PQ PADEMIQ

AI GRCA PRACTITIONER PROGRAMME

Week 1

AI Foundations for Governance

Collins Obasuyi

Lead Consultant

PADEMIQ / AI GRCA Lab

15+

years across product, delivery, quality engineering, SDET, web development and AI analysis.

PUBLIC & PRIVATE SECTOR EXPERIENCE

- Bank of America
- Bloomberg
- Amazon
- Thomas Cook
- HMRC
- UKHSA
- European Banking Authority

SESSION AIM

Understand AI well enough to ask the right questions

By the end of Week 1 you should understand what AI is, how it works at a basic level, why it behaves differently from normal software — and why it needs governance before it is used in real business decisions.

Not the goal

Becoming an AI engineer or building models from scratch.

The goal

Knowing enough to challenge an AI system with confidence.

That is the foundation for everything else in the course.

THE QUESTIONS WE LEARN TO ASK

1. What is this AI doing?
2. What data is it using?
3. Who is using it?
4. Who is affected by it?
5. What can go wrong?
6. Who is responsible?
7. What evidence do we need before trusting it?

The foundation under every framework we will cover

NIST AI RMF
Risk management

OECD AI Principles
Values

ISO/IEC 42001
Management system

ISO/IEC 42005
Impact assessment

EU AI Act
Regulation

UK GDPR
Data protection

COSO
Internal control

AI Assurance
Evidence & testing

WEEK 1 • AI FOUNDATIONS

What AI is, how it learns, where it fails — and who is responsible.

You cannot apply a framework to a system you do not understand. So we start with the basics — in plain language.

By the end of this week, you should be able to...

1. Explain AI in simple language
2. Tell traditional software and AI apart
3. Relate AI, ML, Deep Learning and GenAI
4. Explain supervised learning with everyday examples
5. Explain unsupervised learning with everyday examples
6. Compare classic AI and generative AI
7. Compare small and large models
8. Compare single-purpose and general-purpose models
9. Compare proprietary and open-source models
10. Compare text and multimodal models
11. Explain foundation models
12. Explain Large Language Models (LLMs)
13. Explain hallucination
14. Explain RAG
15. Explain AI agents
16. Explain training vs inference
17. Understand the AI lifecycle
18. Name the main AI actors and their responsibilities
19. Explain why governance starts before launch
20. Complete a simple AI System Description

OPENING MESSAGE

Before we can govern AI, we need to understand what AI is doing.

We are not here to worship AI.
We are not here to pretend AI is magic.
We are not here to use big words to sound clever.
We are here to understand AI well enough to challenge it.

Once AI affects jobs, money, healthcare, insurance, education, public services or people's rights, we need more than excitement. We need:



Ownership



Controls



Evidence

And people in the room confident enough to ask: "What happens if this AI gets it wrong?"

01

Understanding AI

What AI is, what it isn’t, and how it differs from the software we already know.

IN THIS SECTION

01 What is AI?

02 Everyday AI around us

03 Traditional software vs AI

04 The AI family

05 Machine Learning & Deep Learning

06 Generative AI

A simple definition

Artificial Intelligence is software that can perform tasks we normally associate with human intelligence.

THAT COULD INCLUDE

- spotting patterns
- making predictions
- recommending
- recognising images
- understanding language
- writing text
- creating images
- summarising documents
- answering questions
- classifying information
- detecting unusual behaviour
- supporting decisions
- taking actions via tools

IN PLAIN TERMS

AI is software that uses data and patterns to make predictions, create outputs or support decisions.

Most AI systems are very good at recognising patterns and producing outputs based on what they have learned or been given.

That is useful. But it also creates risk.

What AI is not

Not magic

It is maths, data and patterns — not a mystery.

Not a human brain

It does not think or feel the way people do.

Not automatically wise

Clever outputs are not the same as good judgement.

Doesn't understand life

It has no lived experience or common sense.

What it is good at: recognising patterns and producing outputs based on what it has learned or been given.

KEY TEACHING POINT

The clever part of AI is that it can spot patterns. The risky part is that the pattern may be wrong, biased, incomplete, outdated or misunderstood.

EVERYDAY EXAMPLES

You already use AI every day — without calling it AI

Phone camera

Detects a face by its pattern — not by knowing who you are.

Email spam filter

Learns scam patterns: odd links, fake urgency, strange senders.

Netflix · Spotify · YouTube

“People who watched this also watched that.”

Bank fraud alerts

Flags spending that looks unusual for you.

Food delivery app

Learns you like jollof rice, spicy food and late Friday orders.

Fitness app

Uses your pace, sleep and heart rate to suggest a workout.

Customer service chatbot

Answers refund, delivery and account questions.

It is not your friend. It does not truly know you. It is looking at patterns.

EVERYDAY EXAMPLE UP CLOSE

The bank fraud alert

YOUR NORMAL PATTERN

£20 – £100

Usual spending around Chatham, Rochester or London.

SUDDENLY

£900 · abroad · 3am

Your card is used in another country in the middle of the night.

THE AI IS SAYING

"This looks unusual compared with normal behaviour."

Not: "This is definitely fraud."

It can protect people

Stopping a stolen card before real damage is done.

It can also block genuine customers

Someone travelling or buying a gift gets stuck.

SO WHAT?

That is why governance matters: the same pattern that protects one customer can wrongly block another.

IN PLAIN TERMS

AI is like a new employee who has seen thousands of examples

Imagine training someone in a fruit shop. At first they can't tell the fruit apart. After seeing many examples, they start recognising patterns:

Apples	are often round
Bananas	are long
Oranges	are round and orange
Mangoes	are a certain shape
Pears	are a different shape

IN PLAIN TERMS

Show them a new fruit and they say: "That looks like a mango."

They are not using magic. They learned from examples. That is the basic idea behind many AI systems.

KEY TEACHING POINT

The clever part is spotting patterns. The risky part is that the pattern may be wrong, biased, incomplete, outdated or misunderstood.

Rules written by people vs patterns learned from data

TRADITIONAL SOFTWARE

Follows rules written by humans

- If the password is correct, let the user log in. If it is wrong, show an error.
- If the balance is below £0, show an overdraft warning.
- The developer writes the rule. The system does what it was told.

AI SYSTEM

Learns patterns from examples

- A recruitment AI is trained on thousands of CVs and past hiring outcomes.
- It predicts: “This candidate looks like a strong match.”
- No one wrote that rule — the data taught it.

THE PROBLEM

What if the historical hiring data was biased — mostly one type of person hired, some backgrounds overlooked? The AI may learn that pattern and repeat it. Not because it hates anyone, but because the data taught it the wrong lesson.

The simple comparison

Traditional software	AI system
Follows fixed rules	Learns patterns from data
Humans write the logic	The model learns from examples
Usually easier to predict	Can be uncertain
Easier to trace	Harder to explain
Same input usually gives same output	Output may vary
Testing is more straightforward	Testing must consider data, bias, behaviour and context
Rule failure is often obvious	AI failure can be hidden behind confident output



KEY TEACHING POINT

AI is powerful because it learns patterns. AI is risky because learned patterns are not always fair, complete, correct or explainable.

EVERYDAY EXAMPLE

Spotting a ripe banana: rule vs learning

TRADITIONAL RULE

“If banana is yellow, mark as ripe.”

