Department for Science, Innovation & Technology

Government Digital Service

AI Insights

Using AI to manage the digital heap

1. Introduction **3**
1.1 Who this guidance is for 3
1.2 Why this guidance is needed 3
1.3 The case for using AI and data science techniques in lifecycle management 4
1.4 The role of KIM professionals 5
1.5 Examples of successful projects in UK government 5
2. Consider the phases of your AI implementation **6**
2.1 The discovery phase 7
2.2 The alpha phase 8
2.3 The beta phase 10
3. Target your intervention **11**
3.1 Target an environment 11
3.2 Target a point in the life cycle 11
3.3 Target a use case 13
4. Review your retention approach **14**
4.1 Set retention policies 14
4.2 Establish controls at the point of provisioning 14
4.3 Keep contextual information 15
4.4 Carry out first and second reviews 16
5. Use case 1: filtering out unwanted content **16**
5.1 Filtering solutions for legacy content 16
5.2 Filtering solutions in live email systems 17
6. Use case 2: assessing legacy containers **18**
6.1 Assigning containers to retention bands 18
6.2 Grouping similar containers 19
6.3 Assigning containers to a records classification scheme 19
6.4 Making distinctions within containers 20
7. Use case 3: creating summaries of content **21**
7.1 Summarisation as an input into decision making 21
7.2 Summarisation as a potential replacement for low-value content 22
8. Design human control into your AI implementation **22**
8.1 Humans-in-the-loop for legacy content 23
8.2 Humans-in-the-loop for live systems 23
8.3 Ethics and transparency 24
9 Terms in this guide **24**
9.1 Information management terms 24
9.2 Data science terms 26

1. Introduction

1.1 Who this guidance is for

This guide helps you to integrate AI and data science techniques into your organisation's content lifecycle management approach.

It sets out:

- an approach to targeting and planning AI and data science interventions in lifecycle management
- key use cases for AI and data science in lifecycle management
- advice on ensuring human control of AI-enabled lifecycle management

The guide is written from a knowledge and information management (KIM) perspective, and is mainly targeted at government KIM professionals. This guidance will also be relevant to digital and data professionals, including data scientists who have an interest in managing unstructured data.

1.2 Why this guidance is needed

The vast majority of content created by government takes the form of unstructured data.

Unstructured data is the content that your organisation holds in its document repositories, collaboration systems and communication systems, which typically includes documents, posts, chat messages, emails, images and recordings. Efficient management of this data is vital for enabling organisations to operate effectively and account for their actions and decisions.

Government organisations have been accumulating digital content for over 2 decades. This build-up of inactive material is often in the form of legacy:

- shared drives
- collaboration spaces
- electronic document management systems
- email accounts

The content in these repositories will cover a broad range of policy and operational work, including some material that should be retained and some that should be deleted. A common term used for the totality of this content within an organisation is ‘the digital heap’.

You can use this guidance to:

- identify the key steps and decisions you will need to take when tackling unstructured data
- understand how to apply AI and data science techniques at different stages of the lifecycle of unstructured data

After reading this guidance:

- check that content within your live systems has appropriate retention rules
- review legacy repositories to identify and remove redundant, obsolete and trivial material (ROT) whilst retaining material of ongoing value

1.3 The case for using AI and data science techniques in lifecycle management

As a government organisation, you need to follow the Code of Practice on the Management of Records.

This means you must:

- periodically assess the information you hold
- keep a record of decisions that explain why you're retaining or disposing of information

Under the Public Records Act, records of historical value must be transferred to the National Archives or another designated place of deposit within 20 years after their creation.

Applying retention policies consistently:

- improves efficiency by saving storage, information management, maintenance and migration costs
- improves accountability by helping organisations identify and protect the content they need to store, while disposing of what they no longer need. This also supports good freedom of information practice and good data protection practice
- improves information reuse and exploitation by improving an organisation's understanding of the unstructured data it is holding

AI tools exist that can help with making pragmatic and defensible decisions about the content they are targeted on.

These include tools that can:

- filter out unwanted content – for example, by identifying redundant, obsolete and trivial content
- generate summaries of groupings of content to enable KIM professionals to make decisions at scale.

