The AI Ethics Playbook

Implementing ethical principles into everyday business

The GSMA AI for Impact initiative

The GSMA is a global organisation unifying the mobile ecosystem to discover, develop and deliver innovation foundational to positive business environments and societal change. Our vision is to unlock the full power of connectivity so that people, industry, and society thrive. Representing mobile operators and organisations across the mobile ecosystem and adjacent industries, the GSMA delivers for its members across three broad pillars: Connectivity for Good, Industry Services and Solutions, and Outreach.

This activity includes advancing policy, tackling today's biggest societal challenges, underpinning the technology and interoperability that make mobile work, and providing the world's largest platform to convene the mobile ecosystem at the MWC and M360 series of events.

For more information, please visit the GSMA corporate website at **www.gsma.com**.

Follow the GSMA on Twitter: **@GSMA**

The AI for Impact (AI4I) initiative, led by the GSMA, is guided by a task force of 21 mobile operators and an advisory panel of 12 UN agencies and partners. It defines the technical, commercial and ecosystem requirements to deliver viable data-driven products and services that adhere to principles of privacy and ethics. At the national level, the GSMA supports real-world implementations, replicating proven models and delivering market-shaping campaigns to unlock demand and address barriers.

For more information, please visit
www.gsma.com/betterfuture/aiforimpact #AI4I

How to use The AI Ethics Playbook

This playbook is intended as a practical tool to help organisations consider how to ethically design, develop and deploy artificial intelligence (AI) systems. It can be read cover to cover, like a report, or you can find and use the appropriate chapter for your role and purpose. It is designed to cater to organisations with varying levels of maturity with regards to AI adoption and familiarity with ethical principles. Some Chapters may, therefore, be more relevant than others.

1   **An overview of common ethical principles**

This chapter will help introduce some of the ethical issues presented by AI to interested readers throughout an organisation. It can be used to develop a basic level of familiarity with these topics.

2   **A proposed organisational structure for dealing with ethical issues**

This chapter is designed to help organisations to think through the governance of a system. It is useful for people who make decisions about the structure and key resources for an organisation.

3   **A Self-Assessment Questionnaire designed to help establish ethical risks**

A series of tools to help you answer and address issues raised in the questionnaire

This chapter contains a Self-Assessment Questionnaire that can help establish the risks presented by a system, and tools to identify and address those risks. It is most relevant to people working directly with AI systems who are aiming to implement ethical principles. These might include product managers or responsible AI champions.

4   **Key themes and recommendations from the report**

This chapter will be useful for anyone hoping to get a sense of the playbook without reading it in detail. It might be used as an educational tool for people who are less directly involved in AI projects or for senior leadership.
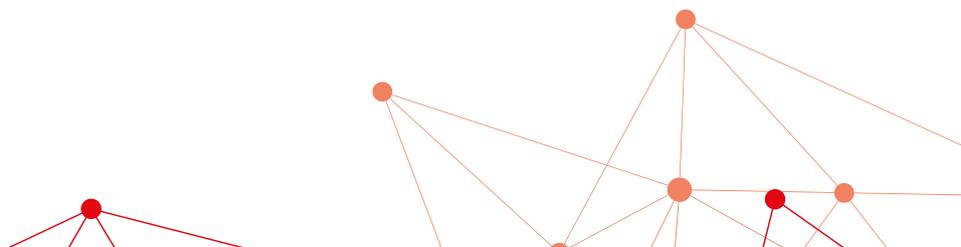
Exercise

Take a few minutes to consider where you are in your ethical AI journey, and what you need to do next, using these step-by-step considerations:

1 Have you established a data governance model?

Notes

2 Have you established AI ethics principles and guidelines?

Notes

3 What is the executive-level understanding of AI ethics? Is change being driven from the top?

Notes

4 Are you connected to external expertise and exchanging best practices?

Notes

5 Have you established an executive committee or advisory panel that drives initial understanding, advises and proposes practices (or policies) for responsible AI?

Notes

6 Do you have a risk management system in place to address risks involved in deploying an AI solution?

Notes

7 How will you ensure awareness of the importance of ethics in developing and deploying AI across the organisation? Have you created an awareness programme on ethical AI?

Notes

8 Do AI product managers understand AI ethical principles, how these translate to policy, and the implications for product design, development and deployment?

Notes

9 Have you empowered AI product managers to act ethically, and provided training and technical tools?

Notes

10 Have you created ethics-focused roles to support AI product managers in developing and deploying ethical AI?

Notes

11 Is there a change management process in place if the adoption of ethical principles represents a departure from business-as-usual?

Notes

12 Are you studying concrete examples and monitoring existing ethics risks on AI systems, as the scale and scope of AI services increase?

Notes

Introduction

AI is a powerful, emerging force that is transforming business and society. The potential of these technologies to unlock benefits for organisations and society is only beginning to be seen[1].

AI can help organisations to improve prediction, optimise operations, allocate resources more efficiently, and personalise digital solutions. PwC estimates AI could contribute $15.7 trillion to the global economy by 2030[2]. However, AI isn't a futuristic technology; it is present in our everyday lives, used across a wide variety of industries.

The mobile industry is no different. AI is at the core of operational and business models for an increasing number of mobile network operators (MNOs). Three common uses are:

As the adoption of AI accelerates, organisations and governments around the world are considering how best to harness this technology for the benefit of people and the planet. AI has the potential to truly change the world, and this represents not only an opportunity but also a risk.

AI depends on large amounts of data, often relating to individuals, and makes inferences based on this data. These inferences may be used to guide decisions that have a significant impact on the things people care about the most – their health, their employment and their access to resources. It is, therefore, vital AI is used in a way that protects our fundamental human rights.

Core business optimisation
MNOs are using AI to improve network optimisation, real-time network monitoring, predictive maintenance and network security through improved efficiency.

Personalised customer interaction
MNOs are improving communication with customers through robotic process automation, virtual assistance, intelligent pricing and B2B sales optimisation.

AI-driven mobile data products for external stakeholders
MNOs are using their data to provide services to third parties, such as predicting customers' media content preferences, providing location-based marketing insights and assisting with new approaches to credit scoring.

"
AI can be a force for good that helps to solve critical challenges. However, it may also impact our fundamental human rights. This creates a responsibility for companies that create and use these technologies to act ethically. The mobile industry is committed to connecting everyone to a better future and is well placed to contribute to this evolving field.

Mats Granryd, Director General, GSMA
"

1   Richard Benjamins: A data-driven company: 21 lessons for large organizations to create value from AI
2   https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html

AI is increasingly an essential element of the infrastructure on which our society is built, playing an active role in financial markets, basic services and international supply chains. We need to be able to trust AI to behave in the way we want it to; and when it doesn't, we need to understand what went wrong and who should be accountable for this.

Ethical AI isn't just the right thing to do; it can also positively impact companies' bottom lines. Organisations that pro-actively choose to act ethically – and tell their stakeholders they are doing so – can generate goodwill, build positive relationships, and increase market share[3]. Research suggests bias can lead to lost revenue[4]. The ethical principles discussed in this playbook can help to ensure AI systems are reliable, reproducible and explainable, which will ultimately increase their value for a company as well as ensure a more positive impact on society.

A huge amount of effort has gone into considering what ethical AI looks like, resulting in many organisations drawing up ethical AI principles. However, putting such principles into practice remains a challenge. With this in mind, a group of MNOs worked with the GSMA to create this practical playbook to support the operationalisation of ethical AI principles into everyday activities.

This playbook is intended as a practical tool for not only the mobile industry but also any other organisation currently grappling with the challenge of designing, developing or deploying AI in an ethical manner. It is intended to be a flexible tool organisations can adapt to their needs.

This is a rapidly evolving space; AI is progressing quickly, and many governments are in the process of drafting regulations relating to AI. This playbook is envisaged as a living document, and we welcome enquiries from organisations that want to help advance it further.

"

Digital leadership is a cornerstone of our T25 strategy, and we're committed to the ethical use of Artificial Intelligence in our operations and in our customer interactions. Our collaboration with the GSMA and other top global mobile operators allows us to work together to protect customers and employees, remove any entrenched inequality and ensure that one of our most advanced technologies operates reliably and fairly for all our stakeholders.

Nikos Katinakis, Group Head, Networks & IT, Telstra Group Executive Networks & IT

"

3   https://www.nist.gov/system/files/documents/2021/09/17/ai-rmf-rfi-0102.pdf
4   https://edpl.lexxion.eu/data/article/17685/pdf/edpl_2021_03-006.pdf

1

AI Ethical Principles

To act ethically, organisations require a guiding framework that explains what good ethical behaviour looks like. As AI has developed, and the potential risks of AI have become clearer, many organisations have written AI ethical principles for this purpose. This chapter explores some of these principles.

? ## How to use this chapter

- This chapter will help you understand common ethical principles.

- It summarises recommendations from the mobile industry, the European Commission and other international organisations.

- Organisations should formulate and select their principles based on organisational norms and values[5], as this is not intended as an exhaustive list.

