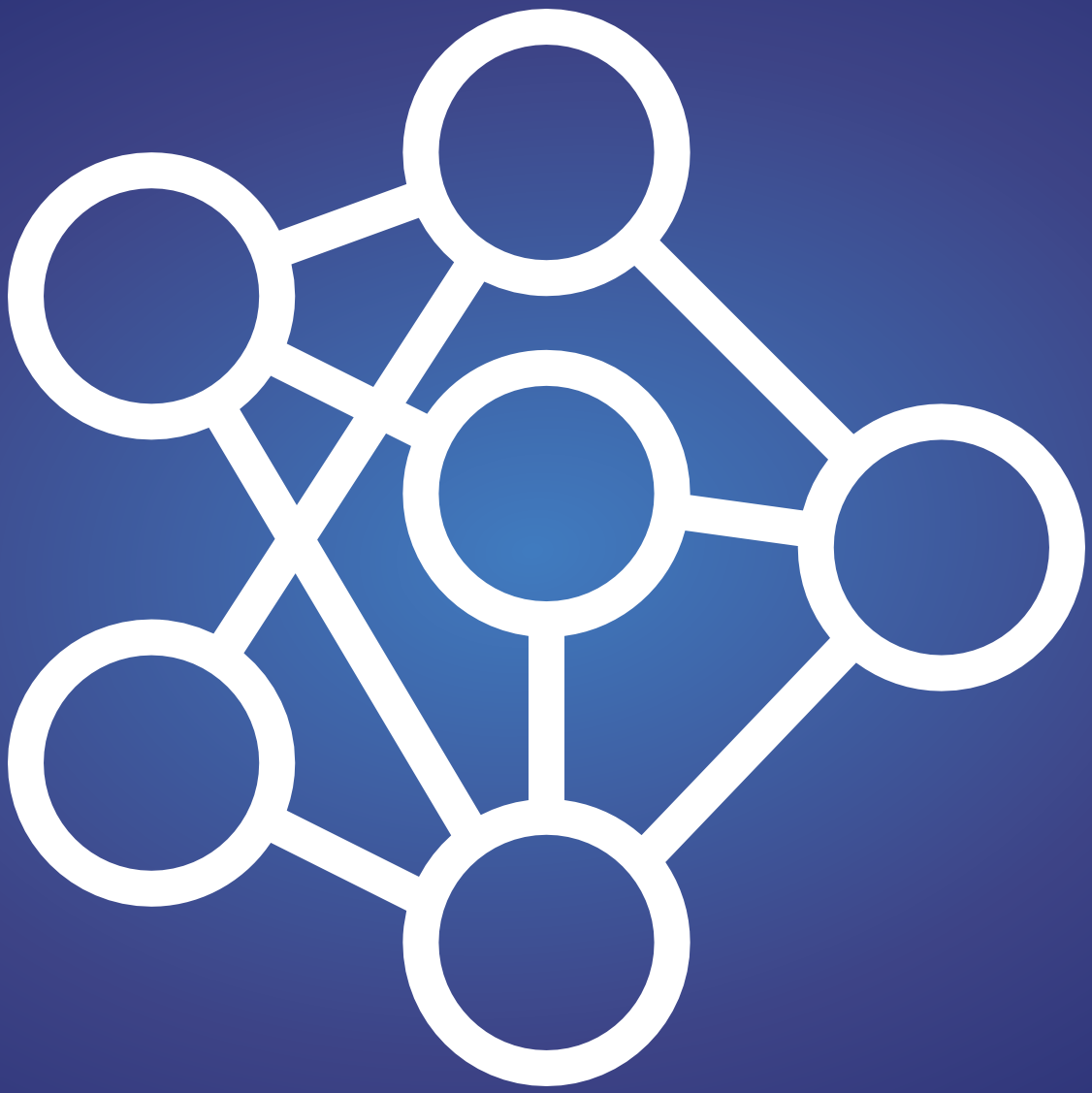
A guide to using artificial intelligence in the public sector



Office for Artificial Intelligence

Government Digital Service

THIS DOCUMENT IS AN EDITED VERSION OF THE FULL ONLINE GUIDE AVAILABLE ON GOV.UK

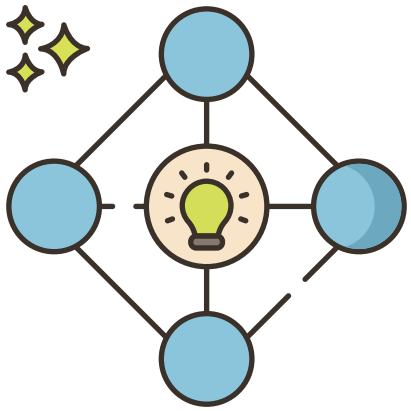
Version: S3BKR13R_V1.6

Contents

Understanding artificial intelligence

Artificial Intelligence (AI) has the potential to change the way we live and work

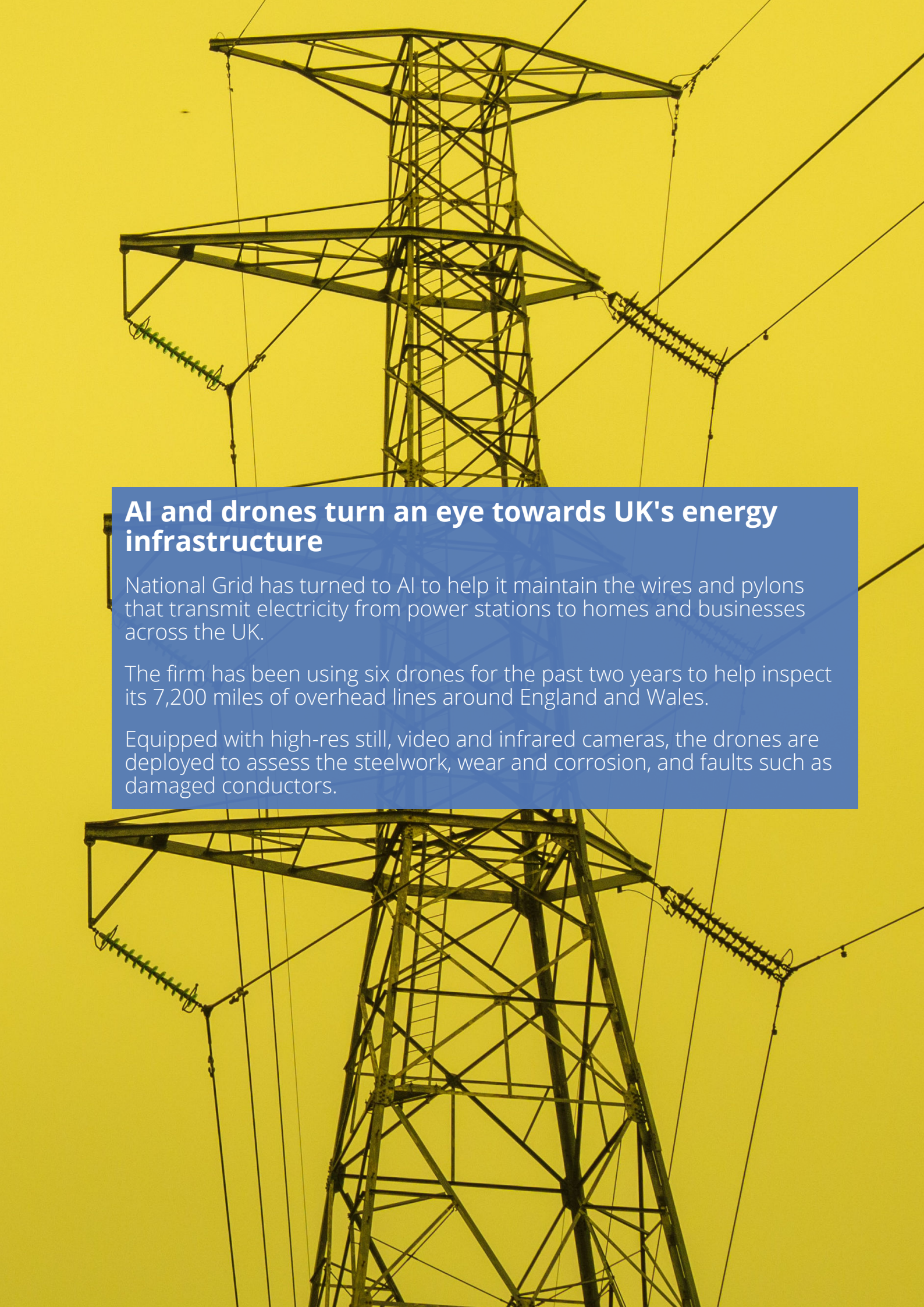
Embedding AI across all sectors has the potential to create thousands of jobs and drive economic growth. By one estimate, AI's contribution to the United Kingdom could be as large as 5% of GDP by 2030.[1]



A number of public sector organisations are already successfully using AI for tasks ranging from fraud detection to answering customer queries.

