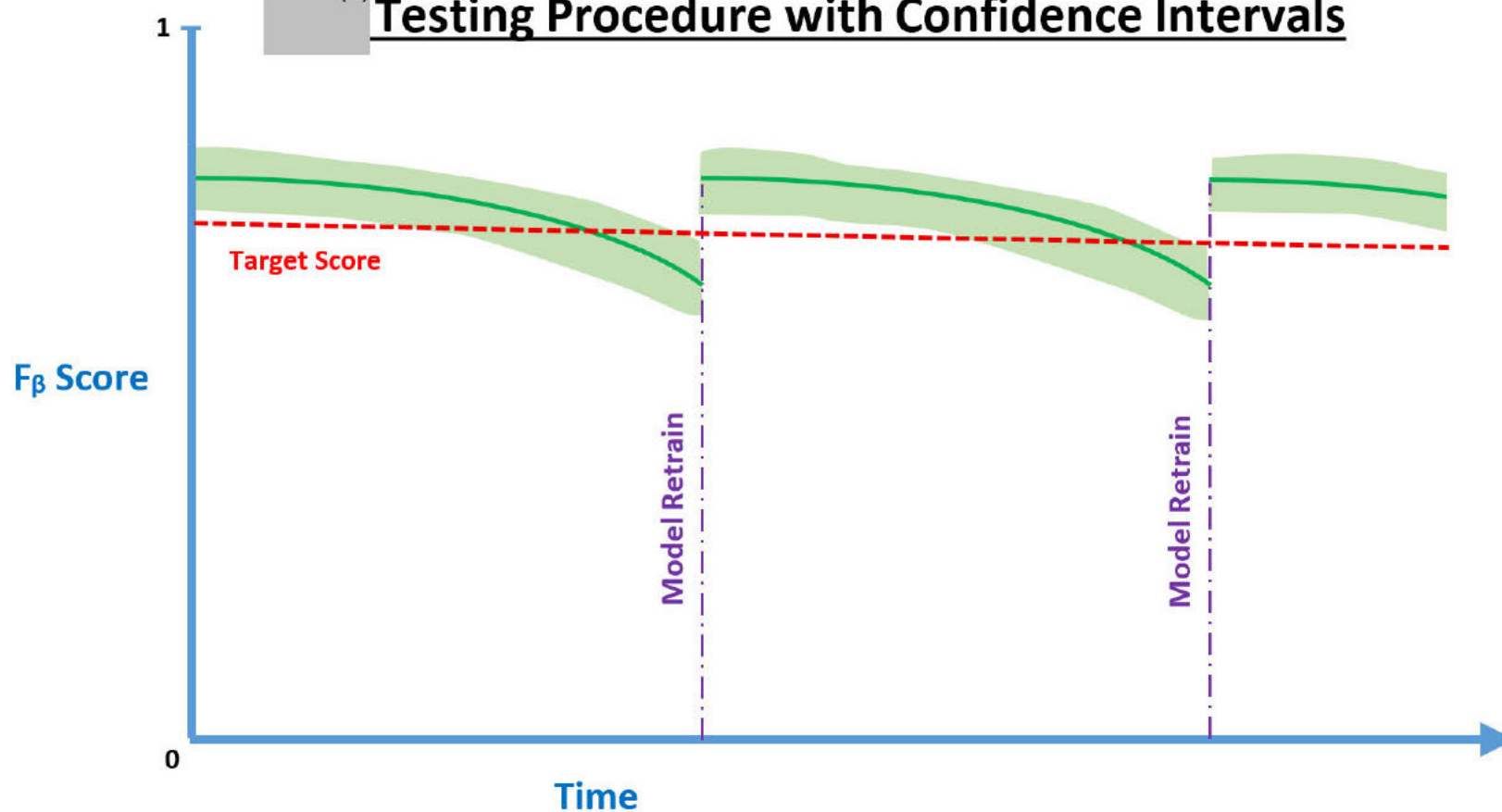
- Prediction results can demonstrate variability, where data diverges from a central average with resampling
- This variability is higher over all data when we have a smaller sample size
- This can be especially problematic when we have class imbalance
- We consider a 95% confidence interval, a range of scores within which we can have 95% confidence that a new sample's score will be between the bounds
- This ensures that we can ascertain when our model is not performing to standard with more than 95% confidence, and confirm that results are *statistically significant*



