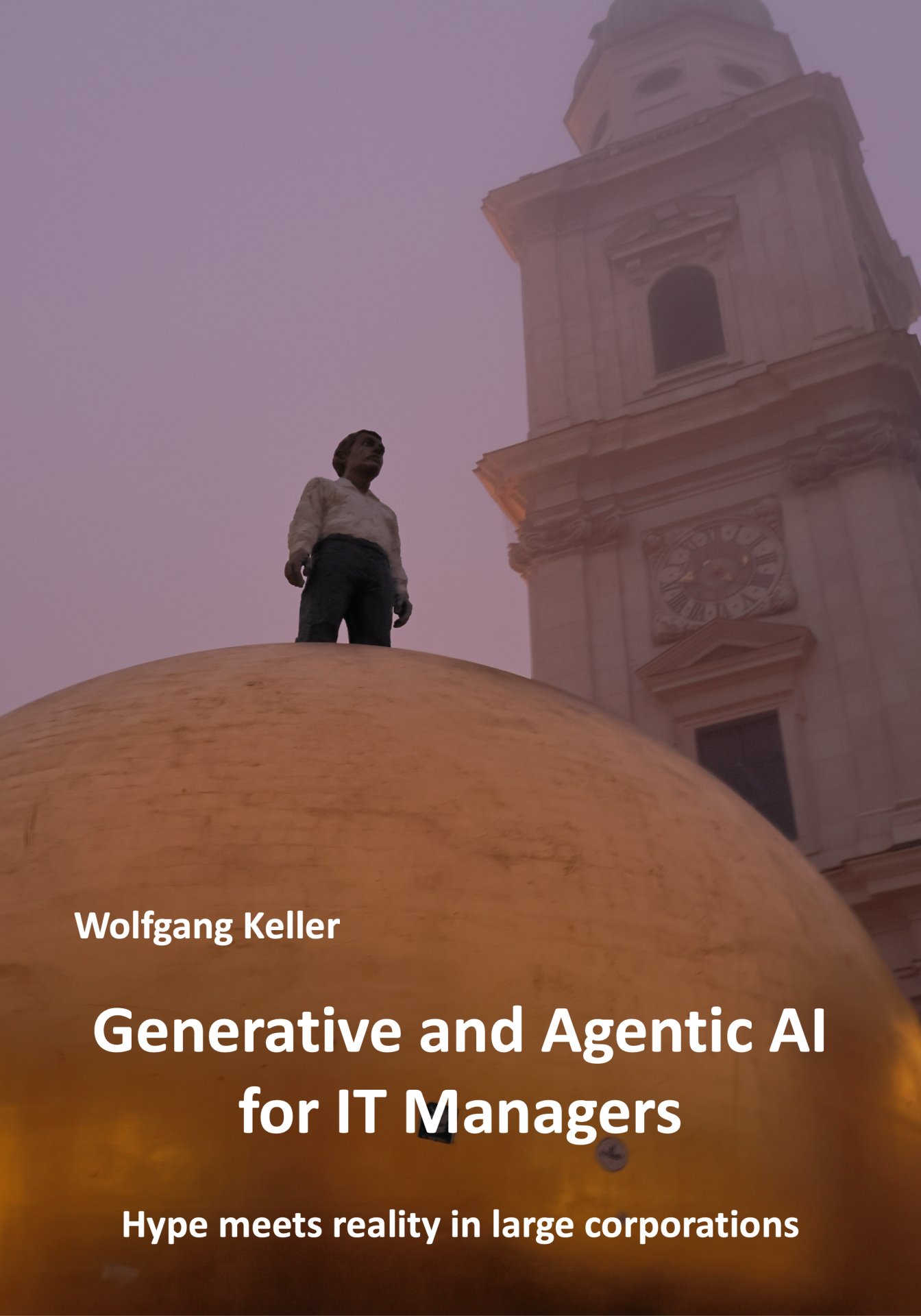

Wolfgang Keller

Generative and Agentic AI for IT Managers

Hype meets reality in large corporations

Generative and Agentic AI for IT Managers

Hype meets reality in large companies

Wolfgang Keller

This book is available at https://leanpub.com/agentai-en-us

This version was published on 2026-06-08

Translated using Claude Opus 4.8

Contents

Before We Begin

About the Author

Wolfgang Keller is an independent consultant specializing in the management of large software projects, enterprise IT architecture, and solution architecture. He has more than 30 years of experience building large, custom application systems as a software engineer, consultant, project manager, chief architect, and line manager of sizable delivery organizations. He is the author of several professional books (see his Amazon author page https://www.amazon.de/Wolfgang-Keller/e/B0043BVFII) and numerous articles in professional journals.

