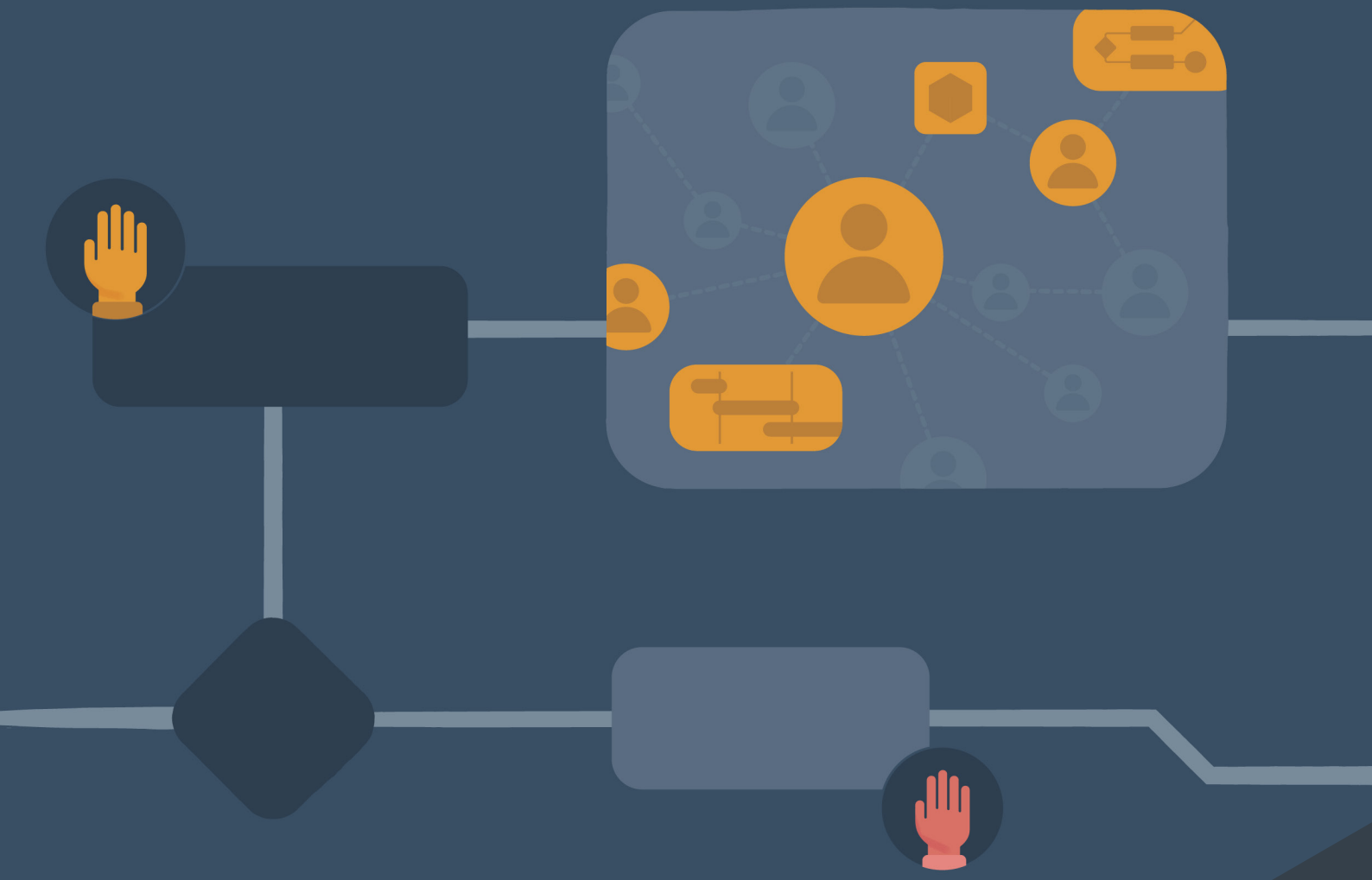
AI Ethics and Governance in Practice Programme

AI Accountability in Practice



Facilitator Workbook

Annotated to support facilitators in delivering the accompanying activities.

The Alan Turing Institute

Acknowledgements

This workbook was written by David Leslie, Cami Rincón, Morgan Briggs, Antonella Maia Perini, Smera Jayadeva, Ann Borda, SJ Bennett, Christopher Burr, Semeli Hadjiloizou, and Claudia Fischer.

The creation of this workbook would not have been possible without the support and efforts of various partners and collaborators. As ever, all members of our brilliant team of researchers in the Ethics Theme of the Public Policy Programme at The Alan Turing Institute have been crucial and inimitable supports of this project from its inception several years ago, as have our Public Policy Programme Co-Directors, Helen Margetts and Cosmina Dorobantu. We are deeply thankful to Conor Rigby, who led the design of this workbook and provided extraordinary feedback across its iterations. We also want to acknowledge Johnny Lighthands, who created various illustrations for this document, and Alex Krook and John Gilbert, whose input and insights helped get the workbook over the finish line. Special thanks must be given to the Digital Office for Scottish Local Government and Open Data Manchester for helping us test the activities and review the content included in this workbook. Lastly, we want to thank Karen Cordier (The Alan Turing Institute) for her meticulous peer review and timely feedback, which greatly enriched this document.

This work was supported by Wave 1 of The UKRI Strategic Priorities Fund under the EPSRC Grant EP/W006022/1, particularly the Public Policy Programme theme within that grant & The Alan Turing Institute; Towards Turing 2.0 under the EPSRC Grant EP/W037211/1 & The Alan Turing Institute; and the Ecosystem Leadership Award under the EPSRC Grant EP/X03870X/1 & The Alan Turing Institute.

Cite this work as: Leslie, D., Rincón, C., Briggs, M., Perini, A., Jayadeva, S., Borda, A., Bennett, SJ., Burr, C., Hadjiloizou, S., and Fischer, C. (2024). *AI Accountability in Practice.* The Alan Turing Institute.

Contents

About the Workbook Series

Key Concepts

Activities

Further Readings

About the AI Ethics and Governance in Practice Workbook Series

Who We Are

The Public Policy Programme at The Alan Turing Institute was set up in May 2018 with the aim of developing research, tools, and techniques that help governments innovate with data-intensive technologies and improve the quality of people's lives. We work alongside policymakers to explore how data science and artificial intelligence can inform public policy and improve the provision of public services. We believe that governments can reap the benefits of these technologies only if they make considerations of ethics and safety a first priority.

Origins of the Workbook Series

In 2019, The Alan Turing Institute's Public Policy Programme, in collaboration with the UK's Office for Artificial Intelligence and the Government Digital Service, published the UK Government's official Public Sector Guidance on AI Ethics and Safety. This document provides end-to-end guidance on how to apply principles of AI ethics and safety to the design, development, and implementation of algorithmic systems in the public sector. It provides a governance framework designed to assist AI project teams in ensuring that the AI technologies they build, procure, or use are ethical, safe, and responsible.

In 2021, the UK's National AI Strategy recommended as a 'key action' the update and expansion of this original guidance. From 2021 to 2023, with the support of funding from the Office for AI and the Engineering and Physical Sciences Research Council as well as with the assistance of several public sector bodies, we undertook this updating and expansion. The result is the AI Ethics and Governance in Practice Programme, a bespoke series of eight workbooks and a digital platform designed to equip the public sector with tools, training, and support for adopting what we call a Process-Based Governance (PBG) Framework to carry out projects in line with state-of-the-art practices in responsible and trustworthy AI innovation.

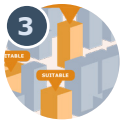
About the Workbooks

The AI Ethics and Governance in Practice Programme curriculum is composed of a series of eight workbooks. Each of the workbooks in the series covers how to implement a key component of the PBG Framework. These include Sustainability, Technical Safety, Accountability, Fairness, Explainability, and Data Stewardship. Each of the workbooks also focuses on a specific domain, so that case studies can be used to promote ethical reflection and animate the Key Concepts.

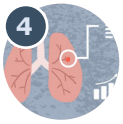
Programme Curriculum: AI Ethics and Governance in Practice Workbook Series

