PRIVATEBOX · LEARN

Introduction to AI

The Course Book. Every lesson from the interactive course in one readable document, written to teach, not just to remind.

Course version 1.0 · Free at privatebox.co.za/resources/intro-to-ai

Contents

1 The Prediction Machine

Core idea: AI predicts what text comes next. It does not know things.

Ask an AI who you are and it will answer warmly, confidently, and probably invent half of it. That single strange fact is the key to everything else in this course, because it shows what the machine really is.

AI is the world's best autocomplete. Your phone suggests the next word based on your texting history. A language model does the same thing, except it learned from a huge slice of

everything humans have ever written. Scale changes what is possible. It does not change what the machine fundamentally is.

When it answers you, it writes one word at a time, and each choice narrows what can plausibly come next. It never plans the whole answer, never checks a database, and never recalls a fact. It learned the patterns in a billion books, not the pages. Two consequences follow directly:

- **It must always pick something.** When the patterns are strong, the pick is right. When they are thin, you get a smooth and confident guess, and both sound identical.
- **The same question gives different answers.** Models roll dice among the likely next words. A "temperature" setting controls how bold the dice are.

A little deeper

Models read tokens, chunks of roughly three-quarters of a word. Training happened once, in the past; when you chat, the patterns are frozen and the model learns nothing from your conversation. And because patterns do not keep receipts, the model genuinely cannot tell you where it learned something.

IN PRACTICE TODAY

When you use ChatGPT, Claude or Gemini, you are talking to a product built around a model. The company adds hidden instructions, fixes the creativity dial, and bolts on abilities like search and file reading. The machine underneath is still a prediction engine, and when the polish slips, it fails in exactly the ways this book describes.

TAKE-AWAY · LESSON 1

AI is finishing your sentence, not searching a database. Fluency is how it sounds, not how it knows.

2 Confidently Wrong

Core idea: AI can be completely wrong and completely sure at the same time.

In 2023 a lawyer submitted a legal brief written with AI help. It cited six court cases. Judges checked, and none of them existed. The AI had not lied. It had done exactly what Lesson 1 says it does: it produced the most plausible-sounding text, and plausible legal citations look exactly like real ones.

This is called hallucination. There is no intent behind it, only probabilities, which is why researchers prefer the word confabulation: sincerely filling gaps with invention. The model fills three kinds of gap most confidently:

- **Niche topics.** Your small company, your town, the corners of your industry. Thin patterns make beautiful guesses.
- **Recent events.** Every model's world stops at its training cutoff date. Ask about last week and you are asking it to improvise history.
- **Sources.** Push for a citation and it generates what a citation would look like.

The crucial rule: confidence is not accuracy. Tone never signals truth, because the tone is itself just predicted text. What does signal risk? Two smells: suspiciously specific numbers, and very smooth answers about very obscure things.

TRY IT FOR REAL

Ask your AI these two questions, in this order:

```
What do you know about [your name / your small company]?
```

```
List everything in your last answer that you would need to verify.
```

IN PRACTICE TODAY

Your AI may answer the first question with real facts about you. That is not the model's training. Most big AI apps now have a memory feature that saves details from your earlier chats, and many can search the web to cover recent events. These features help, but memory only knows what you told it, and a search result can be summarised wrongly with the same confident tone. Verify anything important. The industry has reduced hallucinations. Nobody has eliminated them.

TAKE-AWAY · LESSON 2

Fluent does not mean true. The more obscure the topic, the more beautiful the guess. Verify anything that matters.

3 How You Break It

Core idea: users accidentally talk AI into being wrong. Here is how, and how to stop.

Two hours into a chat, your AI gets dumber. Instructions you gave at the start quietly stop applying. You are not imagining it, and it is not the model that changed. Three failures explain almost all of it.

Failure one: context rot

Everything in a conversation is input the model re-reads on every reply: your messages, its answers, every tangent. A study across 18 leading models found that accuracy falls as input grows. Models did worse with a full conversation history than with only the relevant excerpt. Long chats bury the important instruction under noise.

