Digital Policy Office

Ethical Artificial Intelligence Framework

Quick Reference Guide

(Customised version for general reference by public)

Version: 1.1

July 2024

Amendment History				
Change Number	Revision Description	Pages Affected	Revision Number	Date
1	Change “Office of the Government Chief Information Officer” to “Digital Policy Office”		1.1	25 July 2024

TABLE OF CONTENTS

1. OVERVIEW

1.1 INTRODUCTION

This document is intended to provide readers with quick reference and initial understanding of the Ethical Artificial Intelligence ("**AI**") Framework and procedures that should be carried out to embed ethical elements in planning, design and implementation of AI and big data applications in IT projects or services (hereafter known as "ethical AI").

In this document, the term "AI" is used to refer to analytic operations involving big data analytics, advanced analytics and machine learning that use massive data sets and processing capabilities to find correlations and make predictions. **The term "AI applications" has been used to refer to a collective set of applications whose actions, decisions or predictions are empowered by AI models. Examples of AI applications are IT projects which have prediction functionality and/or model development involving training data.** For IT projects that have AI applications, organisations can make reference to the requirements of the Ethical AI Framework.

1.2 WHAT IS THE ETHICAL AI FRAMEWORK?

The Ethical AI Framework is developed to assist B/Ds in planning, designing and implementing AI and big data analytics in their IT projects and services. It consists of guiding principles, leading practices and AI assessment that should be adopted for B/Ds' AI-powered IT projects. Nonetheless this framework, including guiding principles, practices and assessment template, is also applicable to other organisations in general and this customised version of framework is suitably revised (e.g. removal or adjustment of government specific terms) for general reference by organsiations when adopting AI and big data analytics in their IT projects.

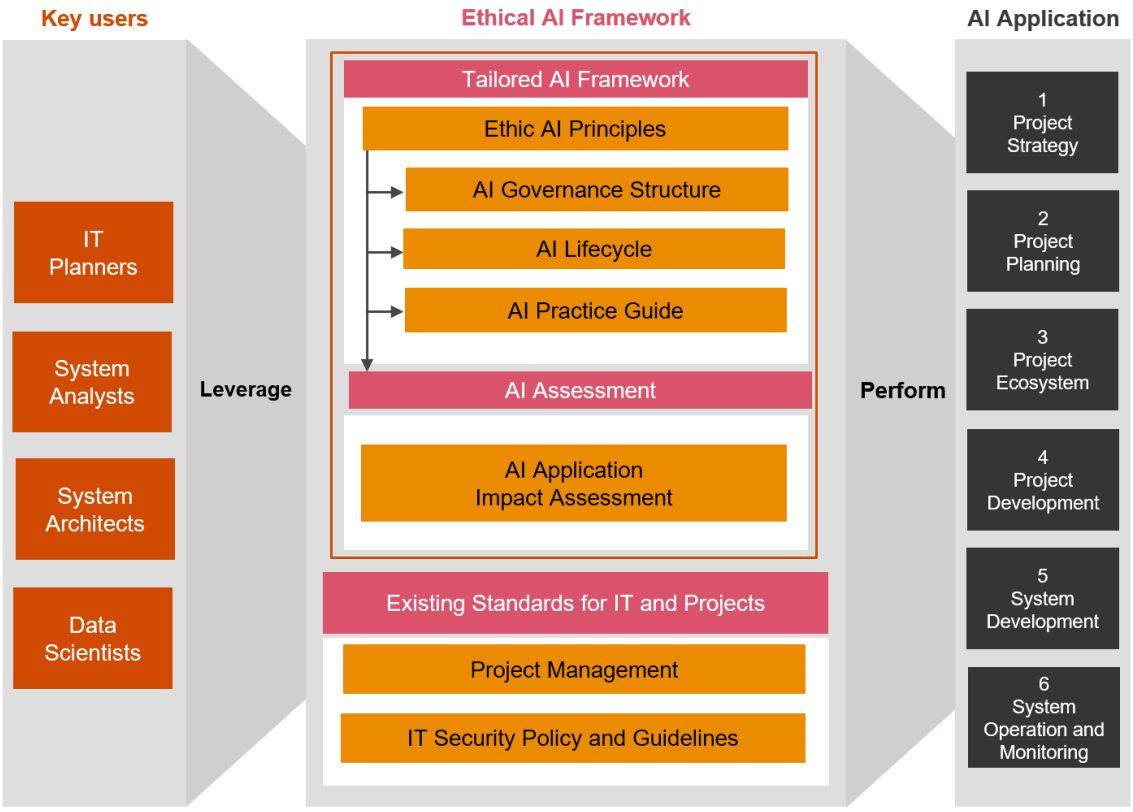

Figure 1: Overview of the Ethical AI Framework

The Ethical AI Framework consists of the following components:

