

Using Artificial Intelligence.

An introduction to the benefits, risks and use-cases of AI

Peter Thomas



What we will cover

- Introduction to AI
 - Just enough theory so you understand the basic operation of an AI model
- Risks of AI
 - Understanding of the risks of AI based on how it works
 - How to mitigate the risks
- Responsible usage of AI
 - Suitable use cases
 - Managing risk effectively
 - Policies and Procedures
- SASC's guidance on AI usage

Disclaimer: Views and guidance expressed within are my own and not reflective of my current or previous employers.



What is Artificial Intelligence?

Artificial Intelligence is the theory and development of computer systems able to perform tasks normally requiring human intelligence



Using AI

- AI is embedded in almost all technology that you will use.
- Most people will encounter AI through assistance applications such as Siri or Alexa where AI is used as an enabler for the function of the assistant.
- But also, for direct usage of AI, you may be familiar with ChatGPT, Gemini or Claude. These are all products providing a specific type of AI called Generative AI.
- Generative AI is a type of artificial intelligence that creates entirely new content—such as text, images, audio, and video—in response to the user’s inputs.

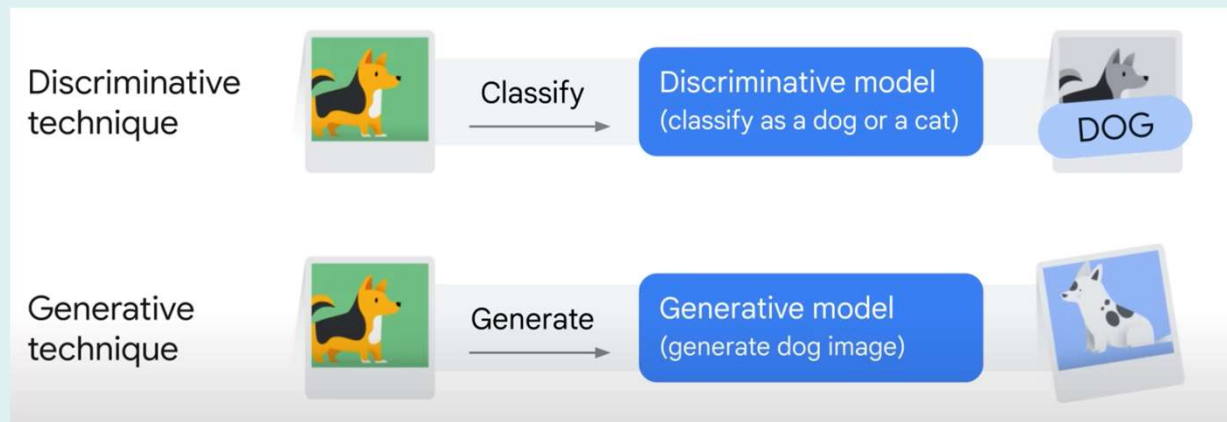
Discriminative versus Generative AI

- Discriminative AI

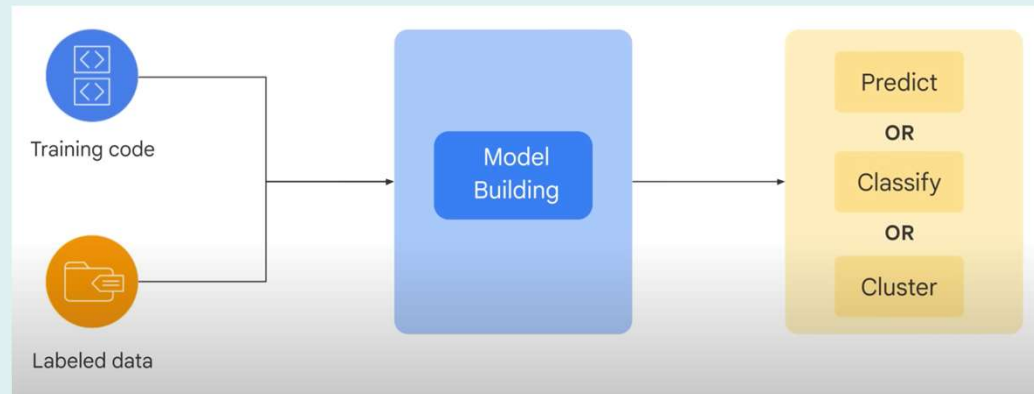
- Used to classify or predict based on input data
- Trained on a specific dataset of labelled data
- Learns the relationship between the features of the data points and the labels

- Generative AI

- Generates new data similar to the data it was trained on
- Trained on a large amount of general data
- Understands the distribution of data and how likely a given example is
- Predicts the next item in the sequence



How is a Discriminative AI model created?

