# ICO - AI and Data Protection Risk Toolkit

## User Guide

The risks and benefits to individuals that arise from personal data processing using artificial intelligence (AI) are heavily context-dependant, and vary significantly across the diverse range of sectors, technologies and organisation types covered by data protection legislation. This toolkit will help you understand some of the AI-specific risks to individual rights and freedoms and provides practical steps to mitigate, reduce or manage them.

Developing AI is generally an iterative process. We have divided the risks and controls by high-level lifecycle stages to help as a guide to what risks and controls you should be considering at each stage. However, you should always ensure your processing is compliant with data protection legislation as a whole. The toolkit can also be divided by risk area. So, for example, if you are struggling to think of how to mitigate risks associated with data minimisation, then you can filter the risk area column to only include information related to data minimisation.

You can use this tool as a way to assess the risks to fundamental rights and freedoms of individuals. By undertaking the practical steps suggested in line with what is expected under the legislation, these risks to fundamental rights and freedoms are reduced and compliance with data protection law becomes more likely. Documenting your assessment of the risk and the steps you take to mitigate them can help you demonstrate compliance with the legislation. We have provided additional cells to illustrate how you could carry out an evaluation.

When scoring risks, we have provided four options, 'high', 'medium', 'low' and 'non-applicable'. The assessment of risks will vary depending on the context, so you should undertake your own assessments of the risks identified.

Using the toolkit is entirely optional and you will not be penalised for not using it. Although this toolkit can complement data protection impact assessments (DPIA) that you are legally required to conduct where processing is likely to result in high risk to individuals, it is not designed to replace them.

Please note that this tool is not designed to be 'one size fits all' and each risk should be assessed in the context in which you are developing and deploying AI. There may also be additional risks that apply to your context that are not included in this toolkit.

| AI Lifecycle Stage | ID | Risk area<br>UK GDPR Reference | Data Protection Risk Statement | Risk Assessment Summary | Inherent Risk Rating | Control | Control Objective | Practical steps to reduce the risk | Further ICO Guidance | Practical Steps Your Organisation Will Take | Control Owner | Current status | Completion date | Residual Risk Rating (following action) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Business requirements and design | 1.1 | Accountability<br>Articles 5(2), 35 and 36 and Recitals 74-77, 84, 89-92, 94 and 95. | The misidentification of risks to individual rights and freedoms caused by not carrying out a risk assessment. As a consequence, an organisation cannot put in place appropriate technical and organisational measures to prevent harms occurring to individuals. | | | Conduct a data protection impact assessment (DPIA) | To identify risks and implement appropriate technical and organisational measures to reduce them. | You must do a DPIA for processing that is likely to result in high risk to individuals.<br>You must, where appropriate, consult with individuals who are likely to be affected by your use of AI.<br>If you identify a high risk that you cannot mitigate, you must consult with the ICO before starting the processing.<br>As part of your DPIA, you should consult with the teams within your organisation who will be involved in your project to identify data protection risks of your AI project. The teams you may want to consult include the engineering team, the legal and compliance team, and any staff who will be part of the decision pipeline.<br>You could consult with domain experts who can advise on what risks you should address. | What are the accountability and governance implications of AI? \| ICO | | | | | |
| Business requirements and design | 1.2 | Accountability<br>Articles 5(2) and 24 and Recitals 39 and 74 | A lack of accountability over risks to individual rights and freedoms created or exacerbated by AI systems is caused by not clearly assigning roles and responsibilities. As a consequence, risks are left unaddressed, and individuals may suffer harm. | | | Assign technical and operational roles and responsibilities and provide clear direction and support on the use of AI systems and the application of data protection law | To make it clear who is accountable for mitigating and managing risks in the AI system. | You should appoint a senior owner or senior process owner to drive accountability.<br>You should put in place operational procedures, guidance or manuals to support AI policies and provide direction to operational staff on the use of AI systems and the application of data protection law | Accountability and governance \| ICO | | | | | |
