# Introduction to AI

## What Is AI

### Definition & Scope

When talking about AI, people generally consider two main topics:

* **Artificial Intelligence (AI)** refers to systems capable of performing tasks that normally require human intelligence - like reasoning, decision-making, and language comprehension.
* **Machine Learning (ML)** is a subset focused on teaching systems to learn from data and improve over time - without being explicitly programmed.

These can often be compared against each other; however, ML is actually considered a subset of AI. There are further subsets of ML including Deep Learning and Generative AI (for example ChatGPT).

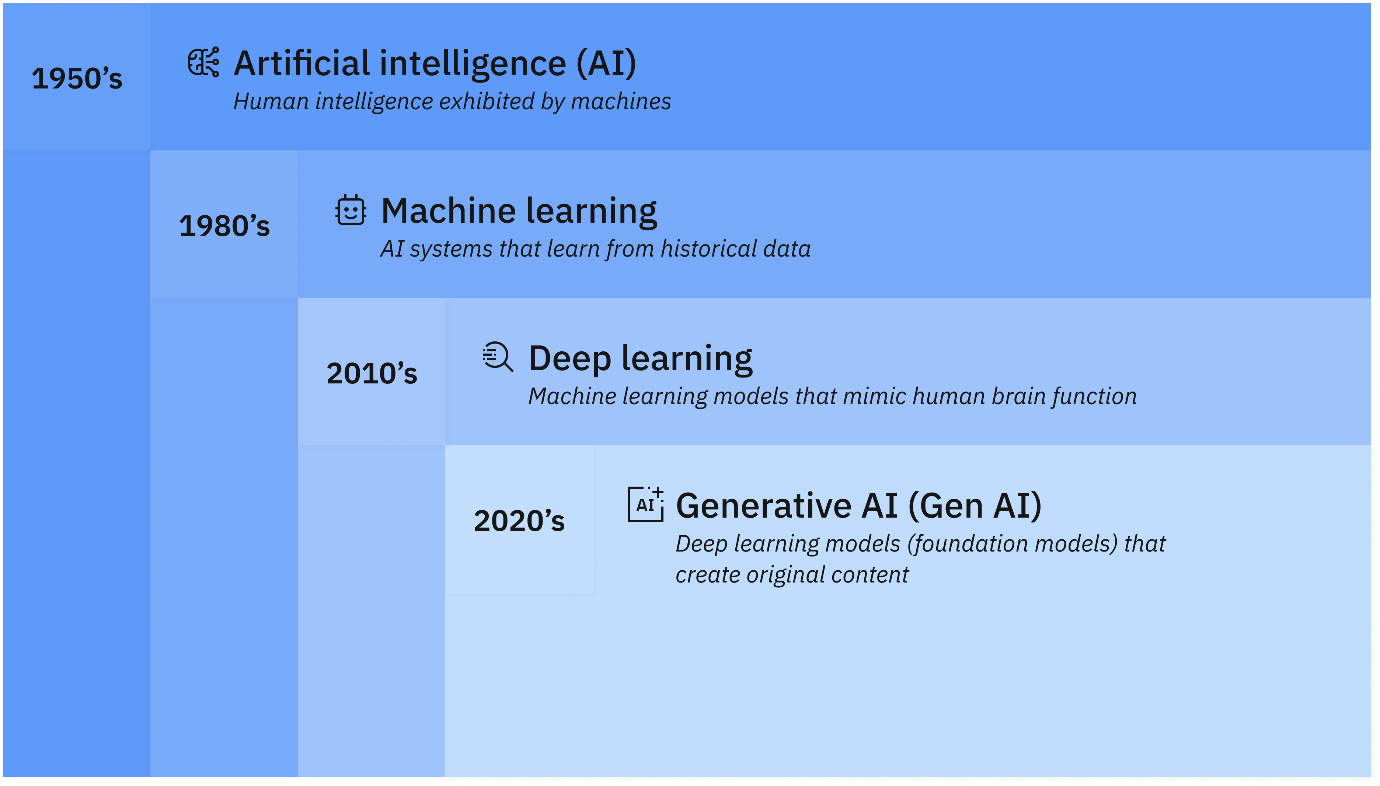


Figure 1 How artificial intelligence, machine learning, deep learning and generative AI are related ([IBM](https://www.ibm.com/think/topics/artificial-intelligence))

In basic terms, AI refers to computers exceeding or matching the capabilities of a human and contains many subsets including ML and Gen AI. Modern AI is usually built using ML algorithms that detect patterns that can be used to make predictions or automate decisions. For example, in credit risk management, machine learning models can be used to analyse borrower financials, market data, and macroeconomic indicators. These models detect early warning signals such as deteriorating leverage ratios, declining cash flow coverage, or unusual covenant breaches. By recognising these patterns ahead of time, the bank can flag counterparties whose risk profile is worsening and adjust credit exposures proactively - reducing potential losses while still enabling timely lending decisions.