[5]   Please see Appendix 4 for examples of AI principles

AI Ethical Principles

Fairness

For an AI system to be fair, it must not discriminate against people or groups in a way that leads to adverse decisions or inferences. Non-discrimination and equality are central features of the major human rights treaties and many countries' laws.

AI is often trained on data that records an unfair world. Care must be given to ensure AI systems do not reinforce human prejudices and inequalities. As far as

possible, organisations should gather and use accurate, complete and diverse datasets to avoid reproducing historical patterns of discrimination. Positive attempts should be made to identify and remedy unintended bias in the data, model design and implementation of the AI system. Consideration should be given to whether data may unintentionally provide a proxy for protected characteristics and discriminate on this basis; otherwise, an AI system may not conform to internationally recognised human-rights frameworks.

Involving diverse stakeholders in the entire development process helps organisations to have a greater awareness and understanding of bias. AI systems should be designed to be as accessible as possible, and the benefits of AI should be available to everyone.

> AI has the power to support our goals around sustainability, diversity and inclusion. However, as a society, we need to be careful not to transfer our human biases to machines. Companies have a responsibility to build AI models that take into consideration diverse groups and go through thorough user testing before being released.
>
> Vinod Kumar, CEO, Vodafone Business

Example

Orange has been working to ensure its AI systems are free from bias wherever possible. The company recently audited its supply chain, signed the International Charter for Inclusive AI[6] and was the first company to receive a GEEIS-AI label[7].

6   https://www.orange.com/en/newsroom/news/2020/1st-international-charter-inclusive-ai-orange-committed
7   https://www.orange.com/en/newsroom/press-releases/2020/orange-confirms-its-commitment-inclusive-artificial-intelligence

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

Human agency and oversight

It's important to determine an appropriate level of human oversight and control of an AI system. As AI directs decision-making, people may become reliant on a system. Organisations must respect human autonomy.

Organisations using AI have three basic choices about the degree to which humans remain in control of the outcomes of the AI system (regardless of whether the system continues to learn autonomously)[8], namely Human-in-the-Loop, Human-on-the-Loop and Human-out-of-the-Loop.

Special consideration needs to be given to high-risk use cases. In many cases, the higher the risk, the higher the need for human oversight. Human oversight will help to establish accountability and provide a safety net to catch unforeseen issues.

Example

STC developed an AI solution for field surveyors. By using an app embedded with a pipeline of computer vision models, a surveyor can take a photo of objects such as buildings, power meters etc and automatically populate a description of the photo. The app design follows a 'human-in-the-loop (HITL)' approach as the surveyor can validate the AI output by accepting the resulting description if it produces a matching result or rejecting or editing if it doesn't.

Privacy and security

AI systems should respect and uphold an individual's right to privacy and ensure personal data is protected and secure. Organisations using AI should pay special attention to any additional privacy and security risks arising from AI systems.

While data protection and privacy are a matter of legal compliance in most countries, it's important to consider the risk of harm to individuals more broadly.

Example

Orange is participating in the EU PAPAYA project[9], which aims to address privacy concerns when data analytics tasks are performed by untrusted third-party data processors. By designing cryptographic modules adapted to the use case, PAPAYA develops dedicated privacy-preserving data analytics solutions that enable data owners to extract valuable information from encrypted data while working with third parties.

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

8   https://link.springer.com/article/10.1007/s43681-020-00012-5
9   https://www.papaya-project.eu/about

Safety and robustness

AI systems should be safe, robust and reliably operated in accordance with their intended purpose throughout the lifecycle.

In some cases, small variations to AI input – the data used to make decisions – may lead a system to make an incorrect inference. For instance, a stop sign with graffiti on it might be incorrectly interpreted by an autonomous vehicle. This has implications for the safety of a system, affecting how it acts when faced with real-world data. Organisations should ensure they are regularly testing, validating, verifying and monitoring a system so it can deal with variations in input. Rigorous processes and planning can help to ensure AI remains safe and organisations are prepared if something goes wrong.

Transparency and explainability

It's important to be transparent about when an AI system is being used, what kind of data it uses, and its purpose. Explainability is the principle of communicating the reasoning behind a decision in a way that is understandable to a range of people, as it is not always clear how an AI system has arrived at a conclusion.

As AI systems are complex, it's worth considering who needs to understand them and how to explain information in a way that makes sense. People who are particularly impacted by a system should be able to understand how and why this is the case.

Example

Telefónica is committed to telling its customers what type of data is used for AI systems, how it is used and when they are interacting with an AI system. Telefónica ensures decisions made by AI systems are understandable; first by ensuring developers understand the logic behind the conclusions AI draws internally, and second by measures to ensure the appropriate stakeholders have the necessary level of understanding. This is further applied to third-party technologies Telefónica uses[10].

> stc is focused on the introduction of innovative AI products and solutions to produce correct, precise and reliable results insights for our customers. We are delighted to collaborate with the GSMA and the other leading operators in developing "International Ethical AI Principles" for use in the telecom industry, and correspondingly embedding into stc's culture. This demonstrates stc's practice of complying to Respect of Privacy, Transparency and Openness, and contributing to international standards.
>
> By adhering to that, we believe it will not only enhance our customers experience and trust, but will also contribute to resource optimisation, improving the efficiency of our products development and deployment, and supporting data/model re-usability. This is an exciting milestone in stc's digital transformation journey, stc will continue to endeavour to always lead the market towards digital transformation and the localisation of AI

Eng. Haithem AlFaraj, Chief Technology Officer, Saudi Telecommunications Company (stc)

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

Accountability

Organisations should have a governance structure that makes it clear who is responsible for reporting and decision-making and is thereby ultimately accountable, throughout an AI lifecycle. Accountability is key to complying with regulatory and legal requirements.

AI systems, particularly those that use machine learning (ML) algorithms, may produce unpredictable outcomes. Their development and deployment commonly involve teams of people. This can make it unclear who or what is responsible for the outcomes of the system. Therefore, it is important, on a practical level, that there is accountability within an organisation based on strong governance structures and rigorous recording of information. This accountability will help organisations to demonstrate compliance with data protection and other regulatory regimes.

It is worth considering how accountability is shared with other actors in the value chain and what the expectations are for them to adhere to ethical principles.

Example

Telstra suggests accountability requires identifying who is accountable at different levels of the organisation for:

- The actions of an AI system
- Implementing the system components correctly
- Setting and balancing the system's objectives.

This applies to AI developed in-house and AI purchased from third parties. When Telstra purchases third-party systems, it remains responsible for its performance. Telstra takes steps to ensure these purchased AI technologies are working in line with its ethical principles.[11]

Environmental impact

AI systems must be designed, developed and deployed in a way that is mindful of environmental impact throughout their lifecycle and value chain. With careful consideration of systemic consequences, AI can help to secure a sustainable future for all.

Example

'Human, social and environmental wellbeing' is Principle 1 of Telstra's AI ethics principles. When analysing potential impacts of an AI system, Telstra looks beyond immediate impacts to broader long-term or indirect outcomes to understand the social and environmental impact of an AI system.[12]

"

Whatever the technological tool, no matter how sophisticated it may be, our first question should always be; "How are we improving people's lives with our services?". I often refer to AI as "augmented intelligence" – it should not replace but complement human intelligence and improve our decision-making capabilities in solving the pressing problems of our planet and humanity.

Kaan Terzioglu, CEO, VEON Group

"

11   https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles
12   https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework/testing-the-ai-ethics-principles/ai-ethics-case-study-telstra

2

Governance models for ethical AI

Applying ethical principles within an organisation requires a strong governance structure and may involve new roles and responsibilities. This chapter discusses how organisational design supports ethical behaviour in everyday activities.

"

At Orange, we are convinced that AI ethics is not negotiable, it is the foundation of our AI strategy. We are now organising ourselves with the support of our group Data and AI Ethics council and per country local AI ethics referent to adapt methodologies and tools.

Steve Jarrett, Data and AI,
Senior Vice President, Orange Innovation

"

? How to use this chapter

- This chapter is designed to help you consider which organisational structure you need to put in place to design, develop and deploy AI systems responsibly.

- Explore one approach to governance, chosen for its flexibility, which illustrates how organisational structure can reduce bottlenecks, provide accountability, and escalate issues when needed.