AI Ethics and Governance in Practice: An Introduction
Multiple Domains

AI Sustainability in Practice Part One
AI in Urban Planning



AI Sustainability in Practice Part Two
AI in Urban Planning

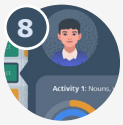


AI Fairness in Practice
AI in Healthcare

Responsible Data Stewardship in Practice
AI in Policing and Criminal Justice

AI Safety in Practice
AI in Transport

AI Explainability in Practice
AI in Social Care



AI Accountability in Practice
AI in Education

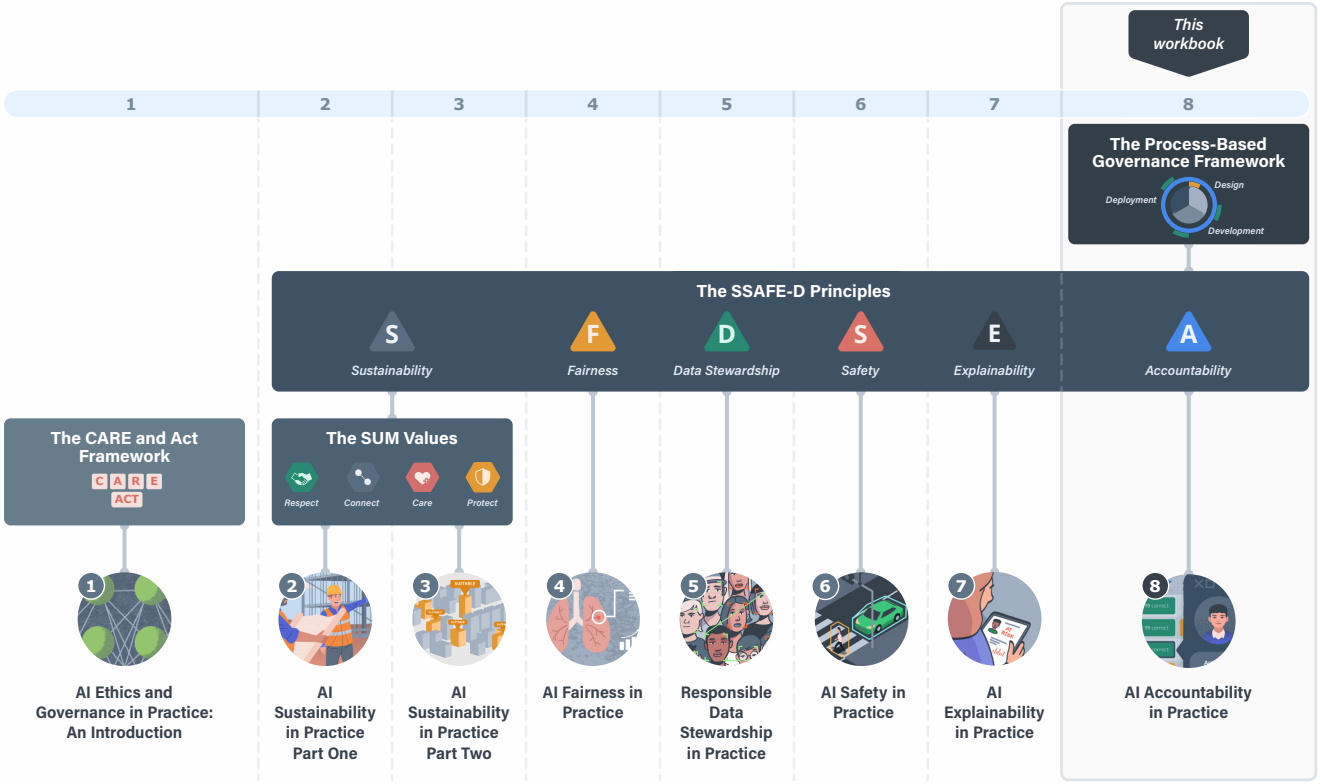
Explore the full curriculum and additional resources on the AI Ethics and Governance in Practice Platform at aiethics.turing.ac.uk.

Taken together, the workbooks are intended to provide public sector bodies with the skills required for putting AI ethics and governance principles into practice through the full implementation of the guidance. To this end, they contain activities with instructions for either facilitating or participating in capacity-building workshops.

Please note, these workbooks are living documents that will evolve and improve with input from users, affected stakeholders, and interested parties. We value your participation. Please share feedback with us at aiethics@turing.ac.uk.

Programme Roadmap

The graphic below visualises this workbook in context alongside key frameworks, values and principles discussed within this programme. For more information on how these elements build upon one another, refer to AI Ethics and Governance in Practice: An Introduction.



Intended Audience

The workbooks are primarily aimed at civil servants engaging in the AI Ethics and Governance in Practice Programme — whether as AI Ethics Champions delivering the curriculum within their organisations by facilitating peer-learning workshops, or as participants completing the programmes by attending these workshops. Anyone interested in learning about AI ethics, however, can make use of the programme curriculum, the workbooks, and resources provided. These have been designed to serve as stand-alone, open access resources. Find out more at aiethics.turing.ac.uk.

There are two versions of each workbook:

- **Facilitator Workbooks** (such as this document) are annotated with additional guidance and resources for preparing and facilitating training workshops.
- **Participant Workbooks** are intended for workshop participants to engage with in preparation for, and during, workshops.

Introduction to This Workbook

This workbook aims to provide resources and training materials to help you and your team establish an end-to-end accountability framework. This will enable you to integrate the ethical values and practical principles, which motivate and steer responsible innovation, into the actual processes that characterise your AI project lifecycle. The workbook is divided into two sections, Key Concepts and Activities:

Key Concepts Section

This section provides content for workshop participants and facilitators to engage with prior to attending each workshop. It provides definitions of key terms, an overview of the components and types of accountability, and introduces the three components to ensure end-to-end operability of good governance practices across the workflow of your AI project. Topics discussed include:

Part One: Introduction to Accountability

Introduction to Accountability

Components of Accountability

Types of Accountability

Part Two: Putting Accountability Into Practice

The Process-Based Governance Log

Template Log

Activities Section

This section contains instructions for group-based activities (each corresponding to a section in the Key Concepts). These activities are intended to increase understanding of Key Concepts by using them.

Case studies within the AI Ethics and Governance in Practice workbook series are grounded in public sector use cases, but do not reference specific AI projects.

Identifying Challenges

Practise identifying challenges to accountability presented by AI projects.

Establishing Proportional Governance Actions

Become familiar with governance actions and understand how they are put in place to address the potential harmful impacts of AI systems.

Mapping Roles and Responsibilities

Practise mapping team roles and responsibilities to governance actions across the AI lifecycle.

Completing Maps

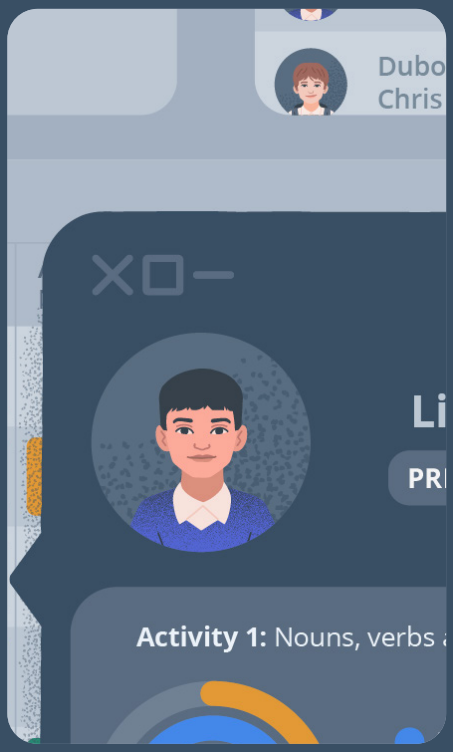
Practise assessing Workflow Governance Maps.

Note for Facilitators

Additionally, you will find facilitator instructions (and where appropriate, considerations) required for facilitating activities and delivering capacity-building workshops.

AI Accountability in Practice

Key Concepts



Part One: Introduction to Accountability

The principle of accountability is an end-to-end governing principle (q.v. definition p. 12). Accountability entails that humans are answerable for the parts they play across the entire AI design, development, and deployment workflow. It also demands that the results of this work are traceable from start to finish. Designers and implementers should be held accountable for:

Fairness - being equitable, reasonable, and just, and not harming anyone through bias or discrimination. More details can be found in AI Fairness in Practice.

Sustainability and Safety - producing AI innovation that is ethical, performant, reliable, robust, and secure in its outcomes and wider impacts. More details can be found in the following workbooks:

- AI Sustainability in Practice Part One
- AI Sustainability in Practice Part Two
- AI Safety in Practice

When considering the role of accountability in the AI project lifecycle, it is important first to make sure that you are taking a 'best practices' approach to data processing that is aligned with the Data Ethics Framework. You should also pay special attention to the new and unique challenges posed to public sector accountability by the design and implementation of AI/ML systems.

Responsible AI project delivery requires that two related challenges to public sector accountability be confronted directly:

KEY CONCEPT

Data Ethics Framework

The Data Ethics Framework is a set of guidelines designed to promote appropriate and responsible use of data within government and the broader public sector. The guidance is aimed at people involved with data in the public sector. This includes data practitioners such as statisticians, analysts, data scientists, policymakers, operational staff, and those contributing to data-informed insights.[3] More details can be found in the Responsible Data Stewardship workbook.

1. Accountability Gap

Human agents can be called to account for their judgements and decisions when they affect the interests of others. However, automated decisions are not self-justifiable. The statistical models and underlying hardware of AI systems are not responsible in the same morally relevant sense.[1] This creates an accountability gap. As such, clear and attributable sources of human answerability should be attached by project teams to decisions assisted or produced by an AI system.[2]

2. Complexity of AI Production Processes

It is not easy to establish human answerability when it comes to the design, development, and deployment of AI systems. This is because the production and use of these systems are complex and involve many agents, including:

department and delivery leads | technical experts | data procurement and preparation personnel | policy and domain experts | implementers | and others

Likewise, AI/ML supply chains can be quite complex and multi-dimensional, with parts or components of AI/ML systems coming from different sources or open source software. This creates a dynamic of distributed responsibility, which can make the attribution of

accountability challenging. Due to this production complexity, it may become difficult to answer the question of who among the parties involved in the production of AI systems should bear responsibility if these systems' uses have negative consequences and impacts.

Meeting the special requirements of accountability, which are born out of these two challenges, calls for a sufficiently fine-grained concept of what would make an AI project properly accountable. This concept can be broken down into two subcomponents of accountability: **answerability** and **auditability**.

Component 1

Answerability

The principle of accountability demands that the onus of justifying algorithmically-supported decisions be placed on the shoulders of the human creators and users of those AI systems.[4] [5] Answerability refers to **establishing a continuous chain of human responsibility across the whole AI project delivery workflow.** For effective end-to-end accountability, there should be no gaps in the answerability of responsible human authorities from the first steps of the design of an AI system to its algorithmically-steered outcomes.

Answerability also demands that explanations and justifications of both the rationale underlying the results of an AI system and the processes behind their production and use be offered by competent human authorities in plain, understandable, and coherent language.[6] These explanations and justifications should be based upon sincere, consistent, sound, and impartial reasons that are accessible to non-technical hearers.

WB 7

More details about explanations and justifications can be found in the AI Explainability in Practice workbook.

Component 2

Auditability

Whilst the component of answerability responds to the question of *who* is accountable for an automation supported outcome, the component of auditability answers the question of how the designers and implementers of AI systems will be held accountable. This aspect of accountability has to do with ***demonstrating*** both the **responsibility of design, development, and deployment practices** and the **justifiability of outcomes**.[7]

Auditability also has to do with traceability. That is, the process by which all stages of the AI innovation lifecycle from Data Collection and Model Selection to System Deployment, Updating, and Deprovisioning are documented in a way that is accessible and easily understood.[8] [9]

Your project team must ensure that every step of the process of designing and implementing your AI project is accessible for audit, oversight, and review by appropriate parties.[10] Successful auditability requires builders and implementers of algorithmic systems to:

- keep records and make available information that enables monitoring of the soundness and diligence of the innovation processes that produced these systems;
- keep track of the accountable parties within an organisation's project team and others involved in the supply chain (where system components are procured);
- keep track of the governance actions taken across the entire AI innovation workflow; and
- keep records and make accessible information that enables monitoring of data provenance and analysis from the stages of collection, pre-processing, and modelling to training, testing, and deploying. This is the purpose of the Dataset Factsheet.