### Key Terms

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Algorithm | A set of rules or instructions that a computer follows to solve a problem or complete a task. In AI, algorithms are designed to identify patterns in data and make predictions. |
| Artificial General Intelligence (AGI) | A theoretical form of AI that can perform any intellectual task a human can. Current AI systems are narrow AI — designed for specific tasks — and do not have AGI’s broad adaptability. |
| Artificial Narrow Intelligence (ANI) | AI systems designed to perform one specific task (e.g., spam filters, recommendation engines). All current AI is ANI. |
| Bias (in AI) | Systematic errors in AI outputs that arise from imbalanced or skewed training data, potentially leading to unfair or discriminatory outcomes. |
| Data Mining | The process of discovering hidden patterns, correlations, or trends within large datasets using techniques such as clustering, classification, and association rule learning. |
| Data Science | A multidisciplinary field that combines statistics, computer science, and domain knowledge to extract insights from structured and unstructured data. |
| Deep Learning | A subset of machine learning that uses multi-layered neural networks to model complex patterns in data. Deep learning powers modern AI systems such as image recognition and language models. |
| Explainable AI (XAI) | AI systems designed to provide transparent reasoning for their predictions or outputs, helping users understand and trust results. |
| Generative AI | AI models that can create new content (text, images, audio, video) based on learned patterns from training data. |
| Hallucination (AI) | When an AI model produces an incorrect, misleading, or fabricated output that appears plausible but is not grounded in data. |
| Human-in-the-Loop (HITL) | A system where humans are actively involved in training, validating, or improving AI models (e.g., reviewing flagged transactions for fraud). |
| Human-on-the-Loop (HOTL) | A governance approach where humans oversee automated systems but intervene only when necessary, ensuring accountability without constant involvement. |
| Large Language Model (LLM) | A machine learning model trained on vast amounts of text data to predict and generate human-like language (e.g., ChatGPT, GPT-4/5). |
| Multimodal Foundation Model (MfM) | An AI model capable of processing and integrating multiple types of data (e.g., text, images, audio) to generate outputs or insights. |
| Neural Networks | Computational models inspired by the human brain, consisting of interconnected layers of nodes (“neurons”). These networks can learn complex relationships in data. |
| Overfitting | When a model learns noise or detail in the training data too closely, performing poorly on new, unseen data. |
| Prompt Engineering | The practice of crafting and refining inputs (prompts) to AI systems to obtain desired outputs. |
| Reinforcement Learning (RL) | A type of machine learning where agents learn by interacting with an environment and receiving feedback in the form of rewards or penalties. |
| Supervised Learning | A machine learning technique where models are trained on labeled data (inputs paired with correct outputs), enabling the model to learn mappings (e.g., predicting credit defaults). |
| Token | A unit of text (word, subword, or character) used as input or output for language models. |
| Transformer | A type of neural network architecture that uses “attention mechanisms” to handle sequences of data, especially text. Transformers power most state-of-the-art LLMs. |
| Trustworthy and Responsible AI | AI that is developed and used in ways that are transparent, fair, explainable, and aligned with ethical and regulatory standards. |
| Unsupervised Learning | A machine learning technique where models find hidden patterns or structures in unlabeled data (e.g., clustering customers by behaviour without predefined categories). |
| Vector Embeddings | Numerical representations of words, sentences, or data objects in a high-dimensional space, enabling semantic search and similarity comparisons. |
| Zero-shot / Few-shot Learning | Techniques where an AI model performs tasks without (zero-shot) or with very few (few-shot) examples during training. |

## What AI isn’t

AI isn’t magical, it’s not a superhuman - it’s pattern recognition plus massive data and computation. Although ML models may appear intelligent, they don’t “think”; instead, they generalise from statistical associations.

As an example a technique called **Gradient Descent** is used to optimise a set of numerical parameters to minimise a loss function. Variants of gradient descent are commonly used to train neural networks. The goal of this is to find the minimum across a range of possible solutions. An illustrative example is displayed below:

A yellow grid with black dots

AI-generated content may be incorrect.