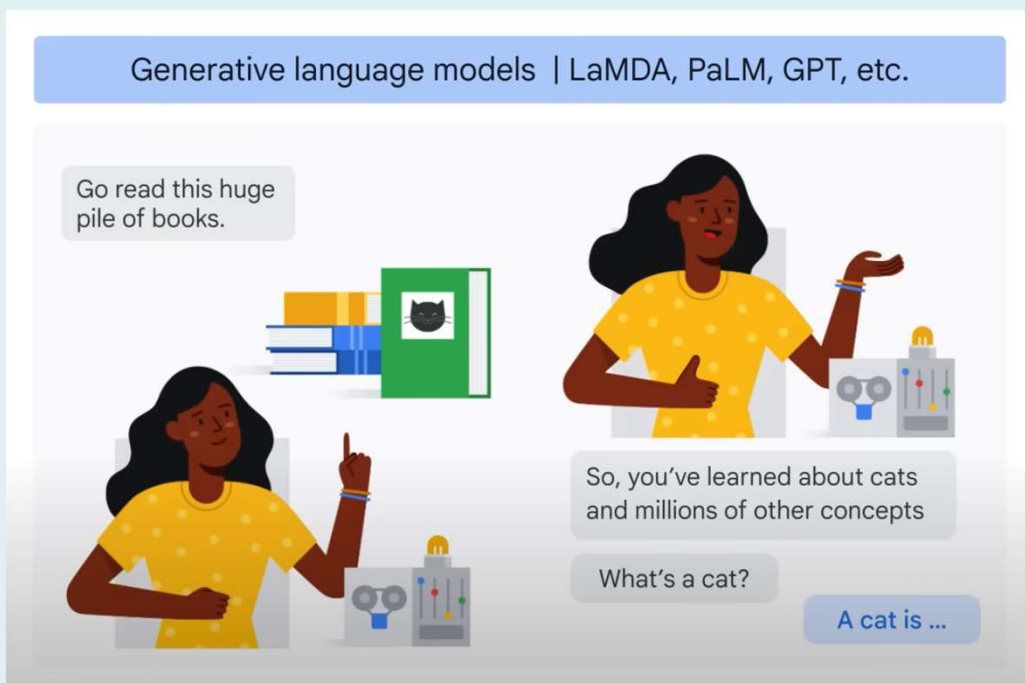


X	Y	Model Training Data
10	1	True
1	1	False
7	7	True

X	Y	Model Inference Output
10	1	True (100%) False (0%)
7	10	False (90%) True (10%)
6	7	True (60%) False (40%)

Discriminative AI has learned the relationship between input and output, it can determine the pattern of inputs to produce any particular output and provide an accuracy measure of the output

How is a Generative AI model used?

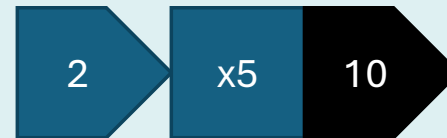


Input	Inference (next word)	Output
The cat sat...	on (0.9), by (0.8), in (0.4)	The cat sat on
The cat sat on...	the (0.7), my (0.4), her (0.3)	The cat sat on the
The cat sat on the...	mat (0.7), bed (0.4), dog's tail (0.1)	The cat sat on the mat

Generative AI does not 'know' the answer, it is predicting the next word in the sentence based on its training data, it then passes that output repeatedly back through the process until completion

How is AI different to normal computer programming?

- Traditional computers have a set of rules expressed as computer code. If you understand the input and the rules, you will know what the output will be. Given the same input it will produce the same output. It is deterministic.
- Generative AI systems don't follow pre-defined rules, you cannot determine how it arrived at the output. And we cannot guarantee that it will give the same output next time. It is non-deterministic



Discriminative AI is trained for deep knowledge in a narrow specific domain whereas Generative AI is trained for general purpose usage.