- The **Tailored AI Framework**, which consists of the following sub-components:
  - The **Ethical AI Principles** define principles to be followed when designing and developing AI applications. The Ethical AI Principles are applicable to all roles as defined in the AI Governance Structure;
  - The **AI Governance Structure** defines standard structures and roles and responsibilities over the adoption process of AI against practices set out in the Ethical AI Framework;
  - The **AI Lifecycle** defines an AI lifecycle model that is used to structure the layout of practices in the AI Practice Guide and the questions in the AI Assessment; and
  - The **AI Practice Guide** defines a set of practices for all stages in the AI Lifecycle. The guiding practices are derived from the Ethical AI Principles.
- The **AI Assessment** consists of the **AI Application Impact Assessment** template which defines questions to be answered by target readers across the stages as part of the AI Lifecycle to assess the impact of AI applications and to ensure that Ethical AI Principles have been considered.

Please refer to Section 3 “Overview of the Ethical AI Framework” in the Ethical AI Framework document for further details.

1.3 WHAT ARE THE KEY ACTIONS TO TAKE TO ADOPT THE ETHICAL AI FRAMEWORK?

The following key actions should be performed for organisations to adopt the Ethical AI Framework.

1. Consider all Ethical AI Principles throughout the project lifecycle;
2. Review the existing project management governance structure to ensure it aligns with AI Governance Structure and set up an optional Ethical AI Committee if necessary; and
3. Follow the AI Practice Guide (and complete AI Application Impact Assessments for each AI application).

2. ETHICAL AI PRINCIPLES

Twelve Ethical AI Principles should be observed for all AI projects. Two out of the twelve principles (1) **Transparency and Interpretability** and (2) **Reliability, Robustness and Security** are "Performance Principles". These fundamental principles must be achieved to create a foundation for the execution of other principles. For example, without achieving the Reliability, Robustness and Security principle, it would be impossible to accurately verify that other Ethical AI Principles have always been followed.

The other principles are categorised as "General Principles", including (1) **Fairness,** (2) **Diversity and Inclusion**, (3) **Human Oversight**, (4) **Lawfulness and Compliance**, (5) **Data Privacy**, (6) **Safety**, (7) **Accountability**, (8) **Beneficial AI**, (9) **Cooperation and Openness** and (10) **Sustainability and Just Transition**. They are derived from the United Nations' Universal Declaration of Human Rights and the Hong Kong Ordinances.

Definitions for the principles are listed in Table 1.

Principle	Definition
Transparency and Interpretability	Organisations should be able to explain decision-making processes of the AI applications to humans in a clear and comprehensible manner.
Reliability, Robustness and Security	Like other IT applications, AI applications should be developed such that they will operate reliably over long periods of time using the right models and datasets while ensuring they are both robust (i.e. providing consistent results and capable to handle errors) and remain secure against cyber-attacks as required by the relevant legal and industry frameworks.
Fairness	The recommendation/result from the AI applications should treat individuals within similar groups in a fair manner, without favouritism or discrimination and without causing or resulting in harm. This entails maintaining respect for the individuals behind the data and refraining from using datasets that contain discriminatory biases.
Diversity and Inclusion	Inclusion and diverse usership through the AI application should be promoted by understanding and respect the interests of all stakeholders impacted.
Human Oversight	The degree of human intervention required as part of AI application's decision-making or operations should be dictated by the level of the perceived severity of ethical issues.
Lawfulness and Compliance	Organisations responsible for an AI application should always act in accordance with the law and regulations and relevant regulatory regimes.