- Learn how mobile operators are embedding an ethical approach to AI into everyday activities.
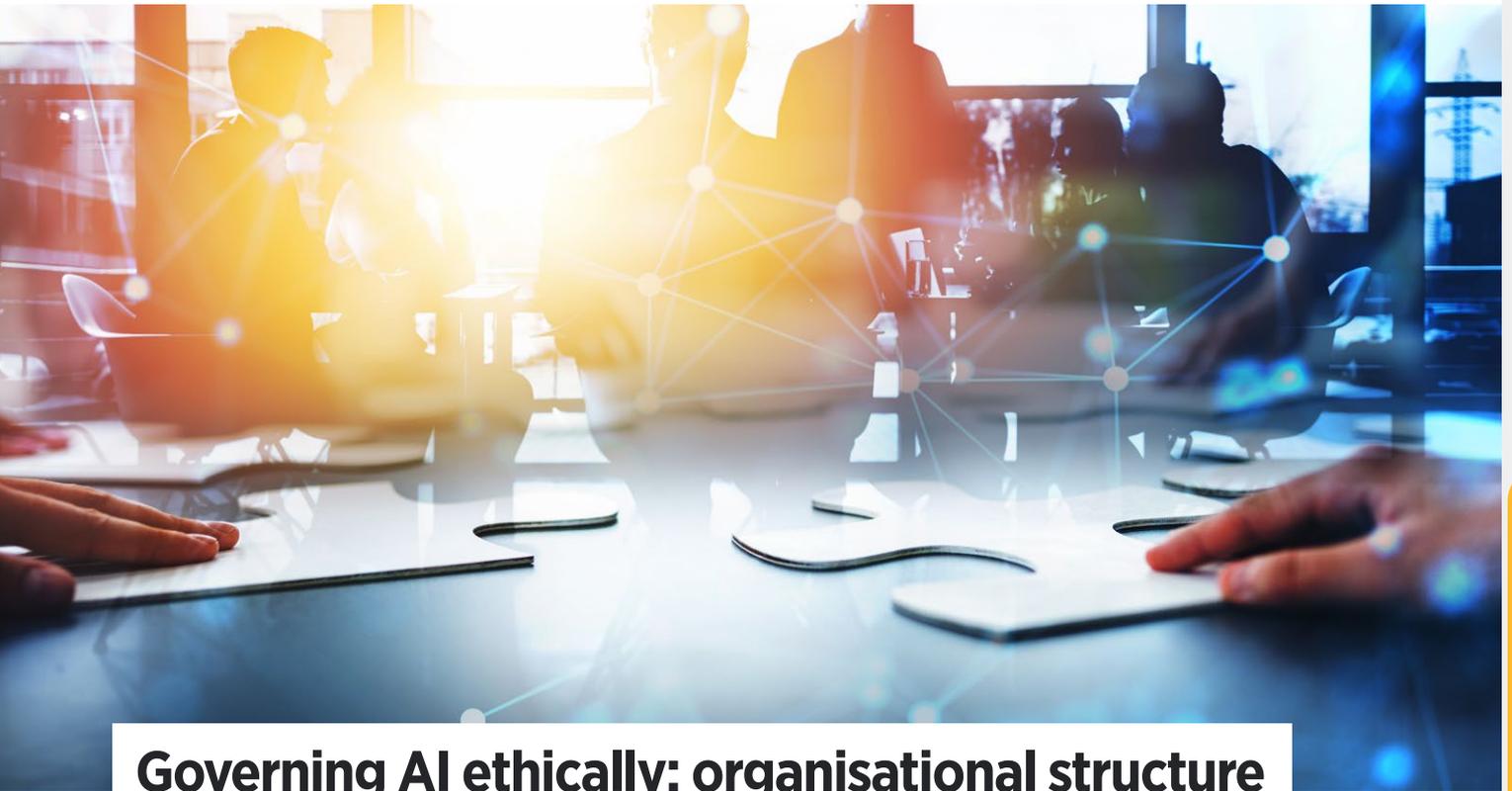
Governing AI ethically: organisational structure

What is AI governance?

AI governance refers to the idea that systems using algorithms to draw inferences and make decisions require policies governing their design, development, and deployment. These policies should be based on ethical principles like those in Chapter 1. Creating policies and ensuring they are followed throughout an organisation requires a robust governance structure with clear roles and responsibilities.

Creating an appropriate structure

An effective AI governance structure needs to consider organisational culture, size, strategy, and AI maturity.

When deciding on an AI governance structure it is important to involve a diverse range of colleagues from within an organisation. A lot of the key skills are likely to already exist within an organisation. For example:

- Appropriate C-suite executives should be involved as project sponsors, as they will often have ultimate responsibility for the AI product development lifecycle

- Legal and regulatory teams are well versed in ensuring compliance
- Product teams understand how systems operate from a technical perspective
- Corporate social responsibility (CSR) professionals are accustomed to considering the societal implications of business activity.

Involving diverse, multidisciplinary stakeholders in the process of creating a governance structure will help to bring a diversity of perspectives.

Existing organisational processes can be leveraged to help implement ethical AI policies. For example, most companies that collect and handle data have established governance systems to ensure best practice and compliance with data protection legislation. It makes sense to build on these when deciding how to best govern AI.

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

Governance model example: designed for escalation of ethical issues

This section explores one approach to creating a governance model based on the experience of mobile operators. This model, chosen for its flexibility, involves a structure of decision-making and reporting that can help to embed accountability for AI systems into organisational culture and improve agility by removing bottlenecks.

As illustrated by the step-by-step considerations at the beginning of the playbook, this governance model will require investment to be successful. This includes strengthening AI ethics awareness across the organisation; training AI product managers; allocating time to document, assess and audit processes; creating new roles or responsibilities; purchasing or creating technical and other tools; and fostering collaboration across departments. Introducing new ethical principles into an organisation requires a change management programme for them to become business-as-usual.

The proposed governance structure has three levels. These levels indicate the process for escalating higher risk or more complex use-cases.

Figure 1   Organisational model designed to provide escalation for ethical issues

LEVEL 3

Executive board

Who is involved: senior leadership, representatives of ethics committee

What is involved: go/no-go decisions about high-risk use-cases

A small number of cases that pose a serious risk are escalated to this level

3

LEVEL 2

Ethics committee

Who is involved: an ethics committee of diverse experts

What is involved: exploration of issues flagged at Level 1, debate how to move forward

More complex or higher risk cases are escalated to this level

2

LEVEL 1

Responsible AI teams

Who is involved: AI product managers, responsible AI champion

What is involved: Self-Assessment Questionnaire and accompanying tools

Most cases are dealt with at this level through assessments, tools and debate

1

Level 1: responsible AI teams

Who is involved: AI product managers, a responsible AI champion

What is involved: the Self-Assessment Questionnaire and accompanying tools

At Level 1 the people who are closest to an AI system, often the AI product managers[13], use a Self-Assessment Questionnaire and accompanying tools (please see Chapter 3) to work out the ethical implications of a particular use-case in line with the principles discussed in Chapter 1.

Many companies create a new role or team dedicated to considering the ethical implications of AI systems,

anticipating unintended consequences, and ensuring accountability for AI-related decisions and actions. Throughout the rest of this document, this role is referred to as the responsible AI (RAI) champion, although alternative names might include AI ethics expert or AI ethics researcher. The potential scope of this role is explored later.

In some instances, the AI product managers and RAI champion may find that AI systems carry risks, or inherent limitations, which mean they may not be appropriate to deploy in some contexts. At this point, the RAI champion will escalate their concerns to an expert board, as discussed later.