Sounds simple. But what about…

- almost-ripe green bananas
- yellow with black spots
- plantain
- poor lighting
- different countries
- small bananas
- overripe bananas
- bananas in a bunch
- a yellow object that isn’t a banana

The simple rule starts to break.

AI APPROACH

Show it thousands of banana photos

- ripe
- unripe
- rotten
- plantain
- different lighting
- in bunches
- on shop shelves
- from different countries

New photo → “This is probably ripe.”

But it can still get it wrong:

- poor lighting or a blurry image
- it is actually plantain
- training examples were not diverse enough
- it has never seen that type of banana before

IN PLAIN TERMS

One sentence explains a lot of AI risk

TRADITIONAL SOFTWARE SAYS

“Follow this rule.”

AI SAYS

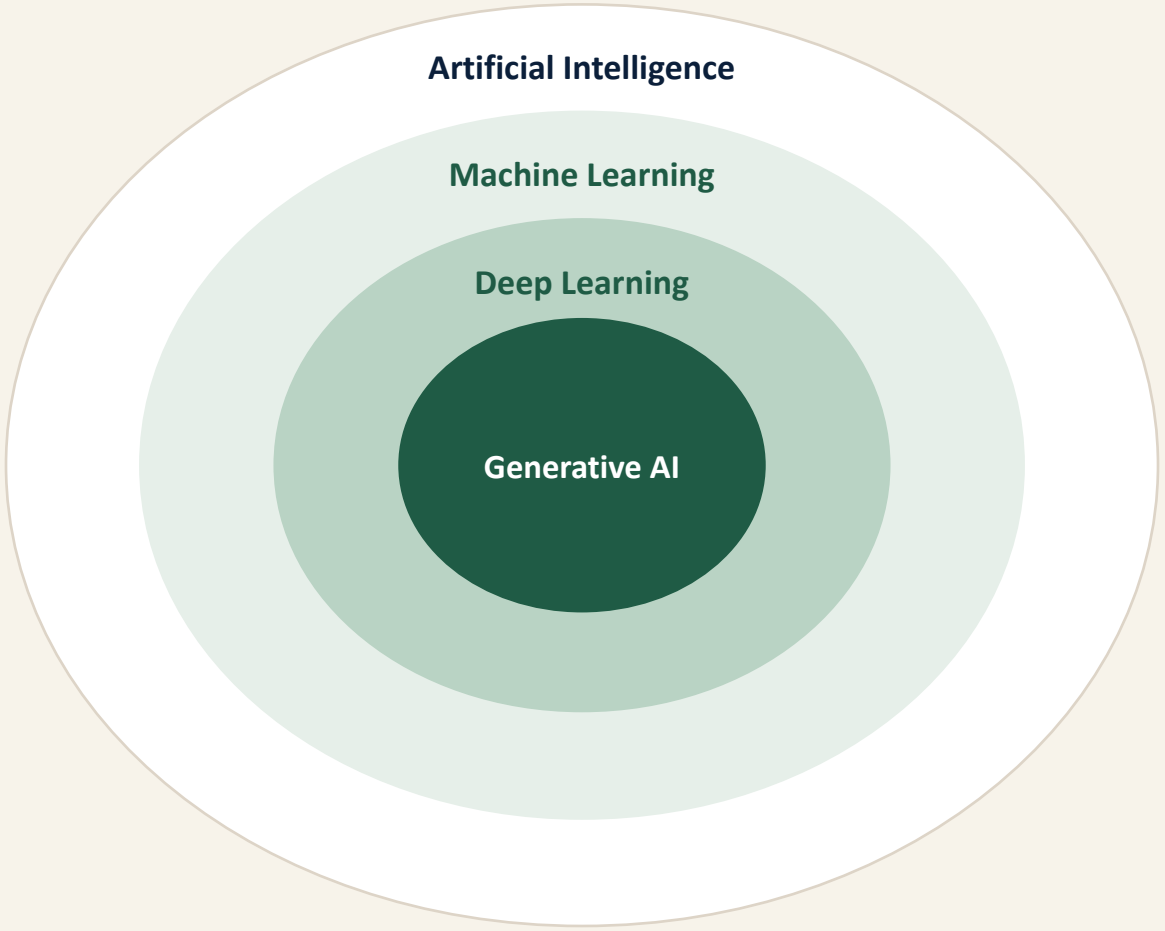
“Based on what I have seen before, this looks likely.”

A rule can be read and checked. A likelihood depends on what the system was shown — and what it wasn’t.

KEY TEACHING POINT

“Likely” is not the same as “right”. That one difference explains a lot of AI risk.

AI › Machine Learning › Deep Learning › Generative AI



1. **AI** is the big umbrella.
2. **Machine Learning** sits inside AI.
3. **Deep Learning** sits inside Machine Learning.
4. **Generative AI** often uses Deep Learning.

IN PLAIN TERMS: “VEHICLE”

Bicycle, car, bus, lorry, train and plane are all vehicles — but they don’t work the same way, carry the same risk, or need the same licence and safety checks. AI is the same.

Same family — different machines, different risks

Spam filter

Fraud detection

Chatbot

Image generator

Self-driving system

All AI-related. But they work differently and carry different risks.

3.1 ARTIFICIAL INTELLIGENCE — THE BROADEST TERM

route planning in Google Maps | face recognition | fraud detection | chatbots | recommendation engines

language translation | voice assistants | image generation | document summarisation | search ranking | CV screening

medical image support | customer support automation

Not every AI system is ChatGPT. Not every AI system generates text. Some only perform one narrow task.

Learning from data instead of being told every rule

SIMPLE DEFINITION

Machine Learning is when a system learns from data so it can make predictions or decisions on new examples.

Spam email

Sees emails labelled spam / not spam, learns the patterns, then predicts for new mail.

Bank fraud

Learns from millions of normal and fraudulent transactions, then flags unusual ones.

Food delivery

Learns what you like ordering and predicts what you may want next.

Gym app

Chest on Monday, legs on Wednesday — it recommends your next workout.

Insurance claim triage

Predicts which claims need closer review — risky if some customers are unfairly flagged.

KEY TEACHING POINT

Machine Learning is powerful because it learns from data. But if the data is poor, biased or incomplete, it may learn the wrong lesson.

Handling complex data: images, speech, video and language

A more advanced type of Machine Learning that uses neural networks. You don't need the maths — just the idea: it is good at messy, complex data.

Face unlock

Recognises you despite new lighting, angle, glasses, hairstyle or facial hair.

Voice assistant

Siri, Alexa and Google Assistant turn sound patterns into text.

Medical image support

Spots patterns in scans. Useful — but health is high impact, so governance must be stronger.

Self-driving vision

Recognises pedestrians, cyclists, cars, lights, signs, lanes and obstacles.

KEY TEACHING POINT

Deep Learning is powerful but harder to explain. If it gives an answer, can we understand why? Acceptable in low-risk settings — serious in high-risk ones.

IN PLAIN TERMS

How a child learns what a dog is

You don’t explain every rule — four legs, tail, ears, fur — because dogs are different:

- Some are tiny
- Some are huge
- Some have long hair
- Some have short hair
- Some look like wolves
- Some look like teddy bears

IN PLAIN TERMS

The child sees many dogs and slowly learns the pattern.