Principle	Definition
Data Privacy	Individuals should have the right to: (a) be informed of the purpose of collection and potential transferees of their personal data and that personal data shall only be collected for a lawful purpose, by using lawful and fair means, and that the amount of personal data collected should not be excessive in relation to the purpose. Please refer to the Data Protection Principles ("**DPP**")**1** "Purpose and Manner of Collection" of the Personal Data (Privacy) Ordinance (the "**PD(P)O**")[1]. (b) be assured that data users take all practicable steps to ensure that personal data is accurate and is not kept longer than is necessary. Please refer to the **DPP2** "Accuracy and Duration of Retention" of the PD(P)O. (c) require that personal data shall only be used for the original purpose of collection and any directly related purposes. Otherwise, express and voluntary consent of the individuals is required. Please refer to the **DPP3** "Use of Personal Data" of the PD(P)O. (d) be assured that data users take all practicable steps to protect the personal data they hold against unauthorised or accidental access, processing, erasure, loss or use. Please refer to the **DPP4** "Security of Personal Data" of the PD(P)O. (a) be provided with information on (i) its policies and practices in relation to personal data, (ii) the kinds of personal data held, and (iii) the main purposes for which the personal data is to be used. Please refer to the **DPP5** "Information to Be Generally Available" of the PD(P)O.
Safety	Throughout their operational lifetime, AI applications should not compromise the physical safety or mental integrity of mankind.
Accountability	Organisations are responsible for the moral implications of their use and misuse of AI applications. There should also be a clearly identifiable accountable party, be it an individual or an organisational entity (e.g. the AI solution provider).
Beneficial AI	The development of AI should promote the common good.
Cooperation and Openness	A culture of multi-stakeholder open cooperation in the AI ecosystem should be fostered.
Sustainability and Just Transition	The AI development should ensure that mitigation strategies are in place to manage any potential societal and environmental system impacts.

Table 1: Ethical AI Principles and Definition

Please refer to Section 3.5.1 "Ethical AI Principles" in the Ethical AI Framework document for further details.

[1] https://www.pcpd.org.hk/english/data_privacy_law/ordinance_at_a_Glance/ordinance.html

3. AI GOVERNANCE

AI governance refers to the practices and direction by which AI projects and applications are managed and controlled. The three lines of defence is a well-established governance concept in many organisations. Figure 2 shows the different defence lines and their roles.

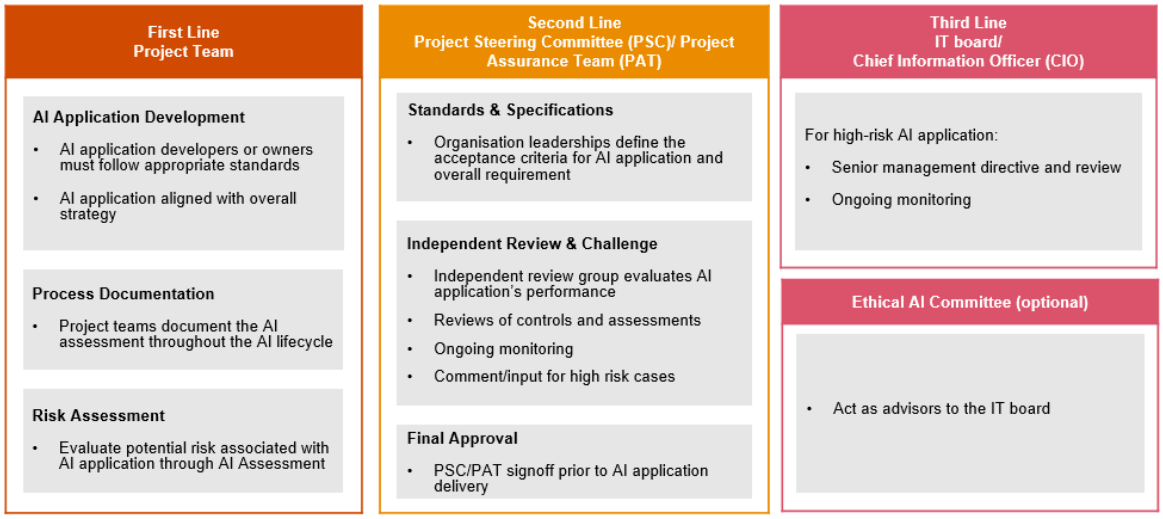

Figure 2: Lines of Defence Model

The governance structure is a board structure with the following setup.