The potential uses for AI in the public sector are significant, but have to be balanced with ethical, fairness and safety considerations.

1 *The economic impact of artificial intelligence on the UK economy* (PwC, 2017)



AI and drones turn an eye towards UK's energy infrastructure

National Grid has turned to AI to help it maintain the wires and pylons that transmit electricity from power stations to homes and businesses across the UK.

The firm has been using six drones for the past two years to help inspect its 7,200 miles of overhead lines around England and Wales.

Equipped with high-res still, video and infrared cameras, the drones are deployed to assess the steelwork, wear and corrosion, and faults such as damaged conductors.

The government has set up two funds to support the development and uptake of AI systems, the:

- **GovTech Catalyst** to help public sector bodies take advantage of emerging technologies
- **Regulators' Pioneer Fund** to help regulators promote cutting-edge regulatory practices when developing emerging technologies

GovTech Catalyst

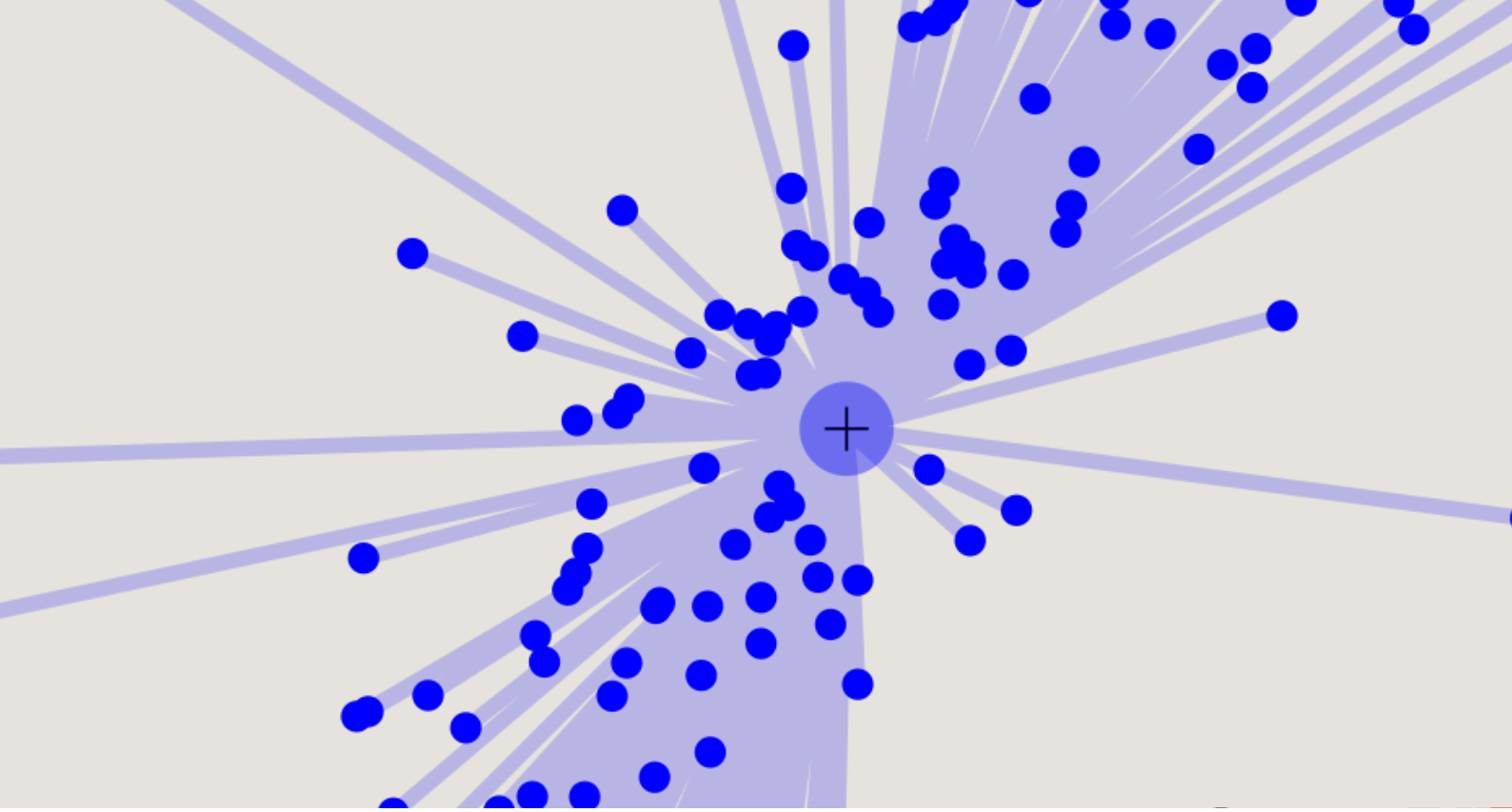
UKRI

AI and the public sector

Recognising AI's potential, the government's Industrial Strategy White Paper placed AI and Data as one of four Grand Challenges, supported by up to £950m in the AI Sector Deal.

The government has set up three new bodies to support the use of AI, build the right infrastructure and facilitate public and private sector adoption of these technologies. These three new bodies are the:

- **AI Council** an expert committee of independent members providing high-level leadership on implementing the AI Sector Deal
- **Office for AI** which works with industry, academia and the third sector to coordinate and oversee the implementation of the UK's AI strategy
- **Centre for Data Ethics and Innovation** which identifies the measures needed to make sure the development of AI is safe, ethical and innovative



Defining artificial intelligence

At its core, AI is a research field spanning philosophy, logic, statistics, computer science, mathematics, neuroscience, linguistics, cognitive psychology and economics.

AI can be defined as the use of digital technology to create systems capable of performing tasks commonly thought to require intelligence.

AI is constantly evolving, but generally it:

- involves machines using statistics to find patterns in large amounts of data
- is the ability to perform repetitive tasks with data without the need for constant human guidance

There are many new concepts used in the field of AI and you may find it useful to refer to a glossary of AI terms.

This guidance mostly discusses machine learning. Machine learning is a subset of AI, and refers to the development of digital systems that improve their performance on a given task over time through experience.