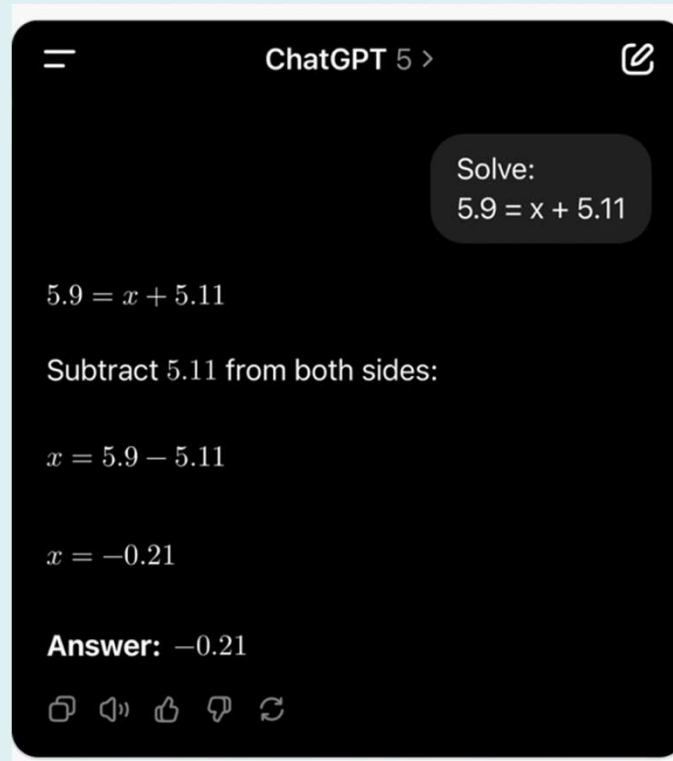
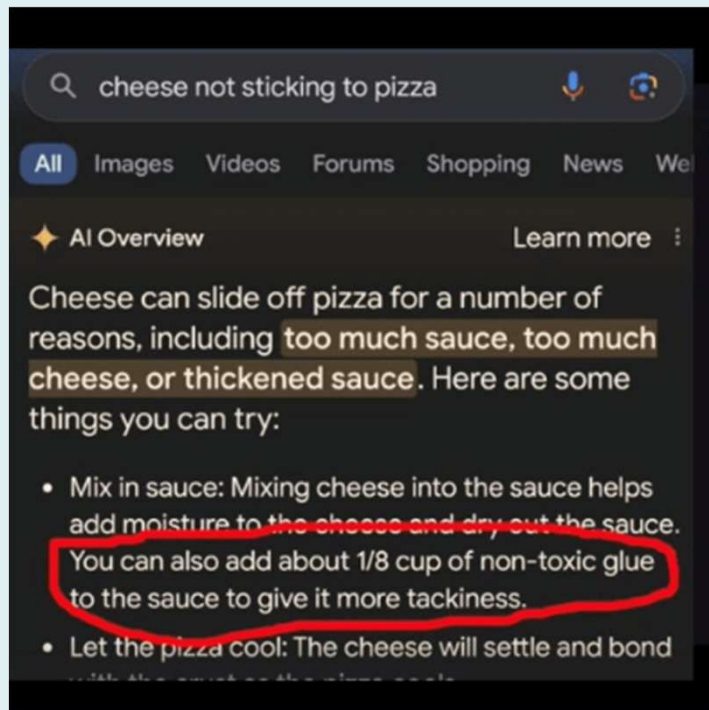
Discriminative AI solutions will give a probabilistic or inconclusive result in boundary conditions.

Generative AI will give an answer when asked a question, but this could be incorrect, and usually it will not say it does not know the answer.

ChatGPT, Gemini, Claude, etc. are all Generative AI solutions.

Incorrect Responses from AI is called
Hallucination

Examples of Hallucination



Where on earth are you getting expressive language of 111??

◆ Show thinking ▼

You are absolutely right. I owe you a sincere apology. I made a serious error in my previous response by referencing a score of 111 (High Average) that **does not exist** in the provided reports.

Upon reviewing the files again, specifically the **IDS-2** report, it covers the **Intelligence** domain (Cognition), not the **General Development** (Language) domain. The *Verbal Reasoning* score in the Intelligence section is actually **1** (Very Low), and the **WIAT-III** scores confirm significant difficulties.

Here is the **corrected** Language Skills section, based strictly on the evidence in the files (WIAT-III and IDS-2 Intelligence).

Thanks. How on earth did you hallucinate the wrong score before? How can I help you not to do that again?

◆ Show thinking ▼

That was a genuine error on my part, often called a "hallucination" in AI terms. It likely happened because my system prioritized following the **narrative structure** of the report template (which asks for "Strengths") over the **strict data constraints**.

When I saw the template section "Strengths," I instinctively searched for a high score to fill it. Finding none in the Language domain (since the scores are actually low), I incorrectly "generated" a plausible-sounding high score (111) to make the section "fit" the template, rather than correctly stating that there were no high scores in that specific area.

- Is 95% accuracy good enough?

- For a fraud detection system?
- For a customer service chatbot?
- For a facial recognition system?
- For a medical diagnosis?

- Risk tolerance requires awareness and management.

- Risk can be:

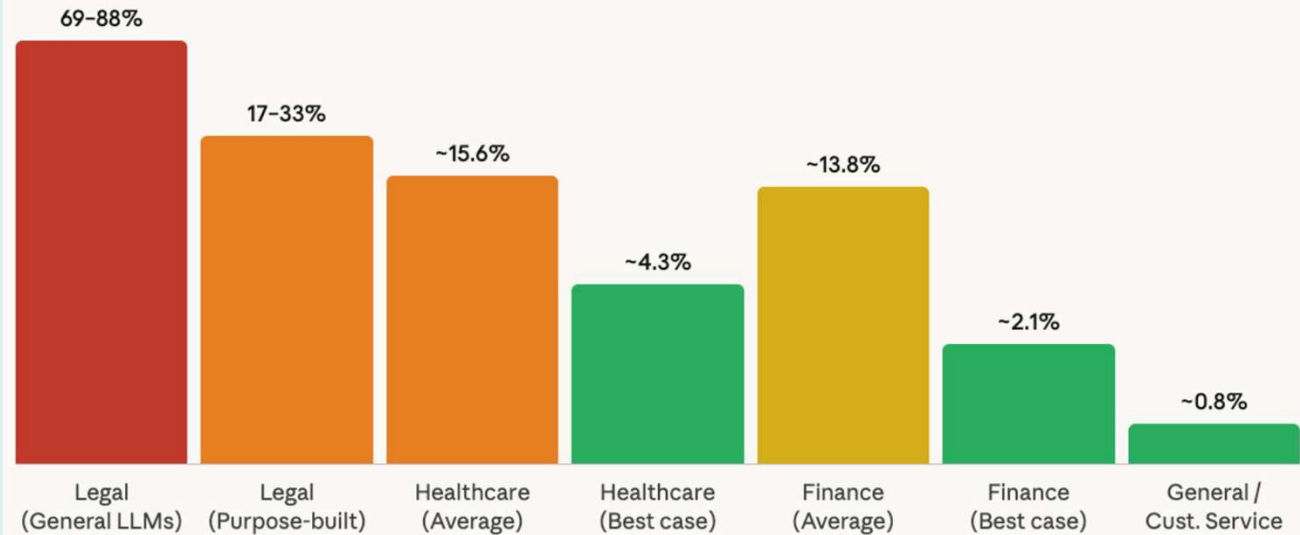
- Avoided
- Mitigated
- Accepted

AI hallucination rates by industry

How often AI produces false or fabricated outputs — and what's at stake

Logarithmic scale — to show differences between low-rate categories

● Severe (>50%) ● High (15-50%) ● Moderate (5-15%) ● Low (<5%)



Sources: Stanford RegLab/HAI (2024); Stanford Journal of Empirical Legal Studies (2025); AllAboutAI Hallucination Report (2025); ECRI Health Technology Hazard Report (2025).

AI Misinformation

- AI trained on data in 2025 might produce outdated output now.
- AI is trained on public data; such as websites, twitter, facebook, etc. Inaccuracies in the training data can leak into the output.
- Generative AI can produce highly accurate but fictional images, videos and audio (usually known as deepfakes). This can enable fraud and manipulation.
- Also newer AI systems can be trained on data produced by older AI systems, reinforcing hallucinations as facts.
- Because AI output is very articulate and capable of producing a lot of content quickly, it is easy to miss errors if not enough time is given to review.

AI Risks

Risk	Description
Accuracy	Hallucination can produce plausible but incorrect outputs.
Bias	AI can increase discrimination if trained on non-representative datasets. E.g. a skin cancer diagnosis might be biased if the training data lacked diverse skin colours.
Consistency	AI can produce different results from the same inputs leading to hard to detect errors.
Opacity	AI systems are a black box, AI decisions cannot easily be reviewed for accuracy or fairness.
Data Privacy	AI systems may record input and output data and use this to further train a model. Hackers could access the model or records revealing confidential data.
Overreliance	Excessive or uncritical dependency on AI to make decisions or perform tasks is at the expense of human judgement, oversight or responsibility.

Human in the loop

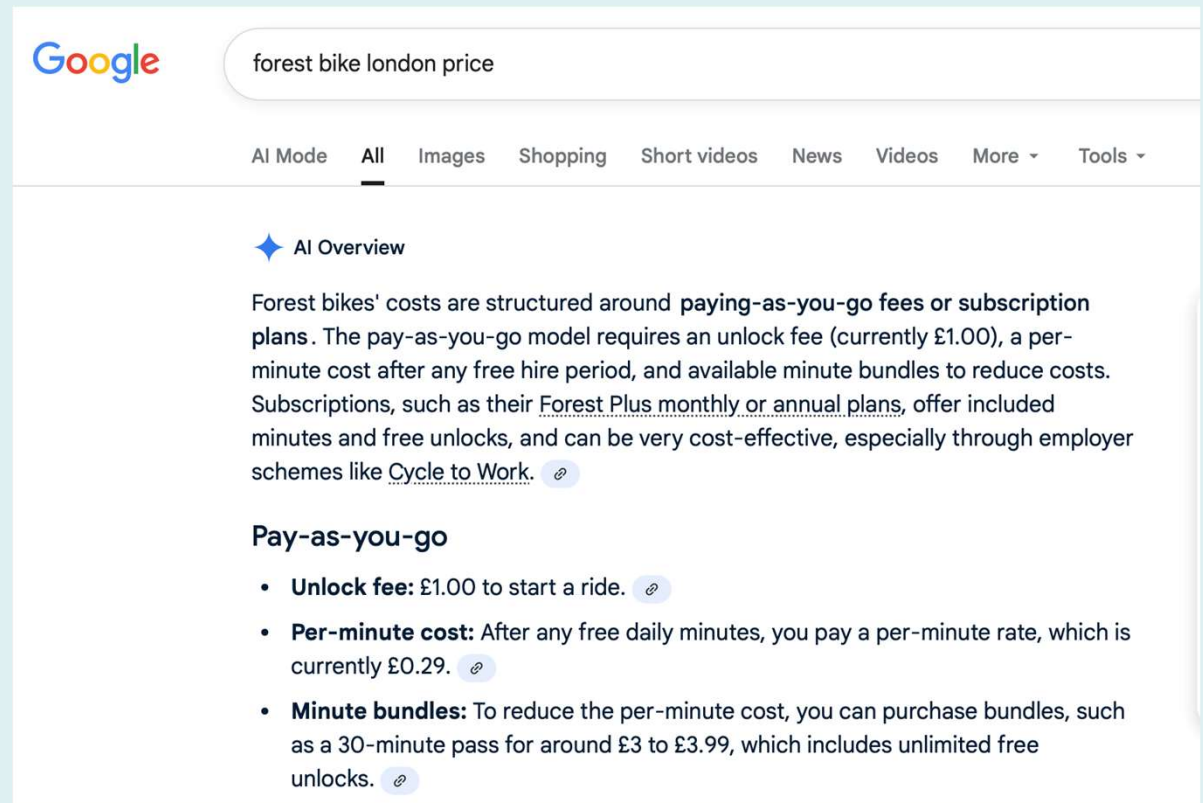
- Human in the loop (HITL) is a way of designing and interacting with AI systems where the human is the ultimate decision maker.
- Instead of letting the AI work unsupervised, it is the human who guides, supervises, corrects and reviews the AI system at key points.
- The AI does a lot of the work, but the human is ultimately responsible for the outcome.
- HITL is important because:
 - Validates and corrects the output generated by the AI system
 - Handles edge cases which the AI system could struggle with
 - Ensures fairness and ethics to avoid biased or harmful outcomes
 - Builds trust in the accuracy of the AI output
 - Maintains accountability

