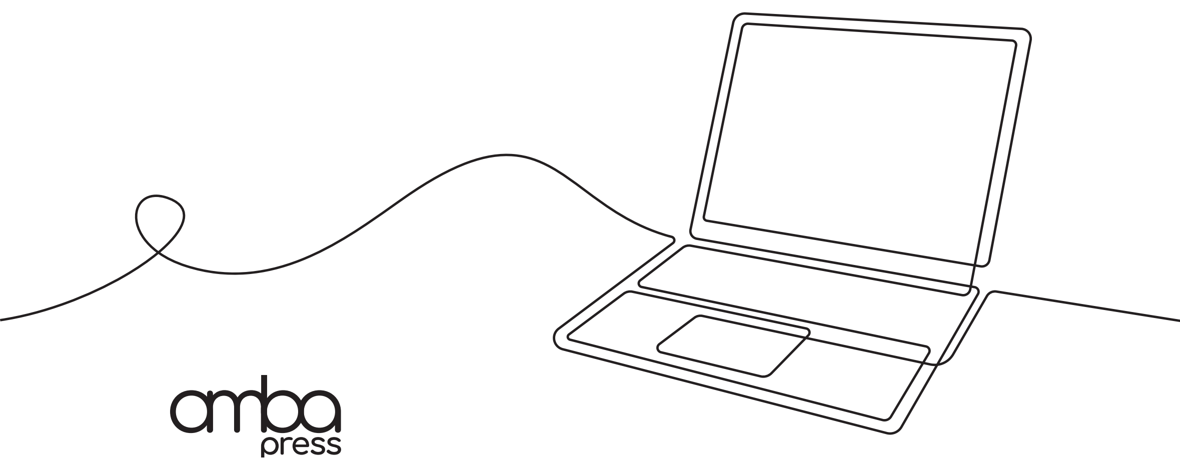
LEON FURZE

Practical AI Strategies

Engaging with Generative AI in Education



amba press

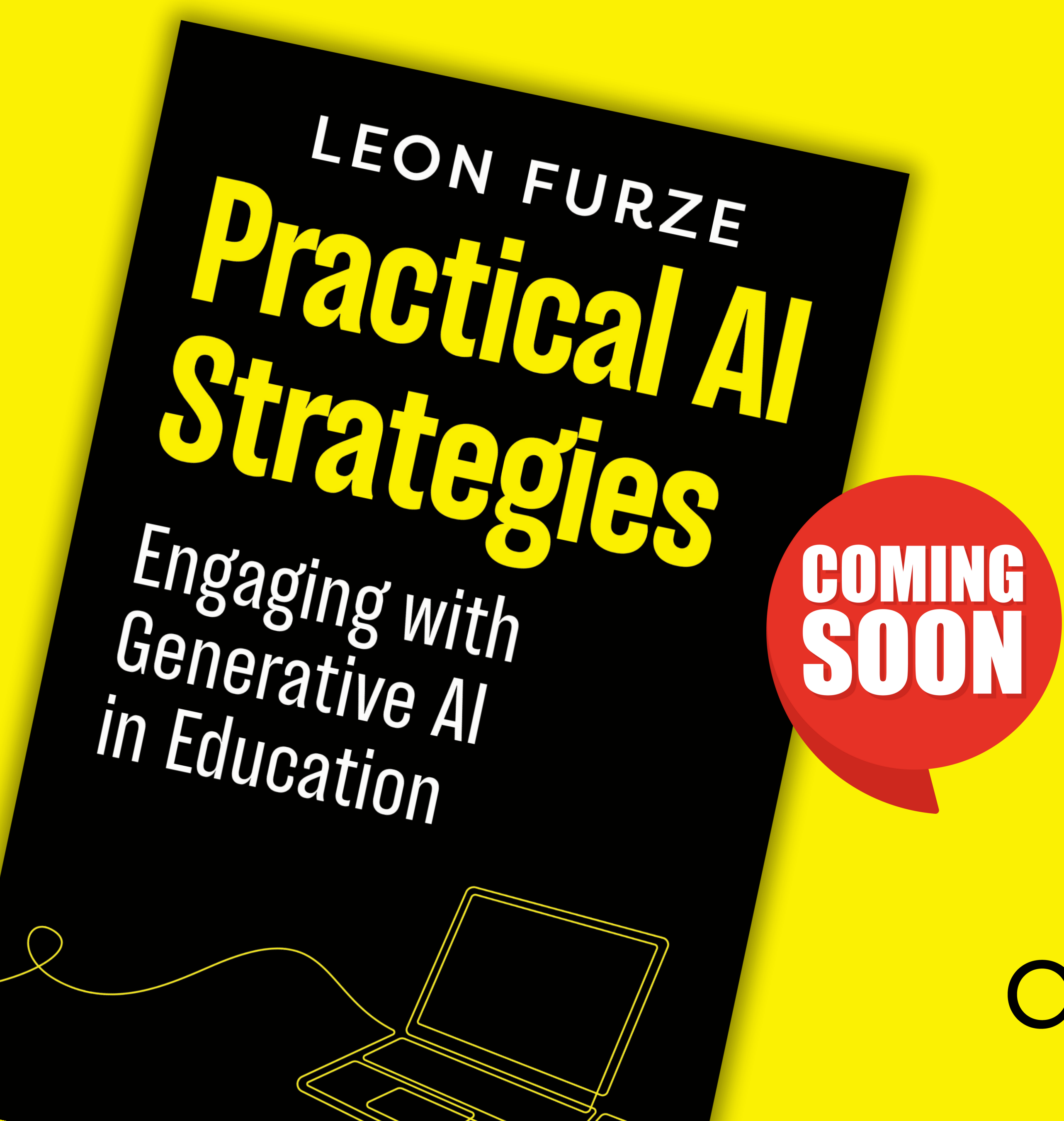
Uncorrected proof

Practical AI Strategies

Engaging with Generative AI in Education

LEON FURZE

PUBLISHING 31ST JANUARY 2024



Contents

Introduction

I applied for my PhD in 2022, interested in how large language models such as OpenAI's GPT-2 and GPT-3 might impact the ways we write texts. At the time, you could use OpenAI models as a developer by building an application on top of them, or through the OpenAI playground, which offered limited examples like a short story creator or grammar corrector.

Just a couple of weeks after the official start date of my PhD in November 2022, OpenAI released ChatGPT. The simple act of placing a chatbot interface on top of one of its most powerful models and releasing it for free to the public created a tremendous shift in how we interact with GenAI. Since then, there has been a lot of hype, but also some impressive developments in the field.

For educators, the release of ChatGPT also caused a fair amount of stress and anxiety. Media reports discussed the threat ChatGPT posed to traditional education, academic integrity and even writing disciplines. Universities in Australia and worldwide moved to pen and paper examinations to cope with the implications of this technology. Departments of education in most Australian states and territories banned ChatGPT, partly due to problematic terms and conditions, but also to address rising concerns over plagiarism and cheating.

In the following months, K-12 and tertiary education continued to grapple with the implications of these technologies for student use, particularly in generating texts for essays, topics and examinations. Yet despite academic integrity concerns, these technologies have potential positive and creative uses, including as assistive tech to support learners. Between November 2022 and November 2023, there were also significant advances in creative GenAI tools for multimodal purposes, including text, audio, video, image and code. This industry has moved incredibly quickly, outpacing education, government and the law.

Rather than focusing on ChatGPT or any other specific apps and services, this book takes a broader view of the technologies and explores ways educators might benefit from artificial intelligence (AI) in their day-to-day work. It aims to help educators learn to use the technology themselves, suggesting that the best way to learn is through experimentation using your own context and subject expertise.