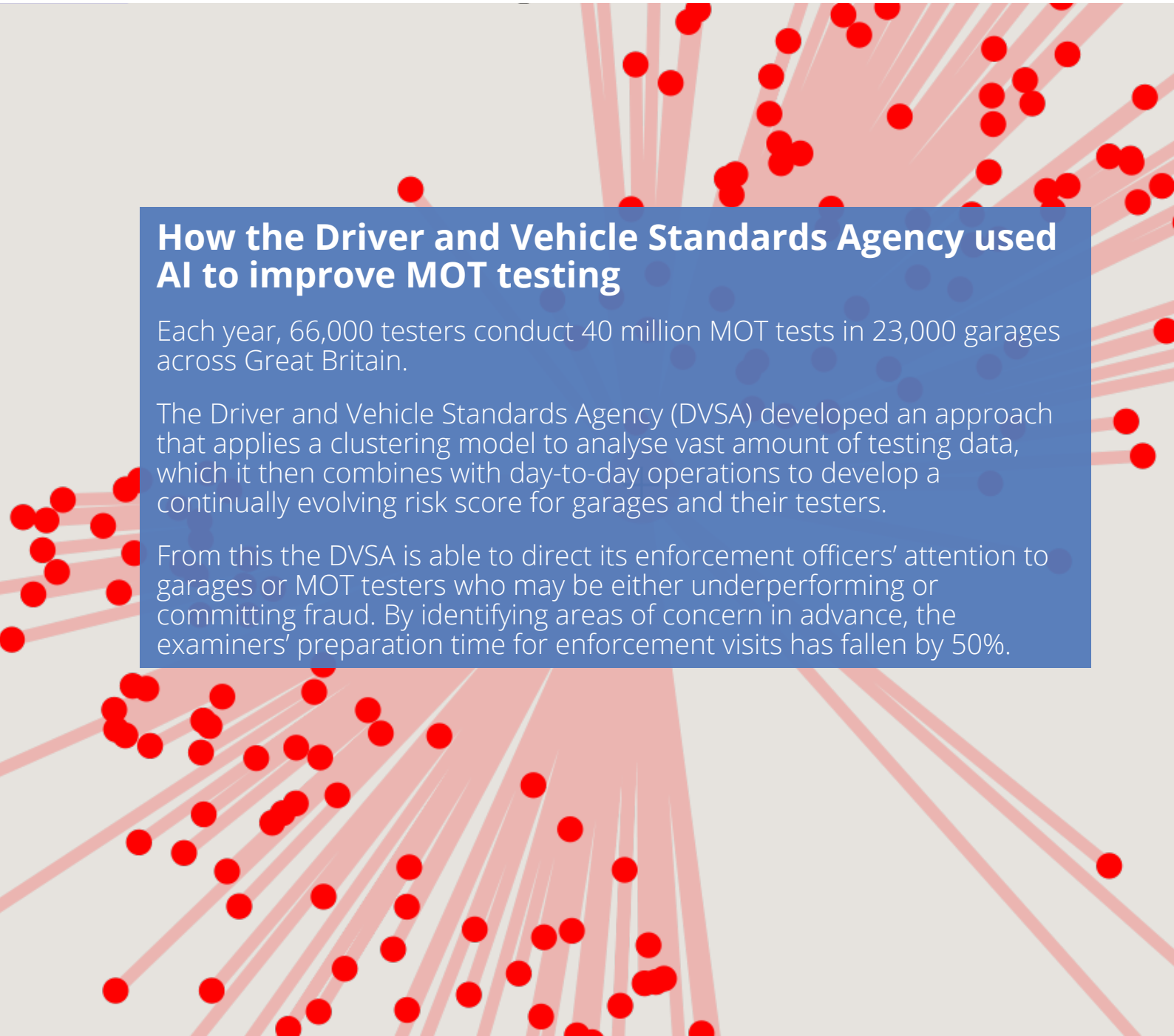
Machine learning is the most widely-used form of AI, and has contributed to innovations like self-driving cars, speech recognition and machine translation.

Recent advances in machine learning are the result of:

- improvements to algorithms
- increases in funding
- huge growth in the amount of data created and stored by digital systems
- increased access to computational power and the expansion of cloud computing

Machine learning can be:

- supervised learning which allows an AI model to learn from labelled training data, for example, training a model to help tag content on GOV.UK
- unsupervised learning which is training an AI algorithm to use unlabelled and unclassified information
- reinforcement learning which allows an AI model to learn as it performs a task



How the Driver and Vehicle Standards Agency used AI to improve MOT testing

Each year, 66,000 testers conduct 40 million MOT tests in 23,000 garages across Great Britain.

The Driver and Vehicle Standards Agency (DVSA) developed an approach that applies a clustering model to analyse vast amount of testing data, which it then combines with day-to-day operations to develop a continually evolving risk score for garages and their testers.

From this the DVSA is able to direct its enforcement officers' attention to garages or MOT testers who may be either underperforming or committing fraud. By identifying areas of concern in advance, the examiners' preparation time for enforcement visits has fallen by 50%.

Using satellite images to estimate populations

The Department for International Development partnered with the University of Southampton, Columbia University and the United Nations Population Fund to apply a random forest machine learning algorithm to satellite image and micro-census data.

The algorithm then used this information to predict the population density of an area. The model also used data from micro-censuses to validate its outputs and provide valuable training data for the model.

How AI can help

AI can benefit the public sector in a number of ways.
For example, it can:

- provide more accurate information, forecasts and predictions leading to better outcomes - for example, more accurate medical diagnoses
- produce a positive social impact by using AI to provide solutions for some of the world's most challenging social problems
- simulate complex systems that allow policy makers to experiment with different policy options and spot unintended consequences before committing to a measure
- improve public services - for example, providing service providers more-accurate and detailed information of citizens with similar needs or interests to provide personalised public services tailored to individual circumstances
- automate repetitive and time-consuming tasks which frees up valuable time of frontline staff

What AI cannot do

AI is not a general purpose solution which can solve every problem.

Current applications of AI focus on performing narrowly defined tasks. AI generally cannot:

- be imaginative
- perform well without a large quantity of relevant, high quality data
- infer additional context if the information is not present in the data

Even if AI can help you meet some user needs, simpler solutions may be more effective and less expensive. For example, optical character recognition technology can extract information from scans of passports. However, a digital form requiring manual input might be more accurate, quicker to build, and cheaper. You'll need to investigate alternative mature technology solutions thoroughly to check if this is the case.

Follow the *Choosing technology: an introduction Service Manual's* guidance on choosing an appropriate technology.