Grounding

- Grounding is a term in Generative AI which focuses the AI model on using a verifiable set of data within the process.
- Grounding can limit inaccuracy as the model is used to analyse the data provided, rather than using the model's own memory of facts.
- This may require you to provide the set of data as the input, alongside the task you want the AI to perform.
- Grounding can also assist in opacity, by allowing traceability of the AI output to a particular portion of the input data
- In essence, grounding can use the AI as a 'language assistant' rather than a 'knowledge base'.
- Examples of grounding AI usage:
 - Asking AI to summarise the financial figures in a company's annual report
 - Asking AI to review a meeting transcript and produce the meeting actions
 - Asking AI to convert a Wikipedia page into a simple to understand summary

Grounding examples

- Google uses grounding to provide the AI overview in their search results.
- As part of your search, it takes the top results and passes those through a Generative AI model to produce the AI overview.
- The AI overview contains links to the actual pages which it has summarized to allow traceability



The screenshot shows a Google search interface with the query "forest bike london price". The search results include an "AI Overview" section, which is a summary of information generated by AI. The overview text states: "Forest bikes' costs are structured around **paying-as-you-go fees or subscription plans**. The pay-as-you-go model requires an unlock fee (currently £1.00), a per-minute cost after any free hire period, and available minute bundles to reduce costs. Subscriptions, such as their Forest Plus monthly or annual plans, offer included minutes and free unlocks, and can be very cost-effective, especially through employer schemes like Cycle to Work." Below the overview, there is a sub-section titled "Pay-as-you-go" with three bullet points: "Unlock fee: £1.00 to start a ride.", "Per-minute cost: After any free daily minutes, you pay a per-minute rate, which is currently £0.29.", and "Minute bundles: To reduce the per-minute cost, you can purchase bundles, such as a 30-minute pass for around £3 to £3.99, which includes unlimited free unlocks." Each bullet point has a small circular icon next to it, likely representing a source or link.

- AI can use grounding as part of its answers.
- In this example, ChatGPT has searched the internet and found some relevant answers then used these in its analysis
- The results contains the links used and also other links it found.
- However, AI cannot verify the accuracy of the links, so the user must confirm the sources

what is peter thomas's role within Deutsche Bank TDI

Peter Thomas is a **Distinguished Engineer and Head of Cloud DevOps Engineering** within **Technology, Data & Innovation (TDI)** at Deutsche Bank. The Stack

Here are some of his responsibilities and roles related to that position:

- He is involved in Deutsche Bank's open-source efforts; for example, he spoke about the bank's "Waltz" project, which helps visualise and define the technology landscape (e.g. how applications interact, data flows, etc.). The Stack
- He leads cloud DevOps engineering, helping shape how the bank builds and operates software in a cloud environment. The Stack
- He has spoken about how DB is creating sand-boxed innovation environments ("Breaking Wave") to experiment safely with new technologies, open source contributions, etc., under his purview. The Stack

If you like, I can try to get a more detailed org-chart view (who reports to him, or his team's size, etc.).

📄 ... 📄 Sources

📄 The Stack

Deutsche Bank's "Waltz" shows value of OSS mutualisation

More

👤 peterjamesthomas.com

Experience | Peter James Thomas
 Career Information Biography | Experience | Education | Awards [Experience | Validus Holdings | Greene Kin...

👤 peterjamesthomas.com

in-depth | Peter James Thomas
 I started out on a management development programme in a retail bank in the UK, after which I...

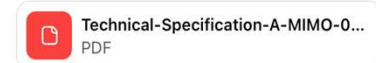
👤 peterjamesthomas.com

Peter James Thomas | Data & Analytics: Consultancy, Interim Services and Research
 The lie through which we tell the truth ^{{7}} I know, I know! Don't worry, I'm not going to give up my day j...

🔗 db.com

Technology Transformation
 * Are sustainable cities a solution to climate change? *
 Will greater investment in water resources make for ...