Deep learning works in a similar way — but at a much bigger scale, and without being able to tell you exactly which clues it used.

KEY TEACHING POINT

The harder a system is to explain, the more evidence we need before trusting it with important decisions.

AI that creates new content

IT CAN GENERATE

text | images | code | audio | video

summaries | reports | emails | presentations

designs | ideas | explanations

Generative AI creates content that can sound confident.

But confident does not mean correct.

“Write a professional email to my landlord.”
A draft appears.

“Create a poster for my AI governance course.”
An image is generated.

“Summarise the key actions from these notes.”
A summary is produced.

A developer asks for help writing or debugging code.
Code suggestions appear.

“Rewrite this LinkedIn post in my tone.”
A new caption is written.

Confident is not the same as correct

GENERATIVE AI MAY…

- invent facts
- misunderstand context
- miss important details
- produce biased wording
- expose sensitive data
- give unsafe advice
- sound more certain than it should

IN PLAIN TERMS

Generative AI is like a very fast assistant.

It can draft quickly, help you think and save time. But you still check the work.

If an assistant drafted a letter to your mortgage lender, you wouldn’t send it blindly. Same with AI: use it to help — don’t switch your brain off.

KEY TEACHING POINT

Generative AI is useful for speed. Governance is needed because speed without checking can spread mistakes faster.

02

How AI learns

Not all AI does the same job — and not all AI learns in the same way.

IN THIS SECTION

01 Classic AI vs Generative AI

02 Supervised learning

03 Unsupervised learning

04 Governance questions for each

Some AI predicts. Some AI creates.

CLASSIC AI

Predicts, scores, ranks

TYPICAL EXAMPLES

- spam filter
- fraud detection
- credit risk scoring
- recommendation engine
- image classification
- demand forecasting
- insurance claims triage
- CV ranking

THE OUTPUT IS OFTEN

- **yes / no**
- **score**
- **category**
- **ranking**
- **probability**
- **alert**

GENERATIVE AI

Creates new content

TYPICAL EXAMPLES

- writes text
- creates images
- summarises documents
- writes code
- drafts emails
- answers questions
- produces reports
- creates video or audio

THE OUTPUT IS OFTEN

- **paragraph**
- **answer**
- **image**
- **code**
- **summary**
- **script**
- **plan**

They fail in different ways

	Classic AI	Generative AI
What it does	Predicts, classifies or scores	Creates new content
Typical output	A label, score or recommendation	Text, image, code or audio
Example	Fraud score	Customer email draft
Main risk	Unfair scoring	Hallucination or unsafe content
Testing	Easier to test in some cases	Harder to fully predict outputs
Often used for	Decision support	Content and assistance

IN PLAIN TERMS

Classic AI sorts items into boxes. Generative AI writes something new.

Both can be useful. Both can be risky. But they fail in different ways.

KEY TEACHING POINT

Before governing AI, ask: what type of AI are we dealing with? The risk of a fraud score is different from the risk of a chatbot answer.

Learning from examples where the answer is already known

We show the system examples and tell it the correct answer. Then it learns the pattern. The labels are the teacher.

1 · LABELLED EXAMPLES

Picture 1	apple
Picture 2	banana
Picture 3	mango
Picture 4	orange

2 · MODEL LEARNS

Over time, it learns the difference between each fruit.

3 · NEW PHOTO → PREDICTION



“This looks like a mango.”

FLASHCARD ANALOGY

Like teaching a child with flashcards: “This is an apple.” “This is a banana.” After enough examples, the child learns.

Supervised learning in real life

BUSINESS

Predicting loan default

- Old customer data with known outcomes: repaid or defaulted.
- The model learns from those labels.
- New applicant → “higher or lower risk of default.”

RECRUITMENT

Screening CVs

- Past data: shortlisted, rejected, hired, performed well later.
- The system learns from those examples.
- If past hiring was biased, the AI may learn that bias.

EVERYDAY APP

Spam detection

- Emails labelled spam / not spam.
- The system learns the patterns behind the labels.
- It predicts whether each new email is spam.

KEY TEACHING POINT

Supervised learning is only as good as the examples and the labels. If the labels are biased, the model may learn the bias.

Governance questions for supervised learning

1. Who labelled the data?
2. Were the labels correct?
3. Were the examples representative?
4. Was the historical data biased?
5. Were some groups missing?
6. Was personal data used lawfully?
7. How was performance tested?
8. Does it work on new data?
9. Does it perform fairly across groups?
10. What happens when it is wrong?



KEY TEACHING POINT

Supervised learning can repeat past mistakes. It learns the teacher's mistakes along with the lessons.

Finding patterns without being given the answer

We give the system data but no labels. It tries to find groups, patterns or unusual behaviour by itself. There is no teacher.

THE FRUIT BASKET

“Group similar fruits together.”

They may group by:

colour | shape | size | texture | smell

Bananas together because they are long. Oranges because they are round and orange. Apples because they are round and shiny.

SORTING LAUNDRY

No categories given.

You start grouping by similarity:

whites | darks | towels | gym clothes | delicates

You worked out the groups yourself by looking at what is similar. That is unsupervised learning.

IN PLAIN TERMS

Supervised learning has a teacher. Unsupervised learning has no teacher — it finds its own groups.

Unsupervised learning in real life

SUPERMARKET

Finding shopper groups

- Nobody tells the AI “budget shopper” or “premium shopper”.
- It finds clusters: baby products, gym food, luxury items, mostly discounts, Friday-evening shoppers.

BANKING

Spotting unusual transactions

- No label for every fraud case.
- Looks for behaviour that doesn’t fit: sudden large amount, new location, strange time, unusual merchant, rapid repeat spending.
- Flagged as unusual — not necessarily fraud. A human may then review.

WORKPLACE

Grouping helpdesk tickets

- Categories aren’t known upfront.
- AI groups tickets: password resets, laptop issues, access requests, software bugs, HR and finance system issues.
- Helps the organisation understand demand.

Governance questions for unsupervised learning

1. What patterns did the system find?
2. Do those patterns make sense?
3. Could the grouping be unfair?
4. Are people being labelled in a way that affects them?
5. Could the system create hidden categories?
6. Could sensitive characteristics be inferred indirectly?
7. Who reviews the clusters?
8. How are the results used?
9. Could grouping affect access to services, pricing or opportunities?

KEY TEACHING POINT

Unsupervised learning can create hidden groupings — and people may be treated differently because of a group they never knew they were in.

Supervised vs unsupervised learning

Type	Simple meaning	Everyday example	Main governance risk
Supervised learning	Learns from examples with correct answers	Fruit pictures labelled apple, banana, mango	Bad labels or biased history
Unsupervised learning	Finds patterns without being told the answer	Grouping shoppers by behaviour	Hidden or unfair grouping
Supervised	Has a teacher	Flashcards	Learns the teacher’s mistakes
Unsupervised	No teacher	Sorting laundry	May create strange or unfair groups



KEY TEACHING POINT

Supervised learning can repeat past mistakes. Unsupervised learning can create hidden groupings. Both need governance.

03

Types of AI models

Before we govern an AI system, we need to know what kind of model we are dealing with. Different model types create different risks.