- The **first line of defence** is the Project Team who is responsible for AI application development, risk evaluation, execution of actions to mitigate identified risks and documentation of AI Assessment.
- The **second line of defence** comprises of the Project Steering Committee ("**PSC**") and Project Assurance Team ("**PAT**") who are responsible for ensuring project quality, defining acceptance criteria for AI applications, providing independent review and approving AI applications. The Ethical AI Principles should be addressed through the use of AI Assessment before approval of the AI application.
- The **third line of defence** involves the **IT Board**, or the Chief Information Officer ("**CIO**") if the IT Board is not in place, and is optionally supported by an Ethical AI Committee, which consists of external advisors. The purpose of the Ethical AI Committee is to provide advice on ethical AI and strengthen organisations' existing competency on AI adoption. The third line of defence is responsible for reviewing, advising and monitoring of high-risk AI applications.

4. AI LIFECYCLE

In order to structure the practices for organisations to follow when executing AI projects/creating AI applications, practices in different stages of the AI Lifecycle have been detailed in the AI Practice Guide (Please refer to Section 4 "AI Practice Guide" in the Ethical AI Framework for further details). A way to conceptualise the AI Lifecycle appears in the following 6-step schematic.

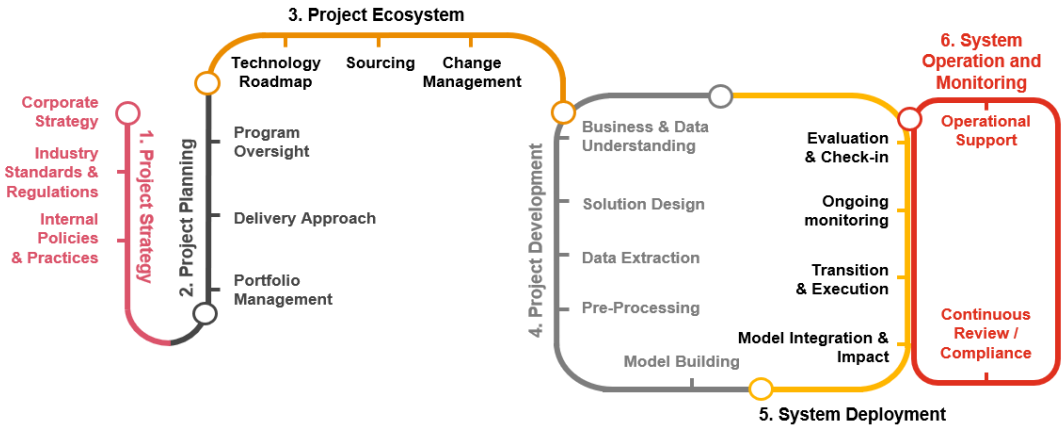

Figure 3: Overview of the AI Lifecycle

The AI Lifecycle shows the different steps involved in AI projects that can

- Guide organisations to understand the different stages and requirements involved; and
- Serve as a reference for the development of AI practices to align with actual stages of how AI is typically developed.

The AI Lifecycle aligns with a traditional software lifecycle model as depicted in Figure 4.