This is a rapidly developing field. There exists other potential uses of AI which organisations may wish to explore to assist with, for example, assessing legacy containers of content.

1.4 The role of KIM professionals

KIM professionals play a vital role in managing unstructured data in their organisation.

They should be involved in any AI implementation that seeks to improve the way that content is managed through its lifecycle. This is because they oversee the management of unstructured data through its lifecycle, including:

    - setting retention policies and access permissions
    - disposing of content that has no value
    - selecting records for permanent preservation and transferring them to the National Archives

KIM professionals also help to manage end-user expectations in regards to how content will be managed.

Your KIM professionals will need support from:

- government digital and data colleagues that have data science knowledge
- senior management who can secure organisational backing and attention

1.5 Examples of successful projects in UK government

The government has already used data science techniques on unstructured data in various ways. We’ve made more progress in legacy systems than in live systems.

Cabinet Office Automated Digital Document Review
The Cabinet Office Digital KIM team has developed an automated document review tool to tackle content in the organisation's legacy repositories. In 2022 the team used the tool to review over 5 million items, and to identify and delete over 1.5 million redundant, trivial and outdated items. Their project is discussed in the section filtering solutions for legacy content.

The team has documented every step of the process. This implementation was one of the first automated government projects to have a published return made under the Algorithmic Transparency Recording Standard.

FCDO Services digital sensitivity review service
Foreign, Commonwealth and Development Office (FCDO) Services introduced a service to help identify sensitive information within documents selected for permanent preservation. This involved developing sets of rules for each exemption in the Freedom of Information Act. Based on these rules, the AI solution can now analyse these documents and identify passages that may be sensitive, noting where there is a relevant FOI exemption.

While digital sensitivity review is out of scope of this guide, there are useful lessons from this service.

In live systems
There has been some experimentation with Microsoft Syntex trainable classifiers to automatically apply labels to content in Microsoft 365. We are not yet aware of mature applications within UK government of this capability.

There is also likely to be some experimentation with Microsoft 365 Copilot. For example, to create summaries of content such as SharePoint libraries.

2. Consider the phases of your AI implementation

The Department for Science, Innovation and Technology (DSIT) have issued a guide to planning and preparing for AI implementation. The guide identifies 3 distinct phases in any AI implementation. These are the discovery, alpha, and beta phases.

2.1 The discovery phase

During the discovery phase, you will explore the underlying problem. You will assess user needs, identify your broad approach to the problem, put the key building blocks in place to enable you to start developing your solution, and establish the benchmarks against which you can evaluate whether the solution is fit for purpose.

Assess your user needs
You should start by thoroughly understanding the problem and your retention requirements. Assess whether AI is the right tool to address those requirements. Understand the content that the AI will be trained on, and how it will be brought into contact with that data.

Identify your broad approach
You should:

- identify where to target your intervention
- review your retention approach
- select the main use case for your AI intervention. You can find examples in the sections for use case 1, use case 2 and use case 3

Set up your project
You'll need to set up your project by establishing the roles and responsibilities for the project. This includes your KIM team, with support from:

  - data science professionals
  - technical specialists
  - the solution provider
  - end users in your organisation (where relevant)
  - a project sponsor

You will also need to ensure that your team has access to the records in question.

Your departmental record officer, (who leads on compliance with the Public Records Act), and the relevant information asset owner (who has responsibilities for ensuring that arrangements are in place to govern and protect the data) should be consulted during the discovery phase, and again during the alpha phase before the project moves into beta.

Establish your benchmarks to support evaluation
AI solutions based on large language models (LLMs) and many traditional machine learning (ML) models are probabilistic. This means that they return a response that is

correct with a high degree of probability, but not with certainty. They are unlikely to achieve 100% accuracy. Some tolerance of error is necessary, just as if humans were performing the task.

Set your benchmarks before you begin the development process. This way, your human reviewers can make objective tests against the benchmarks during the evaluation phase, and adjust the solution until it meets the benchmarks.