Figure 2 Gradient Descent - optimisation, not reasoning

This can be thought of through the following Analogy:

* Imagine a person dropped in a foggy valley trying to reach the lowest point.
* They can’t see the whole landscape; they just feel which direction slopes downward and take a small step that way.
* Eventually, they reach a **local** minimum. This might appear to them as the minimum point, however in actuality it might not be -as can be seen in the 3rd dot in the above image.

Although the solution appears to be the right answer, this is not always the case. As such relying solely on AI creates large risk, as has been observed in the media:

A screenshot of a news article

AI-generated content may be incorrect.

Figure 3 Consequences of blindly using AI ([AFR](https://www.afr.com/companies/professional-services/deloitte-report-suspected-of-ai-invented-quote-from-robo-debt-case-20250825-p5mpjj))

**1.2 Technical Foundations: Pattern Recognition & Predictive Models**

* **Pattern Recognition at the Core**
  + AI systems excel at identifying patterns in complex or high-dimensional datasets—like financial time series or risk indicators. [upGrad](https://www.upgrad.com/blog/what-is-pattern-recognition-and-machine-learning/?utm_source=chatgpt.com)
* **How Predictive Models Work**
  + A brief snapshot of the typical pipeline: Data ingestion → feature engineering → model training → evaluation → prediction. [Saiwa+3Financial Times+3PixelPlex+3](https://www.ft.com/content/6692ba73-fa71-45db-881c-0be221495b3a?utm_source=chatgpt.com)
  + Real-world framing: credit default prediction models assess historical indicators to estimate future risk.

**1.3 Why AI Isn’t a Substitute for Expert Judgement**

* **Limitations of AI**
  + Models may hallucinate or misinterpret inputs—especially when faced with unexpected data.
  + AI decisions are only as good as the training data; biased or incomplete data yields unreliable outcomes.
  + ML systems are not inherently explainable—posing challenges in transparency and compliance. [Deloitte+15Lifewire+15Deloitte+15](https://www.lifewire.com/deep-learning-vs-machine-learning-7495519?utm_source=chatgpt.com)
* **Principles to Safeguard Against Risk**
  + Maintain human oversight—AI is an assistive tool, not an arbiter.
  + Monitor data drift and performance over time (“model operations”).
  + Use governance frameworks (like Deloitte’s “Trustworthy AI”) to ensure fairness, transparency, and resilience. [upGrad+12Deloitte+12Deloitte+12](https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/articles/trustworthy-ethical-ai-thought-leadership.html?utm_source=chatgpt.com)

**1.5 Team Activity: "Your Expectations of AI"**

**Objective**: Foster engagement and surface preconceptions/divergences in expectations.

**Setup**:

* Break into small groups of 3–4.
* Give each group a scenario or question:
  + *"How would you expect AI to improve our monthly credit reviews?"*
  + *"What are the top things AI can't do in our work today?"*

**Deliverables**:

* Each group shares a “Top 3 Expectations” in a Confluence commentary box.
* Instructor collates insights live, then reflects on:
  + Which expectations are realistic?
  + Which may introduce risk?
  + How do they align with AI governance principles?

**1.6 Suggested References & Further Reading**

| **Resource** | **Why It Matters** |
| --- | --- |
| **WIRED: AI vs ML** | Clear differentiation between AI and ML. [Deloitte](https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/articles/trustworthy-ai-governance-in-practice.html?utm_source=chatgpt.com)[WIRED+1](https://www.wired.com/story/machine-learning-ai-explained?utm_source=chatgpt.com) |
| **FT Graphic: How Machines Learn** | Visual summary of model lifecycle. [Financial Times](https://www.ft.com/content/6692ba73-fa71-45db-881c-0be221495b3a?utm_source=chatgpt.com) |
| **Deloitte Insights: Trustworthy AI** | Framework for ethical, governed AI. [Deloitte+1](https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/articles/trustworthy-ethical-ai-thought-leadership.html?utm_source=chatgpt.com) |
| **Intro to Pattern Recognition** | Concise explainer of core AI mechanism. [AI Powered Product Engineering](https://createbytes.com/insights/pattern-recognition-in-ai?utm_source=chatgpt.com)[SaM Solutions](https://sam-solutions.com/blog/pattern-recognition-in-ai/?utm_source=chatgpt.com)[pareto.ai](https://pareto.ai/blog/pattern-recognition-in-machine-learning?utm_source=chatgpt.com) |

Credit Risk Team – AI Training Lesson Plan

Session Title: Harnessing AI for Credit Risk Analysis

Audience: Macquarie Credit Risk team

Duration: 2 hours

Objectives:

- Provide a clear understanding of AI, its capabilities and limitations.