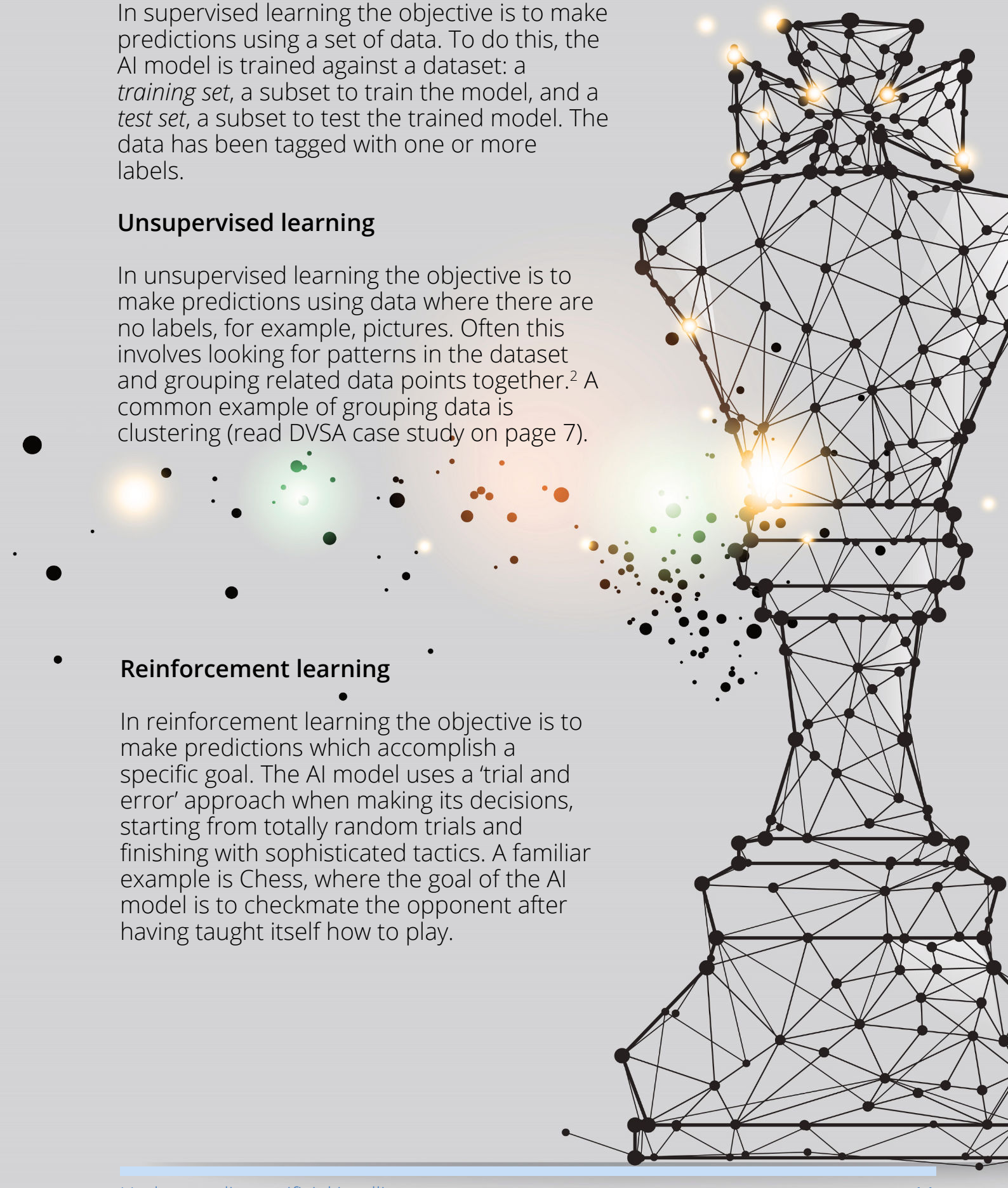
Teaching a machine new tricks

Supervised learning

In supervised learning the objective is to make predictions using a set of data. To do this, the AI model is trained against a dataset: a *training set*, a subset to train the model, and a *test set*, a subset to test the trained model. The data has been tagged with one or more labels.

Unsupervised learning

In unsupervised learning the objective is to make predictions using data where there are no labels, for example, pictures. Often this involves looking for patterns in the dataset and grouping related data points together.[2] A common example of grouping data is clustering (read DVSA case study on page 7).



Reinforcement learning

In reinforcement learning the objective is to make predictions which accomplish a specific goal. The AI model uses a 'trial and error' approach when making its decisions, starting from totally random trials and finishing with sophisticated tactics. A familiar example is Chess, where the goal of the AI model is to checkmate the opponent after having taught itself how to play.

2 Refer to Salian, I. ***SuperVize Me*** (NVIDIA, Accessed: 2 August 2018) for further exploration of this topic www.blogs.nvidia.com/blog/2018/08/02/supervised-unsupervised-learning/

Considerations for using AI to meet user needs

With an AI project you should consider a number of factors, including AI ethics and safety. These factors span safety, ethical, legal and administrative concerns and include, but are not limited to:

- **data quality** - the success of your AI project depends on the quality of your data
- **fairness** - are the models trained and tested on relevant, accurate, and generalisable datasets and is the AI system deployed by users trained to implement them responsibly and without bias
- **accountability** - consider who is responsible for each element of the model's output and how the designers and implementers of AI systems will be held accountable
- **privacy** - complying with appropriate data policies, for example, the General Data Protection Regulations (GDPR) and the Data Protection Act 2018
- **explainability and transparency** - so the affected stakeholders can know how the AI model reached its decision
- **costs** - consider how much it will cost to build, run and maintain an AI infrastructure, train and educate staff and if the work to install AI may outweigh any potential savings

3 *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation* www.ec.europa.eu/newsroom/article29/itemdetail.cfm?item_id=612053

Ensuring your use of AI is compliant with data protection laws

You'll need to make sure your AI system is compliant with GDPR and the Data Protection Act 2018 (DPA 2018), including the points which relate to automated decision making. We recommend discussing this with legal advisors.

Automated decisions in this context are decisions made without human intervention, which have legal or similarly significant effects on 'data subjects'. For example, an online decision to award a business grant.

If you want to use automated processes to make decisions with legal or similarly significant effects on individuals you must follow the safeguards laid out in the GDPR and DPA 2018. This includes making sure you provide users with:

- specific and easily accessible information about the automated decision-making process
- a simple way to obtain human intervention to review, and potentially change the decision

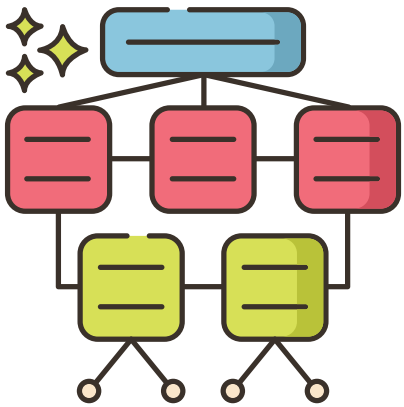
Remember to make sure your use of automated decision-making does not conflict with any other laws or regulations.

You should consider both the final decision and any automated decisions which significantly affected the decision-making process.

Read the *Working Party guidance*[3] on automated individual decision making and profiling for more information.

Assessing if AI is the right solution

AI is just another technology tool to help deliver services. Designing any service starts with identifying user needs. If you think AI may be an appropriate technology choice to help you meet user needs, you will need to consider your data and the specific technology you want to use. Your data scientists will then use your data to build and train an AI model.



When assessing if AI could help you meet users' needs, consider if:

- there's data containing the information you need, even if disguised or buried
- it's ethical and safe to use the data - refer to the *Data Ethics Framework*[4]
- you have a the right sort of data for the AI model to learn from
- the task is large scale and repetitive enough that a human would struggle to carry it out
- it would provide information a team could use to achieve outcomes in the real world

It's important to remember that AI is not an all-purpose solution. Unlike a human, AI cannot infer, and can only produce an output based on the data a team inputs to the model.

4 ***The Data Ethics Framework*** www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework

Working with the right skills to assess AI

When identifying whether AI is the right solution, it's important that you work with:

- specialists who have a good knowledge of your data and the problem you're trying to solve, such as data scientists
- at least one domain knowledge expert who knows the environment where you will be deploying the AI model results

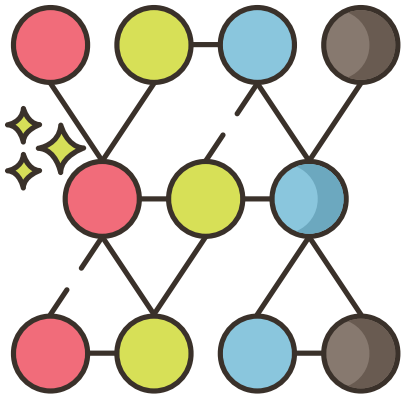
Getting approval to spend

Because of its experimental and iterative nature, it can be difficult to specify the precise benefits which could come from an AI project. To explore this uncertainty and provide the right level of information around the potential benefits, you can:

- carry out some initial analysis on your data to help you understand how hard the problem is and how likely the project's success would be
- build your business case around a small-scale proof of concept (PoC) and use its results to prove your hypothesis

Once you have secured budget, you'll need to allow enough time and resources to conduct a substantial discovery to show feasibility. Discovery for projects using AI can often takes longer for similar projects that do not use AI.

If your organisation is a central government department, you may have to get approval from the GDS to spend money on AI. At this point most AI projects are classified as 'novel', which requires a high level of scrutiny. You should contact the GDS Standards Assurance team[5] for help on the spend controls process.



5 GDS Standards Assurance team contact: www.gov.uk/contact/govuk

Consider your current data state

For your AI model to work, it often needs access to a large quantity of data, and more importantly the right kind of data. Work with specialists who have the knowledge of your data, such as data scientists, to assess your data state. You can assess whether your data is high enough quality for AI using a combination of:

- accuracy
- completeness
- uniqueness
- timeliness
- validity
- sufficiency
- relevancy
- representativeness
- consistency

If your problem involves supporting an ongoing business decision process, you will need to plan to establish ongoing, up-to-date access to data. Remember to follow data protection laws.

Deciding whether to build or buy

When assessing if AI could help you meet user needs, consider how you will procure the technology. You should define your purchasing strategy in the same way as you would for any other technology. Whether you build, buy or reuse (or combine these approaches) will depend on a number of considerations, including:

- whether the needs you're trying to meet are unique to your organisation or you could fulfil users' needs with generic components
- the maturity of commercially available products that meet those needs
- how your product needs to integrate with your existing infrastructure

It is also important to address ethical concerns about the use of AI from the start of the procurement process.

The Office for AI and the World Economic Forum are developing further guidance on AI procurement.[6]

6 ***Guidelines for AI procurement*** www.gov.uk/government/publications/draft-guidelines-for-ai-procurement

Build your AI solution

Your team can build or adapt off-the-shelf AI models or open source algorithms in-house.