For example, if the solution was focused on reviewing documents and removing ROT, the benchmarks could relate to:

- false positives, which are items identified as ROT which actually had historic value. This is a measure of the precision of the model
- false negatives, which are items not identified as ROT which were actually worthless. This is a measure of the recall of the model

You might set a benchmark that the AI solution returned less than 1% false positives to reduce the chance of permanently deleting valuable records. Your benchmark might be a little higher, for example 5%, for false negatives because there's less risk in retaining worthless records.

It's important to consider the context in which the AI solution will be used and the potential consequences of errors when establishing the most appropriate benchmark levels. Also be aware that the tighter you set your benchmarks, the harder it will be to achieve, potentially leading to longer development times or higher costs.

2.2 The alpha phase

In this phase, you'll develop an AI solution or set of data science techniques capable of making judgements on the target content, and you will evaluate it to check that it meets the benchmarks you have established.

Your team will use one or more of the following methods:

- building an AI model by writing a set of custom rules
- training an AI model using supervised, unsupervised or reinforcement learning (or a combination of these)
- developing processes that enable an LLM (perhaps enhanced by internal sources or rules) to carry out tasks that are useful for lifecycle management.

Your organisation needs access to:

- a sufficiently large set of data on which to train or test the model. This could either be a copy of the entire target set, or a copy of a sample of the target set
- a development (or sandbox) environment in which you build and train the AI model or data science technique

Set up your developer environment
Some AI tool providers supply a development environment. When the Cabinet Office developed their automated digital document review, the team copied their target set of content into a specialised development environment.

Within the tool, they were able to build their lexicon and encode their weightings and rules. The development environment also enabled them to run visualisations of how the rule set they were building was classifying content.

There are vendor-neutral sandbox environments where you can run and test your models. The AI Playbook for the UK Government has more information on how to deploy AI securely.

Some projects may not involve any development. For example, if a cloud suite provider develops an AI model that filters, sorts or classifies content in a way that's relevant to retention and disposition decisions. Such tools will still need evaluation prior to use.

Governance during the alpha phase
Make sure you:

- apply version control to different sets of content which the AI model has been run over
- apply version control to the different iterations of the AI model as it progresses
- keep records of evaluations by recording the results of runs of AI models on sample sets or entire copy sets of target data
- securely dispose of sample data sets when you no longer need them for development, testing or recordkeeping purposes

Evaluation before use
During evaluation, you'll carry out tests to establish whether the AI solution is working with sufficient accuracy and precision for you to base disposal actions on its judgements.

Once you have established an AI model that can make judgments on content, it is important to evaluate the model prior to using it to set or change retention rules on content, or to take disposition actions on content. These evaluations should be made with reference to benchmarks set in the discovery phase.

In large repositories, it may not be feasible to check every judgement made by the model. Instead, check a meaningful sample of judgements made by the model.

For example, you could check the judgements made on 10% of the total number of items (or containers) in the collection or, if the collection is very large, 1,000 randomly selected records (or containers).

If the AI solution fails this evaluation, you should either return to the development phase or consider redesigning your solution if there's a more fundamental failure.

Every project must have an evaluation phase regardless of the methods you're using.

Record the tests you conducted against the benchmarks. This will help you explain to your organisation the steps you took to arrive at an effective AI solution.

2.3 The beta phase

This phase is about using the judgements of your AI model to:

- set or change retention rules
- make decisions about which content you'll keep or dispose of

It's important to keep an audit trail of any disposition actions taken on the basis of an AI judgement.

If items are being deleted based on the judgements made by an AI model, it is helpful to keep an audit trail of the action. For example, the audit trail might record:

- the date of deletion
- the titles of the items or containers
- the reason for deletion. Typically, this will be if the AI solution identifies the items as having no ongoing value

You may need to keep an equivalent audit trail for other actions that impact retention, for example, if an item or container was assigned to a different retention rule or band on the basis of an AI judgement.

Evaluation after use
Lifecycle management is not a one-off task. Once you've developed an AI solution for a particular repository or type of content, you'll need to reassess it as the content builds up again.

This might involve applying the solution to different repositories or types of content than those it was originally intended for. Verify whether the solution is still appropriate and decide on new performance benchmarks.

If the solution includes a lexicon or a specialised vocabulary, you may need to update the lexicon to reflect any new business terminology used at the time that the records were created.

3. Target your intervention

When planning an intervention using AI, consider what content you want to target, in which environment and at what stage of the lifecycle.

