Artificial Intelligence
- UK Business Guidelines

Code of Practice
for the implementation of Artificial Intelligence by UK businesses

Automated Analytics

Automated Analytics

Preface

As someone who has been involved in Artificial Intelligence for over a decade and has built a successful business in this space, it is easy for me to forget that for most business people this is a very new area.

The relatively recent developments with Large Language Models like Chat GPT have caused an explosion of interest across business, the press, and society at large.

This can only be a good thing but there is no doubt that it has brought some very new challenges sharply into focus.

The demands from our global client base prompted us to help them navigate the UK regulatory environment in the form of drafting a general purpose AI Code of Practice they could use to understand what regulation is likely to mean to them and the impact it could have on innovation and investment.

The conclusion from that consultation from a wide range of AI experts within their business was calling on Governments in the UK, EU, and the USA to create a regulatory environment that gives their businesses a clear sense of direction and clarity, which they are confident will ultimately encourage growth and investment without crushing innovation.

I have, therefore, presented this first draft as a foundation for further detailing and refinement, inviting further feedback to help us shape each iteration towards a final version of the code that we can help Governments and regulators shape.

Our General-purpose AI Code of Practice is based on our experiences of implementing AI safely and successfully to over 5,000 companies globally, with 3,500 of those based in the USA.

Automated Analytics is unique in its ability to offer insights into AI adoption both here and in the USA. The UK government has spoken openly about Artificial Intelligence as key to growing the UK economy, but to achieve that they must give the AI businesses in every sector, clarity over current and future regulations so investment into the UK AI industry can have confidence in its future development and safely harness this frontier technology.

Mark Taylor
CEO, Automated Analytics

Table of Contents

	Executive Summary		**6**
1.	**Introduction and Scope**		**10**
	1.1	Purpose	12
	1.2	Scope of Application	13
	1.3	Customised Solutions vs Implementing Products	13
	1.4	Legal Framework	15
	1.5	Definitions	15
2.	**AI Within the Organisation**		**16**
	2.1	Organisational Values & AI Policy	18
	2.2	AI Strategy	19
	2.3	Roles and Responsibilities	21
3.	**Safety, and Risk Management**		**22**
	3.1	Risk Management	24
	3.2	Risk Treatment	25
	3.3	AI System Impact Assessment	27
4.	**Compliance and Ethical Standards**		**28**
	4.1	Compliance with Laws	30
	4.2	Non-discrimination and Inclusion	33
5.	**Customer Relations and Transparency**		**34**
	5.1	Honesty in Communications	36
	5.2	Non-discrimination and Inclusion	37
	5.3	A Human Touch	38
	5.4	Handling Complaints	39
6.	**Employee Relations and Development**		**40**
	6.1	Training	42
	6.2	AI Adoption	43
7.	**Environmental and Social Responsibility**		**44**
	7.1	Sustainability	46
	7.2	Social Responsibility	47
8.	**Monitoring, Auditing, and Reporting**		**48**
	8.1	Internal Audits	50
	8.2	Ongoing Monitoring	51
	8.3	Incident Reporting	52
	8.4	Corrective Actions	53
9.	**Review and Continuous Improvement**		**54**
	9.1	Periodic Review	56
	9.2	Feedback Mechanism	57
	Glossary		**58**

Executive Summary

This document aims to provide a set of guidelines for UK businesses, small or large, to approach the application of AI within their organisation. It will help them harness the benefits of AI whilst mitigating potential risks proactively to maximise the gains from this dynamic new technology.

We have defined four overarching sets of values that ensure AI is used with maximum benefit and minimal liability within organisations:

1. Aligned & Integrated
2. Secure & Ethical
3. Responsible & Human
4. Governed & Overseen

To ensure AI applications fulfil these values, we recommend taking different measures for each.

1. Aligned & Integrated

To effectively integrate AI, an organisation will need to define an AI policy that aligns to the overall values and objectives of the organisation itself. From there, an AI strategy should be crafted to clearly identify business areas for AI use and prioritise AI use cases within the organisation as well as defining a clear plan for execution along with needed resources for implementation. To have clear ownership of all AI considerations from strategy all the way through to implementation and oversight, it is recommended that a leader within the organisation takes overall responsibility for the implementation and is fully accountable to the senior leaders.

2. Secure & Ethical

A risk management system is essential to assess the entailed risks properly and identify measures to mitigate them. One way to assess risk of AI systems is by evaluating their potential impact on users or individuals and society as a whole.

AI predictive models are built to predict future outcomes based on historical data. Accuracy measures the reliability of these models, indicating the model's ability to correctly classify or predict an outcome. If a model isn't accurate, the prediction it makes might be less useful, almost like flipping a coin to determine the outcome – which most businesses wouldn't accept. As AI models first start from human intervention, these biases can make their way into AI systems – discriminatory data and algorithms could be 'baked into' AI models, deploying biases at scale and amplifying the resulting negative effects of the model outcome/prediction.

Compliance with existing regulations is a crucial factor in making an AI system secure, not only for users but for the organisation as a whole. Finally, ethical considerations should be taken into account by assessing and mitigating the bias of AI systems where possible. Continual model monitoring (see section 8.2) ensures that the accuracy of the model, and any bias now or in the future, is measured and steps taken accordingly to limit risks.

3. Responsible & Human

To build trust in the technology, AI needs to be responsible and its use needs to put humans at its centre including delivering a humane outcome.

In maintaining good customer relations, honesty in communication and transparency with a human touch are key. Giving customers access to real humans where it adds the most value is important. Additionally, a reliable mechanism for handling complaints needs to be established.

Employee relations and development are crucial for an organisation to thrive. Therefore, it is essential to train all employees on AI basics as well as the established AI policies and mechanisms to give feedback on AI applications, in addition to having a detailed awareness of bias of all types in the results. This builds trust in the technology and the organisation which supports the AI adoption and innovation within the business.

Finally, Environmental and Social Responsibility are an important aspect of responsible AI systems. Integration of AI applications into existing Environmental Social Governance or Corporate Social Responsibility (ESG/CSR) policies and targets where possible is highly recommended. An organisation's social responsibility lies mainly in ensuring the ethical use of the technology. This best practice encompasses safety and security as well as good governance of AI systems.

