Generative AI and responsible use

Ask a chatbot who founded your village's football club, and it answers in a second: a name, a year, a nice sentence about the founders. It sounds certain. Is it true?

Sometimes it is. Sometimes the name, the year and even the "source" it quotes were invented on the spot, written just as fluently as the truth. Companies in the Balearic Islands already use these tools to answer guests, write descriptions and draft code. A hotel chatbot that invents a cancellation policy is not a funny mistake: it is a complaint, maybe a lawsuit. Today you learn how these models work, how to catch them when they are wrong, and which rules Europe has set for them.

What you will learn

- Explain what **generative AI** is and how a **large language model** writes, without magic
- Catch a **hallucination** by checking a model's answers against real sources
- Write rules for the **responsible use** of AI in a company
- Classify AI systems by the four risk levels of the **AI Act**, and know which parts already apply
- Avoid the trap of the day: a model that writes well is not a model that knows

Models that create

> **Generative AI** is AI that **produces new content** (text, code, images, audio, video) from an instruction, instead of only classifying or predicting a value.

Your Teachable Machine model from the last lesson **classified**: correct or defective. A generative model **creates**: a paragraph, an image, a function.

> A **large language model (LLM)** is a very large neural network trained on enormous amounts of text to **predict the next piece of text**. Repeating that prediction again and again, it writes whole answers.

That is the whole trick. It does not look the answer up in a database; it writes the most likely continuation of your text, piece by piece. In the city-builder game it is an advisor that talks with great confidence, even when it has not looked at the map.

Term	What it means
Prompt	The instruction or input text you give the model. More context and more precision give better answers
Token	A piece of text (a word or part of a word) that the model works with. Prices and limits are counted in tokens
Hallucination	A false or invented answer, well written and convincing
Knowledge cutoff	The model only "knows" what was in its training data; anything later it does not know unless you give it, or it searches the web

What it does well	Its limits
Write, summarise, translate and rephrase	On its own it does not check facts: it can invent data, quotes and sources
Explain and generate code, tests and documentation	It can repeat the **biases** in its training data
Pull information out of messy texts	The same question can give different answers
Give you a fast first draft	What you type may leave the company (privacy); it uses a lot of energy

> **Careful: "it sounds sure, so it is right"**
> An LLM writes a correct answer and a wrong answer in exactly the same confident style. The problem is not that it fails: every tool fails. The problem is that it **fails without warning**. That is why every output is a draft until a person checks it.

In class: error hunters

This is the **only activity of the unit where generative AI is authorised**, and only on these terms:

> **What is authorised, and what is not**
> **Authorised:** using the AI assistant your teacher names, **in class, during this activity only**, to ask the questions below and analyse the answers.
>
> **Not authorised:** writing your conclusions with AI, any other class activity, and any part of Deliverable 4. Never type personal data (yours or anyone else's) or internal company data into the assistant.

In pairs. Every answer must be checked against a **real source**: an official website, the press, the language documentation, or by running the code.

Step	Ask the assistant	Check it by
1	Three questions about things you know very well: your village or neighbourhood, your school, a sport or a game you master	Your own knowledge, confirmed with a website
2	Two facts about your project company that you **already cited** in your dossier (year founded, number of hotels or staff…)	Comparing with the source in your dossier
3	A short function in the language you use in Programming	Running it, with a normal case and an edge case
4	The sources for one of its answers	Opening the links: do they exist? Do they say what the model claims?

For each answer, mark **correct**, **wrong** or **cannot be checked**, and paste the link you used to check it. Then name the **type** of error: invented fact, non-existent source, code that does not work, outdated answer, bias.

Rules for responsible use

From the errors you catch, five rules follow. Most companies that use AI write something similar.

Rule	What it means
Check it	Everything a model produces is a draft until someone checks it
Protect the data	No personal data, and no confidential code or documents, in tools the company has not approved
Be transparent	Say when you used AI and what for
You are responsible	Whoever signs the work or publishes the code answers for it, not the tool
Respect authorship	Respect the authorship and licences of content and code

Your group will turn these into rules for **your** company, with at least one rule specific to its sector. In a hotel, for example: “the chatbot never confirms a booking or a price without the booking system”.

This is not only good manners. Since **2 February 2025**, the AI Act requires companies that use AI to make sure their staff have enough knowledge of it, called **AI literacy**.

The AI Act: rules by risk

> The **AI Act** is **Regulation (EU) 2024/1689**, the European law on artificial intelligence. It entered into force on **1 August 2024** and applies in stages.

It mostly affects **providers** (whoever develops or sells an AI system) and **deployers** (whoever uses one professionally) in the European Union. Its idea is a **risk-based approach**: the more harm a system can do to people, the more obligations it has. The European Commission summarises it in four levels.

Level	What it means	Examples
Unacceptable	Prohibited practices	Social scoring of citizens; manipulation that exploits vulnerabilities; emotion recognition at work or at school (except for medical or safety reasons)
High	Allowed with strict obligations: risk management, data quality, documentation, logging, human oversight	Sorting CVs to hire people, assessing students, deciding on loans, critical infrastructure
Limited (transparency)	Transparency obligations	Chatbots must say you are talking to a machine; generated or manipulated content (*deepfakes*) must be labelled
Minimal	No specific obligations	Spam filters, AI in video games

Large language models also have their own obligations, as **general-purpose AI models**: documentation, transparency and respect for copyright, with more obligations for the biggest ones.