More details about the Data Factsheet can be found in the Responsible Data Stewardship in Practice workbook.

In addition, auditability requires your team to enable peers and overseers to probe and to critically review the dynamic operation of the system in order to ensure that the procedures and operations which are producing the model's behaviour are safe, ethical, and fair.[11] Practically transparent algorithmic models must be built for auditability, reproducibility, and equipped for end-to-end recording and monitoring of their data processing. The deliberate incorporation of both of these elements of accountability (answerability and auditability) into the AI project lifecycle may be called **accountability-by-design**:

KEY CONCEPT

Accountability-by-Design

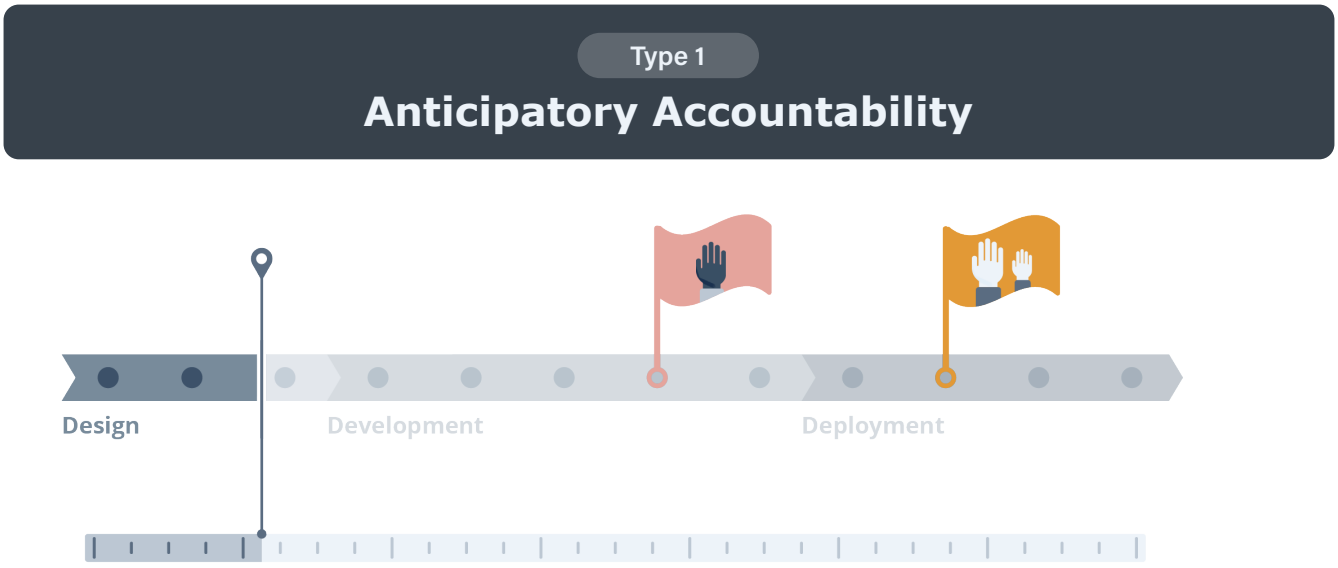
All AI systems must be designed to facilitate end-to-end answerability and auditability. This requires both responsible *humans-in-the-loop* across the entire design and implementation chain as well as activity monitoring protocols that enable end-to-end oversight and review.

Types of Accountability

Accountability deserves consideration across the entire design and implementation workflow. As a best practice, you should actively consider the different demands that accountability-by-design places on you before and after the roll out of your AI project.

We will refer to the process of ensuring accountability during the design and development stages of your AI project as '**anticipatory accountability**'.[12] This is because you are **anticipating your AI project's accountability needs prior to it being completed**. This subtype of accountability is sometimes referred to as ex-ante (or before-the-event) accountability.

Following a similar logic, we will refer to the process of addressing accountability after the start of the deployment of your AI project as '**remedial accountability**'. This is because **after the initial implementation of your system, you are remedying any of the issues that may be raised by its effects and potential externalities**. This subtype of accountability is sometimes referred to as ex-post (after-the-event) accountability, respectively.



Treating accountability as an anticipatory principle entails that you consider first and foremost the decisions made and actions taken by your project delivery team prior to the outcome of an algorithmically supported decision process.

This kind of ex-ante accountability should be prioritised over remedial accountability, which focuses instead on the corrective or justificatory measures that can be taken after the automation supported process has been completed.

By ensuring the AI project delivery processes are accountable prior to the actual application of the system in the world, you will bolster the soundness of the design and implementation processes and consequently more effectively pre-empt possible harms to individual wellbeing and public welfare or other adverse impacts.

Similarly, by establishing strong regimes of anticipatory accountability and by making the design and delivery processes as open and publicly accessible as possible, you will place affected stakeholders in a position to make better informed and more knowledgeable decisions about their involvement with these systems in advance of potentially harmful impacts. In doing so, you will also strengthen the public narrative and help to safeguard the project from reputational harm.[13]

Example of Anticipatory Accountability

During the Preprocessing phase of an AI production lifecycle, technical members of a project team are deciding which features to include, and which to leave out. To safeguard sufficient anticipatory accountability, they make sure to log which team members are involved in making these decisions and record the rationale behind the choices made.

Type 2

Remedial Accountability

While remedial accountability should be seen, along these lines, as a necessary fallback rather than as a first resort for imputing responsibility in the design, development, and deployment of AI systems, strong regimes of remedial accountability are no less important in providing necessary justifications for the bearing these systems have on the lives of affected stakeholders.

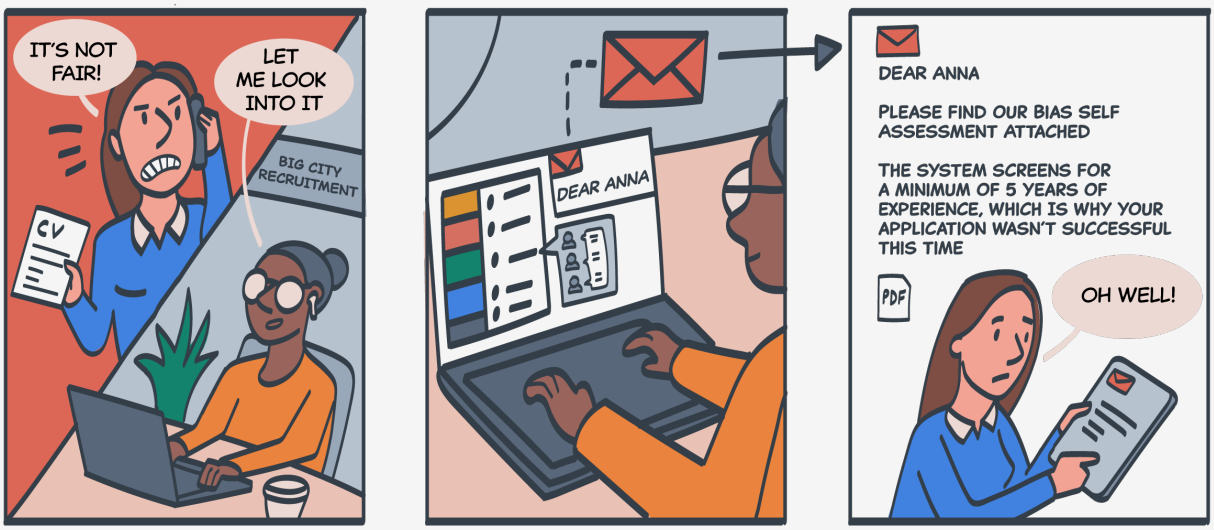
Putting in place comprehensive auditability regimes as part of your accountability framework and establishing transparent design and use practices, which are methodically logged throughout the AI project delivery lifecycle, are essential components for this sort of remedial accountability.[14]

One aspect of remedial accountability that you must pay close attention to is the need to provide **explanations** to affected stakeholders for algorithmically supported decisions.

Offering explanations for the results of algorithmically supported decision-making involves providing decision subjects and other interested parties with an understandable account of the rationale behind the specific outcome of interest. It also involves providing the decision subject and other interested parties with an explanation of the ethical permissibility, the fairness, and the safety of the use of the AI system. These tasks of **content clarification** and **practical justification** are explored in more detail in the AI Explainability in Practice workbook.

Example of Remedial Accountability

After receiving an unfavourable decision in their recruitment process, a job applicant seeks assurance and justification that the resume filtering AI system used in their process was not biased or discriminatory. To safeguard sufficient remedial accountability, the implementers of the system draw on the logged records of bias-mitigation measures and fairness-aware design practices to demonstrate the fair and non-discriminatory practices behind the production of the system. They also provide an explanation of the rationale behind the applicant's negative result, showing that the determinative factors behind the system's output were fair and reasonable.