4. Governed & Overseen

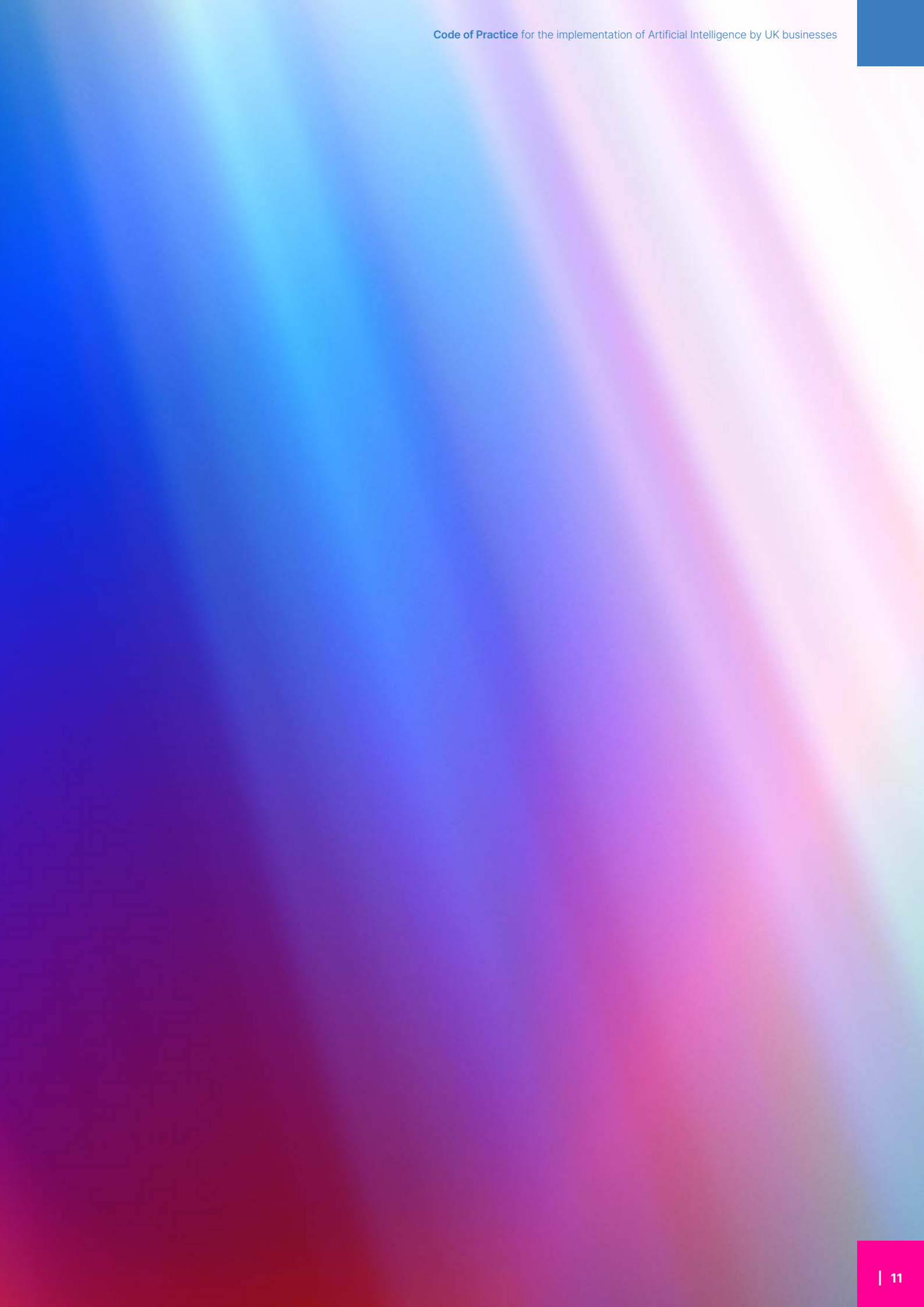
It is essential to establish proper oversight of AI systems through monitoring, auditing and reporting. This includes mechanisms for regular internal auditing, incident reporting as well as corrective actions.

AI applications require regular reviews to enable iterative and continuous improvement. Therefore mechanisms for periodic review and feedback need to be implemented.

Summary

1. Aligned & Integrated	2. Secure & Ethical	3. Responsible & Human	4. Governed & Overseen
AI with the organisation • Organisational values & AI policy • AI strategy • Roles & responsibilities	**Safety & risk management** • Risk management • Risk treatment • AI system impact assessment	**Customer relations & transparency** • Honesty in communication • Transparency • A human touch • Handling complaints	**Monitoring, auditing & reporting** • Internal audits • Incident reporting • Corrective actions
	Compliance & ethical standards • Compliance with laws • Non-discrimination & inclusion	**Employee relations & development** • Training • AI adoption	**Review & continuous improvements** • Periodic review • Feedback mechanism
		Environmental & social responsibility • Sustainability • Social responsibility	

1. Introduction and Scope



1.1

Purpose

The increased use of Artificial Intelligence (AI) in all of its forms over the last decade has seen the excitement of potential opportunities matched by demands for an assessment of the potential risks and in some cases for regulation.

This document is a first attempt to produce a workable code of practice guide for UK businesses. It is clear that well-implemented and managed AI systems will transform all parts of our society but it is also the case that a full and thorough risk assessment must be undertaken at all levels just as it would be with any new technology.

To mitigate the risks of this technology properly, AI governance is crucial. AI governance is the set of rules, policies, and practices designed to ensure AI systems are developed and used safely and ethically to serve human interests. This should fit seamlessly into the existing governance framework of any large business. Smaller businesses have in the past been less able to dedicate resources to wider governance issues, however, the exponential power of AI to enable small businesses to reach a similar scale to a large business means that governance issues are more significant now than ever before. This effectively makes it much more difficult for a small business to keep up on the governance and stewardship side.

The Code of Practice will provide a framework for businesses to evaluate AI within their organisations including setting and aligning AI objectives, determining and managing the risks, establishing the internal processes for managing AI implementation and operation and establishing guidelines throughout the organisation's supply chain.

This document is not intended as a list of rules and regulations, but rather an accessible guide to best practice to be adopted and adapted by UK businesses large and small throughout the UK.

1.2
Scope of Application

This document is focused on UK organisations, or the UK subsidiaries of multinational organisations where appropriate. It does not replace any legal or regulatory requirements on any such businesses.

It is also focused on businesses and organisations that do not have AI as their primary function, and are primarily users and adopters of AI rather than creators. This includes AI developers, designers, and operators, who are working within an organisation. Organisations that are wholly or primarily AI platforms, products or service providers may find elements of this useful but will require substantially more complex guidelines than can be provided here.

1.3
Building customised solutions vs using AI products

A fundamental issue when implementing any AI strategy is the make / buy decision.

Whether to develop a customised solution for your organisation or to simply leverage one of many AI products from broad technologies like ChatGPT or Gemini to niche and industry specific tools that are emerging on a daily basis.

Building customised solutions however, does not mean you are not leveraging existing technology. It means you adapt existing technology to better fit it to your business needs, for example building a customer service chatbot based on ChatGPT or utilising your proprietary information as ‘training data’ and adding further ‘data labelling’ to enable the model to give customised responses to your customers and colleagues.