When making this decision, you should work with data scientists to consider whether:

- your team has the skills to build an AI project in-house
- your operations team can run and maintain an in-house AI solution

Buy your AI solution

You may be able to buy your AI technology as an off-the-shelf product. This is most suitable if you are looking for a common application of AI, for example, optical character recognition. However, buying your AI technology may not always be suitable as the specifics of your data and needs could mean the supplier would have to build from scratch or significantly customise an existing AI model.

Your AI solution will still need to be integrated into an end-to-end service for your users, even if you are able to buy significant components off the shelf.

Choosing AI technology for your challenge

There is no one 'AI technology'. Currently, widely-available AI technologies are mostly either supervised, unsupervised or reinforcement machine learning (refer to page 11 for definitions). The machine learning techniques that can provide you with the best insight depends on the problem you're trying to solve.

Machine learning technique	Description	Examples of machine learning technique
Classification	Learns the characteristics of a given category, allowing the AI model to classify unknown data points into existing categories	• deciding if a consignment of goods undergoes border inspection • deciding if an email is spam or not
Regression	Predicts a value for an unknown data point	• predicting the market value of a house from information such as its size, location, or age • forecasting the concentrations of air pollutants in cities
Clustering	Identifies groups of similar data points in a dataset	• grouping retail customers to find subgroups with specific spending habits • clustering smart-meter data to identify groups of electrical appliances, and generate itemised electricity bills
Dimensionality Reduction or Manifold Learning	Narrows down the data to the most relevant variables to make models more accurate, or make it possible to visualise the data	• used by data scientists when evaluating and developing other types of machine learning algorithms
Ranking	Trains an AI model to rank new data based on previously-seen lists	• returning pages by order of relevance when a user searches a website

Common applications of machine learning

There are certain types of problems for which machine learning is commonly used. For some of these you will be able to buy or adapt commercially available products.

Machine learning application	Description	Examples of machine learning application
Natural language processing (NLP)	Processes and analyses natural language, recognising words, their meaning, context and the narrative	• converting speech into text for automatic subtitles generation • automatically generating a reply to a customer's email
Computer vision	The ability of a machine or program to emulate human vision	• identification of road signs for self-driving vehicles • face recognition for automated passport controls
Anomaly detection	Finds anomalous data points within a dataset	• identifying fraudulent activity in a user's bank account
Time-series analysis	Understanding how data varies over time to conduct forecasting and monitoring	• conducting budget analyses • forecasting economic indicators
Recommender systems	Predicts how a user will rate a given item to make new recommendations	• suggesting relevant pages on a website, given the articles a user has previously viewed

Allocating responsibility and governance for AI projects

When using AI it's important to understand who is responsible if the system fails, as the problem may lie in a number of areas. For example, failures with the data chosen to train the AI model, design of the model, coding of the software, or deployment.

You should establish a responsibility record which sets out who is responsible for different areas of the AI system. It would be useful to consider whether:

- the models are achieving their purpose and business objectives
- there is a clear accountability framework for models in production
- there is a clear testing and monitoring framework in place
- your team has reviewed and validated the code
- the algorithms are robust, unbiased, fair and explainable
- the project fits with how citizens and users expect their data to be used

Depending on your organisation's maturity, it may be useful to set up a dedicated board, committee or forum to handle AI training data and model governance.

Recording accountability

It can be useful to keep a central record of all AI technologies you use, listing:

- where an AI model is in use
- what the AI model is used for
- who's involved
- how it's assessed or checked
- what other teams rely on the technology

National Grid and The Alan Turing Institute improve solar forecasting

The National Grid Electricity System Operator (ESO) balances the electricity system in real time, ensuring the nation's supply always meets demand. This balancing act becomes more challenging as wind and solar power become a larger part of the overall energy mix, as their generation output is hard to predict.

An innovation project between ESO and The Alan Turing Institute used a mix of machine learning prediction methods and computational statistics to achieve a big improvement in forecast accuracy. One result found the solar forecasting system 33% more accurate at day-ahead forecasts. Improved foresting helps ESO run the grid more efficiently, which ultimately means lower bills for households.

Planning and preparing for AI systems implementation

Planning your project

As with all projects, you need to make sure you're hypothesis-led and can constantly iterate to best help your users and their needs.

You should integrate your AI systems development with your wider project phases.

1. **Discovery** - consider your current data state, decide whether to build, buy or collaborate, allocate responsibility for AI models, assess your existing data, build your AI team, get your data ready for AI, and plan your AI modelling phase.

2. **Alpha** - build and evaluate your machine learning model.

3. **Beta** - deploy and maintain your model.

You should consider AI ethics and safety throughout all phases.

Significant time is needed to understand the feasibility of using your data in a new way. This means the discovery phase tends to be longer and more expensive than for services without AI.

Your data scientists may be familiar with a lifecycle called CRISP-DM[7] and may wish to integrate parts of it into your project.

Discovery can help you understand the problem that needs to be solved.

7 CRISP-DM www.ibm.com/support/knowledgecenter/SS3RA7_15.0.0/com.ibm.spss.crispdm.help/crisp_overview.htm

Start your discovery phase

Assess your user needs and data sources

You should:

- thoroughly understand the problem and the needs of different users
- assess whether an AI system is the right tool to address the user needs
- understand the processes and how the AI model will connect with the wider service
- consider the location and condition of the data you will use

Assess your existing data

To prepare for your AI project, you should assess your existing data. Training an AI system on error-strewn data can result in poor results due to:

- the dataset not containing clear patterns for the model to explore when making a prediction
- the dataset containing clear but accidental patterns, resulting in the model learning biases

You can use a combination of accuracy, completeness, uniqueness, timeliness, validity, relevancy, representativeness, sufficiency or consistency to see if the data is high enough quality for an AI system to make predictions from.[8]

When assessing your AI data, it's useful to collaborate with someone who has deep knowledge of your data, such as a data scientist. They will be familiar with the best practice for measuring, cleaning and maintaining good data standards for ongoing projects. Make your data proportionate to user needs and understand the limitations of the data to help you assess your data readiness.

Questions for you to consider with data scientists are:

- do you have enough data for the model to learn from?
- do you understand the onward effects of using data in this way?

8 Refer to ***Data Fairness*** page 79 of ***Understanding artificial intelligence ethics and safety*** (The Alan Turing Institute, 2019) for further exploration of these topics

- is the data accurate and complete and how frequently is the data updated?
- is the data representative of the users the model's results will impact?
- was the data gathered using suitable, reliable, and impartial sources of measurement?
- is the data secure and do you have permission to use it?
- what modelling approaches could be suitable for the data available?
- do you have access to the data and how quickly can you access it?
- where is the data located?
- what format is the data in and does it require significant preparation to be ready for modelling?
- is your data structured - for example, can you store it in a table, or unstructured such as emails or webpages?
- are there any constraints on the data - for example, does it contain sensitive information such as home addresses?
- can you link key variables within and between datasets?
- If you're unsure about your use of data, consult the *Data Ethics Framework guidance*[9] to check your project is a safe application and deployment of AI models.

Build your team for AI implementation

As with other projects, your team should be multidisciplinary, with a diverse combination of roles and skills to reduce bias and make sure your results are as accurate as possible. When working with AI you may need specialist roles such as a:

- data architect to set the vision for the organisation's use of data, through data design, to meet business needs
- data scientist to identify complex business problems while leveraging data value - often having at least two data scientists working on a project allows them to better collaborate and validate AI experiments

9 ***The Data Ethics Framework*** www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework

- data engineer to develop the delivery of data products and services into systems and business processes
- ethicist to provide ethical judgements and assessments on the AI model's inputs
- domain expert who knows the environment where you will be deploying the AI model results - for example, if the model will be investigating social care, collaborate with a social worker

You may not need all of these roles from the very beginning, but this may change as the work progresses. You may want to break up your discovery into smaller phases so you can evaluate what you are learning.