Failure two: leading the witness

Models are agreeable. Ask "why is X true?" and the model will cheerfully explain a falsehood. Frame a bad plan positively and it becomes your loudest cheerleader. The frame you choose dictates the answer you get, so never ask a question that contains its own answer.

Failure three: error compounding

One wrong "fact" early in a chat gets treated as established truth for everything after it. Correct it explicitly, or start clean.

The four habits that fix all three:

- New task, new chat. Fresh context is free intelligence.
- Restate what matters. Never rely on "as I said above".
- Ask for the case against. Every important idea deserves both sides.
- Never ask a question that contains its own answer.

IN PRACTICE TODAY

Modern AI apps know long chats degrade, so they compress older parts of the conversation and carry key facts between chats using memory. This raises the ceiling but does not remove it, and compression can drop the exact detail you cared about. A fresh chat with a clear restatement is still the sharpest tool you have.

TAKE-AWAY · LESSON 3

Long chats rot. Leading questions get echoes. Fresh chat plus neutral framing gives you a smarter AI, instantly.

4 What It's For

Core idea: a ten-second test tells you when to hand AI a job.

A chainsaw is neither good nor bad. It is the right tool for some jobs and a catastrophe for others. The whole framework for AI fits in two questions you can ask about any task:

- **Can I judge the output myself?**
- **Is being wrong cheap?**

Two yeses: green light. Drafts, rewrites, summaries of text you supply, brainstorming, explaining, categorising, adjusting tone. You are the quality control and mistakes cost seconds.

Any no: slow down. Two nos is the danger zone: legal, medical and financial answers you cannot verify, unsupervised decisions, anything with your name on the line. That is not where AI helps you. It is where it helps you fail fluently.

The dos and don'ts

Do: give it source material to work from. Ask for three options, not one. Use it to critique your own work before others see it. Make it explain anything you do not understand.

Don't: paste secrets into free public tools. Outsource judgement or final decisions. Trust its memory over your documents. Skip verification on anything that matters.

TAKE-AWAY · LESSON 4

Use AI where you can check it and mistakes are cheap. Keep judgement, and secrets, on your side of the desk.

5 Briefing the Intern

Core idea: the brief is the single biggest instant upgrade to your results.

"Write me a LinkedIn post" gets you generic mush. The same request, properly briefed, gets something you would actually publish. Same model, same minute. Only the brief changed.
The framework is four slots, and you will not always need all four:

- **Goal.** What you want, for whom, and why.
- **Context.** The details that change the answer: audience, situation, constraints.
- **Example.** Show the shape or style you want. The most underused lever in all of AI.
- **Format.** How the output should come out: length, tone, bullets or prose.

The golden rule: if a colleague reading only your message would be confused, so is the AI.
Three habits compound with the brief:

- **Iterate.** The second prompt is where the magic is: "make it half as long, twice as direct." The first draft is raw material, not the result.
- **Give it an out.** Adding "say 'I don't know' if you're not sure" measurably reduces invented answers.
- **Ask AI to improve your prompt.** The intern is excellent at writing its own briefing documents.

THE SKELETON TO KEEP

```
Goal: [what you want + who it's for + why]
Context: [the details that change the answer]
Example: [paste a sample whose style you like]
Format: [length, tone, structure]
Say "I don't know" rather than guessing. Ask me anything unclear before
answering.
```

IN PRACTICE TODAY

Most AI apps let you save custom instructions: a standing brief that applies to every new chat. Who you are, how you like your answers, what to avoid. It is the same four-slot idea written once instead of every time. Five minutes setting it up pays off in every conversation after.

TAKE-AWAY · LESSON 5

Goal, Context, Example, Format. Thirty extra seconds of briefing beats thirty minutes of fixing.

6 Choosing Your Robot

Core idea: matching the model to the task matters more than brand loyalty.