When building customised solutions, organisations need to ensure that the built systems are safe and ethical. This of course entails higher complexity than just ensuring the safe and ethical use of an AI product.

The overall structure of the approach should be unchanged but the fundamental differences in approach are:

- **Proprietary data:**
  When you leverage your data to customise existing AI technology you need to ensure data quality and privacy.

- **Data quality and lineage:**
  Ensure your data is of high quality and represents the respective use case well. Document which datasets were used for what and how the data was cleansed (munging or wrangling) or augmented (e.g. normalising and repeatability testing) and why.

- **Data privacy:**
  Ensure sensitive data is safely stored according to data privacy regulations and only accessed by those who need access to build and maintain the customised solution. Where possible anonymise personal data. Put transparency and user consent front and centre.

- **Security and Safety:**
  To ensure safety, test the solution on real-world data before deployment. Ensure its robustness against adversarial attacks and adherence to ethical standards (see Section 2). Take measures to prevent cybersecurity vulnerabilities.

- **Documentation and ownership:**
  Assign clear ownership for each stage of the development cycle and document each decision along it.

- **Build inclusive AI systems:**
  AI systems should be designed to produce fair outcomes for all and to avoid bias perpetuation. Implement mechanisms to identify, mitigate, and correct biases. Review training data for fair representation and outcomes for fairness (also see Section 2.3). Hire a diverse team to build and audit the bias in the system.

- **Mitigate bias:**
  As well as impacting inclusivity and fairness, AI bias has been shown to affect every aspect of an AI systems efficacy. Removing this – debiasing – is proving to be one of the most challenging aspects of any AI project. Mitigating bias should include training data bias, algorithmic bias and the cognitive biases of the humans interacting with the AI.

The UK government acknowledges that the challenges posed by AI technologies will ultimately require legislative action in every country. However, it currently relies on existing laws and frameworks, estimating that more time is needed to better understand the risks, opportunities, and appropriate regulatory responses.

(PWC – The impact and necessary response, EU AI Act)

1.4 Legal Framework

The UK Government has made clear its views that we are too early in the development of AI to draft a set of laws.

The EU is working on the EU AI Act which will come into force in 2025 and introduces an impact-based risk classification system ranking AI applications from low risk to high risk and even completely banning some applications (for more details see Section 2.4). This will impact any UK business doing business in the EU or if the outputs of their AI systems are intended for use in the EU (i.e. not part of any commercial transaction).

It is anticipated that most businesses implementing AI solutions will be doing business in the EU either now or in the future so the key aspects of the EU Act should be incorporated into any UK Code of Practice (these are outlined later in the document).

The remaining G7, and to an extent the wider G20 countries, have adopted a framework – The G7 AI Principles and Code of Conduct (AIP&CoC) which is intended to "provide a common reference point for current developments such as the US Executive Order on Safe, Secure and Trustworthy AI, Canada Bill C-27 AI & Data Act, as well as feeding into global efforts such as the UK hosted AI Safety Summit, the Council of Europe Convention on AI, Human Rights, Democracy and the Rule of Law and the UN Global Digital Compact." *(EY – G7 AI Principles and Code of Conduct – 31 October 2023)*.

This document has been prepared with a view of these standards and the international standard on Information Technology – Artificial Intelligence – Management system reference ISO/IEC 42001 (2023).

1.5 Definitions

A full list of definitions is listed in the appendix. This has been sourced from the UK Parliament and is published under the Open Parliament Licence v3.0.

2.

AI within the Organisation

2.1

Organisational Values & AI Policy

All AI applications, whether customised AI solutions or leveraging existing products within an organisation, should reflect the objectives and values of that organisation. These values must be stated explicitly and agreed as the parameters of the AI policy.

Recommendations

1. The leadership team will agree and set the parameters of the AI policy by clearly laying out
    1.1. The objectives of the organisation for which AI should be leveraged to attain those (long and short term)
    1.2. The values of the organisation and the use of AI within it
2. The leadership team will appoint suitably qualified individuals or teams to establish, implement and maintain an AI policy that is fully aligned with the organisation
3. The leadership team will regularly audit all AI applications with respect to adherence and accountability to the established AI policy.