| Business requirements and design | 1.3 | Purpose limitation<br>Article 5(1)(b), Recital 39, Article 6(4) and Recital 50 and Article 30 | Function creep over how personal data is processed is caused by not defining what purpose you will use your AI system. As a consequence, individuals lose control over how their data is being used. | | | Document each purpose for using personal data at each stage of the AI lifecycle, assess whether they are compatible with the originally defined purpose, and schedule reviews to reassess your purposes and whether they remain compatible. | To define what your AI system will be used for, how personal data will be used and prevent incompatible processing taking place. | You must provide clear transparency information to inform individuals about your purposes from the outset. For example, in a privacy notice.<br>You should consider completing a data flow mapping exercise to document the data that flows in, through and out of an AI system to ensure a lawful basis and, if necessary, Article 9 or Article 10 condition (or both) is selected for each purpose. | Principle (b): Purpose limitation \| ICO | | | | | |
| Business requirements and design | 1.4 | Fairness<br>Article 5(1)(a); Recital 71 | AI systems producing unfair outcomes for individuals are caused by insufficiently diverse training data, training data inappropriate for the purpose of the AI system, training data that reflects past discrimination, design architecture choices or another reason. As a consequence, individuals suffer from unjustified adverse impacts such as discrimination, financial loss or other significant economic or social disadvantages. | | | Document an assessment of the different ways your AI system could result in unfairness, which should include appropriate technical and organisational measures you will use to mitigate or manage those risks on a continual basis. | To identify risks associated with fairness and take appropriate preventative action | You should ensure the assessment is conducted by appropriately skilled personnel (this may require a cross-disciplinary approach, eg data scientists working with legal counsel and review boards).<br>You should ensure the assessment initially focuses on the expected outcomes that are experienced by individuals directly affected by your processing.<br>You could engage with stakeholders to draw out the risks that your processing is likely to have. | What do we need to do to ensure lawfulness, fairness, and transparency in AI systems? \| ICO | | | | | |
| Business requirements and design | 1.5 | Transparency<br>Article 5(1)(a); Articles 12-15 | The lack of transparency, interpretability and/or explainability is caused by choices about how an AI system is designed and developed. As a consequence, individuals lack the understanding about how their data is being used, how the AI system affects them, and how to exercise their individual rights. | | | Document and assess the explainability and transparency requirements, considering the domain, sector or use case that your AI system will be deployed in. | To provide clear requirements for transparency and explainability of the AI system that allows for effective product optimisation. | You must provide individuals with privacy information at the time you collect their personal data from them. If you collect their personal data from a source other than the individual it relates, then provide them with privacy information within a reasonable period of obtaining the personal data and no later than one month, before or when the first communication takes place or before or<br>You should assess people's expectations of the content and scope of similar explanations previously offered or researching sector-specific expectations. For example, what are the expectations of individuals in a health context compared to an insurance context. You should consider, for example, the people affected by the decision and the end users.<br>You should assess how well the outcome of the AI system is understood to help you decide how comprehensive your explanation needs to be | Explaining decisions made with AI \| ICO | | | | | |
| Business requirements and design | 1.6 | Security<br>Articles 5(1)(f) and 32-34 and Recital 83 | The unauthorised or unlawful processing, accidental loss, destruction, or damage of personal data is caused by insecure AI systems. As a consequence, individuals can suffer from financial loss, identity fraud and a loss of trust. | | | Document and assess the security risks, and the appropriate technical and organisational measures you will use to mitigate or manage those risks. | To demonstrate that security risks have been thought about from the beginning and that a data protection design approach has been adopted. | You must consider the security risks associated with integrating an AI system with existing systems, and document what controls will be put in place as part of the design and build phase. The level of risk will likely vary depending on the context your AI system will be used in.<br>You should consider processes to report security breaches, and who is responsible for handling and managing them as part of an AI incident response plan<br>You could consult with appropriately skilled technical experts about what the latest state-of-the-art is. | How should we assess security and data minimisation in AI? \| ICO | | | | | |