Professional profile: https://www.linkedin.com/in/wolfgangkeller/

Acknowledgments to the Guest-Chapter Authors

For several chapters I did not have the same first-hand, in-depth knowledge as some of the colleagues in my network. I therefore brought those chapters in as guest contributions, simply because their authors have more genuine expertise in these areas than I do. They are:

Chapter 8: **Coding with LLMs and Agents** by Alexander Hofmann, Dr. Claas Busemann, Dr. Josef Reislhuber, Francesco La Torre, and Tobias Wagner, all of MaibornWolff

Chapter 12: **Why AI Forces Us to Rethink IT Security From Scratch** by Florian Oelmaier, IS4IT

Chapter 14: **AI That Works – Practice, Not PowerPoint** by Dr. Andrea van Aubel and Axel Helmert, msg group

My heartfelt thanks to Dr. Andrea van Aubel and all the authors – not least for delivering at AI speed, and for the genuinely deep perspectives they brought to the book. **THANK YOU!**

Preface to Version 1.0

Going back through the history of artificial intelligence since the 1940s, there is truly no shortage of specialist literature and deeply technical articles on AI in all its facets. As Chapter 2 of this book describes, AI has repeatedly gone through "AI summers" and then AI winters again. As the first version of this book appears, in March 2026, we happen to be in a very hot summer indeed. The ChatGPT moment came back in 2022: from then on, you could ask large language models questions and actually get genuinely helpful answers. But with the arrival of models that can draw up more elaborate plans to solve complex tasks, and the rise of what are called AI agents, the possibilities of these technologies are expanding dramatically once again.

Given today's breakneck pace of development, it is fair to ask whether it even makes sense to write a book on the subject at all. There is certainly no shortage of papers on, say, benchmarks of AI models (is model A better than model B? which benchmarks is model C good at?). Papers like that also go out of date quickly, and this book makes no claim to join that race. But there is a "gap in the market" for writing that helps you, as a decision-maker, get an overview of the technology's potential and its limitations.

A substantial body of regulation has grown up around the topic, and existing regulations have to be observed as well. Regulation and its enforcement move far more slowly than the technology race for the best model or the most efficient way to train one. On top of that, in large companies (and small ones), security has to be taken into account. The issues that fundamentally follow from all this do not age nearly as fast as the latest tool versions, either. And finally, it is worth remembering that AI already has a fairly long – and quite successful – history in many large enterprises. There, you have to think very carefully about which uses of brand-new technology deliver business value and which do not.

As someone who has worked in insurance IT for more than 30 years – in a wide variety of roles – I was simply interested in these questions myself, first and foremost. And whenever I am researching something anyway and cannot find anything "off the shelf," I tend to write it up as I go. That is how the first edition of my book on enterprise IT architecture came about, back in 2006, and it is how the idea for this book came about too. Wherever I felt that people in my network had deeper hands-on experience in a specialized topic than I did, I was able to find very capable authors for a total of three guest chapters.

Now I hope this book proves helpful to you in quickly getting an overview of the key decision areas around generative and agentic AI.

Munich, March 2026

Wolfgang Keller

1: Introduction and Overview

Generative AI and AI agents are not the first hype topic to set companies in motion. But they are clearly one that will develop a transformative force the likes of which history has not seen before.

After the ChatGPT moment in 2022, investors and salaried managers alike—from the CEO to the enterprise designer to the business-unit staffer or software engineer—started asking, among other things, how their company can get the most out of AI.

1.1: The Fog of Hype

Generative AI promises a massive transformation of the way companies work. When it comes to the "how," though, it is easy to end up lost in the fog. That is what the cover of this book is meant to convey: you are standing on a sphere of gold, but the view is a little hazy. New models, agent frameworks, and product announcements appear every day. CEOs ask when the company will finally put "this AI" to work, developer teams push for the latest tools, and consultants sell visions of autonomous systems that make decisions without needing humans to do it.