3.1 Target an environment

Content accumulates in different environments. For example, your organisation might have unstructured data in legacy shared drives, SharePoint sites or email systems.

When determining which environment to target first, you should consider the:

- age of the content, noting the requirement to transfer records of historical value to the National Archives before they become 20 years old
- scale of the content, for example might prioritise larger repositories
- cost of content storage, for example might prioritise repositories that have higher storage costs
- value and accessibility of the content, for example you might prioritise repositories that contain potentially valuable but currently inaccessible content
- sensitivity of the content, for example you might prioritise more sensitive or higher risk content

3.2 Target a point in the life cycle

The lifecycle of digital content has 2 main phases. These are:

- the live phase, which is content in live systems that are still being worked in, such as a live email account or SharePoint site
- the legacy phase,which is content in a system or container that's no longer being added to or used

Content in live systems

Targeting your AI solution in a live system gives you the opportunity to involve its end users.

These users may be able to interact with the AI solution to:

- correct its errors
- reinforce correct distinctions
- help assess its accuracy and effectiveness

However, integrating a custom AI tool into your live environment is challenging. Deploying AI in a live cloud suite environment like Microsoft 365 might require customisation of the end user interface. This risks conflicting with future cloud suite updates.

Training, testing and monitoring the performance of an AI model within live environments, such as email or chat accounts, may present challenges due to information security and data protection concerns.

For content in live cloud suites, it is simpler to use the AI tools offered by that provider within the suite. This includes AI trained by its provider, such as the 'Focused Inbox' feature of Microsoft Exchange. Your KIM professionals should decide whether these tools are sufficiently reliable for lifecycle management.

Records in legacy systems

Targeting your AI solution in a legacy system gives you more freedom to move content to different environments than would be possible in a live system. For example, you could move legacy content to AI-enabled environments for analysis.

In legacy systems, the original end-users of the content may not be available to reinforce or correct the AI model. Your KIM team will need to act as humans-in-the-loop and assess how the AI model runs.

3.3 Target a use case

Ensure an appropriate records retention regime is in place, and then select the relevant use case.

Establish a records retention regime
Put the key elements of your retention regime in place by:

- setting out the key retention principles that your organisation wishes to apply in a policy document
- establishing controls at the point of provisioning new containers to teams and individuals so that all new sites, drives and accounts have default retention rules that are consistent with your retention policy
- establishing review points at which to make disposition and selection decisions on content

Use case 1: filtering out unwanted content
AI can help distinguish valuable content from ROT. This can be done within containers or across entire repositories.

Use case 2: assessing legacy containers
There are many potential uses of AI to assess groupings or containers of content. For example, AI could rank containers in terms of their importance, group similar containers together, assign containers to a records classification or break down large containers by creating clusters within them. These ideas are still largely at an experimental stage.

Use case 3: creating summaries of content
In repositories such as shared drives, the titles of folders do not always give a good indication of the content held within them. AI can generate summaries of the content of containers. This can help KIM professionals better understand the content and make retention decisions at scale.

4. Review your retention approach

4.1 Set retention policies

The Code of Practice on the Management of Records states that you "must define how long to keep information and dispose of it when it is no longer needed". Your organisation should have high-level retention and selection policies in place.

You can refer to these examples:

- HMRC retention policy
- FCDO retention policy
- MoJ retention policy
- Cabinet Office selection policy

These policies help to:

- explain your disposition and selection decisions and actions, and ensure they remain consistent and transparent
- manage stakeholder and staff expectations about how long content will be kept
- comply with legislation and data protection principles
- reduce the cost of storing and processing content

AI can support your application of these policies. Regardless of what AI tools or data science techniques you apply, your organisation needs retention rules to be in place and understood.

4.2 Establish controls at the point of provisioning

Apply retention rules at the highest practical level of content grouping. Containers such as sites, accounts and drives offer a natural point for setting default retention rules.

Set default retention rules at the point that the container is provided to the team or individual that will use it, or as soon as possible afterwards. Set a retention rule that reflects the relative importance of the work of the individual or team in question.