| Business requirements and design | 1.7 | Data minimisation; Storage Limitation<br>Articles 5(1)(c ), 5(1)(e ), 5(1)(b), Recital 39, Article 6(4) and Recital 50 | The excessive and irrelevant collection of personal data is caused by a default approach to collect as much data as possible to design and build AI systems. As a consequence, individuals suffer from unlawful and unfair processing. | | | Document the data you will collect to train the AI system and assess whether it is accurate, adequate, relevant, and limited to your purpose(s). | To demonstrate compliance with the data minimisation principle. | You should include details in your privacy policy and privacy notice about your retention periods and how often you review whether you still need to hold personal will take place. You are in the best position to judge how often these take place. Factors to consider include a change in your overall purpose, signs of<br>You should ensure that you have regular internal discussions about what personal data is needed and why it is required. | How should we assess security and data minimisation in AI? \| ICO | | | | | |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | You could assess what privacy-enhancing technologies would be appropriate for your use case.<br>You could consult with domain experts to ensure that the data you intend on collecting is appropriate and adequate. | | | | | | |
| Business requirements and design | 1.8 | Individual rights<br>Articles 15-22 | The failure to respond adequately to information rights requests is caused by a lack of awareness that data subject rights apply throughout the lifecycle of an AI system wherever personal data is used. As a consequence, individuals become disempowered over how their personal data is used and lose trust in the organisation handling their personal data. | | | Document how you will facilitate individual rights requests throughout the lifecycle of your AI system where personal data will be processed. | To ensure individual rights requests are handled appropriately. | You should index the personal data in your AI system so that it is easier to retrieve when a request is received.<br>You should organise training within the organisation to ensure developers are aware of the need to provide functionality to enable the organisation to respond to individuals exercising their rights.<br>You could conduct user testing to get feedback on how effective the delivery of your privacy information is. | How do we ensure individual rights in our AI systems? \| ICO | | | | | |
| Business requirements and design | 1.9 | Meaningful human review<br>Article 13(2)(f); Article 14(2)(g); Article 15(1)(h); Article 22 | Tokenistic human review of outputs by AI systems may inadvertently cause solely automated decision-making with legal or similarly significant effects. As a consequence, individuals suffer from prohibited processing taking place and inaccurate and/or unfair decisions being made about them, which have legal or similarly significant effects. | | | Document and assess when you will incorporate meaningful human review in the decision pipeline, who will conduct the review, and what additional information they will take into consideration when making the final decision. | To ensure compliance with legal requirements. | You must ensure meaningful human review when your decisions are solely automated and have legal or similarly significant effects unless Article 22 exemptions apply.<br>You must (where Article 22 applies) ensure human reviewers have the authority and ability to challenge and override automated decision-making, and they consider additional factors when making the final decision.<br>You should ensure human reviewers are capable of intervening on the automated decisions and maintain a record of what information the human reviewer saw when making the final decision. You could consider what tools human reviewers need to make a meaningful final decision and how to record that those tools were | How do we ensure individual rights in our AI systems? \| ICO | | | | | |
| Data acquisition and preparation | 2.1 | Lawfulness<br>Article 5(1)(a); Article 6; Article 9 | Failing to choose an appropriate lawful basis causes the unlawful collection of personal data. As a consequence, individuals lose trust over how their data is used and suffer from unfair processing. | | | Identify and document valid grounds for collecting and using personal data. | To ensure your processing is lawful. | You must identify an appropriate lawful basis (or bases) for your processing. If processing special category data or criminal offence data, you must identify a condition for processing this type of data. You should consult with your data protection officer about what is the most appropriate lawful basis (bases) and if required the additional conditions for processing.<br>You must include your lawful basis (plus any additional conditions for processing) in your privacy notice along with the purposes.<br>Where your chosen lawful bases depend on the processing being 'necessary', you must assess whether the processing is targeted and proportionate way of achieving a specific purpose and whether you can achieve the purpose by some other less intrusive means, or by processing less data. | What do we need to do to ensure lawfulness, fairness, and transparency in AI systems? \| ICO<br>Lawful basis for processing \| ICO | | | | | |