Part Two: Putting Accountability into Practice

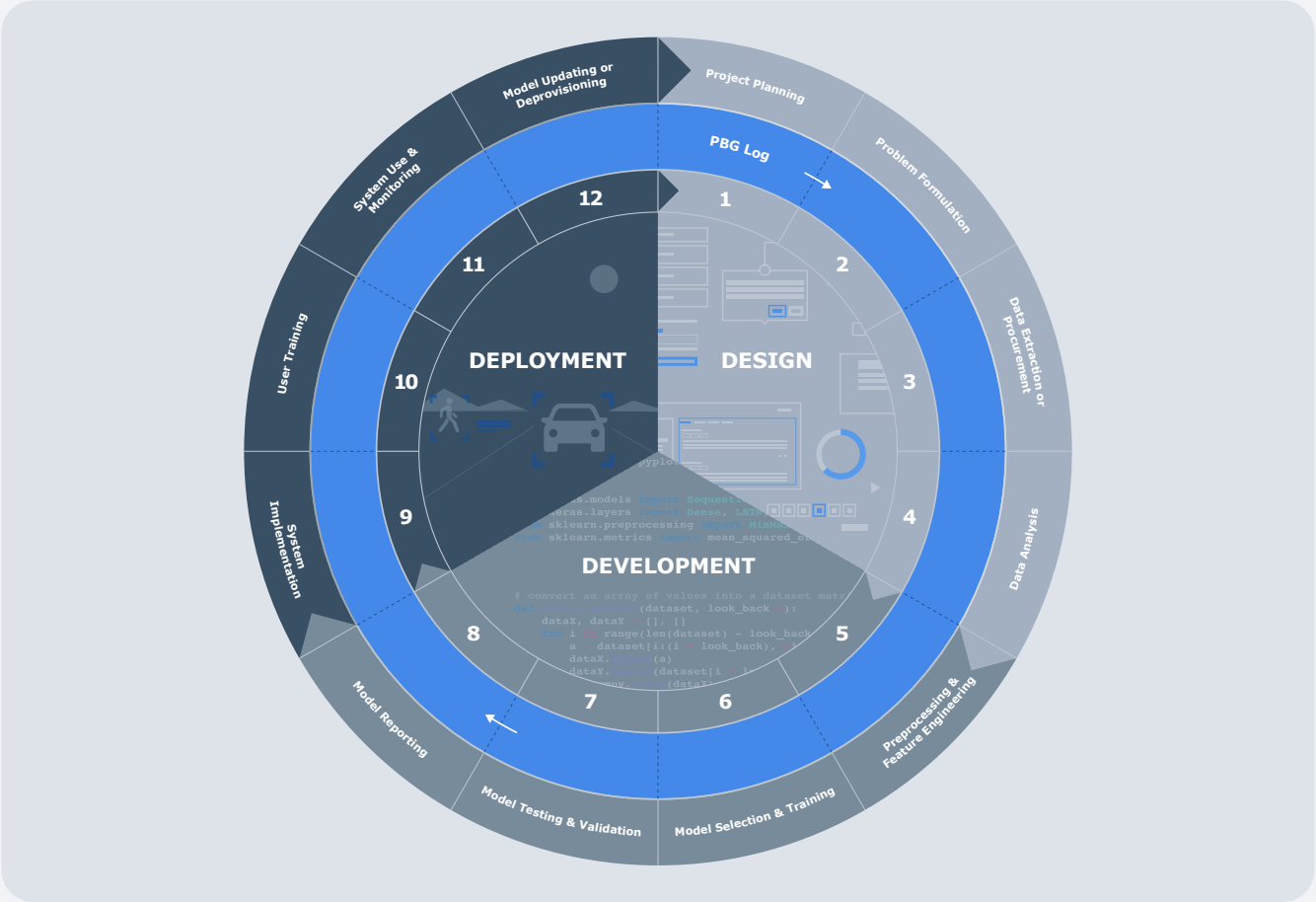
Now that we have explored some of the main aspects of the concept of accountability, we are ready to examine – in greater detail – how elements of answerability and auditability can be put into practice.

The central importance of the end-to-end operability of good governance practices should guide your strategy to embed accountability across the project workflow. Three components are essential to creating such a workflow:

1. Maintaining strong regimes of **professional and institutional transparency**.

2. Establishing and maintaining a clear and accessible **Process-Based Governance (PBG) Framework**.

3. Establishing a well-defined auditability trail for your PBG Framework through robust activity logging protocols that are consolidated digitally in a **Template Log**.



Maintaining Professional and Institutional Transparency

At every stage of the design and implementation of your AI project, team members should be held to rigorous standards of conduct that secure and maintain professionalism and institutional transparency. These standards should include the core values of **selflessness, integrity, honesty, accountability, openness, sincerity, neutrality, objectivity, impartiality, and leadership**. All professionals involved in the research, development, production, and implementation of AI technologies are, first and foremost, acting as **fiduciaries of the public interest** and must, in keeping with these core values of the Civil Service, put the obligations to serve that interest above any other concerns. For helpful guidance on professionalism and good conduct for public office holders, see the Seven Principles of Public Life (also known as the Nolan Principles).

Furthermore, throughout the AI project lifecycle, the design, development, and deployment process should be as transparent and as open to public scrutiny as possible. This might include restrictions on accessibility to relevant information limited to the reasonable protection of justified public sector confidentiality. It might also include limitations of the analytics that could tip off bad actors to methods of gaming public service delivery by exploiting, manipulating or working around current rules and arrangements.

Seven Principles of Public Life[15]

1. **Selflessness -** Holders of public office should act solely in terms of the public interest.
2. **Integrity -** Holders of public office must avoid placing themselves under any obligation to people or organisations that might try inappropriately to influence them in their work. They should not act or take decisions in order to gain financial or other material benefits for themselves, their family, or their friends. They must declare and resolve any interests and relationships.
3. **Objectivity -** Holders of public office must act and take decisions impartially, fairly, and on merit, using the best evidence and without discrimination or bias.
4. **Accountability -** Holders of public office are accountable to the public for their decisions and actions and must submit themselves to the scrutiny necessary to ensure this.
5. **Openness -** Holders of public office should act and take decisions in an open and transparent manner. Information should not be withheld from the public unless there are clear and lawful reasons for so doing.
6. **Honesty -** Holders of public office should be truthful.
7. **Leadership -** Holders of public office should exhibit these principles in their own behaviour and treat others with respect. They should actively promote and robustly support the principles and challenge poor behaviour wherever it occurs.

The Process-Based Governance (PBG) Framework and Log

So far, this workbook series has presented some of the main values and principles necessary for establishing responsible innovation practices in your AI project. Perhaps the most vital of these measures is the effective operationalisation of these practices. The standard ISO 37000 defines governance as 'the system by which the whole organisation is directed, controlled, and held accountable to achieve its core purpose in the long run'.[16] Establishing a diligent and well-conceived governance framework that covers the entire design, development, and deployment process will provide the foundation for your team to effectively establish needed practical actions and controls, exhaustively distribute roles and responsibilities, and operationalise answerability and auditability throughout the AI lifecycle. By organising all of your governance actions into a PBG Framework, you will be better able to accomplish this task.

The purpose of a PBG Framework is to provide a template for the integration of the norms, values, and principles, which motivate and steer responsible innovation, with the actual processes that characterise the AI design and development pipeline. Establishing a PBG Framework creates the baseline conditions to ensure that the goal of instituting an AI innovation process that is accountable-by-design is achieved.

Your Framework should give you a landscape view of the governance actions that are organising the control structures of your project workflow. Constructing a good PBG Framework will provide you and your team with a big picture of:

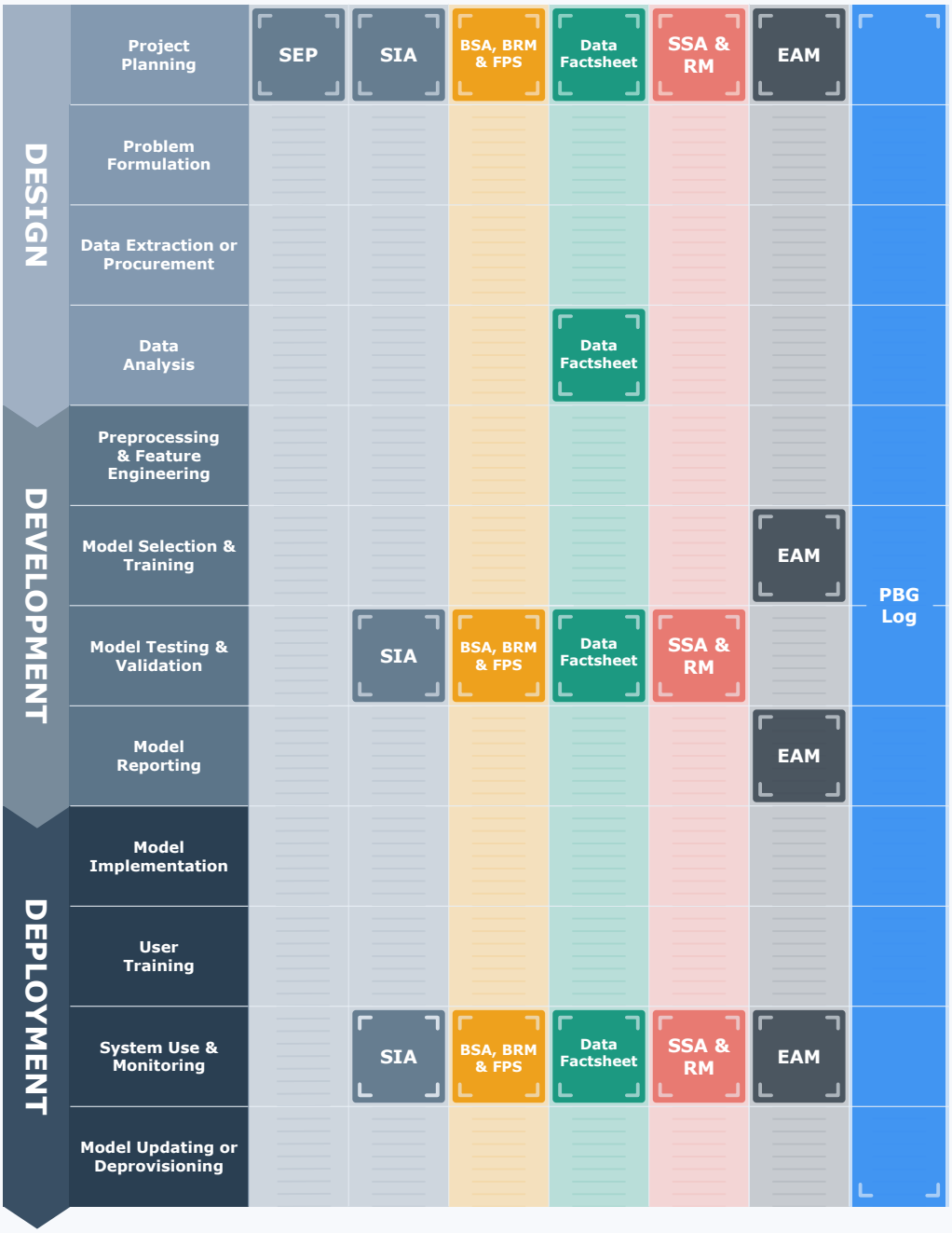
- The relevant stages of the workflow in which actions are necessary to meet governance goals.
- The relevant team members and roles involved in each governance action.
- Explicit timeframes for any necessary follow-up actions, re-assessments, and continual monitoring.
- Clear and well-defined protocols for logging activity and for instituting mechanisms to assure end-to-end auditability and appropriate documentation.

The PBG Framework asks that teams not only outline the governance actions established for individual projects, but also roles involved in each action, timeframes for follow-up actions, and logging protocols.