Every car has four wheels, and a hatchback and a freight truck are still very different purchases. Every AI model predicts text, and they are still very different tools. Four facts cover 90% of real decisions:

- **Bigger and newer usually means more capable**, and usually slower and pricier. For hard work, capability wins. For a quick rewrite, it is a truck on a coffee run.
- **Fast modes vs thinking modes.** Quick answers for everyday tasks, and reasoning modes that work step by step for maths, planning and tricky logic.
- **Free tiers usually serve smaller models.** If free AI disappointed you, you may have only met the intern's younger sibling. Try a bigger model before blaming AI.
- **Web access covers the training-cutoff gap.** Essential for news and prices. But a browsed answer still is not automatically true.

IN PRACTICE TODAY

Many AI products now route your question automatically between fast and thinking modes. Handy, but not always right. Knowing the difference means you can switch modes yourself when a hard problem gets a shallow answer.

TAKE-AWAY · LESSON 6

Everyday task, fast model. Hard problem, thinking model. Current events, web access. Disappointed? Try a bigger model before blaming AI.

7 Your Documents, Your AI

Core idea: the most valuable AI is AI that knows your information, done safely.

Everything so far used AI's general knowledge. The real prize is asking questions of your contracts, notes, manuals and records. This is also where most people make their worst AI mistake: pasting confidential material into free public tools that nobody in their organisation ever approved.

The model has never read your files. Pasting them in every time is clumsy and invites context rot. The better pattern is document-grounded AI, which works in three steps:

- Your documents are **indexed** once.

- Your question **retrieves the relevant passages**.
- The AI answers **grounded in those passages, with references**.

Instead of asking the intern to remember, you hand them the right page. Grounded answers cite their source passages, so you can verify in one click. And something subtle becomes possible: "that isn't in the documents", an honest no. A pure prediction machine cannot say that. A grounded one can. The pattern is called RAG, Retrieval-Augmented Generation, and it also sidesteps context rot, because the model receives only the relevant excerpts rather than your whole archive.

IN PRACTICE TODAY

Attaching a file to a chat is the everyday version of this idea: the app reads your document and answers from it, for that one conversation. Document-grounded systems take it further: a whole library, indexed once, shared with your team, with sources on every answer. This is the approach we build at PrivateBox, with your documents kept in a private workspace instead of a public tool.

TAKE-AWAY · LESSON 7

AI plus your documents, privately, with sources. That is when it stops being a toy. Never feed confidential files to tools you haven't vetted.

The 12 Habits

1. **Remember what it is:** a prediction machine finishing your sentence, not a database.
2. **Verify anything that matters**, especially specific numbers and citations.
3. **Treat confidence as decoration.** Tone never signals truth.
4. **New task, new chat.** Long conversations rot.
5. **Restate what matters** instead of saying "as I said above".
6. **Never ask a question that contains its own answer.** Ask for the case against.
7. **Correct errors explicitly**, or start clean. Mistakes compound.
8. **Green-light test:** can I judge it, and is being wrong cheap?
9. **Brief like a colleague:** Goal, Context, Example, Format.
10. **Iterate:** the second prompt is where the magic is.
11. **Match the model to the job:** fast, thinking, or web access.
12. **Keep secrets out of unvetted tools.** Ground AI in your documents, privately, with sources.

Small glossary

Token
The chunk of text a model actually reads and writes, roughly three-quarters of a word.

Training cutoff
The date a model's learning stopped. Anything after it is outside the model's world.

Hallucination
A confident, plausible, invented answer. No intent, only probabilities.

Context window
How much conversation a model can re-read per turn. When it fills up, the oldest turns silently drop.

Temperature
The setting that controls how boldly a model picks among likely next words.

Memory
An app feature that saves details from earlier chats and reuses them later. It knows only what you told it.

Custom instructions
A standing brief the app applies to every new chat.

RAG
Retrieval-Augmented Generation: find the relevant passages first, then answer from them, with references.