It can be useful for your team to have:

- experience of solving an AI problem similar to the one you're solving
- commercial experience of AI - understanding of machine learning techniques and algorithms, including production deployments at scale
- an understanding of cloud architecture, security, scalable deployment and open source tools and technologies
- hands-on experience of major cloud platforms
- experience with containers and container orchestrations - for example, Docker and Kubernetes
- experience in or strong understanding of the fundamentals of computer science and statistics
- experience in software development - for example Python, R or Scala
- experience building large scale backend systems
- hands-on experience with a cluster-computing framework - for example, Hadoop or Spark
- hands-on experience with data stores - for example, SQL and No-SQL
- technical understanding of streaming data architectures
- experience of working to minimise bias from data

Managing infrastructure and suppliers

When preparing for AI implementation, you should identify how you can best integrate AI with your existing technology and services.

It's useful to consider how you'll manage:

- data collection pipelines to support reliable model performance and a clean input for modelling, such as batch upload or continuous upload
- storing your data in databases and how the type of database you choose will change depending on the complexity of the project and the different data sources required
- data mining and data analysis of the results
- any platforms your team will use to collate the technology used across the AI project to help speed up AI deployment

When choosing your AI tools, you should bring in specialists, such as data scientists or technical architects to assess what tools you currently have to support AI.

Use Cloud First when setting up your infrastructure.[10]

Consider the benefits of AI platforms

A data science platform is a type of software tool which helps teams connect all of the technology they require across their project workflow, speeding up AI deployment and increasing the transparency and oversight over AI models.

When deciding on whether to use a data science platform, it's useful to consider how the platform can:

- provide access to flexible computation which allows teams to have secure access to the power needed to process large amounts of data
- help your team build workflows for accessing and preparing datasets and allow for easy maintenance of the data
- provide common environments for sharing data and code so the team can work collaboratively
- let your teams clearly share their output through dashboards and applications

10 ***Cloud First*** www.gov.uk/guidance/use-cloud-first

- provide a reproducible environment for your teams to work from
- help control and monitor project-specific or sensitive permissions

Preparing your data for an AI model

After you've assessed your current data quality, you should prepare your data to make sure it is secure and unbiased. You may find it useful to create a data factsheet during discovery to keep a record of your data quality.

Ensuring diversity in your data

In the same way you should have diversity in your team, your data should also be diverse and reflective of the population you are trying to model. This will reduce conscious or unconscious bias. Alongside this, a lack of diverse input could mean certain groups are disadvantaged, as the AI model may not cater for a diverse set of needs. You should read the Data Ethics Framework guidance to understand the limitations of your data and how to recognise any bias present.

You should also:

- evaluate the accuracy of your data, how it was collected, and consider alternative sources
- consider the social context of where, when and how the system is being deployed
- consider if any particular groups might be at an advantage or disadvantage in the context in which the system is being deployed

Keeping your data secure

Make sure you design your system to keep data secure. To help keep data safe:

- follow the *National Cyber Security Centre's guidance (www.ncsc.gov.uk)* on using data with AI models
- make sure your system is compliant with GDPR and DPA 2018[11]

As with any other software, you should design and build modular, loosely coupled systems which can be easily iterated and adapted.

11 ***Data Protection*** www.gov.uk/data-protection

Writing and training algorithms can take a lot of time and computational power. In addition to ongoing cost, you'll need to think about the network and memory resources your team will need to train your model.

Using historic data

Most of the data in government available to train our models is within legacy systems which might contain bias and might have poor controls around it. For legacy systems to be compatible with AI technology, you will often need to invest a lot of work to bring your legacy systems up to modern standards.

You'll also need to carefully consider the ethical and legal implications of working with historic data and whether you need to seek permission to use this information.

Evaluate your data preparation phase

When you complete your data preparation phase you should have:

- a dataset ready for modelling in a technical environment
- a set of features (measurable properties) generated from the raw dataset
- a data quality assessment using a combination of accuracy, bias, completeness, uniqueness, timeliness/currency, validity or consistency

Researching the end to end service

During the discovery phase, you should explore the needs of the users of the end to end service. Like other digital services, you'll use this phase to determine whether there's a viable service you could build that would solve user needs, and that it's cost-effective to pursue the problem.

You'll be able to check guidance on how to know when your discovery is finished before moving on to alpha.

Moving to the alpha phase

Plan and prototype your AI model build and service

If you have decided to build your AI model in-house, you should follow these steps.

1. Split the data.
2. Create a baseline model.
3. Build a prototype of the model and service.
4. Test the model and service.
5. Evaluate the model.
6. Assess and refine performance.

Split the data

Your team will need to train the models they build on data. Your team should split your data into a:

- training set to train algorithms during the modelling phase
- validation set for assessing the performance of your models
- test set for a final check on the performance of your best model

Create a baseline model

Your team should build a simple baseline version model before they build any more complex models. This provides a benchmark that your team can later compare more complex models against, and will help your team identify problems in your data.

Build a prototype of the model and service

Once you have a baseline model, your team can start prototyping more complex models. This is a highly iterative process, requiring substantial amounts of data, and will see your team probably build a number of AI models before deciding on the most effective and appropriate algorithm for your problem.

Keeping your team's first model simple and setting up the right end-to-end infrastructure will help smooth the transition from alpha to beta. You can action this by focusing on the infrastructure requirements for your AI pipelines as the same time your team is developing the model. Your simple model will provide baseline metrics and information on the model's behaviour that you can use to test more complex models.

Test the model and service

Your team will need to test your models throughout the process to mitigate against issues such as overfitting or underfitting that could undermine your model's effectiveness once deployed.

Your team should only use the test set on your best model. Keep this data separate from your models until this final test. This test will provide you with the most accurate impression of how your model will perform once deployed.

Evaluate the model

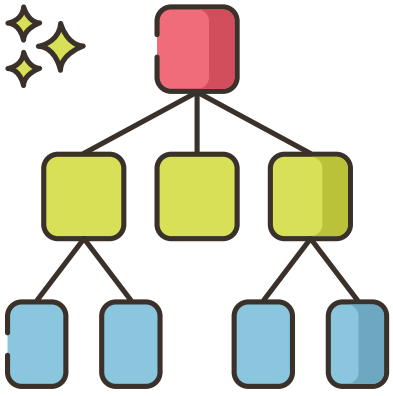
Your team will need to evaluate your model to assess how it is performing against unseen data. This will give you an indication of how your model will perform in the real world.

The best evaluation metric will depend on the problem you are trying to solve, and your chosen model. While you should select the evaluation metric with data scientists, you should also consider the ethical, economical and societal implications. These considerations make the fine tuning of AI systems relevant to both data scientists and delivery leads.

Choose the final model

When choosing your final model, you will need to consider:

- what level of performance your problem needs
- how interpretable you need your model to be
- how frequently you need predictions or retraining
- the cost of maintaining the model



Assess and refine performance

Once you select a final model, your team will need to assess its performance, and refine it to make sure it performs as well as you need it to. When assessing your model's performance consider:

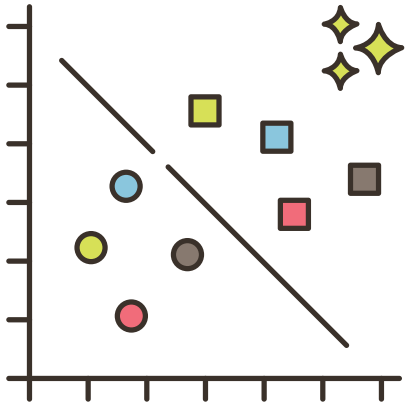
- how it performs compared to simpler models
- what level of performance you need before deploying the model
- what level of performance you can justify to the public, your stakeholders, and regulators
- what level of performance similar applications deliver in other organisations
- whether the model shows any signs of bias

If a model does not outperform human performance, it still may be useful. For example, a text classification algorithm might not be as accurate as a human when classifying documents, however they can perform at a far higher scale and speed than a human.

Evaluate your Alpha phase

When you complete building your AI prototyping phase, you should have:

- a final model or set of predictive models and a summary of their performance and characteristics
- a decision on whether or not to progress to the beta phase
- a plan for your beta phase



Moving to the beta phase

Moving from alpha to beta involves integrating the model into the service's decision-making process and using live data for the model to make predictions on.