Notes
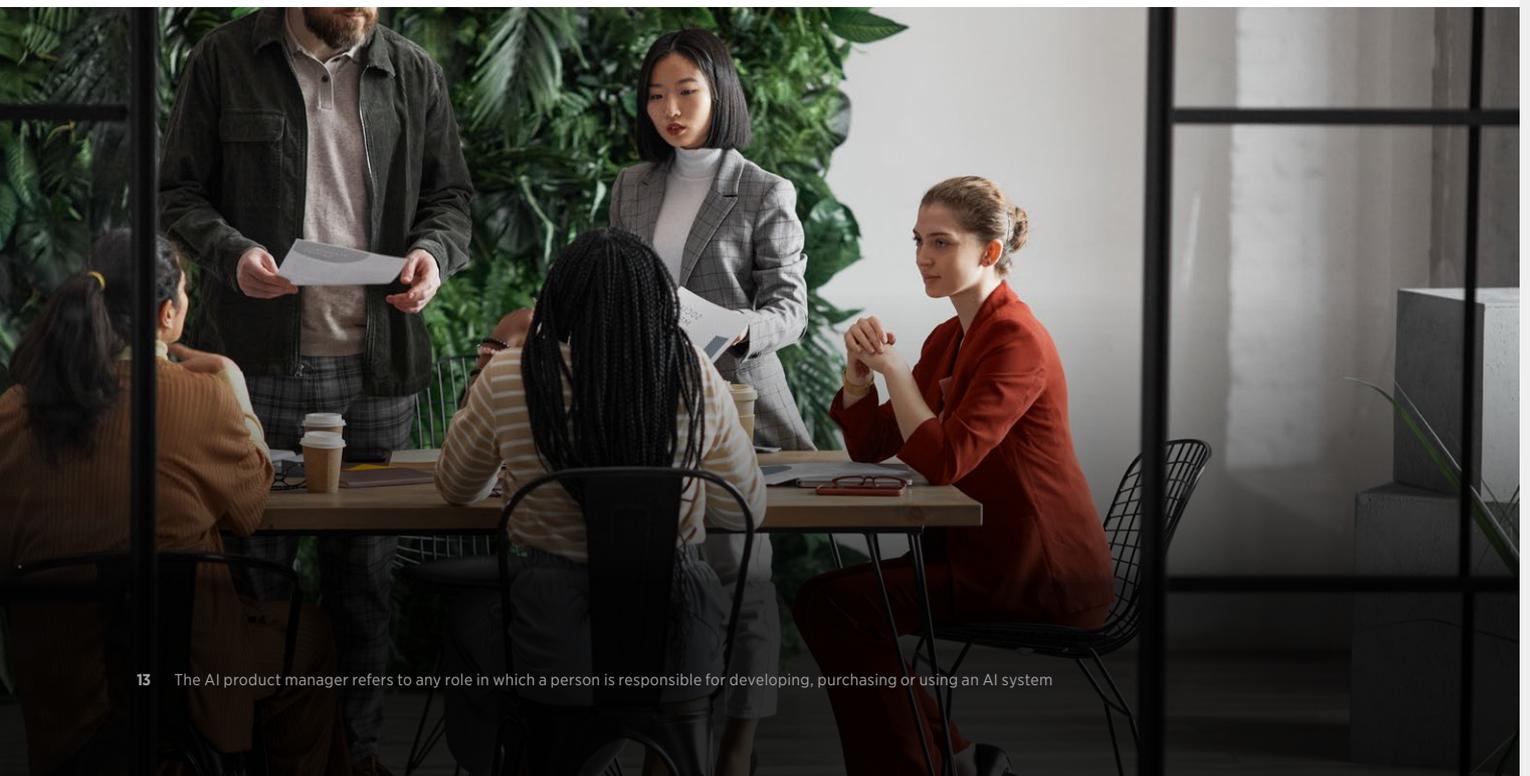
Responsible AI champions

What does an RAI champion do?
An RAI champion brings ethical, social and political perspectives to the design, development and deployment of AI systems[14]. Fulfilling these responsibilities requires a broad range of expertise, including technical and regulatory knowledge, business acumen and industry experience, good communication skills and the ability to work across organisational boundaries. They should anticipate future ways and contexts in which an AI system might be used to identify potential risks. Depending on the organisation, there might be one champion or a team. These champions may be assigned to specific geographic regions or business areas.

Organisations take a different approach to where they place this role. In some instances, the role is within the business unit responsible for data products, in others within the ethics, compliance or legal department. Where you place this person should be guided by your company and AI strategy. For instance, you may have a wider 'responsibility' strategy that means they need close links to your sustainability team or, if you are developing AI systems, you may want to place them within the product development team.

As your use of AI develops, and your governance approach matures, this role may expand or move within the organisation.

The responsibilities of a Telefónica RAI champion[15] include:

- Inform fellow employees about the importance of applying ethics to AI and data to avoid unintended harm and raise awareness of the organisation's AI principles.
- Provide and organise training to the corresponding business unit or region covering the governance model and ways to apply the principles to the product lifecycle. Encourages self-educated experts to form a voluntary community of practice where 'local' employees can get first-hand advice.
- Escalate an issue to a multi-disciplinary group of senior experts if neither the experts of the community of practice nor the RAI champion can address it.

- Coordinate across departments in the organisation. Coordination is needed with the data protection officer (DPO) for privacy-related issues; with the chief information and security officer (CISO) for security-related aspects; with the chief data officer (CDO) for data and AI-related topics; with the CSR officer for reputational and sustainability issues; with the regulatory team for potential future AI regulations; and with the legal team for other legal issues.
- Connect relevant people to form communities of experts on the subject matter. Communities of experts are the first place to go if ethical doubts cannot be solved within a product or project team. RAI champions also form a community among themselves connecting different geographies and business units of the organisation in an active learning and sharing network.
- Drive change and ensure that ethical considerations become an integral and natural part of any business activity touching AI and Big Data, including design, development, procurement, and sales. RAI champions must implement and turn the governance model into business-as-usual.
- More mature organisations also may consider setting up or joining an external RAI champion (or similar) network where experiences and practices are shared with other organisations, either from the same sector or across different sectors.

The future of RAI champions
Many companies, including mobile operators, are leading on developing these new roles within their businesses. However, this is a growing and evolving field, as evidenced by the World Economic Forum's development of a Responsible Use of Technology initiative. This brings together ethical leads from technology companies to learn from one another. Best practice is still being established, by employing an RAI champion organisations can be a part of this cutting-edge conversation.

---

14   https://www2.deloitte.com/content/dam/Deloitte/us/Documents/process-and-operations/ai-institute-aiethicist.pdf
15   https://business.blogthinkbig.com/a-new-organizational-role-for-artificial-intelligence-the-responsible-ai-champion

Industry spotlight

Orange

In March 2021, Orange announced the creation of a Data and AI Ethics Council made up of 11 independent experts and chaired by Stéphane Richard, Orange's Chairman and Chief Executive Officer. Selection was based on member's independence and neutrality, their expertise on the social impact of AI, and the diversity of their backgrounds: ethics specialists, lawyers, philosophers, researchers, professionals from public bodies, and private sector leaders.

This advisory and independent body supports the company's implementation of ethical principles governing the use of data and Artificial Intelligence with the support of a Toolbox consisting of: An executive interview guide; The appointment of a delegate from the general management within the AI Committee; A mission statement and role of the AI Ethics Committee; The job description of an in-house AI Champion; Activity reports of the AI Committee; and A monitoring dashboard of the AI Committee recommendations.

Orange appoints a Data and AI Ethics Officer in each country where it has a presence to manage local ethics assessments and training[16]. These Officers drive change management and report to the group Council.

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

Level 2: ethics committee

Who is involved: an ethics committee of diverse/cross-organisational experts
What is involved: exploration of issues flagged at Level 1, debate about how to move forward

The ethical impacts of AI can be complex and difficult to anticipate and, as with any ethical question, there may be differing viewpoints on the right way forward. For this reason, some issues may require diverse expertise. Level 2, therefore, involves setting up an ethics committee of experts drawn from different disciplines and geographies to explore issues raised by AI product managers and RAI champions.

This ethics committee should include engineers who understand the way AI systems work and can explain this to the rest of the committee, trained ethicists, sociologists and engineers, as well as legal and human rights experts.

Key questions to consider when forming this committee are whether the selection process will lead to a wide variety of viewpoints, how transparent you will be about who has been elected, and why; and what level of decision-making power the committee will be given, and the scope of their role. Consider including representatives of the population that might be affected by an AI system. Finally, there may be a need to communicate the committee's decisions with the support of engagement and advocacy experts.

While the responsibilities of the committee will differ between organisations, they might include:

- Studying concrete scenarios involving the ethical use of data and AI upon request
- Issuing prospective opinions on the use of data and AI technologies
- Drawing up ethics guidelines for the responsible use of data and AI
- Creating recommendations and policies.

For instance, Microsoft Aether[17] formed a committee based on a 'think tank' model, responding to issues that arise with studies, reflection and recommendations. IBM created an AI Ethics Board[18] that supports a centralised governance, review and decision-making process for technology ethics policies, practices, communications, research, products and services.

🔍 Industry spotlight: Microsoft

Microsoft[19] created an Office of Responsible AI (ORA) responsible for putting principles into practice by setting company-wide rules and enacting responsible AI by:

- Defining roles and responsibilities for teams involved
- Enabling teams to adopt responsive AI practices, both within the company and among customers and partners
- Reviewing sensitive use-cases to help ensure Microsoft AI principles are upheld in the development and deployment of AI work
- Shaping new laws, norms, and standards that will be needed to ensure the promise of AI technology is realised for the benefit of society at large.

In addition to ORA, Microsoft also created Responsible AI Strategy in Engineering (RAISE), an initiative accompanied by an engineering team dedicated to enabling and supporting the implementation of responsible AI rules and processes across its engineering groups. RAISE defines and executes the tooling and system strategy for responsible AI across engineering teams and implements compliance tooling for engineering teams to monitor and enforce Microsoft's responsible AI rules and requirements. It operates feedback mechanisms, identifying blockers to integrate responsible AI requirements into everyday development processes.

17 https://www.microsoft.com/en-us/ai/our-approach
18 https://ibmorg-public.s3.us-east.cloud-object-storage.appdomain.cloud/responsibility/cr/pdfs/IBM-2020-CRR.pdf
19 https://www.microsoft.com/en-us/ai/our-approach?activetab=pivot1:primaryr5

— Industry spotlight

— Telstra

In commitment to developing and using AI systems in line with their ethical principles, Telstra has implemented a Risk Council for AI and Data (RCAID). This body is a key enabler of responsible AI at Telstra. The RCAID considers the broader human, societal and environmental impacts of these systems and the decisions they make, along with reviews to check that they accord with the law in all the jurisdictions in which they operate.

The RCAID is responsible for performing thorough risk assessments of the data and AI use-cases brought before it, covering all stages from opportunity through to deployment. Teams go through a robust process covering both internally developed and externally sourced AI and machine-learning solutions. For use-cases carrying higher risk, a further review is conducted by senior executives from Telstra's Data and AI Council.

The framework also ensures that measures to confirm whether a system is performing reliably are planned in the early stages, including AI-bias tooling to monitor the reliability and robustness of the system.