1.2: Who This Book Is For

This book is written for IT-savvy managers—line managers, project managers, enterprise architects, solution architects, business analysts, software engineers—who want to deliver on the promises that generative AI and AI agents are making.

Depending on your industry and how heavily regulated it is, you will run into different challenges. You will find comparatively much in this book about regulated industries such as insurance and banking. Similar constraints—only tighter still—exist in healthcare. All of these industries offer promising approaches, but you cannot simply turn an AI agent loose and naively hope that a golden age will descend on your company as a result.

This book will help you spot challenges that the people trying to sell you products or projects, in many cases, either do not know about or conveniently overlook.

So this is not a book that takes you into the "engine room" of AI: machine learning or the deep questions of agent technology. It goes as deep as it needs to in order to explain the management questions that matter. It is a guide for people who have already lived through more than one hype cycle of the kind that regularly sweeps over companies.

AI in Large and Regulated Companies

One central thesis is this: generative AI is a powerful tool, but not a universal one. In regulated environments, autonomy and nondeterminism collide with requirements such as traceability, reproducibility, and compliance. The question is not "AI, yes or no," but "which AI, where, and under what controls." Many successful AI applications have been running in production in the financial industry for years—entirely without autonomous agents making decisions on their own.

AI Is More Than LLMs and Agents

AI has a history of roughly 70 years, with AI summers and AI winters. Right now we are in a summer, driven by large language models (LLMs) and agentic AI. The next chapter, on agentic AI, therefore opens with a section on how agentic AI and LLMs fit into the larger history of AI.

1.3: Structure: Who Should Read What, and Why?

This book grew out of individual white papers on questions that have to be asked when you deploy generative AI and AI agents—but that often surface only once a project has been running for a while. The next two tables give you an overview of what each chapter covers and which audiences each part is relevant to.

Chapter / Topic	General Management, IT Management	Enterprise Architects	Solution Architects, BAs, other IT Professionals
Part A — Fundamentals			
AI That Works - No New Directors What AI agents are, where you can use them, and where you cannot.	++	++	+
AI and EAM: The Wrong Questions How to use EAM as a lever to advance AI adoption in the enterprise. What happens when you roll out AI without first cleaning up your landscape.	++	++	o
Agent Gateways: The Revenge of SOA in the Age of AI Argues that you need central infrastructure if you do not want to drown in a chaos of point-to-point connections.	+	++	+
In the Land of Lies: LLMs and Hallucinations Explains what hallucinations are and the dangers they pose.	++	++	++
How AI Accesses Enterprise Knowledge How to incorporate your company's knowledge into what LLMs and agents know — without expensive training of neural networks.	+	++	++
It's a Model and It's Looking Good Explains how to proceed when you have to select AI models (LLMs).	o	++	++
Coding with LLMs and Agents How to proceed if you want to make sensible use of generative AI to generate code.	++	++	++

Figure 1. Chapter relevance by audience

Chapter / Topic	General Management, IT Management	Enterprise Architects	Solution Architects, BAs, other IT Professionals
Part B — Compliance and Security			
AI-Relevant Regulation: The Financial Sector as an Example Provides an overview of regulations relevant to the financial industry, especially the EU AI Act, and makes the case for ISO/IEC 42001.	++	++	+
The Inherent Risks of LLMs Introduces risks that are inherent in using LLMs as opposed to conventional software.	++	++	+
Securing Applications Built on LLMs Applications that build in LLMs as a component carry special security risks that did not exist this way in web applications. A considerably broader security strategy is therefore required.	++	++	++
Why AI Forces Us to Rethink IT Security From Scratch Shows how the overall threat landscape worsens dramatically through the use of LLMs.	++	++	+

Chapter / Topic	General Management, IT Management	Enterprise Architects	Solution Architects, BAs, other IT Professionals
Part C — Practice, Not PowerPoint			
Imagine It Is the Age of AI and Nobody Shows Up Looks at change management for the adoption of modern AI in companies.	++	++	+
AI That Works — Practice, Not PowerPoint Shows three key applications of agentic AI in insurance companies and how to deploy agentic AI in a regulatory-compliant way.	++	++	+
Not Every Project Is the Same Discusses different types of AI projects and shows that the approach differs significantly depending on the type.	++	++	++

Figure 2. Chapter relevance by audience

1.4: What Happens Next?