- Demonstrate practical applications of AI tools within credit risk workflows.

- Teach how to structure data and documents to be AI-ready.

- Discuss ethical considerations and risk management when using AI.

Agenda:

1. Introduction to AI – 20 minutes

- Overview of AI and machine learning.

- What AI is (pattern recognition, predictive models) and what it is not (a substitute for human judgement).

- Media caution: discuss Deloitte’s AI communication incident as an example of reputational risk.

- Activity: Group discussion on expectations of AI.

2. Applications of AI – 40 minutes

- Use cases for AI in credit risk: summarising financial reports, extracting ratios, drafting credit notes, brainstorming risk factors.

- Demo: Use an AI assistant to summarise an annual report.

- Activity: Participants use AI to extract key financial metrics from a sample document; compare outputs.

3. Integration & Document Preparation – 40 minutes

- How AI reads documents: importance of clear metadata, headings, and flat tables.

- Guidelines for AI‑friendly templates: consistent headings, explicit labels, flat tables (no merged cells), inline metadata (Company Name, Industry, Date, Author, Document Version, Internal Rating), version history, separation of narrative and data, bullet lists.

- Font selection: use sans‑serif fonts like Arial, Calibri, Verdana, or Atkinson Hyperlegible for clear character differentiation; avoid decorative fonts that confuse letters like "O" vs "0".

- Activity: Review a sample credit report; identify and fix formatting issues to make it AI‑ready.

4. Ethical Considerations & Risk Management – 15 minutes

- Privacy and data security: anonymise client data; do not share sensitive information with external AI services.

- Bias and fairness: discuss how AI may reflect historical biases and require human oversight.

- Hallucinations and verification: emphasise cross‑checking AI outputs against source documents.

- Governance: align with internal policies and regulatory requirements.

- Activity: Case study discussion on potential ethical pitfalls.

5. Conclusion & Next Steps – 5 minutes

- Recap key takeaways: AI is a tool to augment, not replace, human expertise; document structure and metadata are critical.

- Encourage experimentation and continuous improvement; highlight additional resources (internal guidelines, external courses).

- Gather feedback on the session.

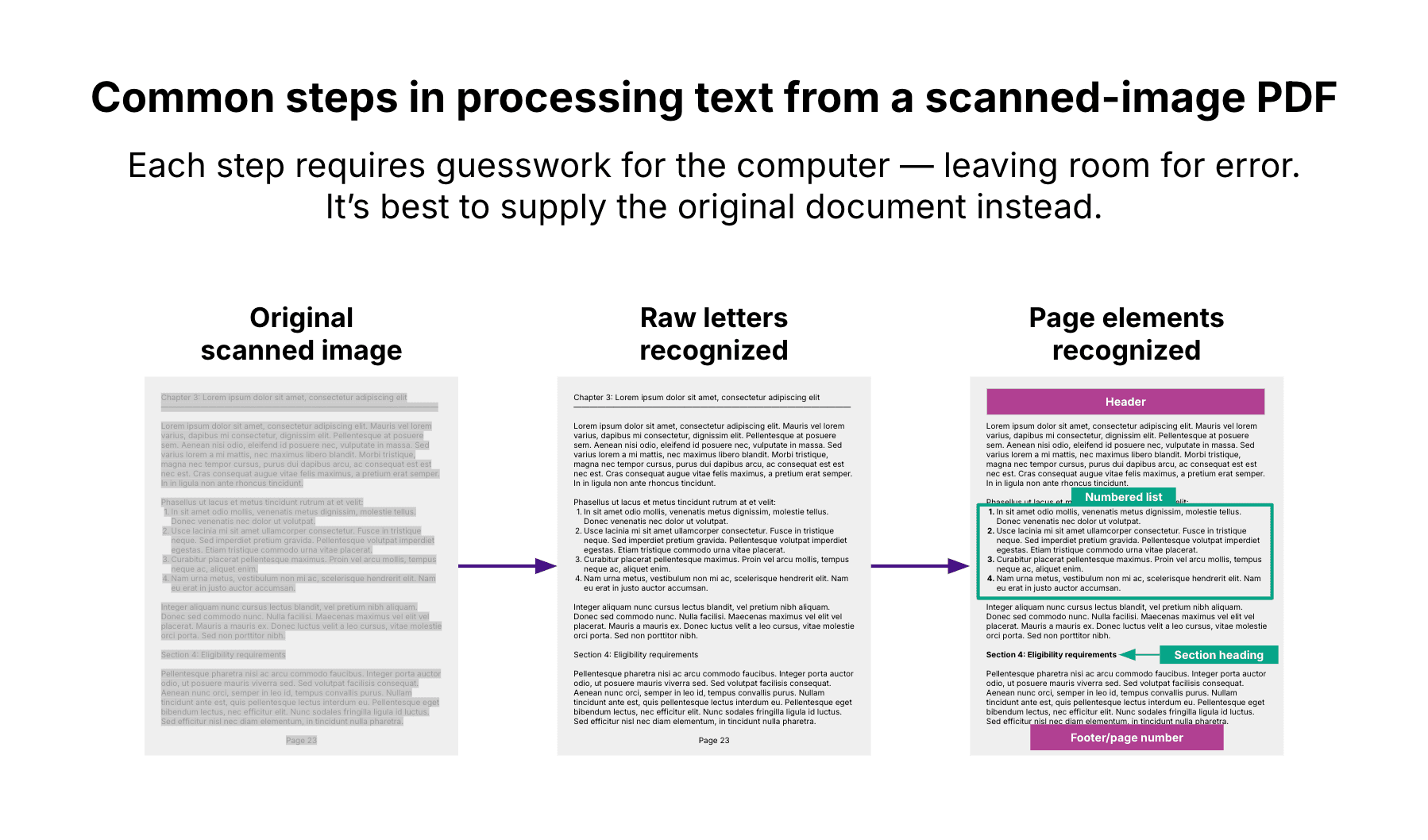
Lesson order:

* 1. Introduction to AI
* 2. Application of AI (how to use it)
* 3. Integration of AI (how to structure data for AI to understand it)

(TBC on the section names)

Section 3: Integration of AI

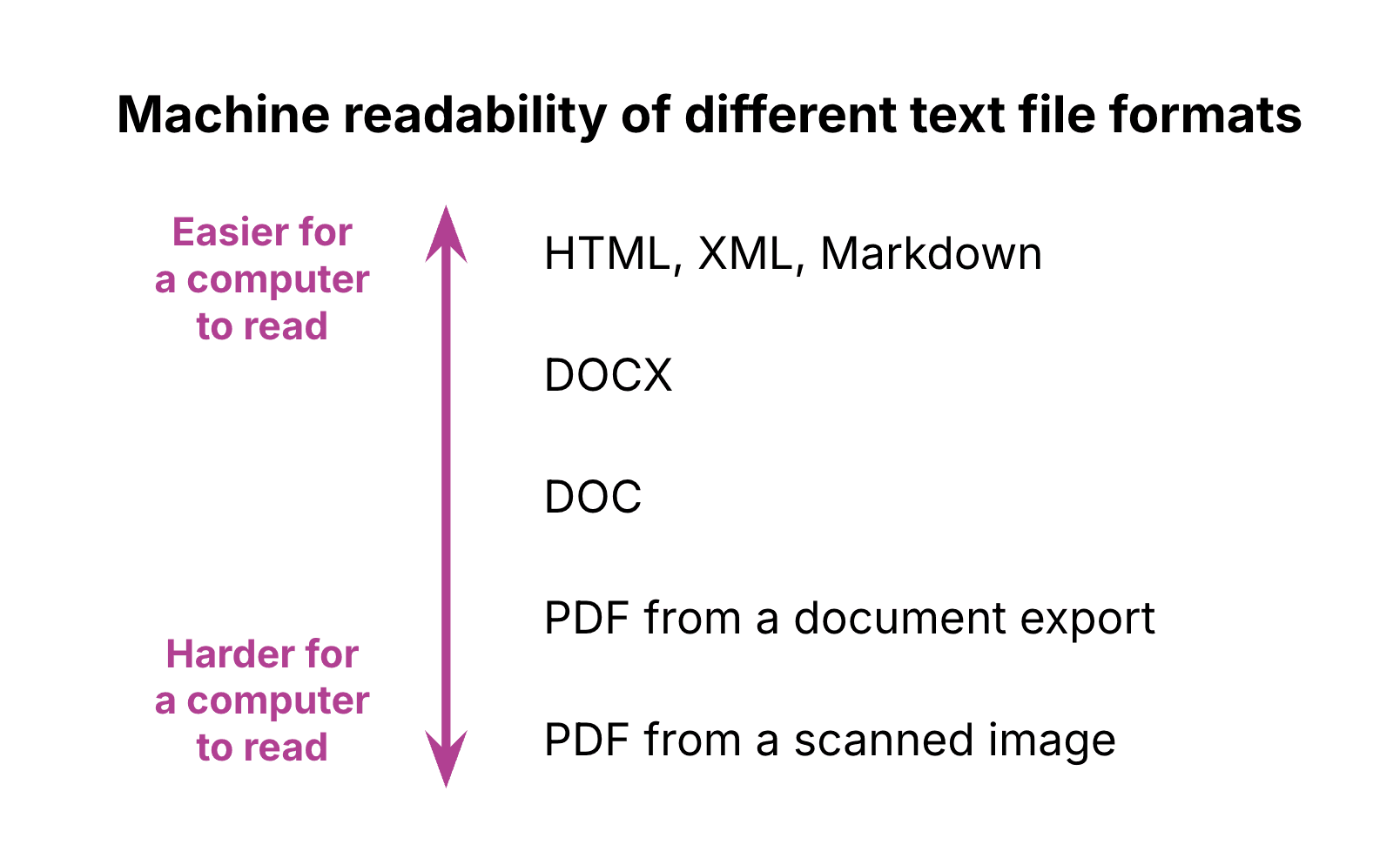
Understanding document structure: After recognizing the individual letters on a page, the computer must differentiate between components on the page — such as page headers and footers, columns, tables, bulleted and numbered lists, headings and subheadings, paragraphs, and image captions. Having a rich understanding of the page elements, rather than seeing just a continuous stream of text, is critical for fully making sense of its contents. But this is a very challenging task for computers, and the results can be a mess. You may have experienced this if you have ever copied text from a PDF or image to a text document. This can result in losing the document’s original formatting and the arrangement of tables, lists, captions, and other elements.