IN THIS SECTION

01 Small vs large

02 Single-purpose vs general-purpose

03 Proprietary vs open-source

04 Text vs multimodal

05 Foundation models & LLMs

06 Hallucination, RAG & agents

Small models vs large models

SMALL MODEL

A specialist tool

Built for a narrow task. Easier to test because the job is specific — but can still cause serious harm in high-impact decisions.

- fraud detection
- spam filter
- loan risk
- customer churn
- delivery time
- CV scoring

Like a screwdriver: made for one job.

vs

LARGE MODEL

A multi-purpose tool

Trained on much broader data and handles many tasks. More flexible — but harder to control because people use it for many purposes.

- large language models
- foundation models
- multimodal models
- general-purpose AI

Like a Swiss Army knife: people may use it in ways it wasn't designed for.

GOVERNANCE QUESTION

Ask: is this model used for one clear purpose, or for many? The broader the use, the stronger the governance needs to be.

Single-purpose vs general-purpose models

SINGLE-PURPOSE

Like a toaster

Designed for one job. The intended use is usually clear.

- identify spam
- predict delivery time
- detect fraud
- classify support tickets
- score insurance claims

It does one job: toast bread.

vs

GENERAL-PURPOSE

Like a kitchen machine

Supports many tasks. Flexible — but teams may start using it everywhere without proper review.

- writing
- summarising
- coding
- searching
- answering questions
- translating
- analysing documents

Chop, blend, mix, knead… you need to know which attachment is in use — and whether it's safe for the job.

KEY TEACHING POINT

The more flexible the AI system is, the more we need to control how it is used.

Proprietary vs open-source models

Proprietary model

Owned and controlled by a company — e.g. a commercial AI API or vendor recruitment AI.

POSSIBLE BENEFITS

- vendor support
- documentation
- service-level commitments
- security reviews
- commercial accountability

POSSIBLE RISKS

- less visibility into training data
- vendor lock-in
- unclear model changes
- contract limits
- data-sharing concerns
- limited audit rights

Like renting a managed office: someone else runs the building — but you depend on their rules.

Open-source model

More openly available — can be self-hosted and customised.

POSSIBLE BENEFITS

- more control
- more transparency (sometimes)
- can be self-hosted
- can be customised
- avoids some lock-in

POSSIBLE RISKS

- security moves to you
- may lack support
- licence misunderstood
- model may be misused
- deployment controls are yours
- updates & monitoring are yours

Like owning your own building: more control — and more responsibility.

Governance questions: who owns and controls the model?

1. Who owns the model?
2. Who controls updates?
3. What data can the provider see?
4. Can the organisation audit the model?
5. What does the licence allow?
6. Who is responsible for security?
7. What happens if the supplier changes the model?
8. Can we stop using it easily?

REMEMBER

Buying a model does not buy away your responsibility. Building on an open model does not remove it either — it moves more of it to you.

Text models vs multimodal models

TEXT MODEL

Reads and writes

Works mainly with written language.

- chatbot
- document summariser
- email drafting assistant
- policy Q&A assistant
- code assistant

vs

MULTIMODAL MODEL

Looks, listens and reads

Handles more than one type of input or output — text, images, audio, video.

- photos
- scans
- screenshots
- voice recordings
- videos
- forms
- ID documents
- handwritten notes

A system that reads a CV is one thing. One that reads a CV, analyses a video interview and interprets a candidate's facial expression is much more sensitive.

KEY TEACHING POINT

More input types can mean more risk. Text, images, voice and video each bring different privacy, fairness and accuracy concerns.

One large model, many uses

A foundation model is a large AI model trained on broad data that can be adapted for many different tasks — the base for many AI applications.

IT MAY BE USED FOR

- summarising documents
- drafting emails
- answering questions
- writing code
- classifying complaints
- translating text
- generating images
- analysing sentiment
- helping with search

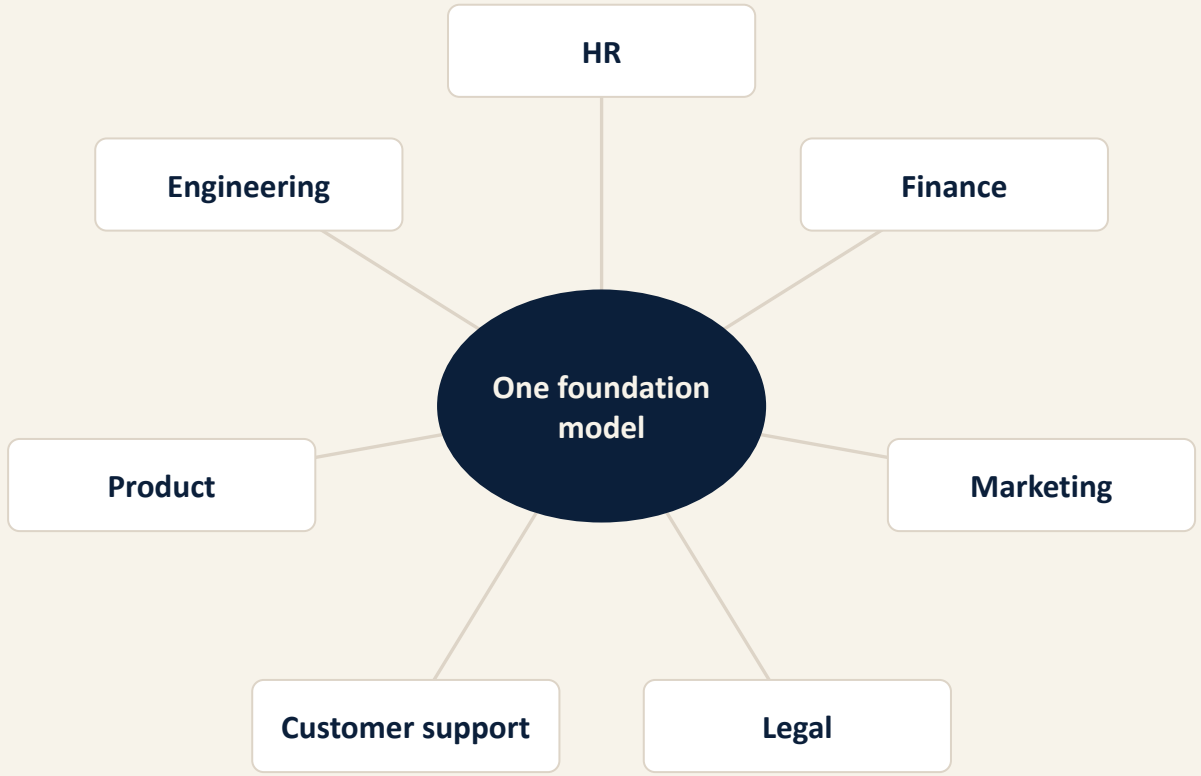
IN PLAIN TERMS

A foundation model is like a Swiss Army knife.

Knife, scissors, bottle opener, screwdriver, file. One tool, many uses. Useful. Flexible. But risky in the wrong situation.

You would not use a Swiss Army knife for surgery. Just because a model can do many things does not mean it should be used for every task.

Foundation models scale risk



If the model has a flaw, bias, security weakness or limitation, that issue can spread across every one of these use cases.

THE GOVERNANCE QUESTION

Where is this model being used, by whom, for what purpose, and with what controls?