PBG Log for
project name

- Sustainability
- Accountability
- Fairness
- Safety
- Data Stewardship
- Explainability

Phase 1: Establishing Proportional Governance Actions



Phase 2: Determining Roles Involved and Timeframes

Action	Roles Involved			Timeframe
	Design	Development	Deployment	
SEP (Stakeholder Engagement Process)				A. ..... B. ..... C. .....
SIA (Stakeholder Impact assessment)				A. ..... B. ..... C. .....
BSA & BRM (Bias Self-Assessment & Bias Risk Management)				A. ..... B. ..... C. .....
FPS (Fairness Position Statement)				A. ..... B. ..... C. .....
Data Factsheet				A. ..... B. ..... C. .....
SSA & RM (Safety Self-Assessment and Risk Management)				A. ..... B. ..... C. .....
EAM (Explainability Assurance Management)				A. ..... B. ..... C. .....
PBG Log				A. ..... B. ..... C. .....

A. First Action Completion Date
B. Last Revision Date
C. Timeframe for Next Revisitation

Phase 3: Outlining Logging Protocols

a. How will governance actions be documented?

b. How will artefacts yielded by governance activities be centralised and organised?

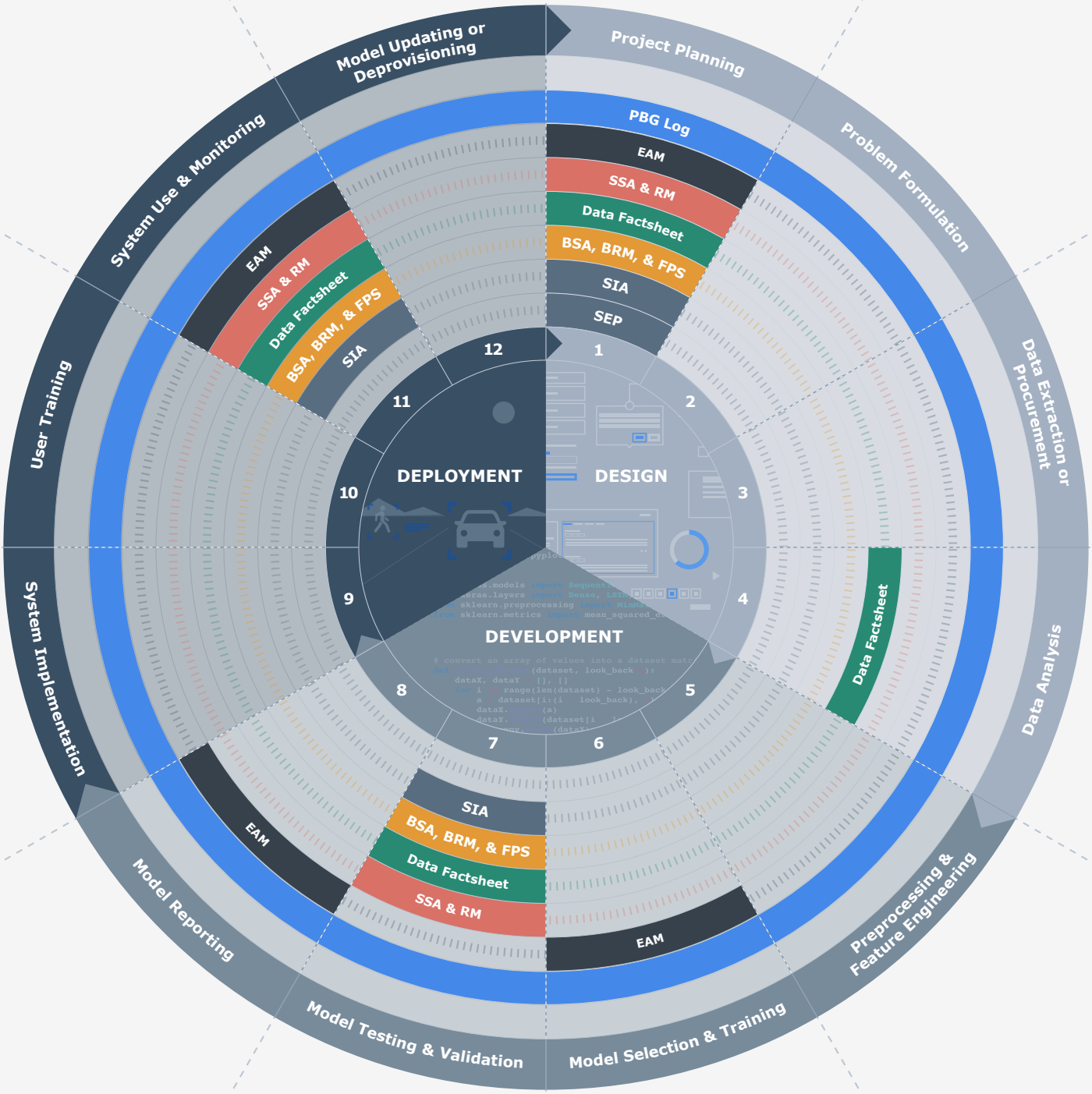
c. Who will be responsible for logging updates to activities and artefacts?

Establishing Proportionate Governance Actions

Just as the determination of proportionate stakeholder involvement requires a risk-based calibration, the establishment of proportionate governance protocols should involve a preliminary assessment of the potential risks and hazards of the model or tool under consideration. Low-stakes AI applications that are not safety-critical, do not directly impact the lives of people, and do not process potentially sensitive social and demographic data may need less proactive governance controls and processes than high-stakes projects. By completing the **Project Summary Report** (found in AI Sustainability in Practice Part One) and **Stakeholder Impact Assessments** (found in AI Sustainability in Practice Part Two), you and your project team will carry out an evaluation of the possible risks that could arise from your project and of the potential hazards it poses to affected individuals and groups. These assessments of the dangers posed to individual wellbeing and public welfare will help you formulate proportionate governance actions to be outlined in your PBG Framework.

Notwithstanding the importance of the need for this reasonable application of proportionate governance actions, you should take care to establish a strong regime of accountability-by-design across your project lifecycle. It may be the case that your assessment of potential risks and adverse impacts does not sufficiently anticipate the full range of possible harms. In instances where such unforeseen harms do arise, you will want to have in place proper mechanisms of anticipatory accountability and corresponding documentation protocols, so that the best practices of your project team are demonstrable.

Becoming familiar with all governance actions will support you and your team in establishing proportional actions for your project. Here is a summary picture of how possible governance actions fit across an AI project workflow:



List of Principles and Governance Actions

Sustainability

Stakeholder Engagement Process

Process facilitating the uptake of proportionate stakeholder engagement and input throughout the AI lifecycle. The SEP enables a contextually informed understanding of the social environment and human factors that may be impacted by, or may impact, individual AI projects.[18] A main output of the SEP is the PS Report, which documents preliminary project scoping, stakeholder analysis, and positionality reflection activities and provides an overview of established stakeholder objectives and methods.

Stakeholder Impact Assessment

Process facilitating the iterative evaluation of the social impact and sustainability of individual AI projects, as well as the corroboration of these potential impacts in dialogue with stakeholders, when appropriate.

Accountability

Process-Based Governance Log

Live document outlining governance actions, relevant team members and roles involved in each action, timeframes for follow-up actions, and logging protocols, for individual AI projects.[17]

Explainability

Explainability Assurance Management

Iterative process aimed to facilitate the implementation and evaluation explainability assurance activities across the project lifecycle and assist in providing clarification of AI system outputs to a range of impacted stakeholders.

Safety

Safety Self-Assessment and Risk Management

Process facilitating the evaluation of how AI projects align with safety objectives through the iterative identification and documentation of potential safety risks across the lifecycle and assurance actions implemented to address these.

Fairness

Bias Self-Assessment, Bias Risk Management & Fairness Position Statement

Process facilitating the evaluation of how AI projects align with the principle of fairness through the iterative identification and documentation of risks of bias across the lifecycle and assurance actions implemented to address these. The FPS is a document establishing the metric-based fairness criteria for individual AI projects, providing an explanation in plain and non-technical language.

Data Stewardship

Data Factsheet

Live document facilitating the uptake of best practices for responsible data management and stewardship across the AI project workflow. The document consists of a comprehensive record of the data lineage and iterative assessments of data integrity, quality, protection, and privacy.

Accountability Across the Workflow

The task of establishing a PBG Framework should be initially undertaken in the project planning step of your project alongside your Project Summary Report (Please refer to the AI Sustainability in Practice Part One workbook for details on the full process of creating your PS Report). The results of your Stakeholder Analysis (particularly, the scoping of potential stakeholder impacts) should inform a proportional selection of governance actions within your PBG Framework. At this stage, your PBG Framework will provide a prospective and provisional plotting of governance actions, roles, and responsibilities for your project. This preliminary outline of governance structures will provide the necessary information for answering the Governance Framework Reflection questions within the PS Report.

In the PS Report, the task of reflecting on your governance framework involves answering the following questions:

Component 5: Map Governance Workflow

a. Do established governance actions proportionally mitigate possible harms to stakeholders posed by this project? If not, how can your PBG Framework be rectified to address these potential harms?

b. Does this distribution of responsibilities outlined in the PBG Framework establish a continuous chain of human accountability throughout the design, development, and deployment of this project? If not, how can any identified gaps or breaks in the chain be rectified in the PBG Framework?

c. How will you ensure that all team members, who are assigned roles/ responsibilities understand the roles/ responsibilities that have been assigned to them?

d. If you are procuring parts or elements of the system from third-party vendors, suppliers, sub-contractors, or external developers, how are you instituting appropriate governance controls that will establish end-to-end accountability, traceability, and auditability for these procured parts or elements?

e. If any data used in the production of the AI system will be acquired from a vendor, supplier, or third party, how are you instituting appropriate governance controls that will establish end-to-end accountability, answerability, and auditability across the data lifecycle?