| Data acquisition and preparation | 2.2 | Fairness<br>Article 5(1)(a); Recital 71 | An AI system that produces unfair outcomes for different individuals or groups is caused by insufficiently diverse training data, training data inappropriate for the purpose of the AI system, or training data that reflects past discrimination. As a consequence, individuals suffer from unjustified adverse impacts such as discrimination, financial loss or other significant economic or social disadvantages. | | | Document and assess what data you need to ensure a representative, reliable and relevant training dataset. | To demonstrate that you have attempted to mitigate risks associated with unfair outcomes caused by low quality datasets. | You should consider appropriate technical approaches to mitigating possible bias, such as re-weighting, or removing the influence of protected characteristics and their proxies.<br>You should factor in risks of past discrimination.<br>You could research the population that your AI system is likely to impact and flag any risks of bias or discrimination. | What do we need to do to ensure lawfulness, fairness, and transparency in AI systems? \| ICO | | | | | |
| Data acquisition and preparation | 2.3 | Fairness<br>Article 5(1)(a); Recital 71 | Individuals suffering discriminatory outcomes are caused by an AI system relying on protected characteristics (or their proxies) to make a decision. As a consequence, individuals suffer from unlawful decisions being made about them and miss out on economic or social benefits. | | | Assess, document, and maintain an index of data sources or features that should not be processed when making decisions about individuals because of direct or indirect discrimination. | To prevent an AI system producing discriminatory decisions based on protected characteristics or their proxies. | You should consider whether to collect these features for the purpose of bias analysis. Note that processing personal data for bias analysis generally carries lower risk than for decision-making purposes that directly affect the individual, but as a separate processing purpose, will require a lawful basis and a condition for processing. | What do we need to do to ensure lawfulness, fairness, and transparency in AI systems? \| ICO | | | | | |
| Data acquisition and preparation | 2.4 | Fairness<br>Article 5(1)(a) and Recital 71 and Article 9 and Recitals 51 to 56 (see also Schedule 1 of the Data Protection Act 2018) | Bias is caused by a lack of, or poorly conducted, bias analysis. As a consequence, individuals suffer from undetected discriminatory outcomes and miss out on economic or social benefits. | | | Assess and document which protected characteristics data you will collect for bias analysis. | To detect and correct an AI system exhibiting bias. | You must identify the lawful bases and additional processing conditions for the bias analysis.<br>You must ensure that individuals are aware of how their data will be used for bias analysis.<br>You should consider whether you need to process additional data to carry out your bias analysis and whether you need to create labels for data you already hold or whether you need to collect more data. This may include special category/protected characteristic data. | What do we need to do to ensure lawfulness, fairness, and transparency in AI systems? \| ICO | | | | | |
| Data acquisition and preparation | 2.5 | Fairness<br>Article 5(1)(a); Recital 71 | Poor labelling of training data is caused by unclear labelling policies. As a consequence, individuals suffer from inaccurate and/or unfair outcomes made about them by AI systems. | | | Document clear criteria and lines of accountability for the labelling of data. | To prevent data labelling that will lead to unfair outcomes for individuals. | You should create labelling criteria that are: easy to understand, include descriptions for all possible labels, examples of every label, cover edge cases.<br>You should produce training manuals for labelling and annotation.<br>You could consult with members of protected groups or their representatives to define the labelling criteria.<br>You could involve multiple human labellers to ensure consistency across multiple rounds of reviewing.<br>You could document statistics on level of agreement reached by human annotators. | What do we need to do to ensure lawfulness, fairness, and transparency in AI systems? \| ICO | | | | | |
| Data acquisition and preparation | | Fairness | Inadequate training datasets leading to overfitting is caused by not collecting | | | Assess and document whether your model is likely to suffer from overfitting. | To detect and correct any features in your training dataset that are likely to | You must remove any features that are likely to result in overfitting. | | | | | | |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2.6 | Article 5(1)(a); Recital 71 | enough features or enough cases. As a consequence, individuals suffer from poor outcomes made by the AI system. | | | | result in your model suffering from overfitting. | You should monitor model performance metrics (eg precision and recall) to determine sources of possible overfitting issues<br>You could collect more data to ensure the training dataset will be representative of the population you will deploy your model on. Although, this should be balanced with individuals' rights to not be subject to excessive, unlawful or unfair processing of their personal data. | What do we need to do to ensure lawfulness, fairness, and transparency in AI systems? \| ICO | | | | | |