The book is organised into six parts:

- Part 1 discusses how these technologies work, emphasising the importance of understanding GenAI, how models are constructed, where data comes from and how user input determines the model's output.
- Part 2 addresses the ethical concerns in GenAI, particularly those relevant to education. It encourages educators to engage in critical and reflective discussions about biases, intellectual property, privacy and data usage.
- Part 3 includes a discussion of guidelines and policies in schools, including a brief overview of the national and international education policy landscape.
- Part 4 introduces practical strategies for GenAI in education, focusing on text-based models like ChatGPT. It includes prompts and ideas applicable across various platforms.
- Part 5 follows this with a focused discussion of image generation, showing how it can be used in education and how it reveals biases and discrimination in AI.
- Part 6 explores multimodal GenAI, covering audio, video and code generation, and discusses the potential futures, challenges, risks and benefits of these technologies.

To get the most out of this book, read it with both an open mind and a critical eye. As with any digital technologies, GenAI is neither inherently positive or negative: it is the humans behind the technology that make ethical (or unethical) decisions. The advice in this book is designed to get you started with using genAI, but also to encourage you to ask important questions about when and why we sometimes *shouldn't* rely on genAI.

PART 1

What is generative artificial intelligence?

CHAPTER 1

Generative artificial intelligence

A brief history of artificial intelligence

Like any complex technology, artificial intelligence (AI) has its roots in several fields. From philosophy to computer science, mathematics to linguistics, tracing the history of AI and automation is a difficult business. The field was officially named in the 1950s, but ideas about automated machines have existed since long before then. This is a brief history of the development of AI from some of its earliest philosophical and theoretical inceptions through to modern-day technology.

Ancient automatons

Our fascination with automatons goes back a long way. Scholars have argued that the Ancient Greeks proposed automatic servants as a utopian alternative to human slaves in one of the earliest examples of technosolutionism I've seen yet. The myths of Daedalus, recalled by Plato and Socrates, describe the inventor creating 'animate statues'. Heron of Alexandria, in 60CE, wrote about steam-powered automata, engines and wind turbines.

Skipping forward a few centuries, in the Byzantine Empire, King Constantinople VII hired craftsmen from Baghdad to create enthralling golden automata to impress his guests. Similarly, the Banū Mūsā brothers, a group of 9th-century Muslim inventors, created several variations of automata, including mechanical birds that could sing and move their wings.

In more recent history, the pace of invention and the passion for automata also accelerated. From Leonardo da Vinci's robotic knight in the 16th century to development of self-driving cars and advanced AI technology in the 21st century, we are obsessed with automation.

Although many of these inventions seem like vanity projects or interesting but useless distractions, the ideas that were generated alongside them – including advances in mathematics, engineering, philosophy and science – were incredibly important. Of all these inventions, one stands out as the obvious predecessor to modern computing: Charles Babbage's Difference Engine.

The impossible machine and the first programmer

Babbage's engines (Difference Engines number 1 and 2, and his Analytical Engine) represent some of the 'greatest intellectual achievements of the 19th century.' Although Babbage found it impossible to build his machines – due to the cost and the materials required – they did inspire the world's first computer programmer.

Ada Lovelace was the daughter of the famous poet Lord Byron, though her mother Annabella Milbanke Byron separated from her father and Ada never knew him. She first met Charles Babbage through a mutual friend in 1833. They exchanged ideas via correspondence and, even though it was never actually built, Lovelace wrote programs for the Analytical Engine, including an algorithm which could be used to compute the Bernoulli numbers.

The first AI summer

Although these early pioneers experimented with automation and even suggested some of the elements of modern computers, it was not until 1950 that we saw the beginnings of the field we now call AI. In his paper, 'Computing Machinery and Intelligence,' Alan Turing proposed a test to determine the likelihood of a machine being capable of intelligence. The 'imitation game' – now more commonly known as the Turing Test – pits a machine against a human. To pass the Turing Test, a machine must be able to engage in a conversation with a human evaluator and convince the evaluator that it is a human, rather than a machine. Turing's work became one of the cornerstones of computer science.