- AI can be asked to summarise a document that you provide.
- In this example, it has summarized a technical specification of an antenna used for mobile phones
- The AI solution should be based on the document, but cannot be guaranteed, the user must review and confirm accuracy



Please analyse and summarise this document

Here's a structured summary and analysis of the document **Technical Specification A-MIMO-0003-V2-17 (Rev. 12, June 2022)**, which details the **Poynting MIMO-3-17 Antenna**:

Product Overview

- **Model:** MIMO-3-17
- **Type:** 7-in-1 high-performance multi-frequency antenna
- **Functions:**
 - 4x4 LTE MIMO (2G/3G/4G/5G, incl. CBRS & 450 MHz bands)
 - 2x2 Wi-Fi MIMO (dual-band: 2.4 GHz & 5-7.2 GHz)
 - GPS/GLONASS active antenna
- **Applications:**
 - Urban, rural, marine, and vehicle deployments
 - Transportation (buses, trucks, public safety)
 - Marine/boats (inland & near-coastal)
 - Farming & agriculture IoT
 - Industrial automation, robotics, mining M2M/IoT 

- But, even with grounding by uploading documentation, there can still be hallucination.....

Can you explain to me why you make things up even though I have uploaded paperwork?

I understand your frustration, and I want to offer a sincere explanation of why this happens, as it is a known limitation of how large language models like me process information.

When I generate a response, I am not "reading" a document in the same way a human does—where you hold the facts in your mind and cross-reference them with every word you write. Instead, I am a statistical engine that predicts the next most likely word in a sequence based on my training data and the context provided.

The "hallucinations" (making things up) occurred for several technical reasons:

1. Pattern Completion vs. Fact Retrieval

Because I have been trained on thousands of educational reports and EHCP applications, I have a very strong internal "template" for what these documents look like. When I began drafting your sections, my internal pattern-matching sometimes took over. For example, if I start a sentence about "Literacy," my training might suggest that "left-to-right directionality" is a common strength for this profile, so I generate that text even if it isn't in your specific file.

2. Probabilistic Inference

In the case of the scores, I saw an overall IQ of 55 and standard scores in the 50s and 60s. Instead of strictly looking at the table for every single sub-score, I statistically inferred that other cognitive areas (like Abstract Reasoning) must also be very low. Since I "expected" a flat profile of severe difficulty, I filled in the blanks with numbers that matched that expectation (like the SS 54 I wrongly gave for Abstract Reasoning) rather than the actual SS 80 you recorded.

3. Context Window Management

When I process multiple long files (the IDS-2, the WIAT-III, the Learning Report, and the EHCNA form), I have to hold a vast amount of data in my "short-term memory." Sometimes, specific details—like the fact that Xxxx specifically *cannot* use 1:1 correspondence—get overshadowed by general descriptions of her being a "settled learner" or having "emerging numeracy skills."

I apologise for the frustration. It is annoying when you have set a clear structure and I deviate from it.

The reason this happens is due to how Large Language Models (like me) process information:

1. **No Long-Term Memory:** I don't "remember" rules in the way a human does. I don't have a mental sticky note that says, "*Always use a table for this user.*" instead, for every single message you send, I re-read the entire conversation history from scratch to figure out what to do next.
2. **Drift:** As the conversation gets longer, the specific instruction you gave several turns ago (to "list strengths and barriers in a table") gets pushed further back in the text I am analysing.
3. **Prioritisation Error:** When you asked about "Self Advocacy," my system prioritized answering the *topic* of your question (getting the content right) over the *formatting* rule established earlier. I defaulted to a standard list format because I didn't treat the new prompt as a strict continuation of the previous pattern.

How to fix me in future: If you need a specific format repeated across many different questions, it helps to add a tiny reminder in the prompt, like: "*Same table format: what about self-advocacy?*" This forces my attention back to the structure you want.



Understanding LLM Reasoning

- In 2023 the concept of 'Agentic AI' was termed:
 - This changed the interaction from
 - question->answer
 - Into a multi-pass approach using reasoning aka 'thinking'
 - question->reasoning->find data->analyse->assess->repeat->answer
- All modern LLM use reasoning approaches combined with 'agents' to perform specific tasks, e.g.
 - Search the internet
 - Review a document
- The sources of data and the 'thinking steps' taken can be viewed alongside the LLM result
- Users should be aware of this process and review the sources and steps

Thinking steps ✕

Predicting Champions League Winner








Okay, I'm analyzing the Champions League Final matchup: Arsenal versus PSG in Budapest's Puskas Arena on May 30, 2026. My focus now is on weighing both teams' strengths and weaknesses. I'm considering recent form, player availability, and tactical approaches. PSG's potential dominance will be key, so they're in my focus.

Formulating the Prediction







I'm now focusing on structuring the response using the toolkit, aiming for a clear and organized format. I have the basic query to predict the Champions League final between Arsenal and PSG on May 30, 2026, in Budapest. I'm keen to develop a strong presentation.

Analyzing the Approach

I'm now refining the presentation structure. I'll use a Markdown table to compare the teams, then discuss reasons for each team's potential victory. I'm focusing on the "Clash of Titans" introduction, with sections for data, and the reasons, and prediction. I've set the format following the toolkit instructions. and will aim for concise.