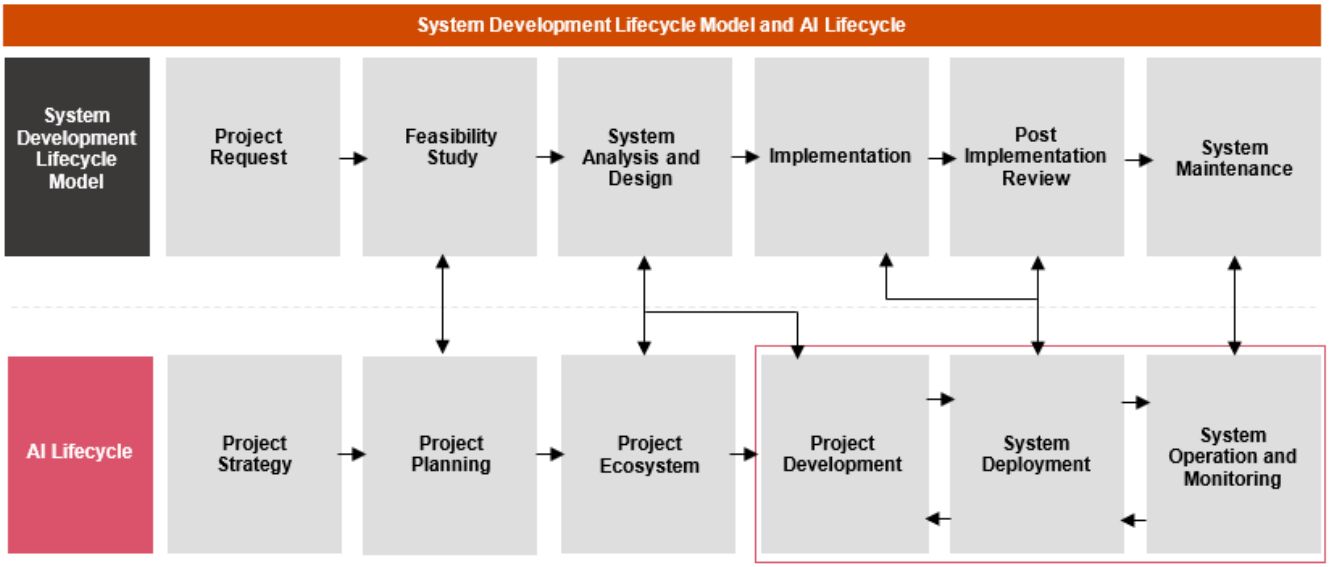

Figure 4: AI Lifecycle aligned with a System Development Lifecycle Model

In a typical development process of an AI application, great emphasis is placed on data because good data are often required for creating a good AI model. Data sourcing and preparation which is part of the project development can be a continuous exercise. This is because an AI model can often benefit from better or more data for iterative model training throughout the development process. The approach of conventional software lifecycle is to program the IT application with a set of instructions for a pre-defined set of events. Thereafter, the IT application will exploit its computing capabilities and other resources to process the data fed into the system. This is different from an AI application where a huge amount of data are fed into the application, which in turn processes all the data resulting in a trained model or AI solution. This trained model is then used to solve new problems.

There is often a continual feedback loop between the development and deployment stages, as well as system operation and monitoring of the AI Lifecycle for iterative improvements making this distinct from a traditional software development lifecycle.

5. AI PRACTICE GUIDE AREAS

Section 4 "AI Practice Guide' of the Ethical AI Framework contains detailed practices to be followed for a number of practice areas. Such practice areas are assessed as part of the AI Application Impact Assessment. A summary of the practice areas in the AI Practice Guide is listed below.

AI Lifecycle	Practice Area	Definition
Project Strategy	**Organisations Strategy, Internal Policies and Practices**	Organisations should be able to explain decision-making processes of the AI applications to humans in a clear and comprehensible manner.
	Industry Standards and Regulations	Relevant regulations and standards require an assessment to ensure that AI and related processes adhere to any relevant laws or standards.
Project Planning	**Portfolio Management**	Portfolio management is performed to ensure that the IT investments embedded in the organisation's processes, people and technology are on course. Assessment of AI projects to ensure that they gainfully address business requirements and objectives.
	Project Oversight and Delivery Approach	Quality control not only monitors the quality of deliverables; it involves monitoring various aspects of the project as defined in the Project Management Plan ("**PMP**").
Project Ecosystem	**Technology Roadmap for AI and Data Usage**	A technology roadmap should enable the organisations to plan and strategise which, when and what technologies will be procured for AI and big data analytics.
	Procuring AI Services	Off-the-shelf products and data can be procured for AI projects. In conducting such procurement exercises, organisations should duly consider the related ethical considerations.
Project Development	**Business & Data Understanding**	Organisations should determine the objectives of using AI and weigh and balance the benefits and risks of using AI in the decision-making process.
	Solution Design	Organisations should assess the AI model within an AI application for suitability compared to the organisation's objectives as well as the appropriate level of human intervention required.
	Data Extraction	Organisations should ensure the data quality, validity, reliability and consistency of information from various sources, within or outside of the organisations.