The book in front of you is a product of "lean publishing." In the end, lean publishing leads to a reputable book, too—one you could take to a traditional publisher if you wanted to. But this medium offers something else: the chance to get fast feedback from your readers.

Early readers were rewarded with a lower entry price. Every reader who wants to can subscribe and will automatically receive emails about updates from leanpub.com. So you will be notified whenever new content is made available here.

You will also find an email address for feedback here:

wolfgang.keller@objectarchitects.de

This book thrives on your questions and suggestions for improvement.

With the version you have here, the book has reached a consolidated state of 15 chapters. More would certainly be conceivable, but in its current form the book already feels complete. Questions and suggestions for additional chapters are very welcome, and there will surely be updates and expansions.

1.5: Why This Book Isn't Free

This is not a quick AI-generated cash grab, not a collection of prompts, and not a tool catalog. It is a strategic professional book for people who carry responsibility. It addresses questions that often surface in projects only once things have already gotten expensive:

- Where do AI agents make sense—and where are they dangerous?
- How do autonomy and regulation collide?
- Which architecture decisions are reversible—and which are not?
- What is hype, and what is sustainable?

You cannot derive the answers from marketing slides. The value of this book lies not in its page count but in the bad decisions it helps you avoid. A single wrong architecture decision, a misapplied agent, or a naive automation in a regulated context can:

- cost millions,
- create compliance problems,
- delay projects,
- destroy trust. If this book helps prevent even one such bad decision, it has paid for itself many times over.

1.6: How AI Was Used for This Book

I have been using generative AI—that is, LLMs—for more than two years now, for a variety of tasks in consulting, enterprise architecture, and solution architecture. My main use has been supporting first review processes and, more recently, supporting the writing of articles or, as in this case, the writing of a book. For this book I primarily used Claude Opus 4.5 and 4.6. Whether that will still be the top model for such tasks three months from now, I do not know—today, in March 2026, it is the one for my purposes.

A popular line in the context of books written with AI's help goes like this:

> If the author can't be bothered to write the book, then I can't be bothered to read it.

That certainly applied to a wave of AI-generated books during the first rush of AI enthusiasm. Things have improved since then, and a book written with AI support can still represent a creative achievement—if you take a closer look at the process:

- First, a book like this needs the right questions. In the case of this book, they come from experience with many other hype cycles and from years of working on large, complex software projects.
- AI is not good at irony, or at writing titles and theses that prompt you to read or to think.
- As the author, you constantly have to weigh where to let the potentially dull lists typical of AI-generated documents stand, and where to rewrite them. AI detectors help with that today—a story of its own, and not one this book sets out to tell.
- AI speeds up literature searches enormously. Granted, those searches do not have the quality of "balanced citations from reputable academic works"—but the world is moving so fast that this is not required for every book anyway. For the goal of this book, it was important to keep up, at least reasonably, with the speed of the hype. And there, even the review cycle of a professional journal is too slow.
- And finally, AI helped quickly transform the first LinkedIn white papers I wrote on the topic into Leanpub's input format. I was already publishing on Leanpub ten years ago, without AI. Generative AI has hugely increased the speed of dull routine tasks and quality checks.

All in all, then, these are the reasons why I see no added value—for readers either—in writing the book "purely by hand." The topic we are pursuing here together is moving far too fast for that.

Part A — Fundamentals

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

2: AI That Works - No New Directors

Why Agentic AI Need Not Be the Optimum in Regulated Environments

Autonomous AI agents are supposed to decide, plan, and act on their own before long. But what if that is exactly what makes them dangerous in regulated environments?

Agentic AI is widely seen as the next evolutionary step of artificial intelligence: systems that formulate goals, develop plans, and act on their own. Together with LLMs, it is driving a wave of hype right now. For marketing, content production, and certain parts of software development, it can deliver enormous productivity gains. But in heavily regulated industries such as insurance, banking, and healthcare, these very properties collide with core requirements: traceability, reproducibility, liability, and compliance.