Powerful because they support many tasks. Need governance because many tasks mean many risks.

LLMs: a foundation model focused on language

An LLM can read and generate text. It can:

- answer questions
- summarise documents
- draft emails
- rewrite text
- classify messages
- translate
- create reports
- help with code
- generate ideas
- explain concepts

EXAMPLE

“What is our company refund policy?”

If the model doesn’t have the real policy, it may still produce a confident, professional, believable answer — that is wrong. That is why organisations use RAG, human review and source checking.

IN PLAIN TERMS

An LLM is like autocomplete on your phone — but much more powerful.

Your phone isn’t thinking when it predicts the next word. It is guessing what is likely to come next.

An LLM does something similar at a much larger scale. That is why it can sound very confident even when it is wrong.

Hallucination: fluent, confident — and wrong

DEFINITION

When an AI gives an answer that sounds fluent but is false, unsupported or made up.

The AI sounds like it knows — but it is guessing.

Why it happens: the model predicts likely words, not verified facts. If it has no real source, it fills the gap with something that sounds right.

EVERYDAY EXAMPLE

“Turn left, then second right — the building is opposite Tesco.”

They sound sure. But they have never been there. That is dangerous — because confidence makes people trust them. LLMs can do the same.

KEY TEACHING POINT

Fluent language is not the same as accurate language. A well-written wrong answer is still wrong.

Governance questions for LLMs

1. What is the model allowed to do?
2. What is it not allowed to do?
3. What data is it using?
4. Can it access sensitive information?
5. Can users rely on its output?
6. What happens if it gives a wrong answer?
7. Is there human review?
8. Are outputs logged?
9. Are users warned about limitations?
10. Has hallucination risk been tested?
11. Is it used in high-impact decisions?
12. Is there a fallback process?

RAG: the AI looks something up before answering

Sounds technical — the idea is simple. RAG connects the AI model to documents, policies, databases or websites, so it answers from real sources rather than general knowledge.



IN PLAIN TERMS: CHECKING THE MANUAL

A bad call-centre worker guesses. A good one says “Let me check the policy”, opens the manual, finds the right page — then answers. RAG gives the AI the manual. That reduces guessing — but does not remove risk.

Looking it up doesn’t guarantee the right answer

RAG CAN STILL GO WRONG. IT MAY…

- retrieve the wrong document
- use an outdated policy
- miss the most relevant source
- access a document the user shouldn’t see
- expose sensitive information
- summarise the right document badly
- combine sources incorrectly
- cite a source that doesn’t fully support the answer

IN PLAIN TERMS

The wrong letter from the drawer.

You ask someone to check your mortgage paperwork. They go to your drawer, pull out the wrong letter — and confidently explain it.

They did look something up. They just looked up the wrong thing. That is the risk with RAG.

Governance questions for RAG

1. What sources can the AI search?
2. Are the documents up to date?
3. Who owns the documents?
4. Are permissions working properly?
5. Can users see only what they are allowed to see?
6. Are answers linked to sources?
7. What happens if the source is wrong?
8. How often is the knowledge base reviewed?
9. Are sensitive documents excluded?
10. Is retrieval tested?
11. Are users told to check important outputs?

KEY TEACHING POINT

RAG can make AI more grounded. But if the source is wrong, outdated or wrongly accessed, the answer is still risky.

A chatbot answers. An agent acts.

An AI agent is an AI system that can plan steps, use tools and take actions — send an email, create a Jira ticket, book a meeting, update a customer record, trigger a workflow.

CHATBOT

“What flights are available to Lagos?”

It replies with options.
You still decide what to do.

A mistake is just advice — you can catch it.

AGENT

“Book me the cheapest flight to Lagos next Friday.”

Searches | Chooses | Books | Pays

A mistake is no longer advice — it becomes an action.

In plain terms: imagine a travel assistant that doesn’t just suggest — it spends your money.

An agent at work: faster — and riskier

WHAT THE AGENT DOES

1. Reads a customer complaint
2. Decides it is urgent
3. Creates a Jira ticket
4. Assigns it to a team
5. Emails the customer
6. Updates the CRM

BUT WHAT IF IT…

- sends the wrong email?
- updates the wrong customer record?
- exposes sensitive data?
- assigns work incorrectly?
- misunderstands the complaint?
- loops and repeats actions?
- acts without approval?

Governance questions for AI agents

1. What actions can the agent take?
2. Does it need human approval first?
3. What systems can it access?
4. What permissions does it have?
5. Can it send messages externally?
6. Can it change records?
7. Can it spend money?
8. Can it make decisions affecting people?
9. Is every action logged?
10. Can the action be reversed?
11. Is there a kill switch?
12. Who is accountable if it acts wrongly?



KEY TEACHING POINT

The more AI can act, the stronger the governance needs to be. A chatbot mistake can be corrected before action. An agent mistake may already have happened.

04

AI in the real world

How a model learns versus how it is used — why outputs can hurt people — and why AI is never “set and forget”.

IN THIS SECTION

01 Training vs inference

02 Why AI outputs can be risky

03 The AI lifecycle

Two stages: learning, then doing

STAGE 1

Training

The model learns from data. It is shown examples, adjusts itself and learns patterns.

Like revising for an exam

Fruit: labelled images of apple, banana, mango, orange.

Recruitment: trained on historical CV and hiring data.

STAGE 2

Inference

The trained model is used on a new, real-world input and produces an output.

Like sitting the exam

Fruit: a new photo → "This is probably a mango."

Recruitment: a new candidate applies → the AI scores the CV.

KEY TEACHING POINT

Bad training data can create bad model behaviour. Weak inference controls can create unsafe real-world use. Governance must cover both.

Governance questions at each stage

AT TRAINING STAGE

1. What data was used?
2. Was it lawful to use that data?
3. Was the data accurate?
4. Was the data biased?
5. Were some groups missing?
6. Who labelled the data?
7. Was the model tested?
8. What assumptions were made?
9. Was sensitive data involved?
10. Was the training data representative?

AT INFERENCE STAGE

1. Who is using the model?
2. What input are they giving it?
3. What output does it produce?
4. Is the output reviewed?
5. Can a human override it?
6. Is the output logged?
7. Is the model still performing well?
8. Is there drift?
9. What happens when the output is wrong?
10. Is it used for its intended purpose?

Clean, professional, confident — and still wrong

AI OUTPUTS MAY BE

- incorrect
- biased
- misleading
- incomplete
- overconfident
- difficult to explain
- inconsistent
- based on poor data
- unsafe in high-impact settings
- based on outdated sources
- misused by humans
- accepted without review

LOW RISK

Music app recommends the wrong song.

You skip it. No big harm.

MEDIUM RISK

Chatbot gives the wrong refund information.

Frustration, complaints, the company has to fix it.

HIGH RISK

Recruitment AI unfairly rejects a good candidate.

Lost opportunity. Legal, ethical, reputational issues.

VERY HIGH RISK

Health triage tool says symptoms aren't urgent — they are.

Someone could be seriously harmed.

Impact on people increases →

KEY TEACHING POINT

AI risk depends on where the AI is used and what happens after the output.