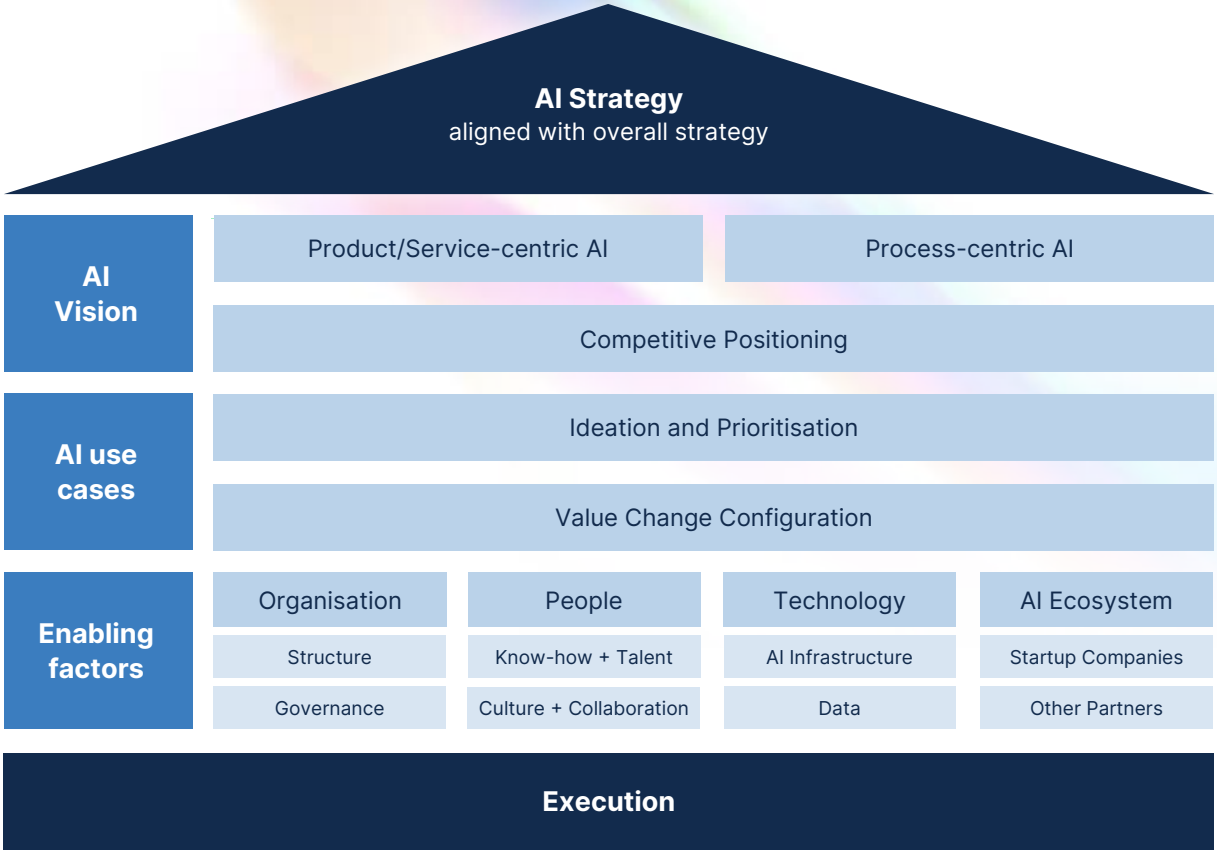
2.2
AI Strategy

The AI strategy should form the heart of all organisations' systematic approach to driving success and achieving their objectives through Atificial Intelligence.

A comprehensive AI strategy consists of three parts: an AI vision, a portfolio of AI use cases, and a clear strategy for the required enabling factors. The AI vision defines in which areas AI is most beneficial to achieve business objectives. The vision then needs to be translated into applications, or a number of concrete prioritised use cases that together form a portfolio. Finally, you need to build a strategy for the enabling factors like people, process, technology and requirements for the organisation that make it possible to execute the strategy and implement the use cases.

Recommendations

1. Identify the areas of the business that can most benefit from AI and assign a prioritisation
2. Outline the scope of projects, so-called use cases, in each of the above areas including the need to develop bespoke tools or buy in existing AI products
3. Provide a full risk assessment for each individual AI project
4. Allocate resources and identify responsibilities for each AI project
5. Include a commitment to continual improvement and monitoring of AI applications in the organisation
6. Communicate all the above with the wider organisation and external stakeholders as directed by the leadership team.



Source: Applying AI: How to find and prioritize AI use cases the applied AI Initiative GmbH www.appliedai.de

2.3
Roles and Responsibilities

It is the responsibility of the leadership team to appoint a top-level leader, the Chief AI Officer, with overall responsibility for the use of AI within the business.

This role should encompass both the commercial opportunities, the new strategic opportunities opened up through AI as well as a thorough and ongoing assessment of the risks. In small businesses, this role can be filled by an existing member of the leadership team instead of hiring a new person.

It is the responsibility of the leadership team to ensure that the Chief AI Officer has access to the resources necessary for the fulfilment of all tasks including:

- Recruitment of competent individuals to implement any AI across the business
- Training & development of key individuals in AI applications in relevant fields
- The provision of hardware and software resources as deemed necessary to any projects
- The allocation of sufficient time to conduct the planning, implementation and ongoing reviews of all AI projects - both customised and off the shelf.

Recommendation

1. The top leadership team appoints an overall AI leader, Chief AI Officer
2. The Chief AI officer oversees the integration of AI into the whole organisation, from strategy all the way through to implementation and maintenance
3. The Chief AI Officer ensures the AI roles are defined and filled with competent individuals
4. The Chief AI Officer is fully accountable to the board of directors for any and all aspects of AI developed by and for the organisation.

3.

Safety, and Risk Management

3.1

Risk Management

Every use of AI within an organisation must be subject to a risk assessment (see Section 3.2). This should include an outline of the intended results and an identification of potential undesirable effects (see Section 3.3).

The undesirable effects should be determined to be either acceptable or unacceptable and if acceptable, a more detailed risk assessment, risk treatment and impact assessment should be undertaken.

This risk management system must also include a periodic review of the risks and their impact. It is crucial to include cybersecurity risks in the list of potential risks.

Recommendations

1. A risk management system must be established to align with the AI Policy & Objectives - i.e. relevant to the culture and objectives of the organisation
2. Each risk should be defined by the potential consequences, the likelihood of those consequences and therefore a perceived level of risk
3. Each risk should be identified as acceptable or unacceptable
4. All acceptable risks should include a risk treatment
5. The risk management system must be reviewed periodically – as determined by the AI policy and the nature of the risk.

3.2
Risk Treatment

Each identified risk should have a corresponding risk treatment. The treatment should be outlined and documented as part of the risk management process.

The risk treatment should cover the entire AI project life cycle and should include elements such as:

- **Resources**
  Both human (the people) and technical (the equipment e.g. hardware and software)

- **Responsibility**
  Internal and external including customers and suppliers

- **Data**
  Acquisition, provenance, quality and preparation

- **Documentation and communication**
  Including to interested third parties

- **Compliance**
  List all regulations that apply and measures needed for compliance.

3.3

AI System Impact Assessment

The AI System impact assessment is a formal written document that outlines the possible consequences of any AI systems in the widest sense.

It should include the deployment, intended use and potential misuse of any AI system and the impact not just on users or individuals but on society as a whole.

It should address legal issues and life opportunities, the physical and psychological well-being of individuals as well as their fundamental human rights. This should be as widespread as to include the rights to privacy, physical safety and financial consequences beyond those of the user but to include the widest definition of society.

This document should be made available to interested parties beyond the organisation itself.

Recommendations

1. Once an AI project and its risks are defined an AI System Impact assessment should be completed
2. This should take the widest view of impact across society and consider typical usage as well as potential misuse
3. As well as political, economic, legal and environmental concerns the impact should consider social norms and traditions particularly where misinformation could cause serious and lasting damage.

The EU AI Act Risk (see Section 4.2) classification system may be adopted to classify each system's risk in terms of impact on individuals.

4. Compliance and Ethical Standards

4.1

Compliance with Laws

Whilst there is no specific legal framework in the UK to cover AI implementation, it is noted that the UK Government believes that the current laws – including health & safety at work, data protection and copyright infringement – will suffice for the time being.