Level 3: executive board

Who is involved: senior leadership, representatives from the ethics committee
What is involved: go/no-go decision-making about high-risk use-cases

The ethics committee may decide the use of an AI system in a specific context presents unacceptable risks, or an AI system has inherent limitations that make it inappropriate for the intended purpose.

The committee will then escalate this to a separate body with executive oversight. Level 3, therefore, involves setting up an executive board that bears ultimate responsibility for deciding whether the development and deployment of an AI system should proceed. Escalation to this level would only happen on the rare occasions where an issue presents a high-level risk to people, society or the environment and, therefore, the reputation of the company.

Notes

Industry spotlight

Telefónica

Telefónica, as part of its Responsible AI by Design methodology, has introduced a three-step escalation process so that everyone understands their level of responsibility and how to manage situations that cannot be resolved individually.

Level 1: During the design phase, product managers receive training on the Telefónica ethical principles and carry out a simple self-assessment of the AI system using an online questionnaire. This self-assessment helps managers to explore potential ethical risks.

If issues arise during the self-assessment, the product team can raise concerns with an RAI champion, a colleague trained on the ethical implications of AI. If the RAI champion is also concerned, and issues are not resolved, then they will escalate the issue to Level 2.

Level 2: The product managers escalate their concerns to a panel of multi-disciplinary experts able to explore ethical implications. If they determine that there is an unacceptable risk, the issue is flagged with the Responsible Business Office.

Level 3: Controversial issues are raised with the Responsible Business Office, an executive-level group of relevant global department directors. These directors assume ultimate responsibility for decisions about risks[20].

"

Telefónica's commitment to ethics and integrity is key in our transformation process"

José María Álvarez-Pallete, Chairman and CEO, Telefónica

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

20   https://www.telefonica.com/en/wp-content/uploads/sites/5/2021/08/ia-responsible-governance.pdf
https://www.telefonica.com/en/commitment/how-we-work/business%20-principles/

3

Operating processes and implementation tools

Implementing AI ethical principles into everyday activity represents a shift from business-as-usual for most organisations. To support this change, Chapter 3 contains tools, information on training and awareness-raising activities, and a brief overview of regulations and compliance.

? How to use this chapter

- Begin by reading a short explanation of the purpose of the Self-Assessment Questionnaire and an overview of how it should be used.

- Many ethical questions require debate and discussion. The section on educational tools also contains resources to help guide these discussions.

- Read through the Self-Assessment Questionnaire in tandem with this chapter – the tools listed in this chapter can help you answer the questionnaire and address some of the issues that may arise.

- This chapter presents tools arranged around each principle that can be used as a basis to consider the need for training or awareness-raising on ethical AI within your organisation. For a list of tools related to ethical principles, with links and descriptions, please check Appendix 1.

- The chapter concludes with a brief overview of compliance and regulatory considerations and lists some related standards and certificates.

The Self-Assessment Questionnaire

The purpose of a Self-Assessment Questionnaire
Bridging the gap between ethical principles and business practice begins with understanding the risks involved in an AI use-case. The GSMA, together with a group of mobile operators, have developed a self-assessment questionnaire that helps establish the risks an AI system may pose and work through steps to mitigate them.

The questionnaire contains a series of baseline questions to determine whether the AI system is being designed, developed and deployed in accordance with each of the AI principles discussed in Chapter 1. Depending on your answers, the questionnaire may propose further actions, ranging from tools used to help answer some of the questions and mitigate potential problems, to more fundamental decisions about how an AI system should be governed. Finally, the questionnaire has an option to record status, evidence and any other notes to support prioritisation and planning. Any AI system, whether developed in-house or by a third party, can be assessed using this questionnaire.

For more information and access to the Self-Assessment Questionnaire, please contact our team: aiforimpact@gsma.com

? Using the Self-Assessment Questionnaire

1 Begin by reading through the Self-Assessment Questionnaire to get a sense of its scope.

2 Given there is no 'one-size-fits-all' approach, the questionnaire should be tailored to address local regulations and legal responsibilities in the location.

3 Cross-departmental collaboration should be sought when customising the questionnaire to capture input from across the organisation. Consider how other organisations have constructed questionnaires[21].

4 Based on your approach to governance, decide who should be responsible for completing the questionnaire. While one person might have ultimate responsibility for the questionnaire it is best answered by a group.

5 Decide who needs to learn the outcome of the questionnaire and how you will communicate with them based on your governance model. Any concerns raised by the questionnaire should be escalated.

6 It's advisable to log answered questionnaires for potential future auditing, reporting and planning of required future actions. The Self-Assessment Questionnaire allows you to also (optionally) save notes, record evidence, outline identified gaps and allocate priorities.

7 Complete, review, and update the questionnaire periodically. It should be used at three stages of the AI system lifecycle at a minimum: during design, development and deployment. Post-deployment reviews are advised; the frequency will depend on the risk level of the AI system use-case.

21   See Appendix 6 for questionnaire examples

Technical tools

The Self-Assessment Questionnaire asks some probing questions to help explore whether an AI system is compatible with the principles examined in Chapter 1. This section describes technical tools aligned with the principles that can help answer these questions by highlighting problem areas, visualising patterns or analysing processes to help interrogate AI systems. This chapter simply gives an overview of the type of tools available; see the Appendix for a list of specific tools with links and descriptions. There is a wealth of ingenuity here, and we strongly recommend examining the full list in the Appendix to discover some novel approaches.

Fairness

Fairness is a complex topic. Creating a tool that measures fairness requires a decision on what fairness means, which can be divisive and nuanced[22]. An appropriate definition needs to be carefully selected or shaped according to the context in which an AI system is used. The tools and references in the Appendix can help you to create a suitable definition of fairness, and to explore the outputs of an AI system with the aim of discovering and reducing bias and improve fairness[23].

Privacy and security

AI systems should respect and uphold an individual's right to privacy and ensure that personal data is protected and secure.

In this context, security specifically refers to types of attacks that don't apply in other areas of computing. For example, extraction attacks aim to copy the model's functionality or recover internal state information, such as trained weights (this is sometimes called inversion); evasion attacks cause the model to misclassify an input; membership inference attacks recover data on which the model was trained; and poisoning attacks involve deliberate contamination of training data[24, 25]. However, as well as exploits from malicious hackers, human error poses a significant threat to privacy. Steps should be taken to protect data even for internal use and to minimise access to reduce the risk of accidental leaks.

Privacy and security issues are usually subject to their own regulatory or contractual obligations. The tools listed in the Appendix should be used alongside existing compliance processes.

Safety and robustness

Tools aligned with the Safety and Robustness principle are designed to analyse and avoid errors when faced with differing inputs. Following strong design principles will offer some protection, as will having a good understanding of various types of error or confusion in the outputs versus the inputs.

Transparency and explainability

There are several commonly used algorithms for quantifying explainability, including SHAP, LIME and Grad-CAM; these can be implemented at low levels using open-source software packages, or via higher-level toolkits that might include visualisations or interactivity.

Environmental impact

AI can help to drive down energy use and improve efficiency, but it also requires significant computing power. As awareness of the environmental impact of AI systems increases, several tools have appeared that help to measure the energy consumption and carbon footprint of computing infrastructure.

22 https://pair-code.github.io/what-if-tool/ai-fairness.html
23 Lee, M.S.A., Floridi, L. "Algorithmic Fairness in Mortgage Lending: from Absolute Conditions to Relational Trade-offs". Minds & Machines 31, 165–191 (2021). https://doi.org/10.1007/s11023-020-09529-4
24 M. Xue, C. Yuan, H. Wu, Y. Zhang and W. Liu, "Machine Learning Security: Threats, Countermeasures, and Evaluations" in IEEE Access, vol. 8, pp. 74720-74742, 2020, doi: 10.1109/ACCESS.2020.2987435.
25 Qiu S, Liu Q, Zhou S, Wu C. Review of Artificial Intelligence Adversarial Attack and Defense Technologies. Applied Sciences. 2019; 9(5):909. https://doi.org/10.3390/app9050909

Training and awareness

Introducing new ethical principles into an organisation has the potential to be disruptive. People need to change their approach to AI and big data tools, challenge their assumptions and follow new processes, and new roles or responsibilities might be introduced as discussed in Chapter 2. A change management programme is required to allow this transformation to happen without disruption.

As with any change process, this experience can be challenging, but it also presents opportunities. Ethical judgements require people to engage in productive debate, a key element of innovation, and consider the tools they work with from a different perspective. (See Appendix for a list of resources designed to help educate and engage organisations about different methods of approaching ethical conversations.)

As an emerging technology, training is often required on what AI is, how it works and why it needs to be managed ethically. Raising awareness of the environmental and societal impact and potential risks of AI is central to creating an organisational culture in which people uphold AI ethical principles. Further, practical training on how to operationalise AI ethical principles in everyday operations is needed. This training could take the form of a course accessible to all employees, a set of workshops, a digital guide or a series of talks from external experts.

An important consideration is how to raise awareness within the wider ecosystem. For instance, Deep Mind's Ethics and Society team plays a dual function: helping technologists put ethics into practice, but also helping society to anticipate and direct the impact of AI so it works for the benefit of all[26].