*Exporting a PDF directly from the source document*— for example, exporting a PDF from Microsoft Word or Google Docs — can be slightly better than scanning a paper document as one of these file types. This is because it can provide clearer text shapes compared to what you might get from scanning a printed document, leaving less room for error during the step of optical character recognition. Still, it’s not ideal.

Meanwhile, with a *file format like HTML, XML, or Markdown*, a computer can directly access the text as it was originally written — eliminating the need for optical character recognition and easing the task of understanding page elements. (Less ideal but still better than a PDF is [DOCX](https://www.onlyoffice.com/blog/2024/03/doc-vs-docx), or alternatively DOC.)

The technical details of this topic can get complex, but generally, you want to use file formats that are higher on the ladder below, rather than ones that are lower:



(Get example with Financial data that includes Note and on where it is whited out)

Fortunately, in many cases you can simply ask for the source document. If you are part of a government agency or working in partnership with one, it should be easy to get a copy of the source document. Remember: the PDF didn’t appear by magic — someone created it from an original document.

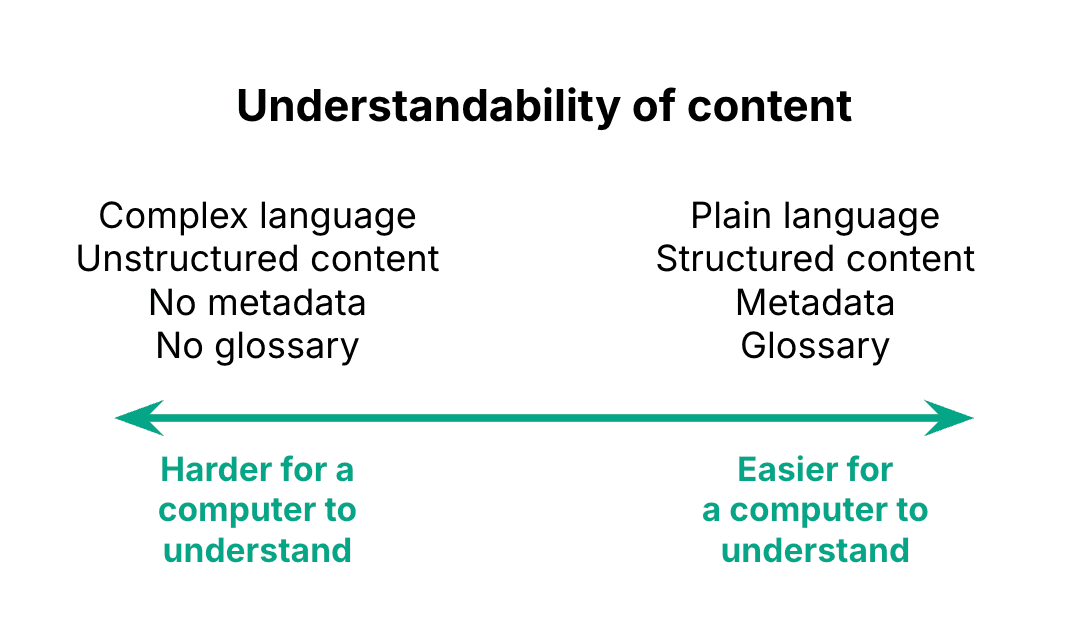
To provide the LLM with as simple of a file as possible, consider exporting as HTML if using Microsoft Word (save as “Web Page, Filtered”) or exporting as a Markdown file if using Google Docs.