This chapter shows why nondeterminism, missing explainability, and autonomy that is hard to control pose not only technical but also regulatory and organizational risks in those settings. Through historical context, an accessible technical analysis of modern AI agents, and concrete real-world examples, one thing becomes clear: many successful AI applications have been running productively for years - entirely without autonomous decision logic.

Instead of "more autonomy at any cost," this chapter argues for a differentiated view: agentic AI is a powerful specialized tool, but not a universal one.

After reading this chapter, you will know where agentic AI has the potential to deliver real value, and when classic, deterministic AI architectures are the better, safer, and smarter long-term choice.

2.1: AI and the Hype

AI Is Much More Than LLMs and Agentic AI

When people talk about artificial intelligence today, most of them immediately think of ChatGPT, Claude, or other large language models. The current hype around agentic AI only reinforces that impression. And yet it is easy to forget that AI has a history spanning more than seventy years - and that much of what works in companies today has nothing to do with LLMs at all.

The story begins in 1950, when Alan Turing posed the question in his famous paper "Computing Machinery and Intelligence": "Can machines think?" [Britannica 2024]. The Turing test he proposed in that context still occupies philosophers and computer scientists to this day. Turing himself was already working on early concepts for neural networks - ideas that would not gain practical relevance until decades later.

In 1956, Dartmouth College hosted the legendary conference at which John McCarthy coined the term "Artificial Intelligence." The participants - among them Marvin Minsky, Claude Shannon, and Herbert Simon - were convinced that within a single generation, machines would be built that could match humans at intellectual tasks [Britannica 2024]. That confidence was not unfounded: in the years that followed, programs emerged that played chess, proved mathematical theorems, and solved algebra word problems. It is hard to overstate the euphoria of those early years - people genuinely believed the secret of intelligence was within reach.

ELIZA, built in 1966 by Joseph Weizenbaum at MIT, simulated a Rogerian psychotherapist and fooled many users into forgetting they were only talking to a program. That was impressive - but it was also an early warning sign of the problems we run into with LLMs today: the ability to sound convincing without truly understanding. Weizenbaum himself was appalled at how readily people confided their most intimate secrets to the program, and he became one of the sharpest critics of his own invention.

Then came the first AI winter in the 1970s. The promises had gone unfulfilled, computing power fell short, and the funders grew impatient. Anyone who has ever watched a hype cycle collapse knows the pattern: first boundless euphoria, then the sobering moment when reality emerges from behind the PowerPoint slides. In the 1980s, expert systems enjoyed a brief boom - programs that encoded the knowledge of human experts in rule sets. MYCIN

diagnosed bacterial infections; DENDRAL identified chemical compounds. These systems worked surprisingly well in tightly bounded domains, but they were brittle: every new situation called for new rules, and translating expert knowledge into formal logic turned out to be more laborious than expected.

The next winter followed at the end of the 1980s. Once again, the problems had proven harder than anyone had thought. AI research retreated into academic niches, and anyone who used the word "AI" in a grant application could brace themselves for skeptical looks.

What many people do not realize: it was during this seemingly quiet phase that the foundations for today's success were laid. The backpropagation algorithms for neural networks were refined. Statistical methods replaced symbolic approaches. And in the financial sector, AI quietly began doing real work - it just wasn't called that, because the term had been burned.

The credit-scoring models that every bank uses today are based on machine learning - even if the marketing department avoids the word. When your credit card suddenly gets blocked because the system has flagged an unusual transaction, that was an AI system that has been catching fraud in a largely automated way since the 1990s. Reinsurers use complex statistical models to assess the risk of natural catastrophes, trained on decades of loss data. PDF extraction and automatic document classification have been standard in the claims handling of large insurers for years - dull, invisible work that excites no one but saves millions in staffing costs.

All of that is AI - just not the kind that makes headlines on LinkedIn. These are specialized systems that do one job well without raising philosophical questions. They do not plan autonomously, they do not generate creative text, and they do not hold conversations. But they work. Reliably. For years now. And in a regulated environment, that is what ultimately counts.