| Data acquisition and preparation | 2.7 | Transparency<br>Article 5(1)(a); Articles 12-15 | Unclear privacy notices are caused by a lack of clear internal definitions about what your AI system will be used for. As a consequence, individuals cannot understand or have control over how their personal information is processed. | | | Assess, document, and publish privacy information about what personal data you will be processing and for what purposes. | To prevent data subjects losing control over how their personal data is processed. | You must tell individuals about your processing in a way that is easily accessible and easy to understand. You must use clear and plain language.<br>You must include information about the purposes of the processing for which the personal data are intended as well as the lawful basis for processing and the categories of personal data concerned.<br>You should test whether the privacy information is presented in a way that is accessible and understandable to the people the AI system will be applied to. | Transparency \| ICO | | | | | |
| Data acquisition and preparation | 2.8 | Data minimisation<br>Article 5(1)(c ) and Recital 39 | The collection of too much personal data is caused by not applying de-identification techniques. As a consequence, individuals suffer from unlawful and unfair processing. | | | Apply de-identification techniques to training data before it is extracted from its source and shared internally or externally. | To prevent the collection of personal data that exceeds the minimum of what is necessary to train your AI model(s). | You should assess what privacy enhancing technologies are appropriate for your use case. | How should we assess security and data minimisation in AI? \| ICO<br>ICO call for views: Anonymisation, pseudonymisation and privacy enhancing technologies guidance \| ICO | | | | | |
| Training and testing | 3.1 | Purpose limitation<br>Article 5(1)(b), Recital 39, Article 6(4) and Recital 50 and Article 30 | Undetected function creep is caused by a learning algorithm developing in an unpredicted way. As a consequence, individuals lose their right to be informed about how their data is being used and lose trust in the organisation handling their personal data. | | | Assess and document whether your current purposes are different from your initial purposes. | To detect any changes in the purposes of how you will deploy your AI system. To correct any identified function creep risks. | You must ensure that if you plan to use or disclose personal data for any purpose that is additional to or different from the originally specified purpose, the new use is fair, lawful and transparent<br>You must update your documentation and your privacy information to reflect your new purpose.<br>You must consider whether any consents you rely on cover the change in purpose. | Principle (b): Purpose limitation \| ICO | | | | | |
| Training and testing | 3.2 | Fairness<br>Article 5(1)(a); Recital 71 | Overfitting is caused by a learning algorithm paying too much attention to the specific features in the training datasets. As a consequence, individuals who aren't similar to the individuals in the training datasets suffer from inaccurate and unfair outcomes. | | | Document and assess whether your AI system can handle data from a wide range of (sub)populations fairly and accurately. | To detect any (sub)populations that will be unfairly treated by your AI system and to optimise the model(s). | You should consider whether you need to collect more data from a subpopulation to yield more accurate results.<br>You should assess and justify your choice between collecting more data to reduce the disproportionate number of statistical errors and not collecting such data due to the risks doing so may pose to the other rights and freedoms of those individuals<br>You could consider using feature engineering techniques to aggregate features for training to avoid using personally identifiable data | What do we need to do to ensure lawfulness, fairness, and transparency in AI systems? \| ICO | | | | | |
| Training and testing | 3.3 | Fairness<br>Article 5(1)(a); Recital 71 | Discriminatory outcomes caused by an algorithm basing decisions on protected characteristics (or their proxies) leads to adverse impacts on individuals such as financial loss or other significant economic or social disadvantages. | | | Test whether your AI system produces similar outcomes for individuals who have different protected characteristics. | To prevent discriminatory outcomes once the AI system is deployed. | You should measure different types of error (eg false positives, false negatives, etc).<br>You should consider whether the testing dataset is adequate.<br>You should record any limitations of the model in the context of statistical inaccuracies<br>You could consider documenting the limitations in a model card. | What do we need to do to ensure lawfulness, fairness, and transparency in AI systems? \| ICO | | | | | |