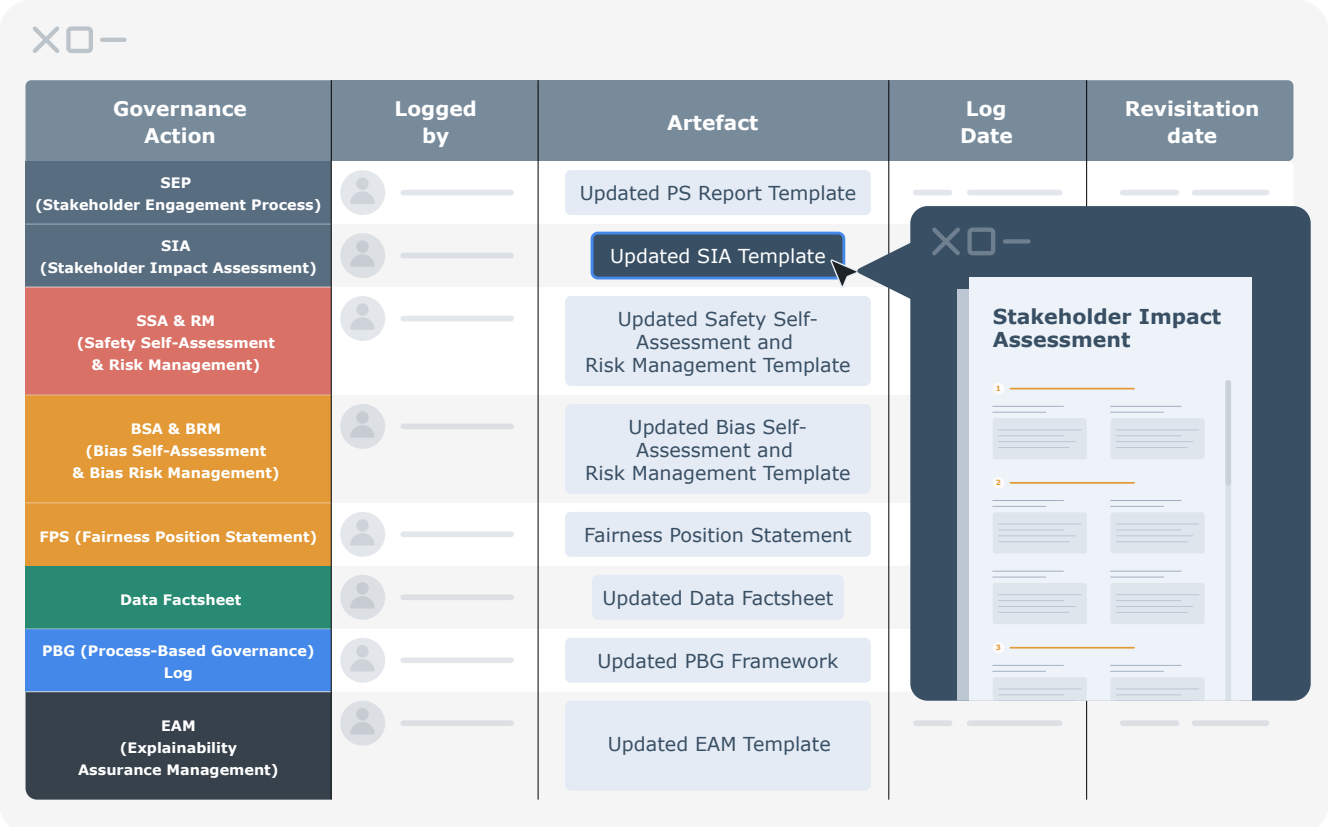
These questions (alongside the rest of the PS Report) are to be revisited and updated as part of completing each iteration of the Stakeholder Impact Assessment, at each point informing any necessary updates to your project's governance structure (and PBG Framework). Your PBG Framework is therefore a live document reflecting a governance structure that responds to the emerging needs across the design, development, and deployment lifecycle. It is to be updated after each revisitation of the PS Report to reflect the project's current governance structure.

The process by which you answer these questions should be as collaborative and inclusive as possible. You should aim to involve all relevant members of your project team (and any other relevant managers, operators, or vendors), so that all the relevant stakeholders in the workflow can share input and come to understand expectations about their roles and responsibilities. Any future revisions or updates to this part of the PS Report should likewise include all affected parties.

Enabling Auditability with a Template Log

With your controls in place and your governance framework organised, you will be better able to manage and consolidate the information necessary to assure end-to-end auditability. A detailed process log and secure repository of artefacts created through your governance activities (e.g. SIA Reports, Data Factsheet) should be maintained and updated across the workflow as PBG actions are conducted, and any associated artefacts created. By centralising your information digitally in a process log, you are preparing the way for optimal process transparency. A process log will enable you to make available, in one place, information that may assist you in demonstrating to concerned parties and affected decision subjects both the responsibility of design and use practices and the justifiability of the outcomes of your system's processing behaviour.

Such a log will also allow you to differentially organise the accessibility and presentation of the artefacts yielded by governance activities conducted throughout your project. Not only is this crucial to preserving and protecting data that legitimately should remain unavailable for public view, but it will also afford your team the capacity to cater the presentation of results to different tiers of stakeholders with different interests and levels of expertise. This ability to curate your explanations with the stakeholder in mind will be vital to achieving the goals of transparent and explainable AI. This is discussed in-depth in the AI Explainability in Practice workbook.



AI Accountability in Practice

Activities

Activities Overview

In the previous sections of this workbook, we have presented the various elements of accountability. In this section, we provide concrete tools for applying these concepts in practice. Activities will help participants identify accountability challenges, gain familiarity with governance actions that will help them embed accountability across the project workflow, and establish an auditability trail for their Process-Based Governance Framework.

We offer a collaborative workshop format for team learning and discussion about the concepts and activities presented in the workbook. To run this workshop with your team, you will need to access the resources provided in the link below. This includes a digital board and printable PDFs with case studies and activities to work through.

Workshop resources for AI Accountability in Practice

A Note on Activity Case Studies

Case studies within the Activities sections of the AI Ethics and Governance in Practice workbook series offer only basic information to guide reflective and deliberative activities. If activity participants find that they do not have sufficient information to address an issue that arises during deliberation, they should try to come up with something reasonable that fits the context of their case study.

Note for Facilitators

In this section, you will find the participant and facilitator instructions required for delivering activities corresponding to this workbook. Where appropriate, we have included considerations to help you navigate some of the more challenging activities.

Activities presented in this workbook can be combined to put together a capacity-building workshop or serve as stand-alone resources. Each activity corresponds to a section within the Key Concepts in this workbook. Some activities have prerequisites, which are detailed on the following page.

We sometimes provide ideas of how a **co-facilitator** can help manage large groups.

Identifying Challenges

Practise identifying challenges to accountability presented by AI projects.

Corresponding Sections

- Part One: Introduction to Accountability (page 10)

Useful Sections to Revise

- Project Summary Report Template from the AI Sustainability in Practice Part One workbook.

Establishing Proportional Governance Actions

Become familiar with governance actions and understand how they are put in place to address the potential harmful impacts of AI systems.

Corresponding Sections

- Part Two: Putting AI Accountability into Practice (page 18)

Useful Sections to Revise

- Part Two: The Sociotechnical Aspect of the AI and Machine Learning Project Lifecycle from the AI Ethics and Governance in Practice: An Introduction workbook
- Project Summary Report Template from the AI Sustainability in Practice Part One workbook

Mapping Roles and Responsibilities

Practise mapping team roles and responsibilities to governance actions across the AI lifecycle.

Corresponding Sections

- Part One: Introduction to Accountability (page 10)
- Part Two: Putting Accountability into Practice (page 18)

Prerequisites

- Activity: Establishing Proportional Governance Actions (page 42)

Useful Sections to Revise

- Project Summary Report Template from the AI Sustainability in Practice Part One workbook

Completing Maps

Practise assessing Workflow Governance Maps.

Corresponding Sections

- Part Two: Putting Accountability into Practice (page 18)

Prerequisites

- Activity: Establishing Proportional Governance Actions (page 42)
- Activity: Mapping Roles and Responsibilities (page 44)

Useful Sections to Revise

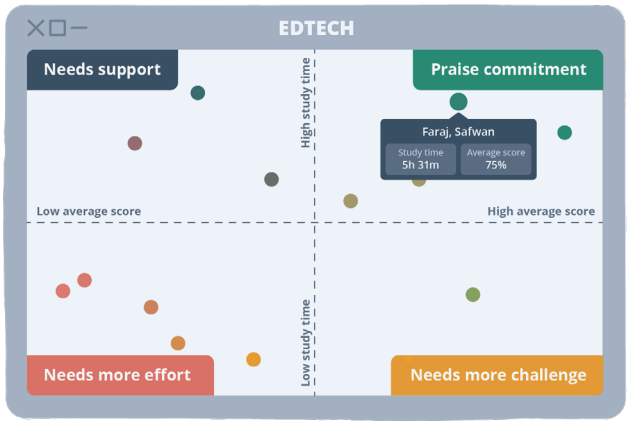
- Project Summary Report Template from the AI Sustainability in Practice Part One workbook

Interactive Case Study: AI EduTech

Your team is the senior management of a Local Education Authority who is working with a private contractor (AI EduTech) to tailor an off-the-shelf web platform that will be trialled across primary schools in your borough.

This platform provides students with a curriculum composed of educational content, activities, and assessments. It aims to benefit students, teachers, and school managers through **an embedded AI system that categorises students based on the type of support they need to deliver satisfactory performance**.

Students will have individual profiles from which they can access learning material and complete activities and assessments.



Student Categorisation

The AI system is to use their time spent in activities, their scores, and other variables (i.e. students' demographics, past academic results, and family history information) that will be tailored to your specific model to categorise students as follows within each class subject:

Needs support

High study time, low average score - needs dedicated learning support.

Praise commitment

High study time, high average score - needs recognition for effort.

Needs more effort

Low study time, Low average score - needs intervention towards more effort.