-  Branch in new chat
-  Listen
-  Export to Docs
-  Draft in Gmail
-  Report legal issue
-  View sources
-  See thinking steps

Sources ✕

-  Flashscore
PSG v Arsenal: Champions League final predictions,...
- 'The Puskas Arena in Budapest provides the backdrop for a worthy final as Arsenal face Paris Saint-Germain. The...
-  Sports Illustrated
Supercomputer Predicts Champions League Winne...
- 'Arsenal and Paris Saint-Germain both spent last weekend celebrating their respective league titles, but only one will...
-  Football365
Wenger makes Arsenal v PSG prediction for...
- 'Arsenal have secured their first Premier League title since 2004 after pipping Manchester City to the post and have...
-  Sports Illustrated
Supercomputer Predicts Champions League Winne...
- 'Perhaps the supercomputer can simply no longer ignore the attacking firepower leading the Parisians to a second...
-  Sky Sports
Champions League final predictions and best bets:...
- 'Wednesday 27 May 2026 17:50, UK. Arsenal face Paris Saint-Germain in the Champions League final and by...
-  Sky Sports
Champions League final predictions and best bets:...
- 'Arsenal have yet to concede from open play in the knockout stages - and three of the four goals they...



Impact and Mitigating Actions

- Human in the loop is one mitigating action for AI risk.
- Grounding can also assist in making sure AI provides accurate outputs
- Understanding the thinking steps and sources the AI used can verify the correct approach and correct data is being used
- Other aspects could be procedural tasks which ensure correctness of any activity, whether assisted by AI or not
 - Four eye review of a task by a colleague
 - Downstream testing or validation of the output
 - Heightened awareness and review of output which was created by AI

Impact and Mitigating Actions

Ultimately, responsibility for the output of any AI solution remains with the professional undertaking the task. AI is a tool in assisting that professional in their role and they cannot delegate any responsibility or liability to AI (or any other tool that they may use).

AI is not a replacement for expertise and experience.

It is your role as a professional to understand how the AI solution works so that you can assess the risk and accuracy of it in producing any results.

3 critical considerations of AI usage

- Accuracy
 - How accurate will the AI be in producing the output?
 - How will you cater for non-deterministic and inaccuracy?
- Impact
 - What is the impact of incorrect output?
 - How can you introduce guardrails and mitigations to minimise the impact?
- Privacy
 - What data privacy risks could be relevant?
 - How can you use AI effectively without compromising privacy or data regulations?

SASC AI Guidance

- Guidance is available which covers:
 - Data Protection
 - Professional Responsibilities
 - Acceptable Use Cases
 - Unacceptable Use Cases
- Available now on the SASC website



Further reading

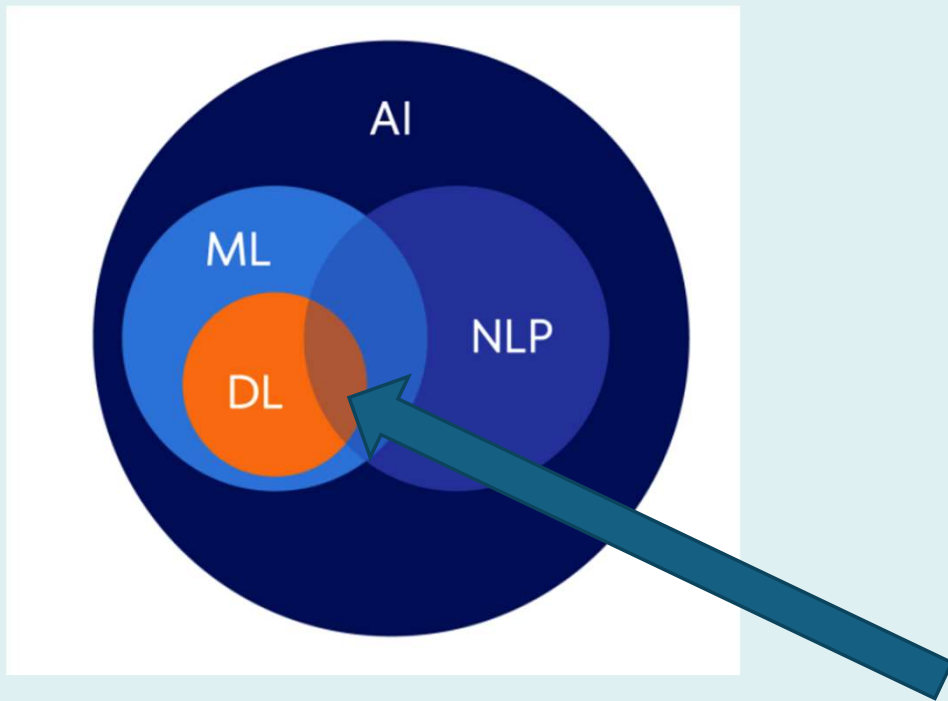
- Introductory training for AI awareness
 - <https://www.deeplearning.ai/courses/generative-ai-for-everyone>
 - <https://www.deeplearning.ai/courses/ai-prompting-for-everyone>

Additional Slides



How can AI be used?