For example, SharePoint is the core knowledge repository of the Department for Energy Security and Net Zero (DESNZ). When a team or official needs a new site, they complete a site creation request form in which they detail the purpose and value of their site. An automated process creates the site and applies the relevant retention label at site level. The label is automatically cascaded down to all data created within the site. This ensures that data is removed when it reaches its 'use by' date.

Consider assigning teams and individuals to retention bands, such as short-term, medium-term and long-term bands. This way, any new site or account provisioned to that team or individual can receive the retention rule relating to that band.

There may be some content within a container that is out of scope of the retention rule you've set on it. For an email account, for example, you should exclude trivial, personal and unsolicited mail that's not needed for business purposes beyond a very short period. You may be able to use AI to identify, separate and eliminate non-business emails faster and more accurately than an end-user could.

4.3 Keep contextual information

Most content created in cloud suites is created in systems that have very flat, modular architectures.

For example, Microsoft 365 enables the creation of new:

- email accounts
- SharePoint sites
- OneDrive accounts

However, at the time of writing, there's no global classification system, file plan or taxonomy that gives context to each new account or site being created. It is therefore useful to capture contextual information about each site or account that you provision.

This contextual information might include the:

- name of the team or individual to whom the site or account has been assigned
- role of the team or individual
- retention band assigned to the site or account
- date the site or account was created
- default retention rule applicable to the site or account

This information would ideally be captured routinely or automatically in your provisioning process.

4.4 Carry out first and second reviews

Your organisation might store legacy content that was not assigned a retention rule when it was created, and for which you have little or no contextual information.

It is useful to carry out a first review on such content to identify material of ongoing business value and to remove material that your organisation no longer needs. Preferably, you should do this between 5 to 10 years after the content was created.

You must consider records for permanent preservation at the National Archives before 20 years have elapsed. One way of achieving this is by carrying out a second review to identify:

- which material has historical value and will be transferred to the National Archives (or another designated place) for permanent preservation
- which material can be destroyed
- which material should be retained within the department for a specific administrative period

5. Use case 1: filtering out unwanted content

Filtering solutions aim to eliminate unwanted content from a container whose core content still has value.

If you no longer require any of the content, there's no need for filtering because you can dispose of the entire container.

Filtering solutions make no attempt to restructure content or change access permissions on content. This reduces the complexity and risk of the AI solution.

5.1 Filtering solutions for legacy content

Your organisation may wish to filter:

- ROT from legacy document repositories
- trivial, personal and social correspondence from legacy email accounts

There's more scope to develop your own, purpose-built AI solution in a legacy system than in a live system. Your KIM team will be in complete control of the development, testing and deployment of the AI model. This should support the explainability and transparency of the implementation.

The Cabinet Office Digital KIM team developed an automated way of reviewing digital documents. There was a strong business need for this because the department had a legacy repository that contained around 12 million documents.

The team used the solution to identify what documents were worthy of ongoing retention. The solution had 2 main elements:

- a lexicon that identified language clues (such as words and phrases within a document or document title) that gave an indication of the purpose and value of a document
- parameters and defaults that weighted contextual clues (such as the document format or location)

It took several years to build a custom algorithmic model. However, the cost of the investment was more than recouped by the time saved, compared with how long it would have taken humans to go through this content.

With this approach you would set defaults appropriate to the environment in which the content resides, and develop a lexicon that is specific to your organisation’s work at the time the content was created.

5.2 Filtering solutions in live email systems

Filtering can be helpful for individual accounts such as those for email and chat. These accounts are likely to contain social and personal correspondence as well as business correspondence.

As discussed in the section on how to target a use case there are obstacles to running a custom AI solution across an entire email system. This might require customisation of the email interface, and customisations may clash with future cloud suite updates.

AI tabbing
Email systems in cloud suites such as Microsoft 365 and Google Workspace include AI models that relegate messages they consider to be trivial or unsolicited. They might do this by moving these emails to another tab, as Microsoft 365 does with the Focused Inbox feature.

These tabbing approaches can be effective. Your KIM team might then choose to explore assigning a shorter retention period to emails moved to the secondary tab than the default retention period you apply to the whole account.

However, tabbing AI does not separate personal and social correspondence from business correspondence. This content is also likely to be of interest to the user, so it tends not to be relegated to a secondary tab.