What applies, and since when

Date	What starts to apply
1 Aug 2024	Entry into force
2 Feb 2025	Prohibited practices and AI literacy
2 Aug 2025	Governance and the obligations of general-purpose AI models
2 Aug 2026	General application, including transparency obligations

Date	What starts to apply
2 Dec 2027	High-risk systems such as hiring, credit or education (postponed by the *AI Omnibus*)
2 Aug 2028	High-risk systems built into regulated products

> **The calendar has already changed once**
> The high-risk dates were first set for 2026 and 2027. An amendment, the **AI Omnibus**, entered into force on **27 July 2026** and pushed them back. When you write about the AI Act, check the current dates on the European Commission's page and cite it: laws move. And do not confuse it with the **GDPR** (Regulation (EU) 2016/679), the data protection law you will see in Unit 5.

In class: which risk level?

In your project groups, classify each case and justify it in one sentence. Then classify the two AI ideas you noted for your own company in the last lesson.

Case	Your level
A chatbot that answers customers' questions on a hotel's website	
A program that sorts CVs and rejects candidates	
The spam filter of a transport company	
A public administration system that scores citizens by their behaviour	
An AI that marks exams and decides who gets into a course	
A promotional video with a person's voice generated by AI	
Cameras that detect the mood of office workers	
A model that decides whether a bank grants a loan	

Hand in on Classroom (in pairs, "S16 · Error hunters and risk levels", by 23:59 the same day): your error table with the link for every check, the five rules adapted to your company, and the classification of the eight cases and of your two ideas. Remember: the AI authorisation ended with the error-hunting activity.

For your project

This lesson feeds the **group part** of Deliverable 4 · AI plan:

Today's work	Becomes	Criterion
Your five rules, adapted to the company	Group part: responsible-use rules	c
The risk level of your own ideas	Group part: the AI Act risk level of each proposal	c
What generative AI can and cannot do	Section C: present and future of AI in the sector	c, f

Key takeaways

- **Generative AI** creates content; an **LLM** writes by predicting the next piece of text, again and again.
- It writes right and wrong answers equally well: **hallucinations** come without warning. Always check against sources.

- Responsible use: **check it**, **protect the data**, **be transparent**, **you are responsible**, **respect authorship**.
- The **AI Act** (Regulation (EU) 2024/1689) follows a **risk-based approach**: unacceptable, high, limited (transparency), minimal.
- Prohibitions and AI literacy apply since 2 February 2025; transparency since 2 August 2026; high-risk obligations were postponed to December 2027 and August 2028.
- In this module, AI is allowed **only** when the teacher authorises it for a specific activity.

Quick check

1 What does a large language model basically do when it answers?

A It looks the answer up in a database of verified facts
B It predicts, piece by piece, the most likely continuation of the text
C It searches the internet and copies the best page
D It runs rules written by experts for every question

2 A chatbot gives you a quote from a report, with the title and the year. You search for the report and it does not exist. What is this called?

A A token
B A prompt
C A hallucination
D A knowledge cutoff

3 Which two actions follow the rules of responsible use?

A Pasting a list of customers with their phone numbers into a public chatbot to write emails faster
B Telling your team that a first draft was written with an AI tool the company approved
C Testing a function generated by an assistant before adding it to the project
D Publishing an AI answer as it is, because the tool is responsible for it

4 Under the AI Act, which level is a program that sorts CVs and rejects candidates?

A Unacceptable: it is prohibited
B High risk: allowed with strict obligations such as human oversight
C Limited: it only has to say it is an AI
D Minimal: no obligations

5 A hotel puts a chatbot on its website. What is the minimum the AI Act requires?

A Nothing: chatbots are minimal risk
B A licence from the European Commission
C Telling users that they are talking to an AI
D Switching it off at night

6 Which of these is a prohibited practice under the AI Act?

A A spam filter
B A video game with AI-controlled enemies
C Cameras that recognise the emotions of workers in an office
D A recommender in an online shop

7 Your group proposes an AI that predicts which machines in a bakery will break down next week, from their sensor data. Which risk level fits best, and what must you still do?

- **A** Unacceptable; you must drop the idea
- **B** Probably minimal: it decides nothing about people, but a person should still check its warnings
- **C** High, because it uses sensors
- **D** Limited, so it must announce it is an AI to the machines

You know how a generative model writes, why you check it, and what Europe asks of it. Now switch sides: you are the company, and you want to know whether AI pays. Next: Automation and profitability.

Solutions

1. B · An **LLM** is a predictor of text. That is why it can write a convincing answer that is false: nothing in the prediction checks the facts.

2. C · An invented but convincing answer is a **hallucination**. Invented sources are one of its most common forms, which is why you always open the links.

3. B, C · **Be transparent** and **check it**. Personal data never go into unapproved tools, and the person who publishes is the one responsible, not the tool.

4. B · Hiring decisions affect people's lives, so the AI Act lists them as **high risk**: risk management, data quality, documentation and **human oversight**.

5. C · Chatbots are in the **limited (transparency)** level: people must know they are talking to a machine. Those transparency obligations apply since **2 August 2026**.

6. C · Emotion recognition **at work or at school** is an **unacceptable** practice (except for medical or safety reasons), prohibited since 2 February 2025.

7. B · A maintenance prediction does not decide about people's rights, so it fits **minimal** risk. Responsible use still applies: someone **checks** before acting, and the data must be good.