Needs more challenge

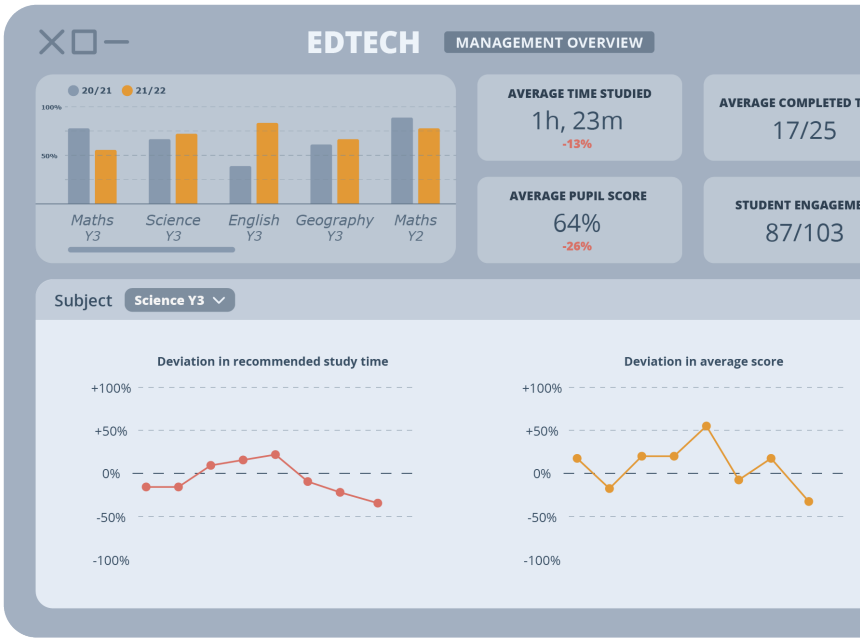
Low study time, high average score - needs more challenging content.

Insights for Teachers

Teachers will have profiles from which they can view class-wide performance metrics and the distribution of students across these categories, allowing them to identify students in need of dedicated learning support or interventions. Teachers will also be able to view individual students' profiles, containing scores, need category, and their subjects of strength and weakness, helping them tailor their support and inform interventions towards satisfying scores.

Insights for Senior Management

Senior management within schools will have access to a dashboard allowing them to view scores and progress across all areas in a given school, including scores per schools, classes, and individual students. This will enable them to deliver well informed school-wide initiatives and interventions towards satisfying performance.



The Project Summary Report contains stakeholder analysis, positionality reflections, and engagement objectives and methods established to follow the Stakeholder Impact Assessment. The initial Project Summary Report takes place during the design phase of the project workflow and is revisited during each subsequent Stakeholder Impact Assessment. This artefact relates to the Stakeholder Engagement Process, which operationalises the principle of sustainability, and is described in further detail in 'Putting the SUM Values Into Practice' in the AI Sustainability in Practice Part One workbook.

Project Summary Report for: AI EduTech

Component 1: Project Scoping and Stakeholder Analysis

1. Outlining Project, Use Context, Domain, and Data

a. Which organisation(s)—yours, other suppliers, or other providers—are responsible for building this AI system?

Responses

- Our team – a local education authority.
- AI EduTech, a private contractor.
- The system will be deployed by school teachers and managers in our borough.

b. Which parts or elements of the AI system, if any, will be procured from third-party vendors, suppliers, subcontractors, or external developers?

Responses

- The platform will be procured from AI EduTech, who will tailor the model to meet the specific needs of our borough.

2. Domain

a. Which, if any, domain experts have been or will be consulted in designing and developing the AI system?

Responses

- A domain expert in public education will be involved in the design and development of this system.

3. Data

a. What datasets are being used to build this AI system?

Responses

- Historic school records including demographic data, graduation rates, subject scores, assessment scores.

b. Will any of the data used in the production of the AI system be acquired from a vendor or supplier? (Describe)

Responses

- The datasets used will be acquired from AI Edutech and include open source data from the Department of Education as well as data acquired from previous uses of the system.

4. Identifying Stakeholders

a. Who are the stakeholders (both individuals and social groups) that may be impacted by, or may impact, the project?

Responses

- Students
- Teachers
- School Managers
- Representatives

b. Do any of these stakeholders possess sensitive or protected characteristics that could increase their vulnerability to abuse, adverse impact, or do they require additional protection require additional protection or assistance with respect to the impacts of the project? If so, what characteristics?

Responses

- Students are children, which is a protected characteristic.
- Considering the scope of deployment being a borough, all protected characteristics are found within stakeholders.

c. Could the outcomes of this project present significant concerns to specific groups of stakeholders given vulnerabilities caused or precipitated by their distinct circumstances? If so, what vulnerability characteristics expose them to being jeopardised by project outcomes?

Responses

- Underperforming students and teachers leading underperforming classes may be particularly impacted by this model, as they will likely be directly influenced by interventions at the teaching or school-wide level.

5. Scoping Potential Stakeholder Impacts

Consider this section when partcipating the Identifying Challenges activity.

a. How could each of the SUM values and their associated ethical concerns values be impacted by the AI system we are planning to build?

Respect
the dignity of individual persons

Ethical Concerns:

- Dignity, autonomy, agency, and authority of persons
- Self-realisation and flourishing of individuals

Responses

- Automated assessments and categorisations undermine teacher's authority in their approach to providing support for their classes.[19]
- The categorisation of students based on their needs might undermine their agency in articulating their preferred avenues for receiving support.

Connect

with each other sincerely, openly, and inclusively

Ethical Concerns:

- Integrity of interpersonal connections
- Solidarity
- Participation-based innovation and stakeholder inclusion

Responses

- The deployment of a learning platform may de-personalise education, harming the development of interpersonal connections between teachers and students.[20] The high level of standardised assessment and oversight of class performance may equally de-personalise the relationships between school managers and teachers by potentially leading to the use of this model for teacher evaluations and jeopardising consideration for individual circumstances.

Protect

the priorities of social values, justice, and the public interest

Ethical Concerns:

- Justice and equity
- Prioritisation of the public interest and common good

Responses

- The use of historic school records risks categorising students based on past performance within historically unequal institutions (as in, unequal resources, class sizes, and student support) thereby discriminating by proxy against those of disadvantaged classes, races, and ethnicities.[21]

Care

for the wellbeing of each and all

Ethical Concerns:

- Beneficence, safety, and non-harm

Responses

- Procuring unvetted and indiscriminate datasets might risk delivering inaccurate categorisations that impact the health, welfare or safety of students. The model must be tailored to the schools in this borough.[22]

Component 5: Map Governance Workflow

a. What roles are involved in each of the project phases?

Design

Deployment

Development

b. What are the responsibilities of each of these roles?

Design Roles

Deployment Roles

Development Roles

c. How are each of these duty bearers assigned responsibility for the system's potential impacts?

- Does this distribution establish a continuous chain of human accountability throughout the design, development, and deployment of this project? If so, how?

d. What logging protocol is established for documenting workflow activities?

- Does this protocol enable auditing and oversight of the design, development, and deployment of this project? If so, how?

e. Can responsible duty bearers be traced in the event that stakeholders are harmed by this system? If so, how does the project's distribution of responsibilities and logging protocol enable this?

f. If you are procuring parts or elements of the system from third-party vendors, suppliers, sub-contractors, or external developers, how are you instituting appropriate governance controls that will establish end-to-end accountability, traceability, and auditability?

Your task

Your team is in the Project Planning phase of the AI lifecycle and has finished drafting the sections of your Project Summary Report pertaining to the Stakeholder Engagement Process. You will now use this information to write up the Governance Framework Reflection section of the report.

25 mins | Participant Instructions

Identifying Challenges

Objective

Practise identifying challenges to accountability presented by AI projects. Participants will brainstorm what accountability means to them and what possible challenges to accountability may be presented by the system under consideration.

Team Instructions

1. Individually take 5 minutes to review the case study for this workshop.
2. Next, your team will be split into groups.
3. Designate a group member to take discussion notes and report back to the team.
4. Have a group brainstorm considering the questions:

Relating to Accountability:

- What does accountability mean to you?
- In what ways is accountability important in schools?

Case Study Gaps:

- What decisions will be automated by this AI system which have traditionally been made by humans?
- Why is it important that there are individuals that are accountable for these decisions?

Complexity in AI Production Processes:

- Take a look at the Scoping Potential Stakeholder Impacts section of the Project Summary report included in the Interactive Case Study. Are there any aspects of this project that present challenges to assigning responsibility to individual people for potentially harmful stakeholder impacts?

25 mins | Facilitator Instructions

Identifying Challenges

1. Give participants a moment to read over the activity instructions, answering any questions.

2. Next, give team members 5 minutes to read over the case study for this workshop.

3. When 5 minutes are up, split the team into groups.

4. Give groups 10 minutes to have a discussion.

5. Facilitators and co-facilitators are to join groups and assist them using the considerations section of this activity.

6. When 10 minutes are up, ask the team to reconvene, giving each volunteer note-taker some minutes to report back on their group's discussion.

7. When all groups have shared, have a discussion about what elements of this AI system the team considers to be the greatest challenge to accountability.
   - **Co-facilitator:** take notes on this discussion in the Team Discussion section of the board.

40 mins | Participant Instructions

Establishing Proportional Governance Actions

Objective

The purpose of this activity is to become familiar with governance actions and understand how they are put in place to address the potential harmful impacts of AI systems.

Team Instructions

1. The team will be split into groups, each assigned a potential impact to stakeholders posed by the system under consideration.

2. In your groups, take a moment individually to review the **Governance Activities**.

3. Next, have a team discussion about what governance actions may help address your assigned impact.

    - Groups can use the example on the table for guidance.

    - For each governance action, consider answering the following question:

        - Would the absence of the governance action negatively affect how the AI project team addresses the assigned impact? If so, how?

4. Having determined a set of activities, place these governance actions on the **Possible Impacts and Activities** section on the board.

5. Reconvene as a team, having a volunteer from each group share their assigned possible impacts and related governance actions.

6. Have a group discussion about how the governance actions map to possible stakeholder impacts for this project.

7. Consider the following questions:

    - Is the incorporation of the governance action into the project workflow proportional to the potential harms posed by this project?

    - How does the governance action map to possible stakeholder impacts for this project?

    - Does the governance action address anticipatory accountability or remedial accountability?

40 mins | Facilitator Instructions

Establishing Proportional Governance Actions

1. Give the team a moment to read over the instructions, answering any questions.

2. Next, split the team into the same groups as the previous activity, letting them know that they will have 15 minutes to brainstorm governance activities that can help address their assigned impact.

3. Discuss the filled-in example on the table to help guide team members.

   - **Group 1:** Respect
   - **Group 2:** Protect
   - **Group 3:** Care

4. **Facilitators** and **co-facilitators** are to join groups and make sure the governance actions discussed are placed in the **Possible Impacts and Activities** section on the board.