6. Use case 2: assessing legacy containers

It is often more efficient to act on containers such as sites, accounts and drives than on individual items.

This section explores potential ways that AI and data science techniques could be used to assess containers, group them, classify them, and make distinctions within them. These are relatively advanced use cases which are designed to inspire further innovation and experimentation.

6.1 Assigning containers to retention bands

Retention bands enable the same retention rule to be applied to broad swathes of content. For example, the SharePoint sites of teams engaged in key policy initiatives might be assigned to a long term retention band. The sites of teams engaged in delivering internal corporate services might be assigned to a medium or short term band.

AI and analytics tools could help assign containers such as SharePoint sites, shared drive folders and email accounts to retention bands. An AI model could assess each container for its

- organisational importance, which is a measure of how central the work was to the core mission or priorities of the organisation
- strategic importance, which is a measure of the work’s influence on the later activities and decisions of the organisation
- financial and legal risk
- level of decision-making responsibility

KIM teams could train the AI using examples of containers that represent high, medium, and low values for each factor. The team could also customise these factors and assign weightings. Content already under specific retention rules or legal hold would be excluded.

This approach can help to categorise containers. For instance, the most important 10-20% could be assigned to a long-term retention band, the least important 40-50% to a short-term band, and the remainder to a medium-term band. These percentages can be adjusted to fit organisational needs.

6.2 Grouping similar containers

AI and analytics tools could suggest groupings of similar containers, such as SharePoint sites, sets of shared drive folders or email accounts based on similarity of:

- contributors
- topics, matters, cases, or vocabulary
- stakeholders dealt with
- content, including occurrences of duplicates

This would allow similar containers to be reviewed or appraised together. It could support the assignment of a grouping of containers to a records classification or a retention band.

It could also support de-duplication. If you have two containers such as two email accounts, sites or shared drive folders for which there exists a high percentage of duplicate content, it would not normally be necessary to retain both containers. Here, you would retain the container that has the higher overall value.

6.3 Assigning containers to a records classification scheme

For more advanced implementations, establishing a corporate records classification scheme enables the consistent application of retention rules across similar types of organisational activities. To be effective, such a classification must include all major business activities. A classification scheme can be single-level or hierarchical, but schemes exceeding two or three levels may become unwieldy.

Classification schemes can be built by human experts or by AI. For example, clustering algorithms could be used to build a classification scheme by:

- clustering similar containers together. KIM professionals could validate each grouping, and give validated groupings a title and a scope
- looking at the clusters created in the first step, and grouping similar clusters together. KIM professionals could validate these higher level groupings, and could give a validated grouping a title and a scope

Another approach involves manually creating a high-level classification scheme (or using an existing classification scheme) and then assigning clusters of similar containers to that scheme.

FCDO Services' digital sensitivity review service had some success in applying clustering techniques and generating classifications for their client government departments. Their service is discussed as one of the case studies in the AI Playbook for the UK Government.

6.4 Making distinctions within containers

Some types of container contain unwanted content or cover a relatively broad range of activities. One way of tackling this is by making distinctions within such containers.

A clustering algorithm could be used to group together similar content within the container. Clustering is especially helpful for email accounts, as they contain a large volume and variety of content.

For example, a clustering algorithm might detect similarities in:

- email senders and recipients
- the content of the emails and any attachments.

It might then seek to unite emails related to particular activities. For example it might clustering together emails from a particular project or type of project.

Ideally, an algorithm like this would run in a live email system where users can engage with it and retrain it. For example the end-user could be given a means to:

- fine-tune the number of clusters that the algorithm creates so that the clustering becomes more or less granular
- reassign items from one cluster to another
- name clusters

Clustering would allow you to remove some less valuable types of content at an earlier point in time than the rest of the container.

Clustering could also be useful in information exploitation. For example, if an individual gives permission for their successor-in-post, and any generative AI acting on their behalf, to access some clusters within their email account.

7. Use case 3: creating summaries of content

Summarising solutions arose with the advent of LLMs that have a capacity to understand language.