The real breakthrough came in 2012, when a neural network called AlexNet won the ImageNet competition by a margin that shook the field [Britannica 2024]. Suddenly deep learning was more than an academic curiosity. The combination of huge data volumes, powerful GPUs - originally built for video games, but perfectly suited to matrix multiplications - and improved algorithms enabled advances that even optimists had not expected.

In 2016, AlphaGo beat the world champion at Go - a game in which the number of possible positions exceeds the number of atoms in the universe. That was no longer brute force; it was something new: a system that seemed

to have intuition. The decisive Move 37 in game two - a move no human player would ever make, and one that proved to be brilliant - became an icon of a new era.

In 2022, ChatGPT hit the market. Suddenly anyone could talk to an AI that no longer sounded like a machine. One that answered questions, wrote text, generated code - and often came across as astonishingly competent. The hype was born, and this time it reached an audience far beyond the tech bubble.

What gets lost in all this: most companies that use AI successfully today do so with little help from ChatGPT or its competitors. They use the dull but proven AI of the past few decades - enriched with modern methods where it makes sense. The claims handler who processes claim notifications every day benefits from automatic document classification running in the background. That it is based on a neural network is, rightly, of no interest to him.

2.2: Agents Explained

What Sets LLM-Based Agents Apart From Other Forms of AI

The term "agent" is nothing new in AI research. In their standard textbook "Artificial Intelligence: A Modern Approach," Stuart Russell and Peter Norvig define an agent as a system that perceives its environment and influences it through actions [Russell/Norvig 2023]. By this definition, a thermostat is already an agent - a very simple one, granted, but one that responds to temperature changes and acts accordingly. That broad definition is philosophically clean, but it does not help us understand what is new about the current developments.

So what do we mean when we talk about "agentic AI" or "AI agents" today? IBM defines it like this: "Agentic AI refers to artificial intelligence systems designed to autonomously perform tasks, make decisions, and adapt to new information with minimal human intervention" [IBM 2025]. McKinsey describes it similarly: "Agentic AI systems can accomplish complex goals with minimal human input by breaking down a task, creating a plan for achieving it, executing each step in that plan, and adapting based on real-time feedback" [McKinsey 2025]. So today's debate means something more specific than Russell and Norvig did: systems based on large language models that can plan autonomously, use tools, and pursue complex tasks without constant human guidance.

The classic distinction between reactive and cognitive agents helps here. A reactive system responds to inputs with preprogrammed answers - like ELIZA, or a chatbot with fixed dialog flows that works on the pattern "if the customer says X, then reply Y." A cognitive agent, by contrast, can set goals on its own, develop plans, and adapt those plans to changing circumstances. When the first path does not work, it tries another - without that alternative path ever having been programmed.

Modern LLM-based agents combine several components, shown in Figure 2.1.

The **planning component** breaks complex tasks down into substeps. If you tell an agent "book me a flight to Vienna and a hotel near the State Opera," it has to understand that this calls for several actions: flight search, hotel search, availability check, booking. The so-called ReAct architecture pattern (reasoning and acting) alternates between thinking and acting - the agent considers what the next step should be, executes it, observes the result, and then plans further. Anyone who has ever done complex project planning will recognize the pattern: you break the big goal down into manageable packages and adjust the plan when reality turns out to be more complicated than expected.

Memory lets the agent retain context across longer interactions. Short-term memory holds the current conversation - what we just discussed, what information has already been exchanged. Long-term memory, often implemented via RAG (retrieval augmented generation) (on RAG, see Chapter 6), can retrieve information from earlier sessions or external knowledge bases. The agent then "knows" that last time you stayed at the Hotel Sacher in Vienna and may be looking for an alternative.

Tool use - known in the jargon as tool calling - lets the agent reach external systems: query databases, call APIs, read and write documents, search the internet. Without that capability, an LLM would be limited to its training knowledge, which ends at a specific cutoff and contains no company-specific information. With tool calling, the agent can pull up current flight prices, check your calendar, and actually make a booking.