β Score: Confidence

s. 47E(d)

Testing Procedure with Confidence Intervals

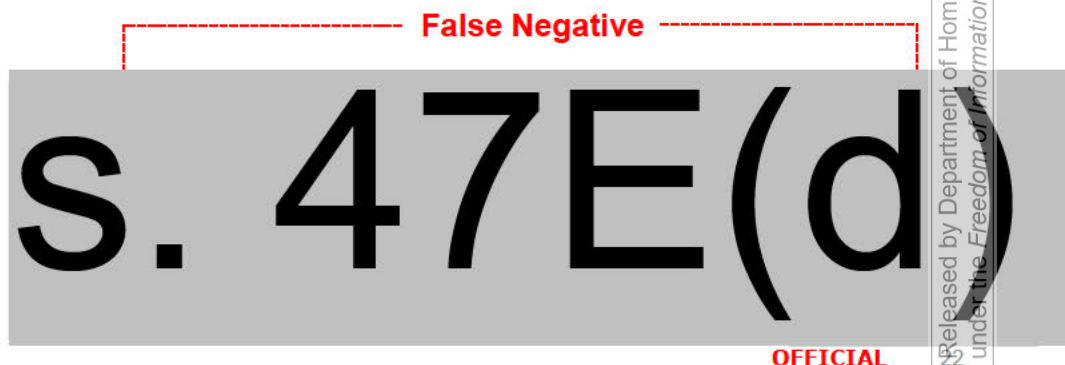


β ?

- β is a tuneable parameter that determines how many times more important recall is than precision
- Analogy: **Think of β as a dial centred at 1. We can turn the dial up to increase focus on recall over precision, or turn it down to prioritise precision over recall.**
- We choose β as a ratio of the weighted average cost of false negatives (*non-compliant applicants that are falsely predicted as streamline*) to false positives (*compliant applicants that are falsely flagged as a positive risk*)
- $$\beta = \frac{\text{weighted average cost of false negatives}}{\text{weighted average cost of false positives}}$$