| Training and testing | 3.4 | Transparency<br>Article 5(1)(a); Articles 12-15 | Individuals being subject to unexplainable decisions are caused by AI systems that use high dimensionality or complex learning algorithms. As a consequence, individuals cannot exercise their right to be informed and may feel disempowered to object to the decision. | | | Assess and document the explainability of your AI system and consider what supplementary tools you can use to help explain decisions made by your AI system to individuals who will be affected. | To prevent an unexplainable model being deployed and negatively impacting a population. | You must provide meaningful information about the logic involved where you are using solely automated decision-making which has legal or similarly significant effects for the individual<br>You should test the effectiveness of your explanations by measuring how well individuals can understand why the model made the decision it did, or how the model output contributed to the decision.<br>You should design explanations that meet the needs of those interacting with the system at different moments e.g. someone interacting with an AI system for the first time vs. someone using an automated result to support a critical decision<br>You could consider when you will need to explain automated results to the different user groups interacting with the AI system and whether there is an easy way to challenge these decisions or obtain human intervention.<br>You could consider textual clarification, visualisation media, graphical representations, summary tables, or a combination as part of your explanation. | Explaining decisions made with AI \| ICO | | | | | |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Training and testing | 3.5 | Security<br>Article 5(1)(f);<br>Articles 32-34 | Inappropriate access to training data, training code, and deployment code is caused by lax security policies. As a consequence, individuals may have their personal data subjected to data poisoning attacks leading to unfair advantages or disadvantages. | | | Document and implement strict controls over who has access to training data, training code, and deployment code. | To ensure there is a clear audit trail of who has access to the training data, training code, and deployment code, and when they have access. | You should consider the principle of least privilege - where a user is given the minimum levels of access or permissions to perform their job functions. Ensure this is regularly reviewed and access is revoked where necessary. You should keep logs of who has access and/or editing rights. | How should we assess security and data minimisation in AI? \| ICO | | | | | |
| Training and testing | 3.6 | Security<br>Articles 5(1)(f) and 32-34 and Recital 83 | Accidental loss of training and testing data is caused by a lack of accountability and documentation. As a consequence, individuals lose control over how their data is processed and lose trust in the organisation handling their personal data. | | | Document clear audit trails of how personal data is moved and stored from one location to another during the training and testing phase. | To prevent a security breach where personal data is accidentally lost. | You should keep an up-to-date inventory of all AI systems to allow you to have a baseline understanding of where potential incidents could occur.<br>You should document security processes and make it freely available for all those involved in the building and deployment of AI systems. This should include processes to report security breaches, and who is responsible for handling and managing them as part of an AI incident response plan. An AI incident response plan should include guidance on how to quickly address any failures or attacks that occur, who responds when an incident occurs, and how they communicate the incident to other parts of the organisation. | How should we assess security and data minimisation in AI? \| ICO | | | | | |
| Training and testing | 3.7 | Security<br>Articles 5(1)(f) and 32-34 and Recital 83 | Undetected security vulnerabilities in an AI system's software stack are caused by a lack of, or poorly conducted, security checks of software. As a consequence, individuals suffer from security attacks and breaches of their personal data. | | | Assess and document the security risks of the software you are using. Implement appropriate technical and organisational measures to reduce risks identified. | To prevent security vulnerabilities from occurring. | You should carry out security testing of your software, either in-house or contract someone external.<br>You could subscribe to security advisories to receive alerts of vulnerabilities and ensure solid patching / updating processes are in place where software is externally maintained. | How should we assess security and data minimisation in AI? \| ICO | | | | | |