There are many resources to help organisations develop the skills and mindsets needed to deal with AI ethically. For a rigorous example, the OECD publication[27] *Tools for trustworthy AI: A framework to compare implementation tools for trustworthy AI systems* lists and assesses technical, procedural and educational tools to help implement trustworthy AI, alongside a framework to compare different tools [28].

🔍 Industry spotlight: Telefónica

Telefónica developed an online training course on Ethics and Artificial Intelligence for all employees, available in three languages (Spanish, English and Portuguese). It explains why artificial intelligence, if used carelessly, can have a negative impact on society and the environment, and what steps can be taken in practice to mitigate that impact.

More than 3,500 employees completed the course in 2021, starting with people with the most relevant roles. The objective is to train more employees as the technology is being democratised.

The course has six modules: Introduction, AI Principles, Implementing AI Principles, Tools, AI Governance, and Practical Examples. Depending on the employee's profile, they can either access a light version, made up of three 10-minute modules, or a longer version that takes around an hour to complete. It also created a quick-reference mini-guide so employees can get a quick overview of how to apply the AI principles for ensuring responsible use of AI.

Each business unit actively engaged in the use of AI participates in dedicated workshops that take a deep dive into the governance model[29].

26   https://deepmind.com/about/ethics-and-society
27   https://www.oecd-ilibrary.org/docserver/008232ec-en.pdf?expires=1638260812&id=id&accname=guest&checksum=1A5B53597044CBEF450036B92D771FE0
28   See the Appendix for a list of procedural and educational tools
29   https://www.telefonica.com/es/wp-content/uploads/sites/4/2021/06/ia-responsible-governance.pdf

Regulations and compliance; standards and certification

Regulations and compliance

As AI policies are evolving and regulations are being drafted, it is vital that organisations monitor the regulatory space in which they operate. Concrete guidance is not possible while regulations are still in flux. However, it is worth considering the role of risk in determining the level of compliance with regulations. In general, higher risk uses of AI can expect to be more heavily regulated.

For instance, a legislative initiative has been tabled by the European Commission that proposes harmonised rules on AI across the EU. The regulation follows a risk-based approach, differentiating between uses of AI that create (i) an unacceptable risk, (ii) a high risk, and (iii) a low or minimal risk. Low- or minimal-risk AI systems providers are encouraged to voluntarily comply with the requirements, while high-risk AI systems are permitted on the European market subject to compliance with certain mandatory requirements and an ex-ante conformity assessment[30]. Unacceptable risk use-cases are prohibited.

It's expected that mandatory requirements for high-risk AI systems will require organisations to record up-to-date information on how AI systems have been developed and how they perform throughout their lifecycle. This information is necessary to assess the compliance of the AI system with relevant regulations. Records should include the general characteristics, capabilities and limitations of AI systems, as well as information about algorithms, data, training, testing, and validation processes, and risk management systems. Further, 'the provider should establish a sound quality management system, ensure the accomplishment of the required conformity assessment procedure, draw up the relevant documentation and establish a robust post-market monitoring system[30].

Another example from the Kingdom of the Saudi Arabia National Data Management Office (NDMO)[31], states: 'Compliance management practices shall accordingly define the following:

1. Scope of the periodic compliance audit exercise
2. Processes to plan for and execute the compliance audits
3. Processes and tools to report compliance audit findings
4. Processes and plans for the remediation and escalation of non-compliance.'

Depending on the scope and scale of AI activities, companies may need to allocate resources to compliance, either by adding roles or expanding existing roles. Mobile operators are well-placed to navigate complex technology regulations, and the industry is often recognised for its commitment to operating responsibly.

Standards and certification

Many organisations and international bodies either already provide or are in the process of launching certifications and standards for ethical AI. You will be able to find examples in the Appendix.

30   https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206
31   https://sdaia.gov.sa/ndmo/Files/PoliciesEn001.pdf

4

Key themes and recommendations

This playbook discusses technical, operational and compliance exercises companies can take to ensure they are designing, developing and deploying AI systems ethically.

Below is a set of key themes from the playbook, followed by a list of recommendations for organisations designing AI systems and for those deploying third-party AI systems.

The ethical AI journey

Organisations may find some chapters more useful than others depending on their level of maturity in AI and ethics. Organisations at the beginning of their journey may find Chapter 1 on 'AI ethical principles' a useful starting point. Chapter 2 gives an overview of how organisational design and governance structures can guide ethical behaviour. Chapter 3 lists practical tools, starting with a Self-Assessment Questionnaire. This chapter contains a series of questions to establish risks and tools that can be used to answer those questions, alongside education and training. It might be more useful for organisations that are developing AI ethical principles or that have principles already in place.

Data quality is fundamental to AI

AI systems are only as good as their data. A fundamental requirement of ethical AI is therefore high-quality, complete, accurate and well-managed data. Organisations should ensure they know where data comes from, and that they can explain this if needed, as AI systems based on incomplete or biased data can lead to inaccurate outcomes that infringe on people's fundamental rights, including discrimination[32].

32 https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-data-quality-and-ai_en.pdf

Clear roles are key to good governance

Clear allocation of roles and responsibilities and defined channels to escalate concerns are key to the good governance of AI systems within an organisation. Chapter 2 proposes a structure of decision-making and reporting that can help to embed accountability for AI systems into organisational culture and improve agility by removing bottlenecks.

Multidisciplinary teams with diverse perspectives, skills and approaches should work together to innovate and deliver AI systems. The mix and the ratio of roles should depend on the use-case.

Organisations also need to strike a balance between individual empowerment and executive oversight. Anyone working with an AI system should be able to raise concerns about the ethical impact of AI systems and be empowered to take decisions related to their role. However, it is also vital that there is ultimate executive oversight and accountability for decisions taken to use (or not use) a system. Executive decisions should be guided by ethical principles, which have been co-created or at least approved, at an executive level.

Adopting a lifecycle approach

Organisations using AI systems should adopt a lifecycle approach to managing risks. Typically, the lifecycle of an AI product has three key phases: design (including product conception and data selection); development (engineering and validation); and deployment (ongoing use and monitoring). Organisations using AI systems must have mechanisms to identify and mitigate risks posed by AI and escalation channels available to raise concerns at each of these stages.

Inclusive design and diversity

Incorporating people with diverse perspectives and lived experiences – including those of underrepresented groups and backgrounds – can help to anticipate the needs and concerns of users who may be impacted by AI systems. This includes people from within an organisation, but also people from outside of the organisation, particularly those that might be impacted by the AI system.

Anticipating future contexts of the application

Organisations using and developing AI systems need to consider not only the present-day application of a given AI system but how its use could evolve in the future and how this may impact people. This covers both the way in which an AI system may change its outputs, but also how the contexts in which it is deployed might change. This is especially important for high-risk use-cases.

Policy, controls and compliance

Organisations need clear mechanisms and controls in place to comply with data and AI ethics guidelines and regulations. It is vital they follow evolving national and global legislation and understand what is required of them. This will also help them to establish a common understanding of the risk level of a specific use-case or industry. The current legal landscape is dynamic, so there may be a need for anticipation. Staying ahead of mandatory requirements will help to build and maintain trust and to ensure a leading position in the ongoing dialogue around AI.

Training and raising awareness

It's important for organisations to establish awareness and training requirements in relation to AI, identifying skill gaps and enabling a consistent approach to upskilling. AI literacy should be built throughout the organisation. Enabling the entire workforce to understand what AI is and how it impacts their jobs, the company and wider society is a part of building confidence in AI and driving adoption and usage.

Recommendations for suppliers and users of AI systems

1 Develop ethical principles that guide the design, development and deployment of AI systems.

2 Build an AI governance structure that establishes accountability and escalation channels.

3 Use a documented process, like a Self-Assessment Questionnaire, covering privacy and human rights, to assess and mitigate potential ethical risks.

4 Maintain an up-to-date register of AI systems, including their intended purpose and subsequent use-cases, allowing for traceability and explainability in situations where AI has been adapted; including supplier instructions where relevant.

5 Provide regular feedback to the AI supplier in instances where AI deviates from its expected outcome, particularly if outcomes cause, or have the potential to cause, harm to end-users to allow for a dynamic and ongoing risk-assessment process.

6 Conduct an awareness campaign with the entire organisation on the ethical principles and up-to-date employee training on AI ethics. Provide clear escalation channels for employees working on AI to raise concerns about ethical dilemmas.

7 Determine exactly how and where a human agent is to be involved in AI decision-making, deciding when a human-in-the-loop or human-on-the-loop model is appropriate.