In 1955, at a Dartmouth College conference, John McCarthy coined the term 'artificial intelligence' to describe a new field which brought together computer science and mathematics, including the work of other conference attendees such as Claude Shannon, famous for Information

Theory, and Marvin Minsky, who co-founded Massachusetts Institute of Technology's Artificial Intelligence Laboratory.

The period 1956–1973 is referred to as the 'first AI Summer' due to an increase in research, funding and government interest in the studies of AI. During the period, there were many notable achievements, including the invention of ELIZA, the first chatbot, and the creation of LISP by John McCarthy, a programming language which is still in use today.

The period came to an end in 1973 when the *Lighthill Report* cast damning aspersions on the potential for AI researchers to achieve some of their grandest claims, including modelling the human mind.

What's in a name?

The term 'artificial intelligence' fell out of favour after the *Lighthill Report*, but that doesn't mean the field disappeared. Research into machine learning, neural networks, natural language processing and other areas continued throughout the 1970s and 1980s.

Although 'AI' had become associated with the hype and unfulfilled promises of these earlier technologies, there were still great strides forward. From advanced mathematics like hidden Markov models to computer science and Kunihiko Fukushima's Convolutional Neural Networks, development continued towards the kinds of AI we are more familiar with today.

In the 1980s, AI entered a brief 'second summer' with the promise of neural networks, but couldn't shake the disappointments of previous eras. Despite efforts by researchers and technology companies, the field entered a second winter and research once again became more conservative.

Garry Kasparov versus the machines

It wasn't mathematics or computer science that ultimately lifted AI out of its second winter and back into the public eye: it was chess. Garry Kasparov's book *Deep Thinking* details the long and complicated journey towards the development of IBM's Deep Blue, the artificial intelligence that was the first to beat a human chess Grand Master.

Since then, Deep Blue hasn't been the only AI to grab headlines by beating humans at their own games. IBM's Watson beat human contestants at *Jeopardy!* in 2011; Google's AlphaGo defeated a human Go champion in 2015; and in 2022, Meta's CICERO model defeated humans in a game of Diplomacy.

The modern age of AI

Over the past two decades, there has been a significant increase in funding and research for AI and ML projects. The rapid advancement of technology, such as smartphones, the internet and social media, has enabled leading tech companies, such as Google, Microsoft, Amazon and Meta, to develop powerful AI systems that drive their predictive engines, search tools and business models. These companies have access to vast amounts of data, which is crucial for ML algorithms.

In the early 1980s, the shift from symbolic methods to neural networks marked the beginning of the current AI revolution. With the availability of massive datasets and mass surveillance, these algorithms have become truly useful. As a result, AI can now be found in a variety of everyday technologies, ranging from cars to refrigerators. Despite controversies surrounding the definition of AI as intelligence, it is unlikely at this stage that we will experience another full AI winter.

The future of AI remains uncertain. While some experts worry about the potential harm that artificial general intelligence (AGI) or artificial superintelligence (ASI) could cause to humanity, others, such as futurist Ray Kurzweil or OpenAI CEO Sam Altman, are optimistic about the positive impact that AI could have on society. There is a growing consensus among experts that finding a balance between the dystopian and utopian perspectives of AI is crucial.

Current AI models such as OpenAI's GPT and Google's PaLM (the basis for its Bard chatbot) are dominating the headlines. Microsoft has built its Copilot products on top of the GPT model, and open-source models like Meta's Llama continue to grow in both popularity and sophistication.

AI has had a long and complex history since the term was coined in the 1950s, but there is no doubt that we are in a time of rapid development and change. The AI arms race between huge corporations like Google and Microsoft will accelerate the pace. As educators, we need to make sure we don't get swept up in another cycle of AI hype and lose sight of the very real ethical and social concerns of these technologies.

What is GenAI?

Generative artificial intelligence (GenAI) is a subfield of AI in which ML is applied to large datasets in order to learn ways to generate new data.