Using your model in your service has three stages.

1. **Integrating your model** - performance-test the model with live data and integrate it within the decision-making workflow. Integration can happen in a number of ways, from a local deployment to the creation of a custom application for staff or customers. This decision is dependent on your infrastructure and user requirements.

2. **Evaluating your model** - undertake continuous evaluation to make sure the model still meets business objectives and the model is performing at the level required. This will make sure the model performance is in line with the modelling phase and to help you identify when to retrain the model.

3. **Helping users** - make sure users feel confident in using, interpreting, and challenging any outputs or insights generated by the model.

You should continue to collect user needs so your team can use the model's outputs in the real world.

When moving from alpha to beta, there are some best-practice guidelines to smooth the transition.

Iterate and deploy improved models

After creating a beta version, you team can use automated testing to create some high-level tests before moving to more thorough testing. Working in this way means you can launch new improvements without worrying about functionality once deployed.

Maintain a cross-functional team

During alpha, you will have relied mostly on data scientists to assess the opportunity and your data state.

Moving to beta needs specialists with a strong knowledge of dev-ops, servers, networking, data stores, data management, data governance, containers, cloud infrastructure and security design.

This skillset is likely to be better suited to an engineer rather than a data scientist so maintaining a cross-functional team will help smooth the transition from alpha to beta.

When you complete your beta phase, you should have:

- AI running on top of your data, learning and improving its performance, and informing decisions
- a monitoring framework to evaluate the model's performance and rapidly identify incidents
- launched a private beta followed by a public end-to-end beta prototype which users can use in full
- found a way to measure your service's success using new data you've got during the beta phase
- evidence that your service meets government accessibility requirements
- tested the way you've designed assisted digital support for your service

Managing your AI systems implementation project

Governance when running your AI systems implementation project

Safety

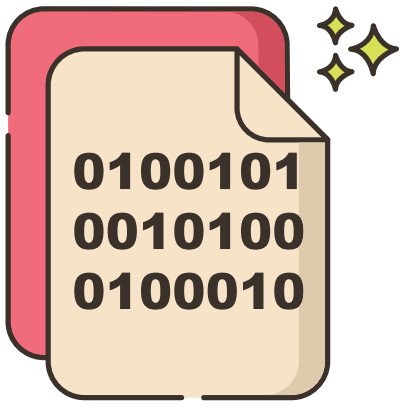
Governance in safety is important to make sure the model shows no signs of bias or discrimination. You can consider whether:

- the algorithm is performing in line with safety and ethical considerations
- the model is explainable
- there is an agreed definition of fairness implemented in the model
- the data use aligns with the Data Ethics Framework
- the algorithm's use of data complies with privacy and data processing legislation

Purpose

Governance in purpose makes sure the model is achieving its purpose/ business objectives. You can consider whether:

- the model solves the problem identified
- how and when you will evaluate the model
- the user experience aligns with existing government guidance



Accountability

Governance in accountability provides a clear accountability framework for the model. You can consider:

- whether there is a clear and accountable owner of the model
- who will maintain the model
- who has the ability to change and modify the code

Testing and monitoring

Governance in testing and monitoring makes sure a robust testing framework is in place. You can consider:

- how you will monitor the model's performance
- who will monitor the model's performance
- how often you will assess the model

Public narrative

Governance in public narrative protects against reputational risks arising from the application of the model. You can consider whether:

- the project fits with the government organisation's use of AI systems
- the model fits with the government organisation's policy on data use
- the project fits with how citizens/users expect their data to be used

Quality assurance

Governance in quality assurance makes sure the code has been reviewed and validated. You can consider whether:

- the team has validated the code
- the code is open source

Managing risk in your AI systems implementation project

Risk	How to mitigate
Project shows signs of bias or discrimination	Make sure your model is fair, explainable, and you have a process for monitoring unexpected or biased outputs
Data use is not compliant with legislation, guidance or the government organisation's public narrative	Consult guidance on preparing your data for AI
Security protocols are not in place to make sure you maintain confidentiality and uphold data integrity	Build a data catalogue to define the security protocols required
You cannot access data or it is of poor quality	Map the datasets you will use at an early stage both within and outside your government organisation. It's then useful to assess the data against criteria for a combination of accuracy, completeness, uniqueness, relevancy, sufficiency, timeliness, representativeness, validity or consistency
You cannot integrate the model	Include engineers early in the building of the AI model to make sure any code developed is production-ready
There is no accountability framework for the model	Establish a clear responsibility record to define who has accountability for the different areas of the AI model

Additional sources and reference

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

Searching for superstars isn't the answer. How organizations can build world-class analytics teams that deliver results (Deloitte, 2018)

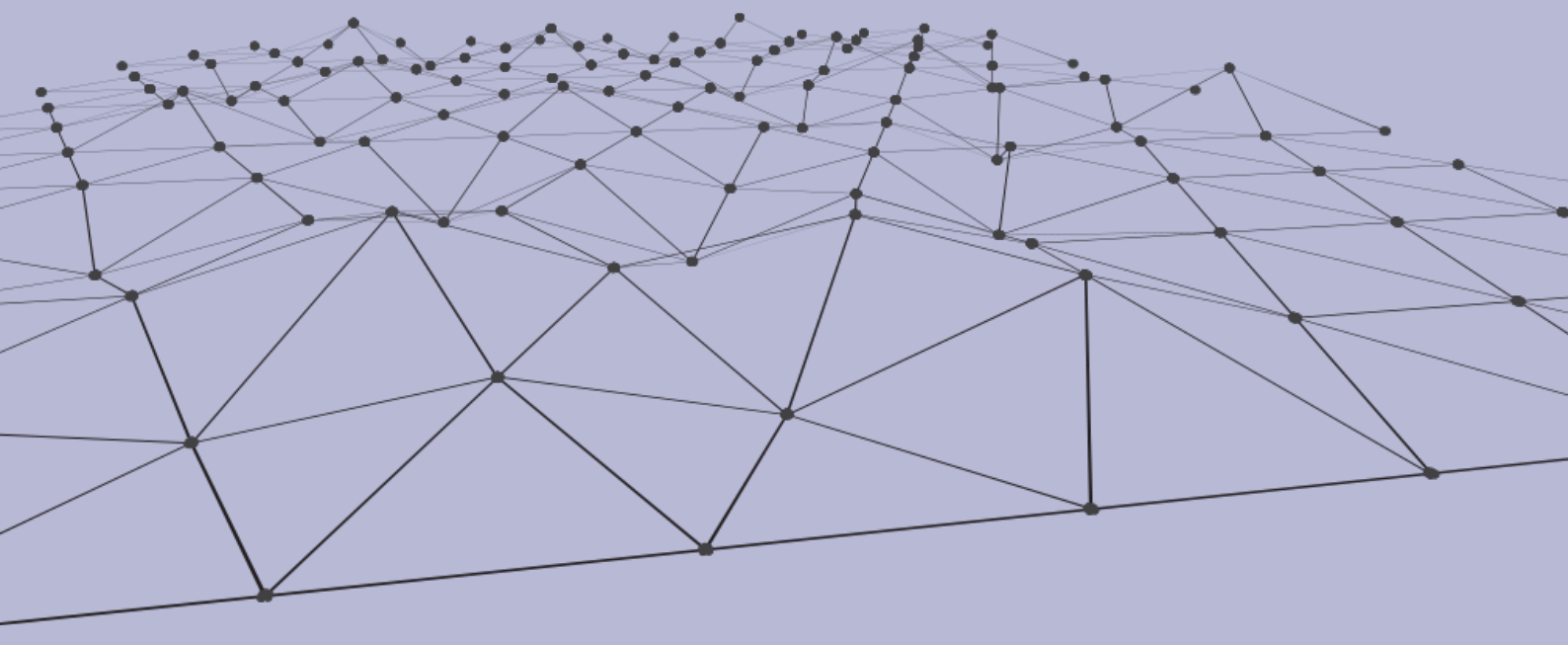
Examples of real-world artificial intelligence use
www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector#examples-of-artificial-intelligence-use

Guidelines for AI procurement
www.gov.uk/government/publications/draft-guidelines-for-ai-procurement

National Cyber Security Centre guidance for assessing intelligent tools for cyber security
www.ncsc.gov.uk/collection/intelligent-security-tools

The Data Ethics Framework
www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework

The Technology Code of Practice
www.gov.uk/government/publications/technology-code-of-practice/technology-code-of-practice



Understanding AI ethics and safety

AI has the potential to make a substantial impact for individuals, communities, and society. To make sure the impact of your AI project is positive and does not unintentionally harm those affected by it, you and your team should make considerations of AI ethics and safety a high priority.