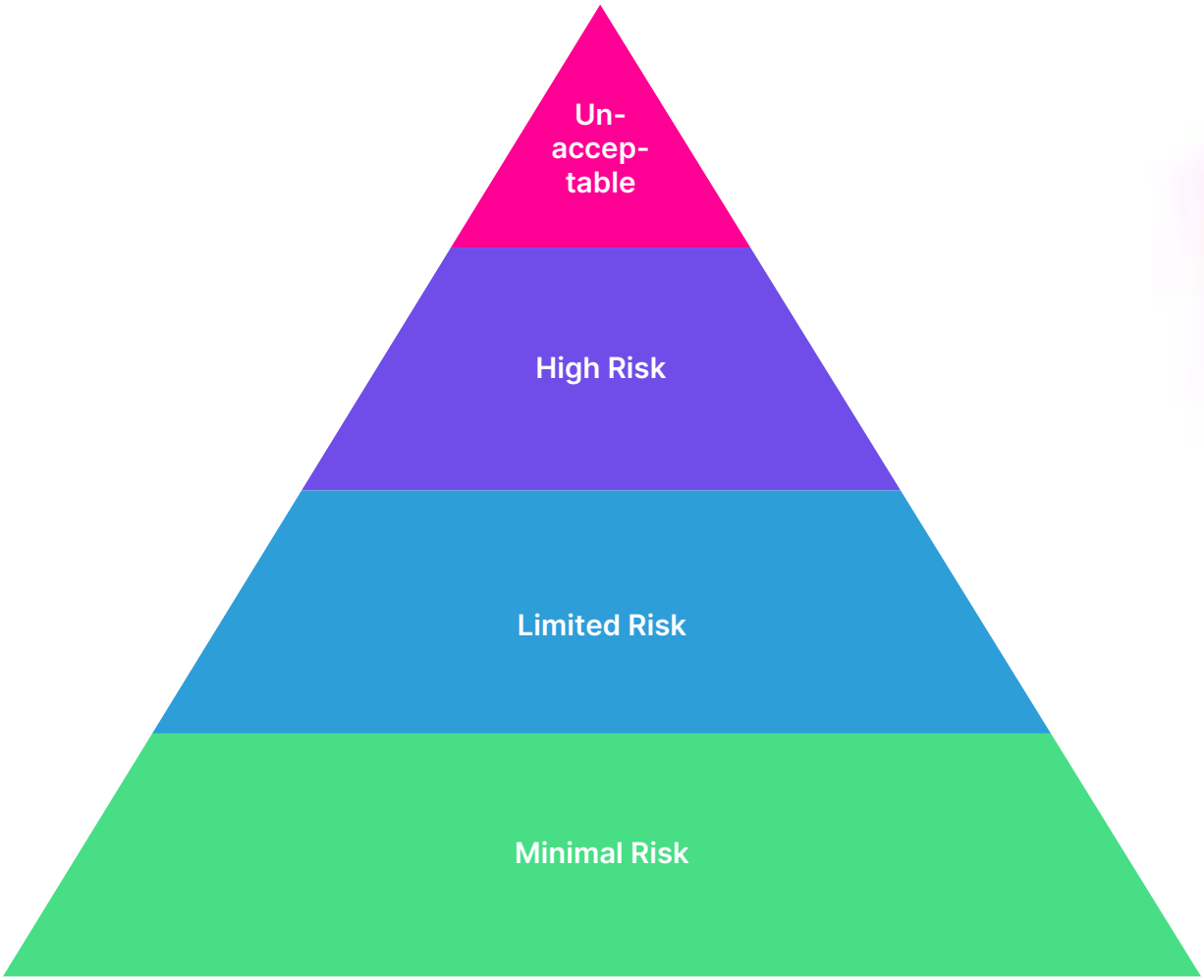
Any legal implications should be addressed as part of the risk assessment in the AI policy and for each specific AI project.

The EU AI Act

As mentioned earlier, the EU is developing the EU AI Act which will be in full effect in 2025 and impact any subsidiaries operating in the EU or any systems where the outputs will be used in the EU.

The EU AI Act is the first comprehensive AI legislation with the goal to specifically addressing the risk of AI applications. At the core of it is the classification of AI systems into 4 risk categories: Prohibited, High Risk, Limited Risk and Minimal Risk.

The EU Act is evaluating what AI is used for instead of looking at the technology itself, taking an impact approach. While all applications considered to be a clear threat to the safety, livelihoods, unimpaired decision-making and rights of people are banned, e.g. social scoring by governments or biometric identification in public spaces for law enforcement, compliance requirements apply to the remaining three categories, increasing from minimal to high.



EU AI Act and AI Governance Risk Levels

High risk systems are AI applications in the fields of critical infrastructure, employment, worker management, access to education and vocational training, product safety, migration and border control as well as administration of justice and democratic processes.

To name a few concrete examples:

- AI-based CV screening for recruitment processes
- AI-based scoring of exams
- Credit scoring denying citizens the opportunity to obtain a loan
- AI application in robot-assisted surgery

For systems falling into the high-risk category a number of compliance requirements apply:

- High security, accuracy & robustness
- Detailed documentation
- Human oversight (with the ability to override the system)
- Activity logging and result traceability
- Clear user information.

Applications with limited risk are those for which a lack of transparency in AI usage is associated, for example, chatbots, deepfakes (audio or video) or other AI-generated content. The EU AI Act introduces transparency requirements for these applications to make humans aware when interacting with AI technology or AI-produced content.

The vast majority of AI applications fall under the **minimal or no risk category**. For these applications, no regulations apply. Examples are AI-enhanced video games, spam filters or text completion.

General Purpose Foundation Models are big models like ChatGPT that can be used for several different purposes and as such are hard to regulate through the use case approach. Therefore the EU AI Act includes separate regulations for the providers of those models. All providers need to provide a summary of the training data and need to comply with copyrights, commercial providers also need to provide instructions for use and detailed technical documentation. In addition, the EU AI Act distinguishes between Providers of GPAI models that pose a systematic risk, defined by needed computing power required to train the model, also need to ensure an adequate level of cybersecurity, regular model evaluation and risk assessment and mitigation as well as track, document and report serious incidents.

4.2
Non-discrimination and Inclusion

AI models are trained on huge datasets, especially models that rely on deep learning such as Large Language Models (ChatGPT, Claude, Gemini etc.) or Image Generation models (DALL-E, Midjourney, Stable Diffusion).

These datasets reflect all the biases and prejudices our society has. As a consequence all our societal inequalities are reflected in the data, often meaning that data on women, people of colour and all underrepresented groups is usually incomplete. The consequence is that AI models do not work or produce unfavourable outcomes for these groups. To mention two examples: Facial recognition technology that was found to perform significantly worse for darker-skinned faces and Amazon's CV screening system downgraded CVs containing the word "women's" and as such discriminating against women who attended women's colleges (due to being trained primarily on CVs of male applicants).

Recommendations

Teams

1. Enable diverse and multi-disciplinary teams working on AI applications
2. Promote a culture where ethical standards are highly valued and adhered to.

AI technology

1. Regularly audit all AI applications with respect to potential bias in outcomes
2. Potentially correct outcomes post-prediction, for example, use guardrails to prevent sending harmful LLM output to the user. This approach can also be taken when using existing AI products
3. When building a customised solution, assess the used dataset with respect to equal representation. Employ data bias and algorithmic bias mitigation when needed.

5.

Customer Relations and Transparency

5.1

Honesty in Communications

All current guidelines about honesty and openness in communications should be extended to include any AI capability.