That data might be in the form of text, code, images, audio or anything else that can be turned into machine-readable content. In practice, that means that GenAI models are *multimodal* and can both read and create data in a variety of modes, such as *text* generation, audio (like speech-to-text), and even video and 3D assets.

While it's not necessary for every educator to have a full technical understanding of how GenAI models work, it is useful to know some of the details so that you can understand the strengths and limitations of the technology. Understanding how models are constructed helps to understand why they fabricate information (often called 'hallucinations') and why they can't be used as search engines. It also highlights why bias and discrimination are features of these models, which is explored more in Part 2.

The AI iceberg

To help understand how large language models (LLM) like GPT are constructed, I use the analogy of the AI iceberg.

Picture an iceberg floating in the ocean. The visible part above the waterline is relatively small compared to the massive structure hidden beneath the surface. Now, imagine that this iceberg represents an LLM, like GPT-3 or 4, with its different components distributed above and below the waterline.

The dataset: underwater bulk

The bulk of the iceberg, hidden underwater, represents the vast dataset on which the LLM is trained. This data forms the bedrock of the model's knowledge and capabilities. It's vast and mostly unseen during any interaction with the model, but it's always there, informing every output.

Different models are trained on different combinations of datasets. For companies like OpenAI and Google, some of that information is proprietary. While we have some information on GPT-3's training data, GPT-4 is more of a mystery, and Google's PaLM is off-limits. But we do know a little about the kinds of data these large models are trained on. We know, for instance, that they contain data from sources like the Common Crawl, the Pile, Wikipedia and coding site GitHub. They are also be trained on social media sites like Twitter and Reddit.

The dataset is *huge*. GPT-4, for example, is trained on around 13 *trillion* tokens – words or parts of words converted to machine-readable format. That's about 1,600 times the population of Earth.