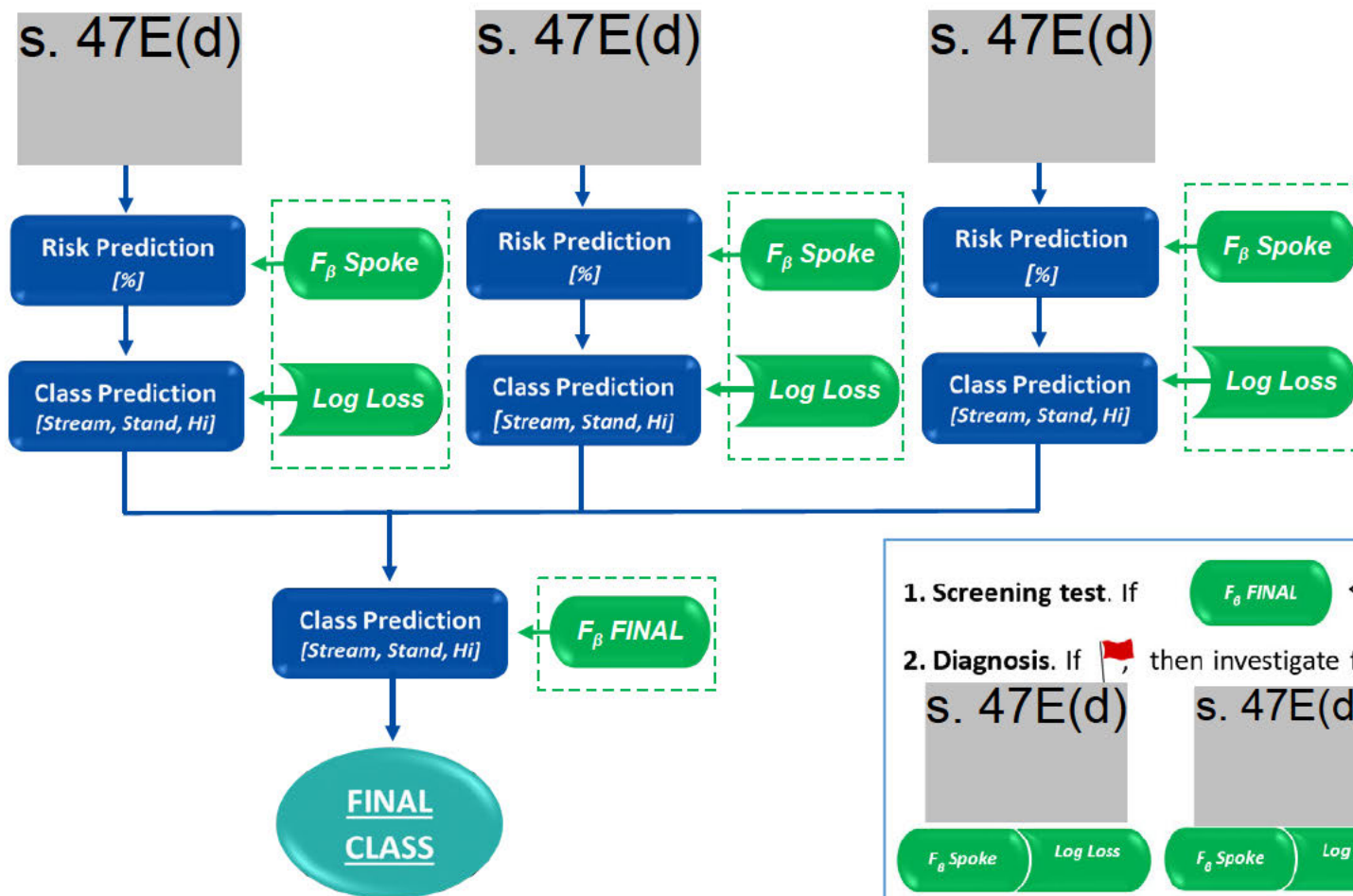


- Constructing the β parameter requires knowing and quantifying the relative costs of each type of misclassification
- Potential methods to extract these include:
 - Directly Asking Scalar Risk Appetites** – *How many s. 47E(d) would you be as tolerant to accept as one s. 47E(d)? How many s. 47E(d) are as tolerable as a false prediction of an adverse outcome for a low-risk applicant?*
 - Bucket Approach** – *Provide business with 100 tokens and a bucket for each misclassification type. Each token represents an applicant that is to be predicted incorrectly. Ask business to distribute the tokens among buckets until they feel that tokens are distributed in a way that balances their risk appetite.*



- F_β only assesses performance after class thresholds are integrated and hard decisions are made (spoke classification as high or low risk)
- We should supplement this with a metric that assesses the original probabilistic predictions of our model
- **Log Loss** is a preferred and established metric for this
- Separates performance of the prediction from the subsequent decision, *ensuring correctness of base predictions and improving diagnosis of decline in performance*
- Ensures stronger calibration of predictions by evaluating confidence levels in hard decisions
- Supplements F_β Score to provide a more informative assessment of model drift and helps to locate sources of decay

Log Loss calculation and explanation: S. 47E(d)



1. Screening test. If

$$F_{\theta} \text{ FINAL} < F_{\theta} \text{ Target} \Rightarrow$$

2. Diagnosis. If , then investigate for each spoke over time:

s. 47E(d)

s. 47E(d)

s. 47E(d)