8 Establish a data governance programme for AI systems that allows for the collection and retention of key data on an AI system's operation and insights on AI decision-making.
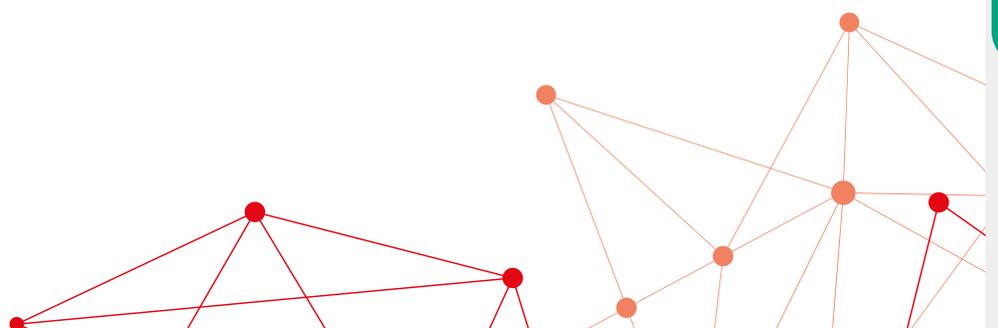
For more information, please go to:

GSMA AI for Impact website
gsma.com/betterfuture/aiforimpact

AI for Impact Digital toolkit
aiforimpacttoolkit.gsma.com

#AI4I

Appendix: further resources

A.1 Technical tools

This section contains an extensive list of technical tools with links and descriptions, aligned around the principles described in Chapter 1. These tools aim to assess fairness, assist bias detection or improve transparency, explainability, robustness, safety and security. They include toolkits, software tools, technical documentation, certification and standards, product development or lifecycle tools, and technical validation tools. Some of the tools listed are commercial while others are integrated into cloud platforms from specific providers, but many are free or open source. Some are high-level, and others are designed to be used by developers. Some are suitable for integration into the design or development process, others are designed for post-hoc analysis. Please note that inclusion in this playbook does not mean that these tools are endorsed by the GSMA.

A1.1 Fairness

• Google What-If Tool[33] allows developers to visually inspect and explore machine learning (ML) model performance and data across multiple hypothetical situationsand across different groups of users. It can also be extended with explainability tools. It is designed for Tensorflow, Python or Google Cloud AI Platform models solving classification or regression tasks on tabular, image or text data. It is available as a pip module and Jupyter notebook extension.

• Google MI-fairness-gym[34] provides a set of components for building simple simulations that explore the potential long-run impacts of ML systems and how their outputs might evolve when deployed in dynamic contexts with active data collection or feedback loops that reinforce existing decision thresholds.

• LinkedIn Fairness Toolkit (LiFT)[35] enables the measurement of fairness according to a multitude of fairness definitions in large-scale ML workflows.

• Microsoft Fairlearn[36] empowers AI developers to assess fairness and mitigate any negative impacts.

• Fairness Flow[37] helps ML engineers detect forms of potential statistical bias in certain types of AI models and labels commonly used at Facebook.

• IBM AI Fairness 360 toolkit (AIF360)[38] helps detect and remove bias in models.

• AT&T's software system to integrate fairness transparently (SIFT)[39] integrates mechanised and human-in-the-loop components in bias detection, mitigation and documentation of projects at various stages of the ML lifecycle.

A1.2 Privacy and security

• Microsoft Counterfit[40] is a command-line tool and automation layer for assessing the security of systems by simulating various types of attack. It is available as open-source software in Python, as a Docker image or for deployment on Azure.

• IBM Adversarial Robustness 360 Toolbox[41] provides tools to evaluate, defend, certify, and verify ML models and applications against the adversarial threats of evasion, poisoning, extraction and interference.

• Microsoft Presidio[42] is an open-source library for data protection and anonymisation for text and images. It uses pre-trained ML models to identify sensitive data. It is available as an open-source Python package, in pip, as a Docker image, or as source code.

• SmartNoise, developed by OpenDP[43], provides a layer between queries and data systems. It adds statistical noise to results and maintains an information budget for sequential queries to limit the total amount of information revealed to any single user. It is available as an open-source Python package, on pip or as a Docker image.

• SEAL[44] is an open-source homomorphic encryption library. It allows computations to be performed on encrypted data while preventing private data from being exposed to cloud operators.

33   https://pair-code.github.io/what-if-tool/
34   https://github.com/google/ml-fairness-gym
35   https://engineering.linkedin.com/blog/2020/lift-addressing-bias-in-large-scale-ai-applications
36   https://azure.microsoft.com/en-us/services/machine-learning/
37   https://ai.facebook.com/blog/how-were-using-fairness-flow-to-help-build-ai-that-works-better-for-everyone/
38   https://developer.ibm.com/open/projects/ai-fairness-360/
39   https://arxiv.org/pdf/2006.06082.pdf
40   https://www.microsoft.com/security/blog/2021/05/03/ai-security-risk-assessment-using-counterfit/
41   https://github.com/Trusted-AI/adversarial-robustness-toolbox
42   https://github.com/Microsoft/presidio
43   https://smartnoise.org/
44   https://www.microsoft.com/en-us/research/project/microsoft-seal/

A1.3 Safety and robustness

- Error Analysis Toolkit[45] enables you to identify cohorts with higher error rates and diagnose the root causes behind them to better inform your mitigation strategies.
- MLOps[46] provides an end-to-end development process to design, build, and manage reproducible, testable and evolvable ML-powered software.
- Facebook's AugLy[47] is an open-source toolkit that helps teach models to be more robust in the face of perturbations of unimportant attributes and focus more on the important attributes of data. It works by applying common perturbations to audio, image, video and text data to augment the training set.

A1.4 Transparency and explainability

- Google Model Card Toolkit[48] provides a schema for communicating the essential facts of an ML model in a structured, accessible way, providing an overview of what the model is intended to do, how it was architected, trained and its limitations.
- IEEE Standard for Transparency of Autonomous Systems[49] is a technical standard that describes measurable and testable levels of transparency so autonomous systems can be assessed, and levels of compliance determined.
- There are several open-source toolkits for implementing explainability and interpretability methods on models. Some examples are IBM AI Explainability 360[50], Captum[51] (built specifically for PyTorch[52]) and Microsoft InterpretML[53].

- Datasheets for Datasets[54] is a process for documenting datasets in detail. The process encourages developers to answer a series of questions detailing the motivation, composition, collection process, preprocessing/cleaning/labelling, uses, distribution and maintenance of each dataset, to encourage transparency and accountability.
- IBM AI Factsheets[55] are collections of relevant information about the creation and deployment of an AI model or service. The idea is modelled on a supplier's declaration of conformity, which shows how a product conforms to standards and regulations.
- SAS Visual Data Mining and Machine Learning[56] includes the ability to generate interpretability reports in models as you build them, using common methods like LIME, ICE, and SHAP

A1.5 Environmental impact

- CodeCarbon[57] estimates the carbon footprint of a piece of code as it is executing on cloud infrastructure, based on the power supplied to the executing hardware and the energy mix of the local grid. It functions as a plugin that integrates into the code itself and produces reports and visualisations. PowerAPI[58], PyJoules, and Scaphandre[59] provide lower-level tools for similar ends. ML CO2 Impact[60] estimates the carbon footprint of computing operations, but functions as a standalone calculator that takes hardware, compute time, platform and region as inputs.

45 https://erroranalysis.ai/
46 https://ml-ops.org/
47 https://ai.facebook.com/blog/augly-a-new-data-augmentation-library-to-help-build-more-robust-ai-models/
48 https://ai.googleblog.com/2020/07/introducing-model-card-toolkit-for.html
49 https://standards.ieee.org/project/7001.html
50 https://aix360.mybluemix.net/
51 https://captum.ai/
52 https://arxiv.org/abs/1803.09010
53 https://pytorch.org/
54 https://github.com/interpretml/interpret
55 https://www.ibm.com/blogs/research/2020/07/aifactsheets/
56 https://www.sas.com/en_sa/software/visual-data-mining-machine-learning.html
57 https://codecarbon.io/
58 http://www.powerapi.org/
59 https://github.com/hubblo-org/scaphandre
60 https://mlco2.github.io/impact/

A2 Training and awareness

You will be able to see below a list of resources designed to help educate and engage organisations about different methods of approaching ethical conversations.

Procedural tools

- The German Trade Union Confederation's *Good Work by Design*[61] gives guidelines for the trustworthy implementation of AI systems in the workplace, with the goal of gaining acceptance among the workforce.
- Google *People + AI Guidebook*[62] shares guidelines to help user experience (UX) professionals and product managers follow a human-centred approach to AI. They are structured based on the product development cycle and contain worksheets to help turn guidance into action.
- IBM *Everyday Ethics for AI*[63] offers practical guidelines for designers and developers of AI solutions, including key questions for team members to consider, and some topical examples of ethical issues.
- IBM *AI Factsheets 360*[64] provides a governance approach to the AI lifecycle and methodology for assembling information about an AI system's important features.

Educational tools

- Denmark Data Ethical Dilemma Game[65] is an educational game to create business awareness around the challenge of developing responsible and ethical AI solutions. The game seeks to stimulate reflections and perspectives on working with data through common ethical dilemmas.