| Training and testing | 3.8 | Data minimisation<br>Article 5(1)(c );<br>Article 5(1)(e ) | Processing more data than is strictly necessary is caused by a lack of a review over what data is needed to effectively train and test the AI system. As a consequence, individuals suffer from unlawful and unfair processing. | | | Reassess and document what data is necessary, adequate, and relevant for training and testing your AI system. Erase any data that is not needed. | To ensure that only personal data that is necessary, adequate, and relevant is processed. | You must consider whether any data has been duplicated or copied during the training and testing phase.<br>You should consider the trade-off between data minimisation and statistical accuracy and whether you can remove some data without significantly affecting the accuracy of your model. | How should we assess security and data minimisation in AI? \| ICO | | | | | |
| Training and testing | 3.9 | Meaningful human review<br>Article 13(2)(f);<br>Article 14(2)(g);<br>Article 15(1)(h);<br>Article 22 | Non-meaningful human review is caused by a lack of training for human reviewers to interpret and challenge outputs made by an AI system. As a consequence, individuals are subject to unlawful solely automated decisions that have legal effects or similar significant effects that are inaccurate and/or unfair. | | | Design and implement appropriate training for human reviewers. | To ensure that human reviewers can interpret, and challenge outputs made by the AI system once the system is deployed. | You must (where Article 22 applies) design your AI system to ensure that human reviewers have meaningful influence over the decision, including the authority and competence to go against the recommendation and take into account other additional factors that weren't included as part of the input data. | How do we ensure individual rights in our AI systems? \| ICO | | | | | |
| Deployment and monitoring | 4.1 | Fairness<br>Article 5(1)(a);<br>Recital 71 | Undetected model drift is caused by irregular system testing. As a consequence, individuals suffer from unfair decisions being made about them and may exclude them from social or economic opportunities. | | | Document and define a testing regime to occur at regular intervals. | To detect and correct model drift in appropriate timeframes. | You must inform individuals about any testing process that involves their personal data.<br>You should establish metrics and thresholds for model drift that trigger review and testing.<br>You should save versions of your model, which can be reverted back to if significant drift occurs.<br>You could consider running a traditional decision-making system and an AI system concurrently and investigate any significant difference in the type of decisions. | What do we need to do to ensure lawfulness, fairness, and transparency in AI systems? \| ICO | | | | | |
| Deployment and monitoring | 4.2 | Transparency<br>Article 5(1)(a);<br>Articles 12-15 | Where the steps set out in 3.4 are followed, non-meaningful explanations may still be caused by failing to update how they are presented following feedback from individuals. As a consequence, individuals cannot be informed about how their data has been used, nor receive meaningful information about the logic involved. | | | Document and define a redress as well as feedback mechanism that allows individuals to comment on the explanations they receive. | To detect and correct ineffective explanations. | You could proactively engage with individuals to see how you can improve the explanations you provide. | Explaining decisions made with AI \| ICO | | | | | |
| Deployment and monitoring | 4.3 | Security<br>Articles 5(1)(f) and 32-34 and Recital 83 | Attacks on AI systems are caused by poor security practices. As a consequence, individuals have their personal data subject to data breaches leading to potential financial losses and/or fraud. | | | Document and define technical and organisational measures that will reduce security risks. | To detect and correct security vulnerabilities. | You should assess the trade-off between explainability of your model and the risk of a security breach.<br>You should proactively monitor your AI system and investigate any anomalies.<br>You should introduce real-time monitoring techniques that can detect anomalies (eg 'rate limiting' which reduces the number of queries that can be performed by a particular user in a given time limit).<br>You could deny anonymous use of your AI system by implementing processes that require user identity.<br>You could employ someone to regularly debug your model. | How should we assess security and data minimisation in AI? \| ICO | | | | | |
| Deployment and monitoring | 4.4 | Data minimisation<br>Article 5(1)(c );<br>Article 5(1)(e ) | Model drift is caused by training data no longer being relevant or adequate. As a consequence, individuals suffer from unlawful and unfair processing. | | | Document and define mechanisms to monitor the performance of your model. Where model drift is identified, assess, and delete (or anonymise) training data that is inadequate or irrelevant to your model's performance. | To detect and correct any inadequate or irrelevant personal data. | You must regularly assess drift and retrain the model on new data where necessary.<br>You should decide and document appropriate thresholds for determining whether your model needs to be retrained, based on the nature, scope, context and purposes of the processing and the risks it poses. | How should we assess security and data minimisation in AI? \| ICO | | | | | |