IN PLAIN TERMS

A wrong answer is not always the same level of problem

Your sat nav takes a slightly longer road	Annoying
An AI wrongly flags you as high fraud risk and blocks your account	Serious
An AI recruitment system rejects you unfairly	Serious
An AI health system gives unsafe advice	Very serious

KEY TEACHING POINT

Context matters. The same kind of mistake can be a shrug in one setting and serious harm in another.

Governance covers the full lifecycle — not just launch

BEFORE LAUNCH

1. Idea
2. Use-case assessment
3. Data collection
4. Design
5. Development
6. Testing

AFTER LAUNCH

7. Deployment
8. Monitoring
9. Change management
10. Incident response
11. Retirement

Keep checking

KEY TEACHING POINT

AI is not “set and forget”. Once it goes live, someone must keep checking it.

One governance question at every stage

1	**Idea**	Should we use AI for this at all?
2	**Use-case assessment**	What decision or process will this AI affect?
3	**Data collection**	Is the data lawful, fair, accurate and suitable?
4	**Design**	Where should humans be involved?
5	**Development**	Are controls being built in?
6	**Testing**	Tested for accuracy, bias, safety, privacy, security, misuse and fairness?
7	**Deployment**	Who approved release — and based on what evidence?
8	**Monitoring**	Is it still behaving safely and fairly?
9	**Change management**	Does this change create new risk?
10	**Incident response**	Who responds, who is informed, and what is fixed?
11	**Retirement**	What happens to data, users, records and evidence?

The first question is never "can we build it?" — it is "should we use AI here at all?"

IN PLAIN TERMS

Running AI is like running a restaurant

You don’t just unlock the door and start serving food.

BEFORE OPENING

- idea
- menu
- ingredients
- kitchen setup
- staff training
- food testing
- hygiene checks
- allergy checks
- pricing
- customer service process

AFTER OPENING — STILL NEEDED

- customer feedback
- food safety checks
- stock checks
- complaint handling
- menu updates
- staff training
- inspections

THE POINT

AI is similar. You do not just launch it and walk away.

05

People, governance and risk

AI governance is not only about the model. It is about the people and organisations around it — and who is accountable.

IN THIS SECTION

01 AI actors and responsibilities

02 Why AI governance exists

03 Introducing AI risk

Who is involved around an AI system?

Developer

Builds the AI system or model — creates it, writes code, tests, fixes bugs.

A technical team builds HireAI.

Provider

Makes the AI system available to others — sells, licenses or provides it as a service.

A software company sells HireAI to employers.

Deployer

Uses the AI system in a real organisation or process.

A company uses HireAI in its recruitment.

User

Directly interacts with the AI system.

A recruiter reviewing candidates in HireAI.

Affected person

Impacted by the output — may never log in or see the AI.

A job applicant whose CV is scored.

Supplier

Provides part of the system: cloud, model, platform, data, integration, testing.

Their work can still create risk.

What deployers must ask — and who they must remember

THE DEPLOYER MUST ASK

1. Is this system suitable for our use?
2. Have we tested it in our context?
3. Are our recruiters trained?
4. Are candidates informed?
5. Do we have human oversight?
6. Do we have evidence?

THIS IS VERY IMPORTANT

The user is not always the person most affected.

A job applicant whose CV is scored by HireAI may never log into the system. They may never see the AI.

But they may be affected by it — and they are the person governance exists to protect.

Ownership roles: governance, risk and control

GOVERNANCE OWNER

Makes sure the process is followed

- policy
- approval
- evidence
- risk assessment
- documentation
- monitoring
- escalation

RISK OWNER

Owns the decision and the risk

- For HireAI: HR director, recruitment lead, business owner or AI governance lead.
- Doesn’t need to be the technical person.
- But must own the decision and the risk.

CONTROL OWNER

Makes sure a control actually works

- human review before rejection
- fairness testing
- access control
- candidate transparency notice
- appeal route
- monitoring dashboard
- Every control needs an owner.

IN PLAIN TERMS

If everyone is responsible, nobody is responsible

Chef	Supplier	Manager	Waiter	Customer	Council
cooks	provides ingredients	runs the restaurant	serves customers	eats the food	inspects hygiene

If something goes wrong, you need to know who was responsible for what. AI is the same.

You cannot govern AI properly if nobody knows who owns the system, the data, the risk, the controls and the final decision.

KEY TEACHING POINT

AI governance needs clear roles. If everyone is responsible, nobody is responsible.

Using AI properly, safely and accountably — with proof

DEFINITION

AI governance is the system of policies, roles, processes, controls and evidence used to manage AI responsibly.

In plain terms: making sure AI is used properly, safely and accountably — with proof.

IT HELPS ORGANISATIONS ANSWER

1. Should we use AI here?
2. Who owns the system?
3. Who owns the risk?
4. Who could be harmed?
5. What data is being used?
6. What laws or regulations apply?
7. What testing has been done?
8. What controls are in place?
9. What evidence exists?
10. Who approves deployment?
11. How will it be monitored?
12. What happens if it goes wrong?

Governance in action: “We want AI to screen job applicants”

THE TEAM SAYS

“We want to use AI to screen job applicants.”

GOVERNANCE SAYS

“Before we do that, let’s understand the risk.”

1. Could it discriminate?
2. What data will it use?
3. Can candidates challenge the result?
4. Is there human review?
5. Is personal data used lawfully?
6. What if it rejects good candidates?
7. Who signs this off?
8. What evidence do we have?

That is governance. It is not blocking innovation — it is making sure innovation does not become harm.

KEY TEACHING POINT

Good governance does not slow AI down for no reason. Good governance helps teams move with confidence.

Risk = Likelihood × Impact

Two questions: How likely is it to happen? How serious would it be if it did?

EVERYDAY	**Forgetting an umbrella** Likelihood: High · Impact: Low You get wet. Annoying, not serious.	**Missing a mortgage payment** Likelihood: Medium · Impact: High Could affect your finances and credit.	**Smoke alarm fails in a fire** Likelihood: Low · Impact: Very high That is why we still test smoke alarms.
AI	**Wrong Netflix recommendation** Likelihood: High · Impact: Low You skip it.	**Wrong recruitment recommendation** Likelihood: Medium · Impact: High Someone could lose a job opportunity.	**Wrong healthcare recommendation** Likelihood: Low · Impact: Very high Even if unlikely — someone could be harmed.

Where AI risk comes from

DATA	poor data; biased historical data
MODEL & OUTPUT	wrong outputs; hallucination; lack of explainability; model drift
PEOPLE & PROCESS	weak human oversight; overreliance on automation; poor testing; lack of monitoring; misuse; wrong deployment context
SECURITY & PRIVACY	privacy breaches; security vulnerabilities
ORGANISATION	supplier failure; unclear ownership; no incident process



KEY TEACHING POINT

AI risk is not only technical. It is human, legal, operational, reputational and organisational. A model can be technically impressive and still be badly governed.

06

Case study: HireAI

A simple AI recruitment screening system — used all programme to make every idea practical.