- Finland AI Course 'Elements of AI'[66] is a free, online course combining theory with practical exercises to encourage people to learn the basics of AI, its impacts and how it is created.
- VIRT-EU Service Package[67] is a capacity-building resource to help develop ethically informed AI systems based on three different ethical frameworks: virtue ethics, care ethics and capabilities.
- Microsoft AI Business School[68] is a free, online course to learn how to drive business impact by creating an effective AI strategy, enabling an AI-ready culture, and innovating responsibly.
- Microsoft Judgement Call[69] is a card game designed to test and disseminate AI ethics principles by cultivating stakeholder empathy.
- Moral Machine[70] allows users to test their ethical principles in a number of scenarios based on the trolley problem, a classic ethical dilemma and thought experiment.
- The Markkula Center for Applied Ethics at Santa Clara University has published a framework[71] for ethical decision-making in the form of a document and an app. It explains five different sets of ethical frameworks and allows users to examine their principles against them.
- EthicalOS[72] provides a toolkit for considering possible unintended consequences of a technological development. It outlines eight risk zones and 14 scenarios to encourage discussion, and seven future-proofing strategies to mitigate potential risks.

61 https://www.dgb.de/downloadcenter/++co++b794879a-9f2e-11ea-a8e8-52540088cada
62 https://pair.withgoogle.com/guidebook/
63 https://www.ibm.com/design/ai/ethics/everyday-ethics/
64 https://aifs360.mybluemix.net/
65 https://xd.adobe.com/view/c783a856-4feb-460b-9a89-58ef02107c29-92d0/
66 https://www.elementsofai.com/
67 https://www.virteuproject.eu/servicepackage/?fbclid=IwAR1Unln2Fi9aZHOByTJRqL0vHPPVp-npShq95-tvgDGirP79AbUGtqvCS_8
68 https://www.microsoft.com/en-us/ai/ai-business-school
69 https://docs.microsoft.com/en-us/azure/architecture/guide/responsible-innovation/judgmentcall
70 https://www.moralmachine.net/
71 https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/
72 https://ethicalos.org/

A3 Standards and certification

Many organisations and international bodies either already provide or are in the process of launching certifications and standards for ethical AI. Some of the examples are:

- IEEE 7000TM-2021 Standard[73] addresses ethical concerns by integrating ethical and functional requirements to mitigate risk and increase innovation in systems engineering design and development.
- IEEE Ethics Certification Program for Autonomous and Intelligent Systems[74] is a certification programme to create specifications for AI processes that advance transparency, accountability, and reduction in algorithmic bias.
- ISO/IEC JTC 1/SC 42/WG 3[75] addresses components of trustworthiness in AI systems.
- German BSI – BSI 8611:2016[76] *Guide to the ethical design and application of robots and robotic systems* provides additional guidelines to eliminate or reduce the risks associated with ethical hazards to an acceptable level, and covers safe design, protective measures and information for the design and application of robots.

- Digital Access to National Institute of Standards and Technology (NIST) U.S. Department of Commerce[77] Technical AI Standard promotes innovation and public trust in systems that use AI.
- Linux Foundation for AI[78] is a global group working on policies, guidelines, tools, and use cases to ensure the development of trustworthy AI systems. Future direction includes a badging or certification process for open-source projects that meet the Foundation's Trusted AI policies and guidelines.
- Laboratoire National de Metrologie et d'Essais (LNE)[79] offers a certification standard in collaboration with various AI developers and users. The objective of this certification is to demonstrate that the AI supplier has a mastery of the AI life cycle, allowing them to achieve results in line with the customer's expectations, in terms of performance but also regulations, ethics, transparency, or other ethical concerns.

73   https://engagestandards.ieee.org/ieee-7000-2021-for-systems-design-ethical-concerns.html
74   IEEE Ethics Certification Program for Autonomous and Intelligent Systems
75   https://www.iso.org/committee/6794475.html
76   https://shop.bsigroup.com/products/robots-and-robotic-devices-guide-to-the-ethical-design-and-application-of-robots-and-robotic-systems?pid=000000000030320089&_
      ga=2.238084480.1221993653.1635320069-1777755661.1635320046
77   https://www.nist.gov/artificial-intelligence/technical-ai-standards
78   https://lfaidata.foundation/projects/trusted-ai/

A4 AI principles

To act ethically, organisations require a guiding framework that explains what good ethical behaviour looks like. As AI has developed, and the potential risks of AI have become clearer, many organisations have written AI ethical principles for this purpose. In Chapter 1 we explore a list of AI ethical principles, which are informed by the research listed below.

Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI* Berkman Klein Center for Internet & Society, 2020. http://nrs.harvard.edu/urn-3:HUL.InstRepos:42160420

Independent High-Level Expert Group on Artificial Intelligence Set up by the European Commission. *Trustworthy AI Assessment List* https://ec.europa.eu/ newsroom/dae/document.cfm?doc_id=60440

Arborus and Orange *International Charter for Inclusive AI"* https://charteia.arborus.org/en/

Facebook *Facebook's five pillars of Responsible AI* https://ai.facebook.com/blog/facebooks-five-pillars-of-responsible-ai/

Google AI *Building responsible AI for everyone* https://ai.google/responsibilities/

GSMA. *Mobile Big Datya Analytics and AI for a Better Future: AI Ethics Principles* https://www.gsma.com/ betterfuture/resources/aiethicsprinciples

House of Lords Select Committee on Artificial Intelligence *AI in the UK: ready, willing and able?* https://publications.parliament.uk/pa/ld201719/ ldselect/ldai/100/100.pdf

IBM *AI Ethics* https://www.ibm.com/artificial-intelligence/ethics

Microsoft *Microsoft AI Principles* https://www.microsoft.com/en-us/ai/responsible-ai?a ctivetab=pivot1%3aprimaryr6

OECD Council Recommendation on Artificial Intelligence *OECD Principles on AI* https://www.oecd.org/going-digital/ai/principles/

Telefonica *Business Principles* https://www.telefonica.com/en/web/responsible-business/our-commitments/ai-principles

Telekom *Telekom's AI Guidelines* https://www.telekom.com/de/konzern/ digitale-verantwortung/details/ki-leitlinien-der-telekom-523904

Telia *AI Ethics* https://www.teliacompany.com/en/ about-the-company/public-policy/ai-ethics/

The Japanese Society for Artificial Intelligence Ethical Guidelines http://www.ai-elsi.org/wp-content/uploads/2017/05/JSAI-Ethical-Guidelines-1. pdf

Vodafone *Artificial Intelligence Framework* https://www.vodafone.com/what-we-do/public-policy/policy-positions/artificial-intelligence-framework

A5 AI working groups and strategies

For further thinking, research and potential connection to leading organisations and working groups, we are sharing below a list of AI working groups and strategies.

Cercle InterElles *Women and AI Pledge for an accountable and gender fair AI* http://www.interelles.com/wp-content/uploads/2021/06/Pacte-Femmes-IA-EN-2021-def35348.pdf

Council of Europe. *Ethical frameworks* https://www.coe.int/en/web/artificial-intelligence/ethical-frameworks

Expert Group on Architecture for AI Principles to be practised *AI Governance in Japan Ver. 1.1* 2021 https://www.meti.go.jp/press/2020/01/20210115003/20210115003-3.pdf

KPMG *The shape of AI governance to come* 2021. https://assets.kpmg/content/dam/kpmg/xx/pdf/2021/01/the-shape-of-ai-governance-to-come.pdf

Personal Data Protection Commission Singapore. *Model Artificial Intelligence Governance framework 2nd Edition* https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf

PWC. *A practical guide to Responsible Artificial Intelligence (AI)*" https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai/responsible-ai-practical-guide.pdf

UNESCO *Recommendation on the ethics of artificial intelligence*." https://en.unesco.org/artificial-intelligence/ethics

Artificial Intelligence Ethics & Governance Body of Knowledge https://ai-ethics-bok.scs.org.sg/

Government of Canada R*esponsible use of artificial intelligence (AI)* https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html

Smart Dubai *AI Ethics Principles & Guidelines* https://www.smartdubai.ae/docs/default-source/ai-principles-resources/ai-ethics.pdf?sfvrsn=d4184f8d_6

World Economic Forum *Global AI Action Alliance* https://www.weforum.org/projects/global-ai-action-alliance

The Global Partnership on Artificial Intelligence (GPAI) https://gpai.ai/projects/responsible-ai/

Data Ethics *Questionnaire for Data Ethics Impact Assessment* https://dataethics.eu/questionnaire-for-data-ethics-impact-assessment/

Digital Dubai *AI System Ethics Self-assessment Tool* https://www.smartdubai.ae/self-assessment

High-Level Expert Group on Artificial Intelligence (AI HLEG) *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*". 2020. https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

KOSA *Responsible AI Self-Assessment* http://www.kosa.ai/resources

A6 Self-Assessment Questionnaire examples

Further to Chapter 3, see below a list of other questionnaires available.

Data Ethics *Questionnaire for Data Ethics Impact Assessment* https://dataethics.eu/questionnaire-for-data-ethics-impact-assessment/

Digital Dubai *AI System Ethics Self-assessment Tool* https://www.smartdubai.ae/self-assessment

High-Level Expert Group on Artificial Intelligence (AI HLEG) *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*". 2020. https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

KOSA *Responsible AI Self-Assessment* http://www.kosa.ai/resources