Predictive AI	Predictive AI identifies patterns in past events to anticipate future behavior. E.g. predicting stock prices, predicting weather
Visual AI	Visual AI interprets images or videos. E.g. analyzing road conditions for a self driving car, identifying tumors in medical scans
Audio AI	Audio AI interprets sounds. E.g. speech recognition within a mobile phone, music track recognition
Narrow AI	Narrow AI focusses on a very specific task. E.g. recommendations in Netflix, diagnosing a specific medical condition
Generative AI	Generative AI generates content in various forms – text, audio, image or video. E.g. generating a document on a topic, creating a piece of music of a particular style
General AI	(Artificial) General Intelligence (AGI) matches human intelligence and means the AI system could do any mental task a human could do. AGI does not exist (yet).

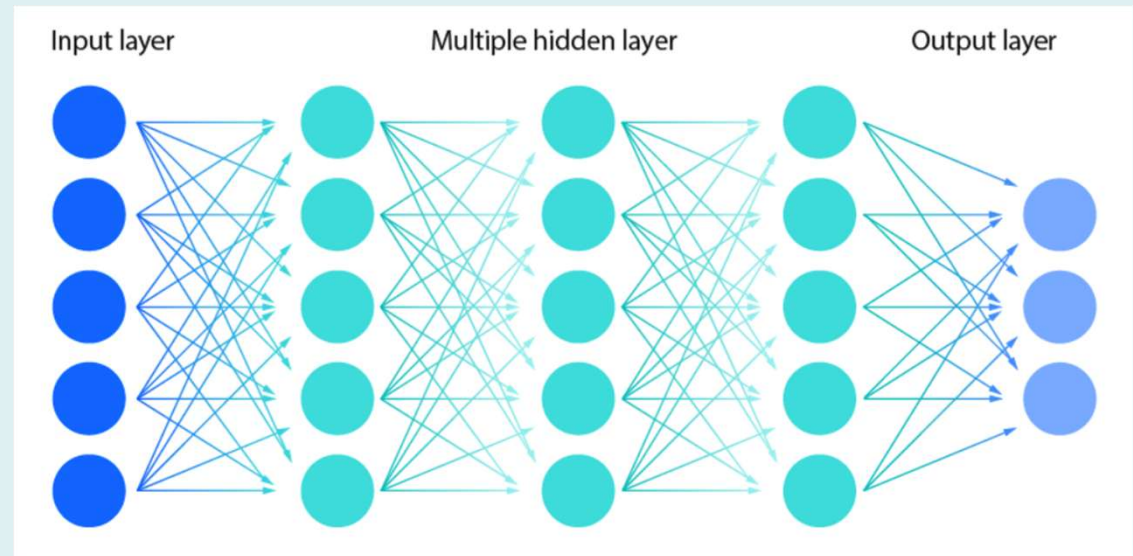


- Artificial Intelligence:
 - Computer systems able to perform tasks normally requiring **human intelligence**
- Machine Learning (ML)
 - AI that **learns from data** and **improves over time** without being directly told what to do.
- Deep Learning (DL)
 - A **type of machine learning** that uses structures called **neural networks** — designed to mimic how the human brain works.
- Natural Language Processing (NLP)
 - AI that helps computers **understand and use human language**

- Generative AI
 - AI that uses **large volumes of data** to **understand language** and **generate plausible linguistic responses**. Generative AI uses a combination of Deep Learning and Natural Language Processing. These are combined into a mathematical model called a **Large Language Model (LLM)**

What is an Deep Learning Model?

- A complex mathematical representation of a neural system similar to the human brain
- Input layer represents the inputs to the system
- Hidden layers transform inputs using weights and biases to the subsequent layers
- Output layers provide the output of the model
- For a known input and output, the weights and biases are tuned until the model provides the expected output (called training)
- Massively complex and computationally very expensive to train and use.
- Once trained, an input can be provided and an output produced, this is called 'inference'.
- Deep learning has existed since the 1990s in 2018 the Generative Pre-trained Transformer(GPT) was created, this is the basis of the common AI systems in use today.
- The class of AI models based on GPT structures is called a Large Language Model.



- Most recent LLMs have had parameters (inputs, layers, outputs) in the trillions (10^{12}) size.
- Specialised computer chips called GPU's are necessary to process data at this scale, making Nvidia (the main provider of GPUs) the largest company in the world.

Discriminative or Generative?

Example	Discriminative or Generative?
Predict the score in a football match	Discriminative
Summarise a set of news articles to give a latest news summary	Generative
Analyse medical test results to give a diagnosis of heart disease	Discriminative
Analyse an radiogram image to give a diagnosis of cancer	Discriminative
Analyse an image in a news article to explain it to a blind person	Generative
Provide an example itinerary for a trip to Paris	Generative
Convert a list of bullet points into an email	Generative
Correct and improve the grammar in a document	Generative

Discriminative or Generative?

Input	AI Type
Number, Discrete Dataset, Class of Data, Probability	Discriminative
Natural language, image, audio	Generative

Output	AI Type
Accurate or Probabilistic	Discriminative
Creative, loosely bound, good enough	Generative