	Pre-processing	Sensitive data containing an individual's information require extra care during solution development to prevent data leakage as well as breaches of privacy and security.
	Model Building	Model building should aim at mitigating common AI application errors such as inaccurate model assumptions, input variable selection, model overfitting and adversarial attacks.
System Deployment	**Model Integration & Impact**	Verification, validation and testing is the process of ensuring the AI applications perform as intended based on the requirements outlined at the beginning of the project. This would ensure proper integration of the AI application.
	Transition & Execution	Assuming the AI application would fail, mitigation steps should be incorporated to minimise damages in the case of failure prior to deployment of the AI application.
	Ongoing Monitoring	Provide feedback to monitor/maintain model performance and robustness of the AI application.
	Evaluation & Check-in	Traceability, Repeatability and Reproducibility of the AI application are required to ensure the AI application is operating correctly and to help build trust from the public and key stakeholders.
System Operation and Monitoring	**Data and Model Performance Monitoring**	AI models (as part of AI applications) should be continuously monitored and reviewed due to the likelihood of the AI models becoming less accurate and less relevant.
	Operational Support	Upon AI applications deployment, ongoing operational support should be established to ensure that the AI applications performance remains consistent, reliable and robust.
	Continuous Review/Compliance	The Project Manager (or Maintenance Team), PSC/PAT (or Maintenance Board) and IT Board/CIO (or its delegates) are responsible to monitor risks of AI such as non-compliance with applicable new/revised laws and regulations.

Table 2: Practice areas covered in the AI Practice Guide within the Ethical AI Framework

Please refer to Section 4 "AI Practice Guide" in the Ethical AI Framework document for further details.

6. AI APPLICATION IMPACT ASSESSMENT

The AI Application Impact Assessment should be conducted on an AI application at different stages of the AI Lifecycle. The AI Application Impact Assessment introduces a systematic thinking process for organisations to go through different aspects of considerations of individual applications for their associated benefits and risks whilst highlighting the need for additional governance activities and identifying follow-up actions to ensure necessary measures and controls required for implementing ethical AI.

The AI Application Impact Assessment template used for this assessment is in Microsoft Word format with sections for providing qualitative answers. Please refer to Appendix C "AI Application Impact Assessment Template" in the Ethical AI Framework document for details.

The AI Application Impact Assessment has the following components:

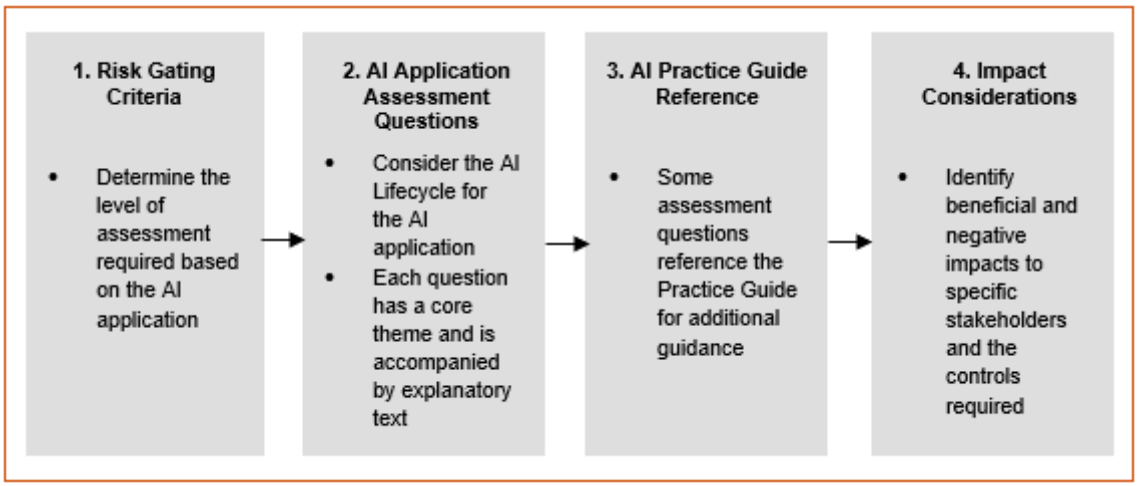

Figure 5: AI Application Impact Assessment Components

An AI Application Impact Assessment should be conducted regularly (e.g. annually or when major changes take place) as AI projects progress and when the AI application is being operated.

The stages of the AI Lifecycle where AI Application Impact Assessment should be reviewed are shown in Figure 6.