IN THIS SECTION

01 Meet HireAI

02 What could go wrong

03 Who could be harmed

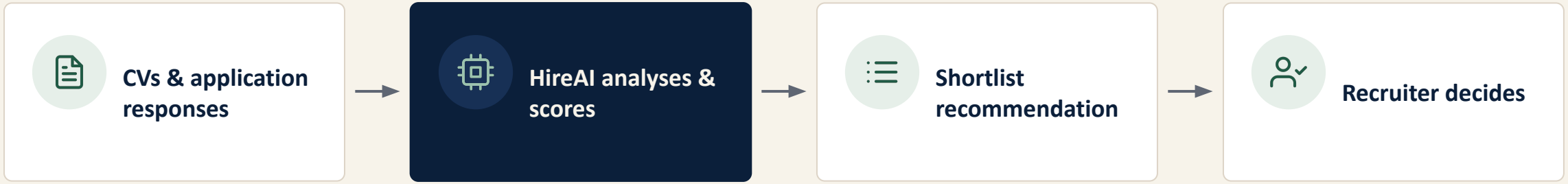
04 Should AI decide?

05 Workshop

06 AI System Description

Meet HireAI

HireAI is an AI recruitment screening system. It analyses CVs and application responses, then recommends which candidates should be invited to interview.



THE ORGANISATION WANTS TO

- save recruiter time
- screen applications faster
- improve consistency
- reduce manual workload
- support hiring managers

At first, this sounds useful.

NOW THINK LIKE A PRACTITIONER

HireAI affects real people and real opportunities. If it gets it wrong, someone may lose the chance to interview. That is not the same as recommending the wrong film.

What could go wrong?

- It may learn bias from historical hiring data
- Candidates from certain backgrounds may be ranked lower
- CV gaps may be misunderstood
- Non-traditional career paths may be penalised
- Disabled candidates may be disadvantaged
- Older candidates may be unfairly scored
- Foreign names or qualifications may be treated unfairly
- It may overvalue certain keywords
- Recruiters may trust the score too much
- Candidates may not know AI was used
- There may be no proper appeal route
- Nobody may know who is accountable

Who could be harmed — and what data could cause problems?

WHO COULD BE HARMED

- job applicants
- disabled candidates
- older candidates
- career changers
- candidates with employment gaps
- candidates with foreign qualifications
- recruiters
- hiring managers
- the organisation
- HR and legal teams
- the company’s reputation

DATA THAT COULD CREATE PROBLEMS

- historical hiring decisions
- CV text
- application forms
- interview scores
- performance reviews
- education history
- employment gaps
- postcode
- nationality indicators
- names
- gender indicators
- age indicators
- disability-related information
- scraped LinkedIn / profile data
- biased job descriptions
- incomplete records

Should AI make the final decision?

THE SIMPLE ANSWER

No.

AI may support screening, help recruiters prioritise and identify matches. But it should not quietly become the final decision-maker. The human review must be meaningful — a rubber stamp is not enough.

A HUMAN SHOULD STILL CONTROL

1. Final shortlist decision
2. Final rejection decision
3. Ability to override the AI score
4. Review of edge cases
5. Explanation to candidates
6. Handling of appeals
7. Review of fairness concerns
8. Sign-off before deployment
9. Ongoing monitoring
10. Decision to pause or stop the system

KEY TEACHING POINT

AI can support a decision. But when people's opportunities are affected, accountability must stay with humans.

Workshop: first AI risk thinking

USING HIREAI, ANSWER THESE FIVE QUESTIONS

1. What is the AI system doing?
2. Who uses the system?
3. Who is affected by the system?
4. What could go wrong?
5. What should a human still control?

GUIDANCE

Use plain, practical language. Don’t try to sound like a lawyer or a data scientist. The goal is to understand the system clearly.

INSTEAD OF WRITING

“Potential adverse algorithmic impact on protected subpopulations.”

WRITE

“The system may unfairly reject some candidates because it learned from biased hiring history.”

Example answers

1	**What is the AI system doing?**	Analyses CVs and application responses, then recommends which candidates should be invited to interview.
2	**Who uses the system?**	Recruiters and hiring managers.
3	**Who is affected by the system?**	Job applicants — especially candidates rejected before interview.
4	**What could go wrong?**	It may rank candidates unfairly because historical hiring data contains bias. It may overvalue keywords, degrees or polished CV formatting and miss strong candidates from different backgrounds.
5	**What should a human still control?**	The final shortlist and rejection decisions. The recruiter should be able to override the AI score, review edge cases and explain decisions.

Your first evidence item: the AI System Description



You cannot govern what you cannot describe.

If a team cannot clearly explain what the AI system does, they are not ready to deploy it.

THE 14 FIELDS

1. System name
2. Business purpose
3. Users
4. Affected people
5. AI capability
6. AI model type
7. AI actors
8. Data used
9. Decision supported
10. Human role
11. Deployment context
12. Potential harms
13. Key assumptions
14. Governance questions

HireAI System Description (1 of 2)

1 **System name**	HireAI — CV Screening Assistant
2 **Business purpose**	To speed up first-pass CV review and help recruiters shortlist candidates more consistently.
3 **Users**	Recruiters and hiring managers.
4 **Affected people**	Job applicants, especially those screened out before interview.
5 **AI capability**	Analyses CVs and application responses, ranks candidates against role requirements, and produces a candidate score and shortlist recommendation.
6 **AI model type**	Likely a machine-learning or AI scoring system, possibly using natural language processing on CV text. If an LLM is used, extra risks include hallucination, inconsistent explanations and data leakage.
7 **AI actors**	Provider: company supplying HireAI · Deployer: employer using it · User: recruiter / hiring manager · Affected: job applicant · Supplier: cloud, model or platform provider · Risk owner: HR or recruitment lead · Control owner: responsible for fairness testing, human review and monitoring.

HireAI System Description (2 of 2)

8	**Data used**	CV text, application answers, job descriptions and historical hiring outcomes. Possibly interview notes, recruiter comments, qualification and employment history.
9	**Decision supported**	Who progresses to interview. It should not decide who is hired, or silently reject candidates without meaningful human review.
10	**Human role**	A recruiter reviews the AI recommendation before any rejection is sent, can override the score, and has the information, time and authority to challenge the AI.
11	**Deployment context**	First-pass screening only. Not for final hiring or salary decisions. Not used without candidate transparency and governance review.
12	**Potential harms**	Unfair rejection, reduced opportunity, bias against certain groups, lack of transparency, poor explanation, reputational damage, legal risk, overreliance by recruiters.
13	**Key assumptions**	Historical hiring data is reliable and not systematically biased; recruiters meaningfully review; candidates are informed; the model is fair across groups; the supplier can provide evidence. These must be tested.
14	**Governance questions**	Tested for bias? Do candidates know AI is used? Meaningful human review? Can rejections be challenged? Who owns the system, the risk and the controls? What testing, evidence and post-launch monitoring exist?

07

Wrap-up

Check your understanding, take the homework away, and see where we go next.

IN THIS SECTION

01 Knowledge check

02 Glossary

03 Homework

04 Recap

05 Next week

Can you answer these? (1 of 2)

1. What is AI in simple language?
2. How is AI different from traditional software?
3. What is Machine Learning?
4. What is Deep Learning?
5. What is Generative AI?
6. What is supervised learning?
7. What is unsupervised learning?
8. Classic AI vs generative AI — the difference?
9. Small model vs large model — the difference?
10. Single-purpose vs general-purpose model?
11. Proprietary vs open-source model?
12. What is a multimodal model?
13. What is a foundation model?
14. What is an LLM?
15. What is hallucination?