Summarisation has at least 2 potential roles to play in lifecycle management:

- summaries as an input into decision making
- summaries as replacements for relatively low-value, high-volume content

7.1 Summarisation as an input into decision making

Using AI to generate summaries of containers can help KIM professionals to make decisions at a higher level than on individual items.

This can be particularly useful for shared drives where the titles of folders are often insufficiently informative to help in making high-level decisions. Once AI has generated a content summary, it should be easier to tell whether the folders contain a coherent set of content that's worth keeping.

Summaries can also help:

- support those who authorise disposal actions, by providing a summary of the content accompanied by a reason for the recommended action
- support the transparency of disposal actions. When containers are destroyed or otherwise disposed of, a summary of the container might be retained along with a justification for why it was disposed of

If you're considering this solution, make sure you test it thoroughly. The LLM and its prompting strategy should produce summaries that are sufficiently complete and accurate for the types of container you're making decisions on.

These tests are important because LLMs have known weaknesses, such as producing hallucinations.

7.2 Summarisation as a potential replacement for low-value content

Summarisation may be an appropriate solution in applications such as Microsoft Teams Chat where content accumulates quickly, and where it's difficult to separate business conversations from social or personal conversations.

It may be possible to retain a daily summary of an individual's conversations rather than the conversations themselves. An LLM could generate this summary and you could give the LLM instructions to exclude personal and social messages.

You could then routinely delete many messages without a significant loss of business information, and without asking users to filter or move their messages.

This idea is in its infancy and is included here to encourage further innovation, experiment and testing.

8. Design human control into your AI implementation

Most AI implementations require a human-in the-loop to validate that the model is working accurately enough to justify the decisions you're making based on the model's judgement.

You need a human-in-the-loop whenever your AI model is making decisions that your organisation would normally expect to account for, such as any implementation that sets or changes retention rules on records.

Ensure you have meaningful human control at the right stages. This means deciding:

- who should be a human-in-the-loop
- how the human-in-the-loop will be aware of the AI solution's judgements, and able to provide feedback

- at what stage in the process of developing and deploying the AI model the human-in the-loop should act, and over what scale

The typical humans-in-the-loop in your organisation will be KIM professionals. In some circumstances, end users may be able to act as additional humans-in-the-loop for AI running in live systems.

8.1 Humans-in-the-loop for legacy content

KIM professionals are the only viable humans-in-the loop for AI running on legacy records. You cannot use end users as an additional human-in-the-loop layer because legacy content does not have any end users. Experts who are currently working in that policy or business area can provide additional human assurance.

8.2 Humans-in-the-loop for live systems

When AI is deployed in live systems to inform retention rule decisions for content, the KIM team should serve as the primary humans-in-the-loop.

End users may act as additional humans-in-the-loop for AI running on their individual accounts. This will add security and scalability. It’s more secure because these users are interacting with the AI model in an account they can already access. It's also more scalable because it uses local input to help the KIM team oversee the content. End users can provide a subjective view on the judgements the model is making on the content in their own account.

In some circumstances, end users can help to assess the accuracy of the AI model’s judgement. For example, in a cloud suite email system, if the AI model was incorrectly identifying emails as trivial or unsolicited, users could turn off the model. However, if they chose to leave the model running, this might suggest that the model was mainly making correct judgements.

Your KIM team should still be responsible for deciding when an AI model is relevant and accurate enough to justify basing retention decisions on it.

Privacy- enhancing technologies
In the future, KIM teams may be able to play a greater role in monitoring how an AI model operates in a live email account. Emerging privacy-enhancing technologies might one day help KIM teams to observe the model without being able to actually see into the accounts.

8.3 Ethics and transparency

It is vital to ensure that your AI implementation is built ethically and responsibly.

You must consider all of these themes from the AI Playbook for the UK Government:

- safety, security and robustness
- transparency and explainability
- fairness, bias and discrimination
- accountability and responsibility
- contestability and redress
- societal wellbeing and public good

9 Terms in this guide

9.1 Information management terms

Access regime
This covers all the measures an organisation takes to ensure that it shares internal content appropriately within the organisation and that it protects internal content from inappropriate access.

An access regime includes policy statements which set out what information the organisation seeks to protect and how it seeks to protect it. It will also include default access rules on systems or containers, as well as any access rules applied specifically to individual items.