There should be no attempt to mislead, intentionally or otherwise, about the use of AI, the benefits and any potential risks.

All communications both internal and external should be legal, decent, honest and truthful.

Recommendations

1. The communications strategy around any AI project should integrate with any existing communications policies by being legal, decent, honest and truthful

2. The language used should not be overly technical in either a data or legal sense and easily understood by users.

5.2
Non-discrimination and Inclusion

Beyond traditional communications, there is a further need when dealing with AI projects to make clear the full nature of the project.

Most notably, but not exclusively, the data sources being used, the acquisition, provenance and processing of such data.

It should also be made clear when users are interacting with an AI system.

The wide reaching nature of AI may mean that this is far from obvious to the casual observer; it is therefore incumbent on the organisation to outline the full transparency of the project. This is particularly relevant to data and privacy issues but should not be limited to that.

Recommendations

1. Transparency should be addressed in all communications touchpoints throughout a user journey. This should include before the journey has begun and whenever a user is interacting with an AI system.

5.3

A Human Touch

The objective of any AI system is to enhance and accelerate a user experience not simply to replace all human interaction.

Organisations should look to enhance the AI system with human engagement at those points where it will add the most value to the user journey. This may be to reassure, to supplement or simply as an alternative option where possible.

5.4
Handling Complaints

Within the ongoing review and assessment programme should be a formal mechanism for the capture, recording and processing of complaints.

A complaint being any feedback which evidences a negative result, either planned or unintended. As such complaints should be welcomed as feedback that can be used to train any AI-based project. All complaints should be added to other feedback as part of the review process.

The existence of a complaints process should form part of the communications plans around any specific AI project as well as the AI policy as a whole. The complaints procedure should aim to be as simple to use as possible and ideally no more complex than the system for which the feedback is being sought.

Recommendations

1. Have a fully documented complaints procedure
2. Ensure the procedure is well communicated and easy to complete
3. Complaints, as a definition of negative feedback, should be welcomed as learning experiences for all AI projects.

6.

Employee Relations and Development

6.1

Training

The development of AI is incredibly fast moving and it is impossible for anyone to maintain awareness of all aspects.

It remains the responsibility of the organisation to invest in suitable training at various levels for employees. This includes strategic-level reviews and updates for the leadership team as well as specific sector-relevant training at an operational level. This can include technical, data science or AI-related project management.

A training covering AI basics as well as the AI policy, risks and incident reporting is strongly recommended for all employees. This builds trust in the technology and supports the development of new AI use cases and innovation. The training should be part of the onboarding process for every new employee.

Given the speed and nature of the development of AI, the definition of training could be expanded to include a Continual Professional Development (CPD) approach to include conference attendance, independent research and blogs/journals in a wider definition of training.

Recommendations

1. A top-level commitment to AI-based training and development is essential to maintain an AI policy
2. Create an introduction to AI workshop for all staff
3. A wider CPD approach to training and development should be undertaken rather than restricting employees to formal training formats.

6.2
AI Adoption

In the context of employee relations, it is fair to say that AI is, at the time of writing, more feared than welcomed. This must be addressed as part of an AI implementation.

A change management programme that takes time to listen to employee concerns and address them as far as possible is an essential step. The process should be a dialogue and should begin as the AI policy is being drafted not as a launch event to announce an AI decision that has already been made.

Whilst specific concerns will be unique to each organisation they will generally include:

- **Safety concerns** – where relevant e.g. self-driving cars etc, safety will always be the top concern.
- **Job loss & financial issues** – negative changes to financial status is probably more widespread. Media stories about AI doing to white collar jobs what automation did to blue collar jobs in the end of the 20th century do not help.
- **Privacy and data protection** – again, scare stories do not help but the concerns are valid and should be addressed.
- **Fear of change** – a general fear of change is common to all change management processes and again should be planned for.

Recommendations

1. Plan a change management programme around any large-scale AI projects – especially if you are developing a bespoke system
2. Begin the dialogue before you launch the plan as AI can create more perceived fear than other IT projects
3. Update company HR documents to align with new AI technology
4. Maintain the dialogue post-project implementation and build it into the feedback process.

7.

Environmental and Social Responsibility

7.1

Sustainability

Environmental concerns form a major challenge to any AI programme. This is particularly the case when developing a customised AI solution for an organisation.

Where possible the environmental issues should already fit clearly within the overall ESG/CSR objectives of the business. In some cases, an AI project will negatively impact the organisation's environmental policies, particularly in relation to timescales for achieving net zero. In the case of public companies, this reporting will have additional legal and governance issues depending on the specific nature of the organisation.

In the case that the AI opportunities will impact the environmental position and governance of the organisation, this must be addressed by the leadership of the organisation who must make a decision and outline the reasoning as part of the AI Strategy. Where possible the organisation should look to mitigate any additional energy consumption issues for example by using carbon offsetting if appropriate.

Recommendations

1. Integrate all AI projects into existing ESG/CSR policies and targets where possible
2. Where the Environmental impact of the AI project will have a negative impact on existing ESG/CSR targets, reevaluate the cost/opportunity decision
3. Publish the results and reasoning for any changes to your ESG/CSR targets due to planned or ongoing AI projects
4. Any revisions should be communicated to all stakeholders and interested parties, including any legal implications where necessary.

7.2
Social Responsibility

When it comes to AI, an organisation's social responsibility lies mainly in ensuring the ethical use of the technology.

This encompasses safety and security as well as governance of AI systems including measures for non-discrimination and inclusion as discussed in Section 4.2. All the recommendations in this document aim at establishing the socially responsible use of AI.

8. Monitoring, Auditing, and Reporting

8.1

Internal Audits

Regular internal audits form a part of the code of practice to ‘close the feedback loop’ on the specific implementation for each organisation.

The internal audit is focused on two elements:

1. Are the objectives set by the organisation being met and are they still the correct objectives?
2. Is the code of practice and AI policy - as defined by the organisation – being followed and maintained?

The organisation needs to identify an independent team to conduct the audit, as well as deciding the frequency, methods and reporting parameters.

Recommendations

1. Internal audits must be established at a frequency determined by the AI leadership team
2. The scope of the internal audit should be limited to adherence to the policy and the validity of the objectives
3. Internal audits must be carried out by an independent team and given the resources necessary to complete their task
4. The final reports should be documented and communicated to all interested parties within the organisation.

8.2
Ongoing Monitoring

Every deployment should include a comprehensive set of affirmative signals, which absence (or inclusion) will trigger a response from the maintainers of the system.

Monitoring model performance should be treated similarly to how one might monitor systems and servers.

The impact of this, or a failure to implement this, would cause significant impact to the performance of the specific AI projects, the business as a whole as well as wider ethical and societal implications.