5. When 15 minutes have passed, ask the team to reconvene.

6. Give team members a moment to individually review the **Possible Impacts and Governance Actions** section and ask for volunteers from each team to take a few minutes to discuss why their team chose the governance actions that they did.

7. After each group has had a discussion, give the team a moment to provide feedback, editing results as needed.

8. Next, facilitate a discussion on each of the governance actions, going through each of them and asking the following questions:

   - Is the incorporation the governance action into the project workflow proportional to the potential harms posed by this project?
   - How does the governance action map to possible stakeholder impacts for this project?
   - Does the governance action address anticipatory accountability or remedial accountability?
   - **Co-facilitator:** Place selected actions within their respective places on the table on **Mapping Roles and Responsibilities** section.

9. When all groups have shared, your team will have a map of Possible Impacts and their associated governance actions, as well as a PBG Framework with selected governance actions, which is visualised in the table on the Mapping Roles and Responsibilities activity.

40 mins | Participant Instructions

Mapping Roles and Responsibilities

Objective

Practise mapping team roles and responsibilities to governance actions across the AI lifecycle.

Team Instructions

1. In this activity, your team will determine what roles will be involved in each established governance action.

2. Individually, take a moment to read over the **Roles and Responsibilities** section and review the **PBG Framework** section.

3. Next, return to your groups from the previous activity. Each group is to map out what roles will be involved in PBG activities within a specific phase of the workflow.

    - **Involved roles** are team members who are assigned with successfully completing a governance activity.

4. As a group, read over the Roles and Responsibilities table on your Group's section of the board. For every governance activity in each phase in the lifecycle, consider what roles should be involved in each governance action by evaluating role descriptions.

5. Move roles to their respective section of the **PBG Framework** on the board.

6. The team will reconvene and engage in a discussion aimed at consolidating the assignment of responsibilities across the project workflow.

Roles and Responsibilities

PBG Framework

40 mins | Facilitator Instructions

Mapping Roles and Responsibilities

Note

This activity should help participants establish roles and responsibilities within the PBG Framework and answer the following questions within the Governance Framework Reflection section of the PS Report:

- Does this distribution of responsibilities establish a continuous chain of human accountability throughout the design, development, and deployment of this project? If not, how can any identified gaps or breaks in the chain be rectified in the PBG Framework?

Facilitator Instructions

1. Give the team a moment to read over the activity instructions, answering any questions.

2. Next, give team members 5 minutes to individually read over the Roles and Responsibilities table and review the PBG Framework.

3. Then split the team into the same groups from the previous activity, letting them know that they will have 20 minutes to fill in their **Roles and Responsibilities** tables.

    - **Facilitators** and **co-facilitators** are to join groups throughout this time, providing support using the considerations section of this activity.

4. When 15 minutes have passed, let the team know that they have 5 minutes to finish filling in their table.

5. When 20 minutes have passed, ask the team to reconvene.

6. Give the team 3 minutes to individually look over each groups' Roles and Responsibilities tables.

7. Next, lead a group discussion by asking the groups to share their experiences mapping roles and responsibilities. If the team struggles starting conversation, consider the following questions:

    - Does this distribution of responsibilities establish a continuous chain of human accountability throughout the design, development, and deployment of this project? If not, how can any identified gaps or breaks in the chain be rectified?

    - In the event that this system inflicts a specific harmful impact to stakeholders (out of the potential harms discussed in the previous activity), does this assignment of responsibility allow one to trace the responsibility for this event back to specific roles?

45 mins | Participant Instructions

Completing Maps

Objective

The purpose of this activity is to practice assessing Workflow Governance Maps. This activity will help answer the the questions in the PS Report.

Team Instructions

1. In this activity, your groups will address final questions pertaining to the established governance actions, roles, and responsibilities for this project. Your group will start by answering one question and then provide feedback to the answers of the questions addressed by other groups in your team.

2. Your team will be split into groups, each assigned a question from the Governance Workflow Map template.

   Group 1: How will you ensure that all team members, who are assigned roles/ responsibilities understand the roles/ responsibilities that have been assigned to them?

   Group 2: If you are procuring parts or elements of the system from third-party vendors, suppliers, sub-contractors, or external developers, how are you instituting appropriate governance controls that will establish end-to-end accountability for these procured parts or elements?

   Group 3: If any data being used in the production of the AI system will be acquired from a vendor, supplier, or third party, how are you instituting appropriate governance controls that will establish end-to-end accountability across the data lifecycle?

3. In your groups, have a 10-minute discussion about your assigned question, coming up with a proposed approach for addressing it.

   - Use sticky notes to write your answers, placing them in your group's section of the board.

4. After 10 minutes, groups will be assigned a different question. In this part of the activity, groups will have 5 minutes to review proposed answers, use sticky notes to write any feedback or additions that they would incorporate to better address the questions.

5. When time is up, groups will be assigned the final question, conducting another 5 minute review.

6. After the final review, groups are to return to their original question and discuss the feedback provided, taking 5 minutes to incorporate feedback and strengthen their approach.

7. The team will reconvene, taking a moment to read over the final answers to each question and engage in a discussion.

45 mins | Facilitator Instructions

Completing Maps

1. Give the team a moment to read over the activity instructions, answering any questions.

2. Split the team into three groups, each assigned a question from the Governance Workflow Map template.

3. Let the groups know that they will have 10 minutes to come up with answers to their questions.

    - **Facilitators** and **co-facilitators** are to join in group and aid in discussion.

4. Let the groups know when they have 2 minutes left for this section of the activity.

5. When 10 minutes have passed, assign each group to a different question, letting them know that they will have 5 minutes to provide feedback.

6. When 5 minutes have passed, assign groups to a final question, again letting them know that they will have 5 minutes to provide feedback.

7. When 5 minutes have passed, assign groups to their original question, letting them know that they will have 5 minutes to incorporate feedback.

8. When 5 minutes have passed, ask the team to reconvene, giving them a moment to read over the final answers to each question.

9. Next, lead a 15-minute group discussion Considering the question:

    - In what ways do the established governance actions, roles, and responsibilities help address the challenges to AI accountability posed by this project?

10. Ask participants to refer to the challenges identified in the Mapping Roles and Responsibilities activity when engaging in this discussion.

Endnotes

1 Coeckelbergh, M. (2020). Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Science and Engineering Ethics. 26,* 2051–2068. https://doi.org/10.1007/s11948-019-00146-8

2 Goetze, T. S. (2022). Mind the Gap: Autonomous Systems, the Responsibility Gap, and Moral Entanglement. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 390-400). https://doi.org/10.1145/3531146.3533106

3 Central Digital & Data Office (2020). Data Ethics Framework. https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020

4 Bryson, J. J., Diamantis, M. E., & Grant, T. D. (2017). Of, for, and by the people: the legal lacuna of synthetic persons. *Artificial Intelligence and Law, 25*(3), 273-291. https://doi.org/10.1007/s10506-017-9214-9

5 Coeckelbergh, M. (2020). Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. Science and Engineering Ethics. 26, 2051–2068. https://doi.org/10.1007/s11948-019-00146-8

6 Leslie, D. (2020). Explaining Decisions Made with AI. The Alan Turing Institute and ICO. http://dx.doi.org/10.2139/ssrn.4033308

7 Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. *Fordham Law Review, 87,* 1085-1139. https://dx.doi.org/10.2139/ssrn.3126971

8 Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92. https://doi.org/10.1145/3458723

9 Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In Proceedings of the conference on fairness, accountability, and transparency (pp. 220-229). https://doi.org/10.1145/3287560.3287596

10 Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020, January). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 33-44). https://doi.org/10.1145/3351095.3372873

11 Wagner, B. (2020). Accountability by design in technology research. *Computer Law & Security Review, 37,* 105398. https://doi.org/10.1016/j.clsr.2020.105398

12 see Anderson, B. (2010). Preemption, precaution, preparedness: Anticipatory action and future geographies. *Progress in Human Geography, 34*(6), 777-798. https://doi.org/10.1177/0309132510362600

13 Reisman, D., Schultz, J., Crawford, K. and Whittaker, M. (2018). Algorithmic impact assessments: a practical framework for public agency accountability. AI Now Institute. https://ainowinstitute.org/publication/algorithmic-impact-assessments-report-2

14 Reisman, D., Schultz, J., Crawford, K. and Whittaker, M. (2018). Algorithmic impact assessments: a practical framework for public agency accountability. AI Now Institute. https://ainowinstitute.org/aiareport2018.pdf

15 Committee on Standards in Public Life (1995) The Seven Principles of Public Life. https://www.gov.uk/government/publications/the-7-principles-of-public-life/the-7-principles-of-public-life--2

16 International Organization for Standardization (2021). Governance of organizations - Guidance (ISO 37000:2021). ISO. https://www.iso.org/standard/65036.html

17 Leslie, D. (2019). *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector.* The Alan Turing Institute. https://doi.org/10.5281/zenodo.3240529

18 Leslie, D., Katell, M., Aitken, M., Singh, J., Briggs, M., Powell, R., Rincon, C., Perini, A. M., & Jayadeva, S. (2022). *Data Justice in Practice: A Guide for Policymakers.* SSRN. https://dx.doi.org/10.2139/ssrn.4080050

19 see Cardona, M. A., Rodríguez, R. J., & Ishmael, K. (2023). Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations. U.S. Department of Education: Office of Educational Technology. https://www2.ed.gov/documents/ai-report/ai-report.pdf

20 see Guilherme, A. (2019). Considering AI in Education: Erziehung but Never Bildung. In: Knox, J., Wang, Y., Gallagher, M. (eds) Artificial Intelligence and Inclusive Education. Perspectives on Rethinking and Reforming Education. Springer, Singapore. https://doi.org/10.1007/978-981-13-8161-4_10

21 see Perry, A. M., & Lee, N. T. (2019). AI is coming to schools, and if we're not careful, so will its biases. Brookings. https://www.brookings.edu/articles/ai-is-coming-to-schools-and-if-were-not-careful-so-will-its-biases/

22 see Pedro, F., Subosa, M., Axel, R., & Valverde, P. (2019). Artificial intelligence in education: challenges and opportunities for sustainable development. United Nations Educational, Scientific and Cultural Organization (UNESCO). https://unesdoc.unesco.org/ark:/48223/pf0000366994

To find out more about the AI Ethics and Governance in Practice Programme please visit:

aiethics.turing.ac.uk

Version 1.2

The Alan Turing Institute