Note that for structured data, like what you find in a spreadsheet, it’s best to supply it in a format such as CSV, JSON, or XML — not as a text document.

Recent advances in AI have improved the ability of computers to recognize elements of document structure, but supplying original machine-readable documents remains the better way to provide information.

## Tactic 2: optimizing content design

LLMs have a much easier time reading content that is clear, well-structured, and semantically rich.



There are a number of things you can do to make content easier for a computer to understand:

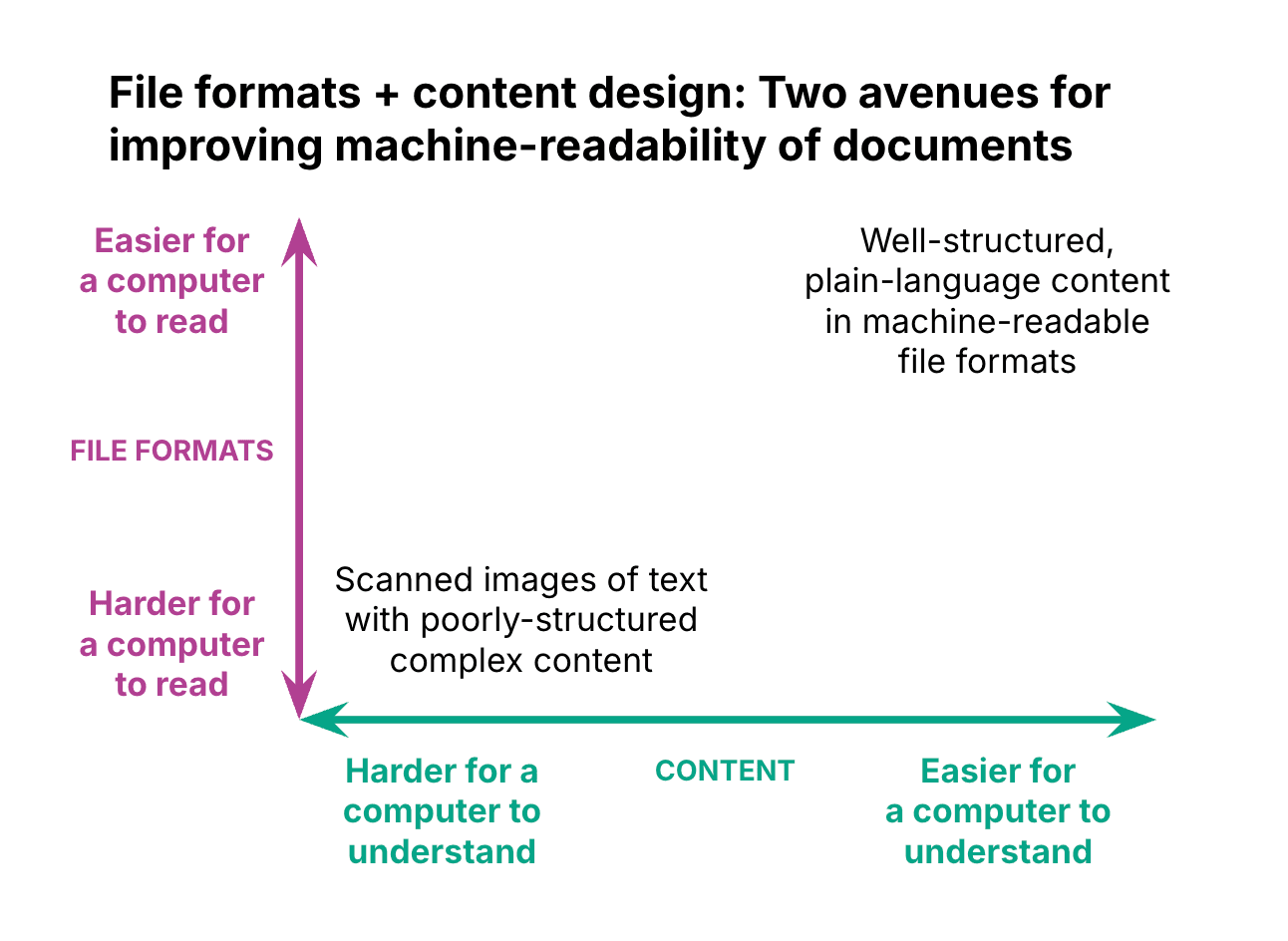
* Write using [*plain language*](https://www.plainlanguage.gov/)
* Use frequent [*headings*](https://www.section508.gov/blog/accessibility-bytes/document-headings/) to describe sections of the text, and create a clear hierarchy of headings and subheadings
* Organize lists of items as *bulleted or numbered lists*, rather than as a run-on sentence or paragraph
* *Define key terms*, and consider including a *glossary*
* Provide document *metadata*, which can include information like important dates (when the document was first written, when it was published) and topic tags that describe the content