Recommendations

1. Ongoing monitoring must be established with every deployment.

8.3

Incident Reporting

Create a formal process for reporting non-compliance or ethical concerns.

As highlighted earlier AI implementations can be approached in the manner of a health and safety programme as much as a technical or IT implementation project.

As such formal incident reporting is essential.

Like Health and Safety, this process should be well communicated and feedback should be encouraged, as it forms an essential element of training any AI system.

Incidents should also be well documented so that they can be reviewed as part of any external audit and communicated, along with corrective actions, to all stakeholders.

Recommendations

1. Establish an AI incident reporting system – in the manner of a Health and Safety system
2. Communicate the existence of the system and encourage feedback to help train the AI
3. Document the incidents to allow for further review and communication at a later stage.

8.4
Corrective Actions

Where the incident reporting in 8.3 highlights a need for a corrective action this should be implemented in a timely manner based on the risk management and system impact reports.

Again, using Health and Safety models to ascertain high, medium and low risk and then attribute timescales and reporting accordingly. These will be dependent on the organisation and sector, for example where a risk to life has been identified, then this would constitute a major or high risk incident and require immediate corrective action. This would additionally be governed by Health and Safety laws in advance of any code of practice or guideline.

Most corrective action should be lower level and can be prioritised as part of the day-to-day project management.

Recommendations

1. Establish a corrective action programme alongside the incident reporting
2. Use a prioritisation method appropriate to your organisation and attribute response times accordingly
3. Document and communicate any corrective actions as part of the ongoing feedback and change management programme.

9.

Review and Continuous Improvement

9.1

Periodic Review

In addition to an internal audit in 8.1, a periodic management review of the Code of Practice should be undertaken to ensure its relevance and effectiveness, particularly as AI systems evolve.

The management review process should encompass all of the elements above including:

- AI objectives and validity
- Efficacy of the current AI policy and its implementation
- Incidents and corrective actions
- Internal audit report
- Review of risk and impact assessment

This review should be held at the top level of the organisation and have an open brief to review all aspects of the AI policy across all of the organisation.

Recommendations

1. A periodic review should be held at the most senior leadership level of the organisation
2. This review should have the freedom to review all aspects of the AI policy, its efficacy, the risks and system impacts as well as unintended consequences
3. The review should encompass any change external to the organisation including political, legal and any interested third parties
4. The review should aim to implement a policy of continual improvement and look for additional opportunities to improve the implementation of AI within the organisation.

9.2
Feedback Mechanism

As stated earlier feedback is more critical to an AI project application than almost all other projects undertaken across an organisation.

Feedback must be actively encouraged not simply facilitated and the need for feedback must be communicated to all stakeholders at every stage of the project.

Recommendations

1. Create channels for stakeholders to simply and easily provide feedback on AI practices and the code
2. Communicate all of these feedback channels as widely as possible
3. Publish and communicate any improvements gained as a result of feedback to further encourage continual improvement.

Glossary

Automated Analytics

Glossary
Contains Parliamentary information licensed under the Open Parliament Licence v3.0

Algorithm
A set of instructions used to perform tasks (such as calculations and data analysis) usually using a computer or another smart device.

Algorithmic bias
AI systems can have bias embedded in them, which can manifest through various pathways including biased training datasets or biased decisions made by humans in the design of algorithms.

Artificial intelligence (AI)
The UK Government's 2023 policy paper on 'A pro-innovation approach to AI regulation' defined AI, AI systems or AI technologies as "products and services that are 'adaptable' and 'autonomous'." The adaptability of AI refers to AI systems, after being trained, often developing the ability to perform new ways of finding patterns and connections in data that are not directly envisioned by their human programmers. The autonomy of AI refers to some AI systems that can make decisions without the intent or ongoing control of a human.

AI governance
AI governance is the set of rules, policies, and practices designed to ensure AI systems are developed and used safely and ethically to serve human interests. This includes establishing guidelines for AI development, defining accountability measures, and creating oversight mechanisms that involve multiple stakeholders from government, industry, academia, and civil society.

Artificial general intelligence
Sometimes known as general AI, strong AI or broad AI, this often refers to a theoretical form of AI that can achieve human-level or higher performance across most cognitive tasks. See also Superintelligence.

Artificial neural network
A computer structure inspired by the biological brain, consisting of a large set of interconnected computational units ('neurons') that are connected in layers. Data passes between these units as between neurons in a brain. Outputs of a previous layer are used as inputs for the next, and there can be hundreds of layers of units. An artificial neural network with more than 3 layers is considered a deep learning algorithm. Examples of artificial neural networks include Transformers or Generative adversarial networks.

Automated decision-making
A term that the Office for AI, within the Department for Science, Innovation and Technology, refers to in an Ethics, Transparency and Accountability Framework for Automated decision-making as "both solely automated decisions (no human judgement involved) and automated assisted decision-making (assisting human judgement)." AI systems are increasingly being used by the public and private sector for automated decision-making.

Compute
Compute is defined by the Independent Review of the Future of Compute as 'the systems assembled at scale to tackle computational tasks beyond the capabilities of everyday computers. This includes both physical supercomputers and the use of cloud provision to tackle high computational loads.' Compute is a driver of AI developments.

Computer vision
This focuses on programming computer systems to interpret and understand images, videos and other visual inputs and take actions or make recommendations based on that information. Applications include object recognition, facial recognition, medical imaging analysis, navigation and video surveillance.

Deep learning
A subset of machine learning that uses artificial neural networks to recognise patterns in data and provide a suitable output, for example, a prediction. Deep learning is suitable for complex learning tasks and has improved AI capabilities in tasks such as voice and image recognition, object detection and autonomous driving (PN 633).

Deepfakes
Pictures and videos that are deliberately altered to generate misinformation and disinformation. Advances in generative AI have lowered the barrier for the production of deepfakes.

Disinformation
The UK Government defines disinformation as the "deliberate creation and spreading of false and/or manipulated information that is intended to deceive and mislead people, either for the purposes of causing harm, or for political, personal or financial gain". Advances in generative AI have lowered the barrier for the production of disinformation, misinformation, and deepfakes.

Educational technology
Technologies specifically developed to facilitate teaching and learning which may or may not encompass AI. See PN 712 for further details.

Fine-tuning
Fine-tuning a model involves developers training it further on a specific set of data to improve its performance for a specific application. For more details see PB 57.

Foundation models
A machine learning model trained on a vast amount of data so that it can easily be adapted for a wide range of general tasks, including being able to generate outputs (generative AI). See also large language models.