| Deployment and monitoring | 4.5 | Meaningful human review<br>Article 22 | Non-meaningful human review is caused by automation bias or a lack of interpretability. As a result, individuals may not be able to exercise their right to not be subject to | | | Document and define measures to ensure human review remains meaningful. | To ensure human reviewers are able to carry out their function meaningfully. | You could periodically test whether a human reviewer identifies an intentionally inaccurate decision. | How do we ensure individual rights in our | | | | | |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 4.5 | | solely automated decision-making with legal or similarly significant effects. | | | | | You could maintain a log of all automated decisions that were overridden by a human reviewer and the reasons why. | AI systems? \| ICO | | | | | |
| Deployment and monitoring | 4.6 | Purpose limitation<br>Article 5(1)(b), Recital 39, Article 6(4) and Recital 50 and Article 30 | Incompatible or repurposed processing is caused by a shift in how the AI system is deployed. As a result, individuals' lose control over how their data is used, become uninformed and lose trust in the organisation handling their personal data. | | | Assess any changes in the purpose of your AI system and ensure any changes meet legal requirements. | To detect any changes in purposes and to ensure that processing remains lawful. | You must ensure that if you plan to use or disclose personal data for any purpose that is additional to or different from the originally specified purpose, the new use is fair, lawful and transparent.<br>You should regularly review your processing, documentation and privacy notices to check that your purposes have not evolved over time beyond those you originally specified | Principle (b): Purpose limitation \| ICO | | | | | |

# Glossary

| | |
|---|---|
| Accuracy (see the ICO's guidance on the accuracy principle for further information) | ‘Accuracy’ in a data protection context is a fundamental principle requiring you to ensure that personal data is accurate and, where necessary, kept up to date. It requires you to take all reasonable steps to make sure the personal data you process is not ‘incorrect or misleading as to any matter of fact’ and, where necessary, is corrected or deleted without undue delay. |
| De-identification techniques | De-identification is the process used to prevent a person’s identity from being connected with information. There are various techniques to apply de-identification. In most cases, deidentification is not anonymisation, but it’s still useful as a data minimisation technique. |
| Edge case | An edge case is a problem or situation that occurs only at an extreme (maximum or minimum) operating parameter. In programming, an edge case typically involves input values that require special handling in an algorithm. |
| Fairness (see the ICO's guidance on the fairness principle for further information) | In a data protection context, ‘fairness’ means handling personal data in ways people reasonably expect and not use it in ways that have unjustified adverse effects on them. |
| High dimensional data | High dimensional data refers to a dataset in which the number of features, p, is larger than the number of observations, N. |
| Meaningful human review (see the ICO's guidance on rights related to automated decision-making including profiling for further information) | Key considerations for meaningful human review from the ICO are:<br>1) human reviewers must be involved in checking the system’s recommendation and should not just apply the automated recommendation to an individual in a routine fashion;<br>2) reviewers’ involvement must be active and not just a token gesture. They should have actual ‘meaningful’ influence on the decision, including the ‘authority and competence’ to go against the recommendation; and<br>3) reviewers must ‘weigh-up’ and ‘interpret’ the recommendation, consider all available input data, and also take into account other additional factors. |
| Model debugging | Model debugging attempts to test ML models like code and to probe sophisticated ML response functions and decision boundaries to detect and correct accuracy, fairness, security, and other problems in ML systems |
| Model drift | Model drift occurs when the accuracy of predictions produced from new input values “drifts” from the performance during the training period. Two main categories of model drift are:<br><br>Concept drift: When the statistical properties of the target (dependent) variable change<br>Data drift: When the statistical properties of the independent variables change (example: feature distributions, correlations between variables) |
| Overfitting | In statistics, overfitting is the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably. |
| Re-weighting | Weighting is applied to input data that is fed into an artificial neural network to produce related outputs. Re-weighting is where the initial weighting is altered so that it produces a different related output. |
| Statistical accuracy | Broadly, statistical accuracy refers to how often an AI system guesses the correct answer, measured against correctly labelled test data. In many cases, the outputs of an AI system are not intended to be treated as factual information about the individual, but statistically informed guesses as to something which may be true about the individual now or in the future. |