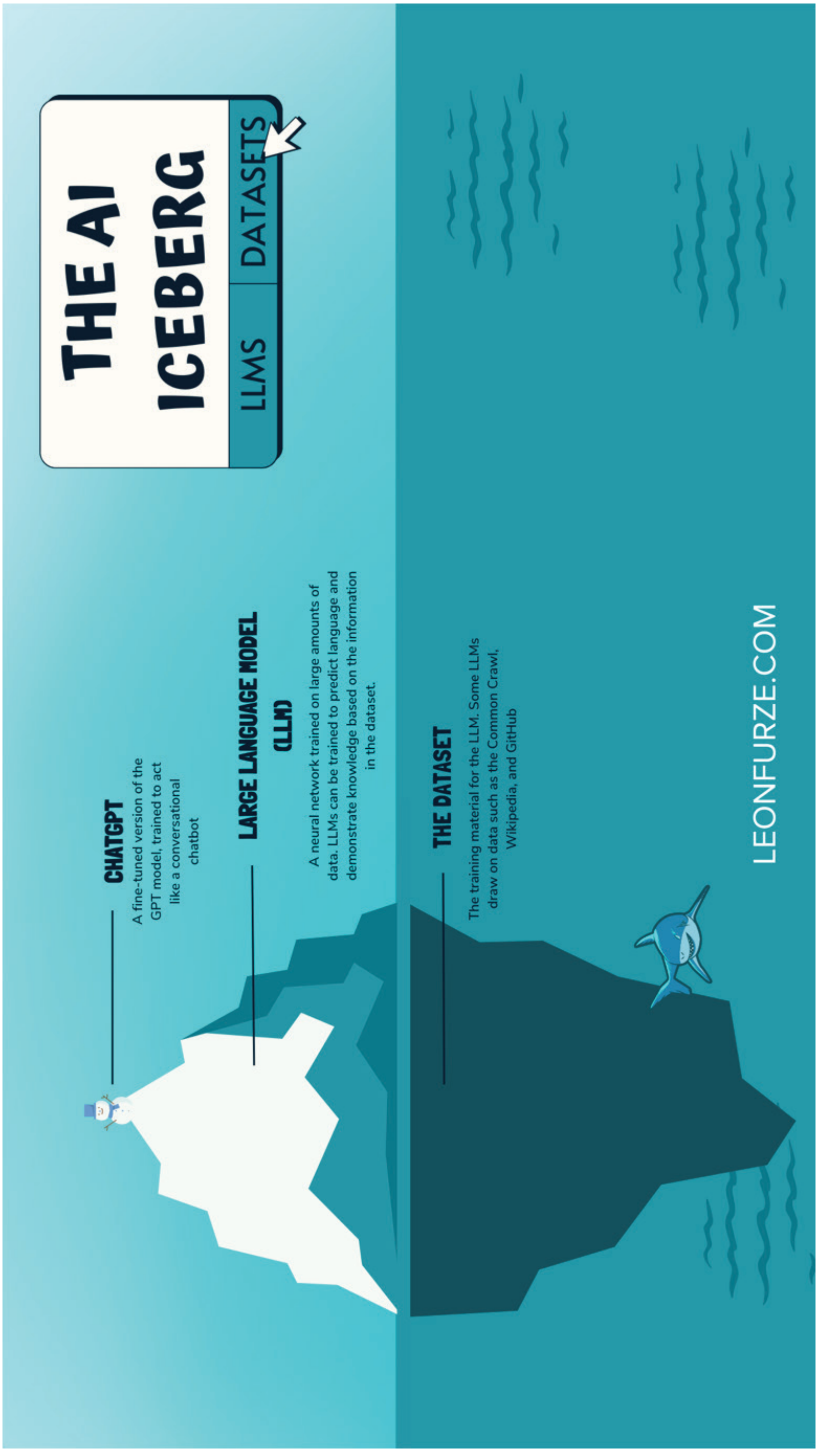


Figure 1: The AI Iceberg

The LLM: above the waterline

Emerging above the waterline is the LLM itself, the result of the training process fuelled by the vast dataset beneath. This visible portion is what we interact with when we use applications built on top of the LLM.

A LLM is an AI system that has been trained to understand and generate human language, including multiple languages and dialects, and code. These models are designed to predict the likelihood of a certain word given the words that came before it in a sentence or text. This ability allows them to generate coherent and contextually appropriate sentences, paragraphs and even entire texts.

Here's a *very* brief overview of some of the key processes:

1

Tokenisation

The text in the dataset is *tokenised*, which means it is broken down into machine-readable numerical code.

2

Weighting

The connections between tokens in the neural network are weighted, assigning numerical values for the probability of them occurring near to one another.

3

Transformer architecture and the attention mechanism

The Transformer architecture, introduced by Google researchers in 2017, allows the model to 'zoom out' and pay 'attention' to larger groups of tokens, representing entire phrases, sentences,and so on.

During this process, the model 'learns' the rules of language by associating words with other words, and predicting what the output should be based on the user's input (the *prompt*).

The most talked-about LLM right now is OpenAI's GPT, currently in version 3.5 (free) or 4 (subscription or access via Bing and the developer API). But there are many more out there, some open source, and some owned by companies like Google.

Applications like ChatGPT: the snowman

Finally, picture a snowman sitting on top of the iceberg. This represents an application like ChatGPT, which is built on top of the general LLM. The snowman is a more specialised figure carved from the raw material of the iceberg, just as ChatGPT is a version of the GPT model that has been fine-tuned specifically for conversational tasks.