F_{θ} Spoke

Log Loss

F_{θ} Spoke

Log Loss

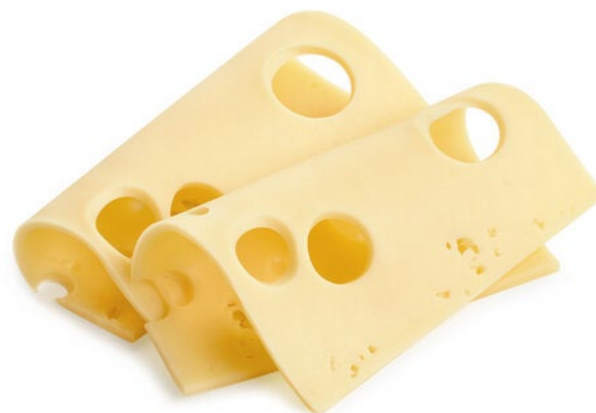
F_{θ} Spoke

Log Loss

1. Get the scalar costs of each false positive and false negative misclassification through collaboration with business
2. Determine the F_β Score target by calculating on business-determined ideal frequencies for each confusion matrix quadrant for all spokes
3. Calculate F_β Score and Log Loss over time for each spoke using s. 47E(d)
4. Calculate the final F_β Score and broad confidence interval as part of the same scripts
5. Confirm that the performance target is less than the upper bound of the 95% confidence interval of our final F_β Score
6. If not, then flag and undertake time-based, spoke-by-spoke analysis of F_β Score and Log Loss to diagnose sources of model degradation at the prediction and decision levels of each spoke
7. Consider model retraining to improve predictive performance to an acceptable level

Within this presentation, I have proposed a comprehensive methodology for testing the effectiveness. I believe that this methodology captures:

- Sensitivity to cohort-specific misclassification costs and business risk appetite
- Robustness at handling class imbalance
- Need to define exact parameters for acceptable model performance
- Framework that supports systematic analysis of all spokes and decisions
- Use of a clear methodology and presentation media to support interpretability by relevant stakeholders and businesses without a machine learning background
- Potential for industrialisation and expansion to other predictive modelling systems in immigration systems



F-Score	Log Loss
✓ More focussed on correctly identifying positives ✓ More intuitive to stakeholders ✓ Balances Precision and Recall according to misclassification cost	✓ Stronger measure of fit ✓ Punishes more confident misclassifications more harshly by considering numerical predictions
✗ Performs poorly when positive is the majority class ✗ Sensitivity to class distributions limits comparability across models ✗ Not as informative as to overall predictive performance (No TN)	✗ Harder to explain to stakeholders ✗ Does not consider decisions at class thresholds

Why consider both F-score AND log loss:

- Scoring rules are considered a stronger measure of fit than binary classification
- More complete and descriptive for diagnosis of final performance issues
- Separation of prediction and classification phases may help indicate when classification threshold selection may be the cause of poor performance (Good log loss, poor F-Beta)
- Can negate the limitations of the other metric (e.g. log loss not considering misclassification cost, F-score bias w/ positive-class majority imbalance)

By considering both:

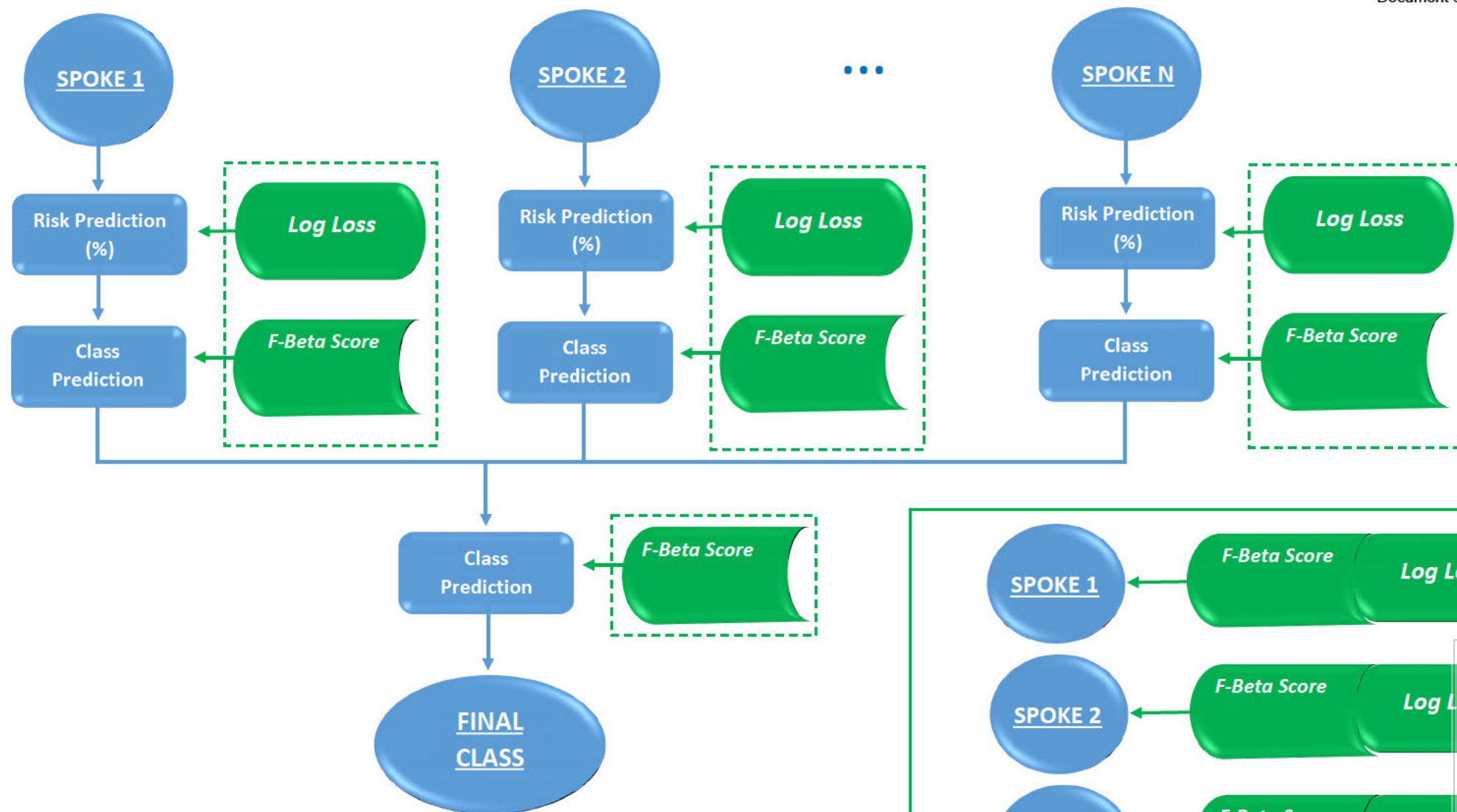
- F-score tells us *how well-calibrated and accurate the pure predictive outputs are of the spoke's random forest model*
- F-score tells us *how well hard decisions are being made with the consideration of classification thresholds*

Proposal:

- 1) *Determine target F-Beta score from business misclassification costs*
- 2) *Calculate final class F-score and confidence interval to determine overall risk decision produced by the entire model and compare over time*
- 3) *Determine if significant decay in performance or model score confidence interval does not intersect target F-Beta score. If so;*
 - 4) *Perform spoke-by-spoke log loss to determine predictive performance at each spoke*
 - 5) *Perform spoke-by-spoke f-Beta score calculation to determine hard decision performance accounting classification thresholds at each spoke*
 - 6) *Consider metrics from 4) and 5) to diagnose poor performance – which spokes, and pertinence to predictive outputs or translation decisions*

Potential next steps:

- 1) Hone implementation of full methodology to dummy input data
- 2) Determine methodology for extracting cost coefficients
- 3) Create accessible medium to explain confusion matrices, F-Beta Scoring and model drift to stakeholders
- 4) Industrialise and apply across different cohorts



$$\text{LogLoss} = -\frac{1}{n} \sum_{i=0}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

$$F_{\beta} = \frac{(1 + \beta^2) \cdot \text{true positive}}{(1 + \beta^2) \cdot \text{true positive} + \beta^2 \cdot \text{false negative} + \text{false positive}}$$