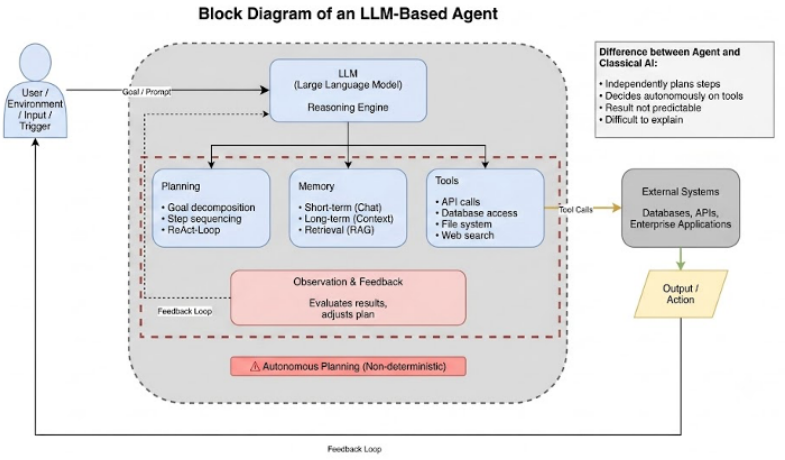

Figure 3. Block diagram of an LLM-based agent

The decisive feature that sets this apart from earlier AI is autonomous planning. A classic expert system follows predefined rules - if A and B, then C. A supervised machine learning model classifies inputs according to learned patterns - this image probably shows a cat. An LLM agent, by contrast, can find paths that were never explicitly programmed. When the direct route is blocked, it tries a detour. It can improvise.

That sounds great - and for many use cases it is. But it comes with two fundamental properties that become a problem in certain contexts.

Nondeterminism: the same prompt can produce different answers. More than that: the same agent with the same goal can develop different plans and carry out different actions. That is not a bug, it is a feature - it enables flexibility and creativity. But it also means you cannot predict what the agent will do. For a software developer used to the same input always producing the same output, that is a paradigm shift. For compliance reviews, it can be a showstopper.

Lack of explainability: why did the agent make this decision and not another? With a neural network that has billions of parameters, you cannot reconstruct it. You can see the result, but not the reason. That makes debugging hard - when something goes wrong, you do not know which screw to turn - and compliance nearly impossible, as we will see in a moment.

These two properties lead straight to the core question of this chapter: in

which environments is autonomous planning an advantage - and in which does it become a risk?

2.3: Directors Not Wanted

Why Autonomy and Nondeterminism Can Be a Problem in Regulated Environments

Picture the following insurance claim: a storm damages the roof of a single-family home. The policyholder reports the loss through an online form with photos. What happens next? In a hypothetical "pure agentic AI" world, it would look like this:

- an AI agent receives the notification,
- analyzes the photos,
- estimates the loss amount,
- checks coverage under the policy,
- asks follow-up questions and requests additional documents where needed,
- decides on the claims settlement in the end,
- and authorizes the payment.

Fully autonomous, with no human intervention, and done in minutes instead of days. That sounds tempting - faster, cheaper, no waiting for the customer. But let us play out what can go wrong.

Escalating hallucinations: LLMs hallucinate - they generate statements that sound plausible but are factually wrong. This is described in detail in Chapter 5 on hallucinations [Rippling 2025]. In a chat with an LLM, that is annoying; you notice the error and correct it. With an autonomous agent that builds on its own outputs, it can turn catastrophic. The agent estimates the loss at 15,000 euros - based on a hallucination about current contractor rates in the region, which the LLM "remembers" from its training knowledge, but which is two years old. That wrong number becomes the basis for its next decision. It approves the payment without human review, because it falls below the threshold. The error manifests itself in the real world as a real payment to a real bank account.

The Rippling security analysis aptly calls this phenomenon "Cascading Hallucination Attacks" [Rippling 2025]: a hallucination in step one triggers an automated action in step two, and so on - a chain reaction of faulty data running through the entire system. This is not a theoretical scenario but a documented risk with multi-step agents.

Missing traceability: BaFin (Germany's financial regulator), EIOPA (the EU insurance regulator), and other supervisory authorities require that decisions be documented in a traceable way. When a customer complains, or an auditor asks: why was this claim settled this way and not another? With a rule-based system, I can answer that: "Because Section 12, Paragraph 3 of our policy terms prescribes this calculation method, and the input values were X, Y, and Z." With an LLM agent, the honest answer is: "The agent developed and executed a plan whose exact logic cannot be reconstructed." That is not an acceptable answer in front of a regulator.