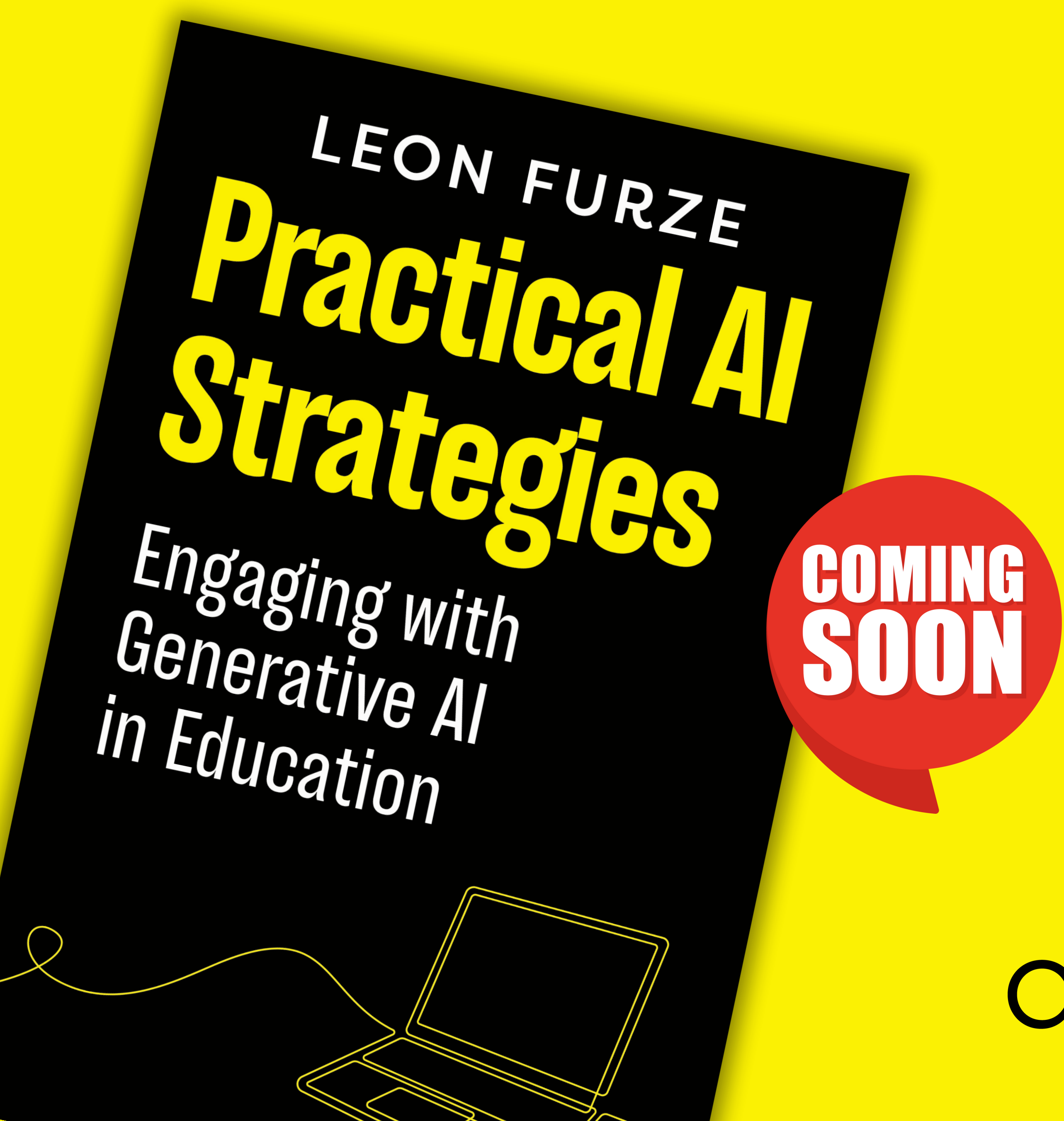
The chatbot is trained to act as a 'helpful assistant' through various processes, including *system messages*, which are the rules passed to the model for every interaction, and reinforcement learning from human feedback (RLHF), where humans train the model and judge its output as positive, negative, harmful, correct, incorrect, and so on. There are ethical complications with RLHF, which I'll discuss in Part 2.

Practical AI Strategies

Engaging with Generative AI in Education

LEON FURZE

PUBLISHING 31ST JANUARY 2024



amba press