Container
In the context of this guide, a container is a grouping of content to which access or retention rules are applied. In the digital world, containers might take the form of sites, accounts, drives or top level shared drive folders.

Digital asset
Digital content that an organisation wishes to retain to support its business activities.

Digital heap
The sum total of all the digital information held by an organisation across its repositories and databases, on live and decommissioned (legacy) systems.

A digital heap comprises digital assets and digital liabilities, spread across systems that are organised in multiple ways, including:

- collaboration systems, document management systems and shared drives
- case file systems
- individual accounts in email systems, chat systems and document management system
- structured databases
- web content management systems

Digital liabilities
Digital content that an organisation has no need to retain, which are also known as redundant, obsolete or trivial (ROT). This might be because:

- the content in question does not relate to the organisation's business activities
- the content in question relates to a business activity that the organisation no longer needs to account for

Legacy content and repositories
Legacy content refers to containers of any kind that are no longer being added to by the team or individual that used them.

Legacy repositories refers to entire repositories of content that are no longer being added to.

Note that the usage of 'legacy' in this document does not refer to legacy IT.

Records classification
Any structure or taxonomy used to group containers of content together and apply retention rules to them.

Records management
A discipline for creating and capturing records of an organisation's work, and managing the records based on rules and policies that reflect the organisation's business needs and its legal and compliance obligations.

Retention regime
A records retention regime covers all the measures an organisation takes to ensure that content is retained for an appropriate period of time.

These measures include policy statements, such as retention schedules and any selection policies that identify what records the organisation will select for permanent preservation. They also include default retention rules applied to systems and containers within systems, as well as any retention rules applied specifically to an individual item, and any processes for approving proposed disposals of information.

Unstructured data

Examples of unstructured data, are documents, recordings, messages, posts and emails that your organisation holds in its live and legacy:

- document repositories, including shared drives, electronic document management systems and Microsoft SharePoint
- collaboration systems, such as Microsoft Teams
- communication systems, such as an email system

9.2 Data science terms

Artificial intelligence (AI)

Algorithms and models that allow machines to perform tasks typically requiring human intelligence. For example, tasks that involve making judgements to inform decisions or actions. These include traditional ML models that learn from labelled data, spot patterns in unlabelled data or learn from having their guesses confirmed or reinforced. LLMs that support generative AI capabilities are another example.

The following definition of Artificial intelligence has been adopted by OECD countries, and is used by the AI Playbook for the UK Government:

> "An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment."

AI implementation

The use of AI in a real-world situation. An implementation consists of one or more AI models, together with the policies and processes that determine how the AI model is deployed, tested and monitored, and how its outputs influence actions.

Data science techniques
Statistical, mathematical or algorithmic methods used to extract insights from data and inform or automate decisions or actions, or to take action on data. They encompass various approaches, including traditional statistical analyses, machine learning, data wrangling and data visualisation.

Traditional Machine Learning models
The AI Playbook for the UK Government defines ML as “the branch of AI that learns from data. It does this by extracting features from data and learning the relationships between those features.

ML models can be trained to perform a particular task or make a particular judgement, distinction or classification. They can be trained in any (or any combination of) the following ways:

- in a supervised way through the provision of curated training sets of accurately labelled data, allowing it to learn the relationships between inputs and corresponding outputs
- through unsupervised learning where the ML model explores unlabelled data to identify patterns or structures on its own
- through reinforcement learning where the model’s actions are assessed by feedback from the environment, including humans, enabling the model to iteratively refine its judgements based on rewards or penalties.

The advantage of ML models is that machines can be successfully trained to automate tasks even when humans are unable to codify a set of rules supporting those tasks.

Large language models
LLMs are advanced computational models trained on extensive datasets to understand and generate human-like text. Unlike traditional machine learning models that are typically designed for specific tasks, LLMs possess general language capabilities, enabling them to respond to diverse user prompts in natural language and perform a variety of functions, such as text completion, summarisation, translation, and conversation.

Pre-programmed rule set
An algorithmic model for automated decision making in which all the rules are written by human beings. These models tend to be more predictable and explainable than ML models. Pre-programmed rule sets can be used with LLMs to add context and controls to their outputs.