Can you answer these? (2 of 2)

16 What is RAG?

17 What is an AI agent?

18 What is training?

19 What is inference?

20 Why can AI outputs be risky?

21 What is the AI lifecycle?

22 Why should governance start before launch?

23 Who is a provider?

24 Who is a deployer?

25 Who is a user?

26 Who is an affected person?

27 What is AI risk?

28 Why does context matter?

29 What belongs in an AI System Description?

30 Why can’t we govern what we cannot describe?

Key terms (1 of 2)

Artificial Intelligence	Software that performs tasks normally associated with human intelligence — recognising patterns, generating content or supporting decisions.
Machine Learning	A type of AI where systems learn patterns from data instead of only following fixed rules.
Deep Learning	Machine learning using neural networks; strong at complex data like images, speech and language.
Generative AI	AI that creates new content such as text, images, audio, video or code.
Supervised Learning	Learning from examples where the correct answer is known. e.g. fruit photos labelled apple, banana or mango.
Unsupervised Learning	Finding patterns or groups without being given the answer. e.g. grouping customers by shopping behaviour.
Classic AI	AI that predicts, classifies, detects, scores or recommends. e.g. fraud detection or spam filtering.
Foundation Model	A large AI model trained broadly and adapted for many tasks.
Large Language Model	A foundation model focused on reading and generating text.
Hallucination	A fluent but false, unsupported or made-up AI output.
RAG	Retrieval-Augmented Generation: the model retrieves information from documents or data sources before answering.

Key terms (2 of 2)

AI Agent	An AI system that can plan, use tools and take actions — not just reply.
Training	The stage where a model learns patterns from data.
Inference	The stage where a trained model produces an output for a new input.
AI Lifecycle	The journey from idea, design and testing through deployment, monitoring, change, incidents and retirement.
Provider	The organisation or person that makes the AI system available.
Deployer	The organisation using the AI system in a real process.
User	The person directly interacting with the AI system.
Affected person	A person impacted by the AI system, even if they do not use it directly.
AI Governance	Policies, roles, processes, controls and evidence used to manage AI responsibly.
AI Risk	The possibility that an AI system could cause harm or create negative consequences.

Find the AI hiding in your own apps

Pick one AI feature you already use, then complete a short AI System Description using the 14 fields.

IDEAS TO CHOOSE FROM

- bank fraud alert
- Netflix recommendation
- Spotify recommendation
- email spam filter
- Google Maps route suggestion
- social media feed ranking
- work chatbot
- phone face unlock
- smart reply in Gmail or Outlook
- AI writing assistant
- customer service chatbot
- fitness app recommendation
- food delivery recommendation

IN PLAIN TERMS

It is okay to write “unknown”.

That is part of the learning.

Often the unknown fields are the most important — they show where transparency, ownership or evidence may be missing.

AI System Description — your template

1. **System name**
   [Write the name]
2. **Business purpose**
   [Why does this AI exist?]
3. **Users**
   [Who uses it?]
4. **Affected people**
   [Who could be affected by the output?]
5. **AI capability**
   [Prediction, recommendation, generation, classification, recognition, action]
6. **AI model type**
   [Classic AI, GenAI, LLM, RAG, agent, multimodal, unknown]
7. **AI actors**
   [Provider, deployer, user, affected person, supplier]
8. **Data used**
   [What data does it probably use?]
9. **Decision supported**
   [What decision or action does it influence?]
10. **Human role**
    [Does a person review or approve the output?]
11. **Deployment context**
    [Where and how is it used?]
12. **Potential harms**
    [What could go wrong?]
13. **Key assumptions**
    [What are we assuming is true?]
14. **Governance questions**
    [What would you ask before trusting it?]

Worked example: bank fraud detection (1 of 2)

1	**System name**	Bank fraud detection system
2	**Business purpose**	To detect suspicious transactions and protect customers from fraud.
3	**Users**	Bank fraud team, customer support team and automated transaction monitoring systems.
4	**Affected people**	Bank customers.
5	**AI capability**	Detects unusual transactions and produces a fraud risk score or alert.
6	**AI model type**	Classic AI / machine learning classification or anomaly detection model.
7	**AI actors**	Provider: bank or fraud-detection vendor · Deployer: bank using the system · User: fraud analyst or support agent · Affected: bank customer · Supplier: payment processor, cloud provider or fraud platform vendor.

Worked example: bank fraud detection (2 of 2)

8	**Data used**	Transaction amount, location, merchant, time, customer spending history and device information.
9	**Decision supported**	Whether a transaction should be approved, blocked or sent for review.
10	**Human role**	A fraud analyst or support process may review flagged transactions. The customer may be contacted to confirm whether it is genuine.
11	**Deployment context**	Used during payment processing and fraud monitoring.
12	**Potential harms**	A genuine transaction may be blocked wrongly — the customer may be embarrassed, delayed, financially affected or unable to access money.
13	**Key assumptions**	Unusual behaviour may indicate fraud. But unusual is not always fraud — a customer may be travelling, buying a gift or making an emergency payment.
14	**Governance questions**	How often are genuine transactions wrongly blocked? Can customers resolve blocks quickly? Are vulnerable customers protected? Is there human review? Are false positives monitored? Who owns the risk? What evidence shows it works fairly?

What we learned today

- AI spots patterns, makes predictions, generates content or supports decisions.
- AI is a family: Machine Learning, Deep Learning and Generative AI.
- Traditional software follows fixed rules; AI often learns from data.
- Supervised learning uses labelled examples.
- Unsupervised learning finds patterns without labels.
- Classic AI predicts, classifies, scores or recommends.
- Generative AI creates new content.
- Small and large models carry different risks.
- Proprietary and open-source models carry different responsibilities.
- Text and multimodal models bring different privacy and safety risks.
- LLMs can produce fluent but wrong answers.
- RAG helps AI look things up — but the source can still be wrong.
- AI agents are riskier because they can take actions.
- Training is when the model learns; inference is when it is used.
- AI risk depends on context.
- Governance must start at the idea stage.
- AI governance needs clear actors and responsibilities.
- **You cannot govern an AI system properly if you cannot describe it clearly.**

CLOSING MESSAGE

Before we ask whether an AI system is compliant, we first need to understand what it is doing.

Not to turn everyone into engineers. Not to bury people in technical language. But to give you the confidence to ask proper questions.

That is where AI governance begins.

- What is this AI doing?
- Who uses it?
- What type of AI is it?
- Who is affected by it?
- What data is it using?
- Who could be harmed?
- Who built it?
- Who owns the risk?
- Who provides it?
- What controls exist?
- Who deploys it?
- What evidence do we have?

NEXT WEEK

Week 2 — Responsible AI, Trustworthy AI and Harm

WE WILL COVER

- responsible AI
- trustworthy AI
- fairness
- bias
- transparency
- explainability
- accountability
- privacy
- safety
- security
- human oversight
- automation bias
- types of AI harm
- stakeholders and affected people
- harm mapping

Bring your homework: one AI System Description for an app you use.



HireAI continues

We keep using HireAI to make every idea practical — mapping who could be harmed, and how.

PQ PADEMIQ

Thank you

Questions, challenges and real examples welcome.

Collins Obasuyi
Lead Consultant
PADEMIQ / AI GRCA Lab