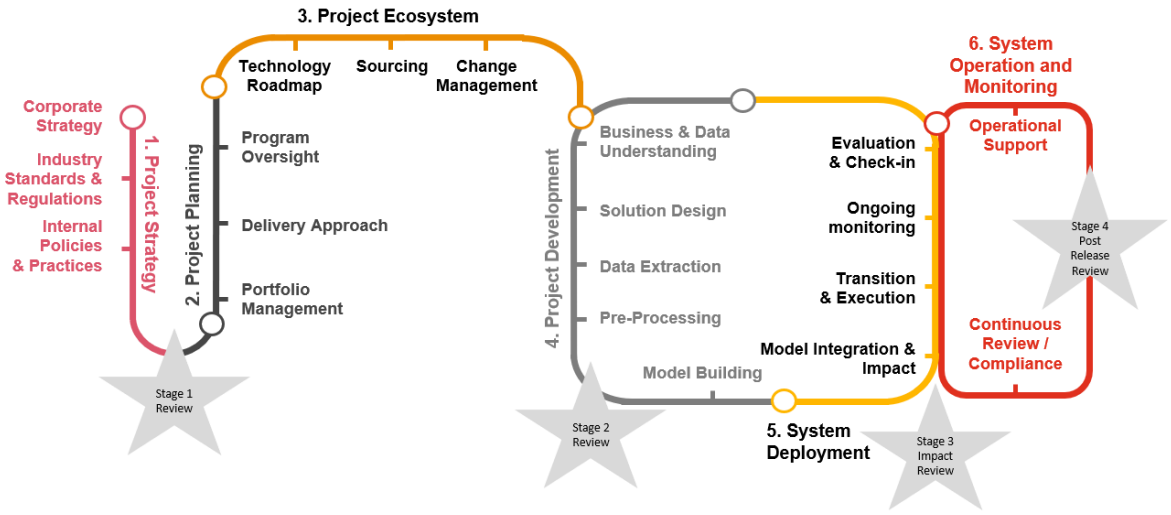

Figure 6: Stages for AI Application Impact Assessment

The AI Application Impact Assessment can be used as a 'live' document throughout the AI Lifecycle, but the associated AI Application Impact Assessment should be reviewed at 4 key stages of the AI Lifecycle with a copy of the AI Application Impact Assessment being retained for historical records. The mapping of these reviews to the system development lifecycle with responsible parties and actions to be performed is summarised below.

Lifecycle stage	Responsible party	Actions to be performed
SDLC: Project Request **AI Lifecycle:** Project Strategy	Project Team	Categorise the project as "high-risk" or "non high-risk" based on answers to the "Risk Gating questions" Conduct AI Application Impact Assessment • Answer questions 1-8, 9(i), 41-49, 57-58
	PSC/PAT	Review and endorse the assessment for non-high-risk project
	IT Board/CIO	Review and endorse the assessment for high-risk project
SDLC: System Analysis and Design **AI Lifecycle: Project Ecosystem** Project Development	Project Team	Conduct AI Application Impact Assessment • Answer questions 9(ii), 10-14, 18-19, 22-29, 30(i)(ii), 31-34, 37(i), 52-56 • Review and update the answers of questions 1-8, 9(i), 41-49, 57-58 according to the latest position • If third-party technology or data is used, complete and review questions 15-17, 48 <u>before procurement</u>
	PSC/PAT	• Review and endorse the assessment

Lifecycle stage	Responsible party	Actions to be performed
SDLC: Implementation (before rollout) **AI Lifecycle:** System Deployment	Project Team	Conduct AI Application Impact Assessment • Answer questions 20, 21, 30(iii), 35, 36, 37(ii)(iii), 38-40, 50, 51 • Review and update the AI Application Impact Assessment according to the latest position (questions 1-14, 18-19, 22-29, 30(i)(ii), 31-34, 37(i), 41-49, 52-58) • If third-party technology or data is used, review questions 15-17, 48
	PSC/PAT	• Review and endorse the assessment
SDLC: System Maintenance (annually or when major changes take place) **AI Lifecycle:** System Operation and Monitoring	Maintenance Team	• Review and update the "AI Application Impact Assessment" according to the latest position • Escalate any significant issues
	Maintenance Board IT Board/CIO (or its delegates)	• Handle escalation • Monitor high-risk projects

Table 3: Actions to be performed for completing and reviewing the AI Application Impact Assessment