When creating headings, it’s important to *use the “paragraph styles” feature* of programs like [Microsoft Word](https://support.microsoft.com/en-us/office/video-using-styles-in-word-9db4c0f4-2754-4294-9758-c14a0abd8cfa) or [Google Docs](https://support.google.com/docs/answer/116338?hl=en&co=GENIE.Platform%3DDesktop#zippy=%2Cchange-the-text-style), rather than simply making the text larger or bold. This automatically applies a distinctive style to all headings in the document, saving you formatting work. More importantly, if you’re exporting the document in a machine-readable format as discussed above (e.g. HTML, XML, Markdown, DOCX), the file will internally tag the headings as such. This way, a computer reading the document will see them marked as headings and not merely body text that happens to have a different visual appearance.

These are good practices for designing content even if you’re not planning on having an LLM read it. Improving content design can also*improve skimmability and*[*accessibility*](https://www.w3.org/WAI/fundamentals/accessibility-intro/) forhumans, including for people using screen readers.

For more information on content design, check out [this introduction](https://hub.innovation.ca.gov/content-design/principles/index.html).

**Combining both tactics**



* **🔹 Suggested Order & Additions**

**1. Introduction to AI**

* ✅ Keep this as your opener. Good to cover:
  + What AI *is* (large language models, pattern recognition, not “magic”).
  + What AI *isn’t* (it doesn’t “know,” it predicts; it isn’t authoritative).
  + Risks and cautions (confidentiality, hallucinations, reputational issues – e.g. Deloitte in the media).
  + Internal usage guidelines (e.g. anonymise data, don’t paste sensitive info unless tools are approved).

**2. Applications of AI (Practical Use Cases)**

* ✅ Great middle section. Anchor it in **real, work-relevant tasks**:
  + Summarising financial or regulatory documents.
  + Drafting first-pass credit notes, templates, or reports.
  + Extracting tables or ratios from messy PDFs.
  + Brainstorming questions, checklists, or scenario analysis.
* ⚠️ Important: stress that outputs must be checked against source documents (human oversight).

**3. Integration with Workflows & Documents**

* ✅ Perfect as the final “how to use it properly” section:
  + How AI parses documents (headings, metadata, tables).
  + How to structure Word/Excel docs for machine-readability (your AI template guide fits here).
  + How to integrate AI into Confluence, Word templates, or Excel macros.
  + Emerging integrations (Teams, Outlook Copilot, etc.).
* **🔹 Optional Extras You Might Add**

1. **“AI in Our Context” (short section)**
   * A 5-minute overview of how your **bank/RMG Credit** is using or exploring AI.
   * Ties the lesson directly to team goals → makes it relevant rather than abstract.
2. **“Hands-On Demo / Exercise”**
   * Show a *before/after*: an unstructured credit assessment vs. an AI-ready version.
   * Or give a prompt and show how to refine it iteratively.
3. **“Limitations & Ethics”** (could sit inside Introduction or as a short conclusion)
   * Bias, transparency, explainability, hallucinations.
   * Why it’s *assistive* not *decisive*.

* **🔹 Recommended Flow**

1. **Introduction to AI** → definitions, risks, context.
2. **Applications** → practical work examples.
3. **Integration** → technical guidance on documents and workflows. (note that I am creating a document for this)
4. (Optional) **Future & Ethics** → wrap up with limitations and opportunities.

📌 **Bottom line:**  
Your structure is good, but I’d put **Applications before Integration**. That way people see *what it can do for them* before you dive into *how to structure documents for it*. Adding a **short “context in our team” intro and a live demo** will make it even more powerful.