This section introduces AI ethics and provides a high-level overview of the ethical building blocks needed for the responsible delivery of an AI project.

The following guidance is designed to complement and supplement the *Data Ethics Framework.* The Framework is a tool that should be used in any project.[12]

Ethical considerations will arise at every stage of your AI project. You should use the expertise and active cooperation of all your team members to address them.

Understanding what AI ethics is

AI ethics is a set of values, principles, and techniques that employ widely accepted standards to guide moral conduct in the development and use of AI systems.

The field of AI ethics emerged from the need to address the individual and societal harms AI systems might cause. These harms rarely arise as a result of a deliberate choice - most AI developers do not want to build biased or discriminatory applications or applications which invade users' privacy.

12 ***The Data Ethics Framework*** www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework

The main ways AI systems can cause involuntary harm are:

- **misuse** - systems are used for purposes other than those for which they were designed and intended
- **questionable design** - creators have not thoroughly considered technical issues related to algorithmic bias and safety risks
- **unintended negative consequences** - creators have not thoroughly considered the potential negative impacts their systems may have on the individuals and communities they affect

The field of AI ethics mitigates these harms by providing project teams with the values, principles, and techniques needed to produce ethical, fair, and safe AI applications.

Varying your governance for projects using AI

The guidance summarised in this chapter and presented at length in The Alan Turing Institute's further guidance on AI ethics and safety is as comprehensive as possible. However, not all issues discussed will apply equally to each project using AI.

An AI model which filters out spam emails, for example, will present fewer ethical challenges than one which identifies vulnerable children. You and your team should formulate governance procedures and protocols for each project using AI, following a careful evaluation of social and ethical impacts.

Establish ethical building blocks for your AI project

You should establish ethical building blocks for the responsible delivery of your AI project. This involves building a culture of responsible innovation as well as a governance architecture to bring the values and principles of ethical, fair, and safe AI to life.

Building a culture of responsible innovation

To build and maintain a culture of responsibility you and your team should prioritise four goals as you design, develop, and deploy your AI project. In particular, you should make sure your AI project is:

- ethically permissible - consider the impacts it may have on the wellbeing of affected stakeholders and communities
- fair and non-discriminatory - consider its potential to have discriminatory effects on individuals and social groups, mitigate biases which may influence your model's outcome, and be aware of fairness issues throughout the design and implementation lifecycle
- worthy of public trust - guarantee as much as possible the safety, accuracy, reliability, security, and robustness of its product
- justifiable - prioritise the transparency of how you design and implement your model, and the justification and interpretability of its decisions and behaviours

Prioritising these goals will help build a culture of responsible innovation. To make sure they are fully incorporated into your project you should establish a governance architecture consisting of a:

- framework of ethical values
- set of actionable principles
- process based governance framework

Start with a framework of ethical values

You should understand the framework of ethical values which support, underwrite, and motivate the responsible design and use of AI. The Alan Turing Institute calls these 'the SUM Values':

- respect the dignity of individuals
- connect with each other sincerely, openly, and inclusively
- care for the wellbeing of all
- protect the priorities of social values, justice, and public interest

These values:

- provide you with an accessible framework to enable you and your team members to explore and discuss the ethical aspects of AI
- establish well-defined criteria which allow you and your team to evaluate the ethical permissibility of your AI project

You can read further guidance on SUM Values in The Alan Turing Institute's comprehensive guidance on AI ethics and safety.

Establish a set of actionable principles

While the SUM Values can help you consider the ethical permissibility of your AI project, they are not specifically catered to the particularities of designing, developing, and implementing an AI system.

AI systems increasingly perform tasks previously done by humans. For example, AI systems can screen CVs as part of a recruitment process. However, unlike human recruiters, you cannot hold an AI system directly responsible or accountable for denying applicants a job.

This lack of accountability of the AI system itself creates a need for a set of actionable principles tailored to the design and use of AI systems. The Alan Turing Institute calls these the 'FAST Track Principles':

- fairness
- accountability
- sustainability
- transparency

Carefully reviewing the FAST Track Principles helps you:

- ensure your project is fair and prevent bias or discrimination
- safeguard public trust in your project's capacity to deliver safe and reliable AI

Fairness

If your AI system processes social or demographic data, you should design it to meet a minimum level of discriminatory non-harm. To do this you should:

- use only fair and equitable datasets (data fairness)

- include reasonable features, processes, and analytical structures in your model architecture (design fairness)
- prevent the system from having any discriminatory impact (outcome fairness)
- implement the system in an unbiased way (implementation fairness)

Accountability

You should design your AI system to be fully answerable and auditable. To do this you should:

- establish a continuous chain of responsibility for all roles involved in the design and implementation lifecycle of the project
- implement activity monitoring to allow for oversight and review throughout the entire project

Sustainability

The technical sustainability of these systems ultimately depends on their safety, including their accuracy, reliability, security, and robustness.

You should make sure designers and users remain aware of:

- the transformative effects AI systems can have on individuals and society
- your AI system's real-world impact

Transparency

Designers and implementers of AI systems should be able to:

- explain to affected stakeholders how and why a model performed the way it did in a specific context
- justify the ethical permissibility, the discriminatory non-harm, and the public trustworthiness of its outcome and of the processes behind its design and use

To assess these criteria in depth, you should consult The Alan Turing Institute's guidance on AI ethics and safety.

Build a process-based governance framework

The final method to make sure you use AI ethically, fairly, and safely is building a process-based governance framework. The Alan Turing Institute calls it a 'PBG Framework'. Its primary purpose is to integrate the SUM Values and the FAST Track Principles across the implementation of AI models within a service.

Building a good PBG Framework for your AI project will provide your team with an overview of:

- the relevant team members and roles involved in each governance action
- the relevant stages of the workflow in which intervention and targeted consideration are necessary to meet governance goals
- explicit time frames for any evaluations, follow-up actions, re-assessments, and continuous monitoring
- clear and well-defined protocols for logging activity and for implementing mechanisms to support end-to-end auditability

You may find it useful to consider further guidance on allocating responsibility and governance for AI projects.

AI REVIEW FOR GOVERNMENT DELIVERY TEAM

OFFICE FOR ARTIFICIAL INTELLIGENCE

Jacob Beswick
Sébastien A. Krier

GOVERNMENT DIGITAL SERVICE

Emily Ackroyd
Bethan Charnley
Pippa Clark
Lewis Dunne
Breandán Knowlton
Matt Lyon
Nick Manton
Gareth Reilly
Clive Richardson
Nicky Zachariou

GET IN CONTACT

Email ai-guide@digital.cabinet-office.gov.uk if you:

- want to talk about using AI in the public sector
- have any feedback on the AI guidance
- would like to share an AI case study with us

About Office for Artificial Intelligence

The Office for Artificial Intelligence is a joint BEIS-DCMS unit responsible for overseeing implementation of the AI and Data Grand Challenge.

Its mission is to drive responsible and innovative uptake of AI technologies for the benefit of everyone in the UK. The Office for AI does this by engaging organisations, fostering growth and delivering recommendations around data, skills and public and private sector adoption.

www.gov.uk/officeforai Twitter: @officeforai

BEIS 1 Victoria Street, London, SW1H 0ET
DCMS 100 Parliament Street, London, SW1A 2BQ

About GDS

The Government Digital Service (GDS) is leading the digital transformation of government.

Its aim is to make world class digital services based on user needs and create digital platforms fit for the civil service of today.

www.gov.uk/gds Twitter: @GDSTeam

The White Chapel Building, 10 Whitechapel High Street, London, E1 8QS

Published January 2020