Frontier AI
Defined by the Government Office for Science as 'highly capable general-purpose AI models that can perform a wide variety of tasks and match or exceed the capabilities present in today's most advanced models'. Currently, this primarily encompasses a few large language models including ChatGPT (OpenAI), Claude (Anthropic) and Gemini (Google)

Generative AI
An AI model that generates text, images, audio, video or other media in response to user prompts. It uses machine learning techniques to create new data that has similar characteristics to the data it was trained on. Generative AI applications include chatbots, photo and video filters, and virtual assistants.

General-purpose AI
Often refers to AI models that can be adapted to a wide range of applications (such as Foundation Models). See also narrow AI.

Generative adversarial networks
Generative adversarial networks are a driver of recent AI developments (PB 57). These are made up of two sub artificial neural networks: a generator network and a discriminator network. The generator network is fed training data and generates artificial data based on patterns in training data. The discriminator network compares the artificially generated data with the 'real' training data and feeds back to the generator network where it has detected differences. The generator then alters its parameters. Over time the generator network learns to generate more realistic data, until the discriminator network cannot tell what is artificial and what is 'real' training data and the AI model generates the desired outcomes. See also artificial neural networks and transformers.

Graphical processing units
These are similar to central processing units, found on a typical home computer. Graphical processing units have been used since the 1970s in gaming applications and have been designed to accelerate computer graphics and image processing. In the past decade, graphical processing units have been increasingly applied in the training of large machine learning models after they were found to be effective in processing large amounts of data in parallel.

Hallucinations
Large language models, such as ChatGPT, are unable to identify if the phrases they generate make sense or are accurate. This can sometimes lead to inaccurate results, also known as 'hallucination' effects, where large language models generate plausible sounding but inaccurate text. Hallucinations can also result from biases in training datasets or the model's lack of access to up-to-date information.

Interpretability
Some machine learning models, particularly those trained with deep learning, are so complex that it may be difficult or impossible to know how the model produced the output. Interpretability often describes the ability to present or explain a machine learning system's decision-making process in terms that can be understood by humans. Interpretability is sometimes referred to as transparency or explainability.

Large language models
A type of foundation model that is trained on vast amounts of text to carry out natural language processing tasks. During training phases, large language models learn parameters from factors such as the model size and training datasets. Parameters are then used by large language models to infer new content. Whilst there is no universally agreed figure for how large training datasets need to be, the biggest large language models (frontier AI) have been trained on billions or even trillions of bits of data. For example, the large language model underpinning ChatGPT 3.5 (released to the public in November 2022) was trained using 300 billion words obtained from internet text. See also natural language processing and foundation models.

Machine learning
A type of AI that allows a system to learn and improve from examples without all its instructions being explicitly programmed. Machine learning systems learn by finding patterns in training datasets. They then create a model (with algorithms) encompassing their findings. This model is then typically applied to new data to make predictions or provide other useful outputs, such as translating text. Training machine learning systems for specific applications can involve different forms of learning, such as supervised, unsupervised, semi-supervised and reinforcement learning.

Misinformation
The UK Government defines misinformation as "the inadvertent spread of false information". Advances in generative AI have lowered the barrier for the production of disinformation, misinformation, and deepfakes.

Narrow AI
Sometimes known as weak AI, these AI models are designed to perform a specific task (such as speech recognition) and cannot be adapted to other tasks. See also general-purpose AI.

Natural language processing
This focuses on programming computer systems to understand and generate human speech and text. Algorithms look for linguistic patterns in how sentences and paragraphs are constructed and how words, context and structure work together to create meaning. Applications include speech-to-text converters, online tools that summarise text, chatbots, speech recognition and translations. See also large language models.

Open-source
Open-source often means the underlying code used to run AI models is freely available for testing, scrutiny and improvement.

Reinforcement learning
A way of training machine learning systems for a specific application. An AI system is trained by being rewarded for following certain 'correct' strategies and punished if it follows the 'wrong' strategies. After completing a task, the AI system receives feedback, which can sometimes be given by humans (known as 'reinforcement learning from human feedback'). In the feedback, positive values are assigned to 'correct' strategies to encourage the AI system to use them, and negative values are assigned to 'wrong' strategies to discourage them, with the classification of 'correct' and 'wrong' depending on a pre-established outcome. This type of learning is useful for tweaking an AI model to follow certain 'correct' behaviours, such as fine-tuning a chatbot to output a preferred style, tone or format of language. See also supervised learning, unsupervised learning and semi-supervised learning.

Responsible AI
Often refers to the practice of designing, developing, and deploying AI with certain values, such as being trustworthy, ethical, transparent, explainable, fair, robust and upholding privacy rights.

Robotics
Machines that are capable of automatically carrying out a series of actions and moving in the physical world. Modern robots contain algorithms that typically, but do not always, have some form of Artificial Intelligence. Applications include industrial robots used in manufacturing, medical robots for performing surgery, and self-navigating drones.

Semi-supervised learning
A way of training machine learning systems for a specific application. An AI system uses a mix of supervised and unsupervised learning and labelled and unlabelled data. This type of learning is useful when it is difficult to extract relevant features from data and when there are high volumes of complex data, such as identifying abnormalities in medical images, like potential tumours or other markers of diseases. See also supervised learning, unsupervised learning, reinforcement learning and training datasets.

Superintelligence
A theoretical form of AI that has intelligence greater than humans and exceeds their cognitive performance in most domains. See also artificial general intelligence.

Supervised learning
A way of training machine learning systems for a specific application. In a training phase, an AI system is fed labelled data. The system trains from the input data, and the resulting model is then tested to see if it can correctly apply labels to new unlabelled data (such as if it can correctly label unlabelled pictures of cats and dogs accordingly). This type of learning is useful when it is clear what is being searched for, such as identifying spam mail. See also semi-supervised learning, unsupervised learning, reinforcement learning and training datasets.

Training datasets
The set of data used to train an AI system. Training datasets can be labelled (for example, pictures of cats and dogs labelled 'cat' or 'dog' accordingly) or unlabelled.

Transformers
Transformers have greatly improved natural language processing, computer vision and robotic capabilities and the ability of AI models to generate text (PB 57). A transformer can read vast amounts of text, spot patterns in how words and phrases relate to each other, and then make predictions about what word should come next. This ability to spot patterns in how words and phrases relate to each other is a key innovation, which has allowed AI models using transformer architectures to achieve a greater level of comprehension than previously possible.
See also artificial neutral networks and generative adversarial networks.

Unsupervised learning
A way of training machine learning systems for a specific application. An AI system is fed large amounts of unlabelled data, in which it starts to recognise patterns of its own accord. This type of learning is useful when it is not clear what patterns are hidden in data, such as in online shopping basket recommendations ("customers who bought this item also bought the following items"). See also semi-supervised learning, supervised learning and reinforcement learning and training datasets.

Automated Analytics

For more information about how to incorporate AI safely into your business, scan this QR code to contact the team at Automated Analytics.

automatedanalytics.co