In August 2025, EIOPA published a comprehensive Opinion on AI Governance and Risk Management that addresses precisely these points [EIOPA 2025]. The core message is unmistakable: insurers must implement human-in-the-loop safeguards and ensure that staff can understand, question, and override AI-generated outputs. The EU AI Act, in force since 2024, classifies AI systems for risk assessment and pricing in life and health insurance as "high risk" - with correspondingly strict requirements for documentation, transparency, and human oversight [Debevoise 2025].

Limited ability to intervene: when an agent works autonomously, it works autonomously - that is tautological, but the consequence is often overlooked. When the agent heads down the wrong path, how do you stop it? With a multi-step plan, the first step may already be done before anyone notices something is going wrong. Emails have been sent, bookings made, payments authorized. Recalling an email you have already sent is embarrassing; canceling a booking costs fees; but clawing back a wrong claim payment is a legal nightmare.

Technical risks: on top of that come technical risks such as circular calls and uncontrolled resource consumption. In multi-agent systems, agents can call other agents - that is by design, in order to tackle complex tasks. But what happens when agent A delegates a task to agent B, which passes it to agent C, which in turn asks agent A for help? Or when an agent gets stuck in a loop because it keeps trying the same action that keeps failing? This is not a theoretical problem; it happens in practice, and the costs for cloud-based LLM API calls can escalate fast in the process. A few hundred euros for a single claim that spirals out of control may still be bearable, but across thousands of cases a day, it adds up.

Liability: who is actually liable when an autonomous agent makes a wrong decision? The company that deploys it? The vendor of the LLM? The developer of the agent framework? The EIOPA Opinion is clear on this point: insurers remain responsible for the systems they deploy, even when those systems were built by third parties [DLA Piper 2025]. You cannot outsource liability by saying "the agent decided that." The decision to deploy the agent was yours.

2.4: How Claims Processes Already Work Today

With AI, but Without Agentic AI

You do not need autonomous agents to use AI successfully in claims handling. The insurance industry has been doing it for years - with classic, deterministic AI methods inside a well-defined business process that builds in human control at the decisive points.

Let us look at the AI elements you can already find in a claims process today, even without agentic AI.

- **Image analysis:** the customer submits a claim notification with photos and a description. A computer vision model - trained on tens of thousands of historical loss photos - automatically classifies the type of loss. Is it storm damage to a roof, water damage in a basement, fire damage? The system makes this call with high reliability, because it has trained on this one task thousands of times over.

- **Text analysis:** in parallel, an NLP model extracts structured information from the free-text description - date of loss, rooms affected, measures already taken.
- **Coverage check:** in the next step, a rule-based system checks against the policy data whether the reported loss is covered in principle. There is no interpretation here, only clear if-then rules derived directly from the policy terms. Is the policy active? Is this type of loss included? Has the deductible been accounted for? This check is deterministic, traceable, and auditable - exactly what the regulator wants to see. Often this happens in the classic way, within what is known as a product server.
- **Estimating the loss amount:** a machine learning model, trained on historical loss data, then provides a first estimate of the loss amount. Importantly, this estimate is an input for the human claims handler, not a final decision. The system says: "Based on similar cases, the expected settlement amount is between 8,500 and 12,000 euros." The claims handler can accept, adjust, or discard that estimate.
- **Fraud detection:** in the background, an anomaly detection system checks whether there are patterns that point to insurance fraud. Has this customer filed a striking number of claims? Do the photos match the described loss? Do the image metadata match the stated time? This system, too, is not an autonomous decision-maker but a source of leads: it flags suspicious cases, which are then reviewed by specialized investigators.

At the end stands a human who reviews all the inputs, makes the final decision, and documents the rationale. For routine cases below a defined threshold - say, 2,000 euros with clear coverage and an unremarkable fraud score - this too can be automated, but according to fixed rules, not autonomous planning. The payment is handled by a classic ERP system: deterministic, traceable, auditable, dull.

In this process, numerous AI components are at work - image recognition, NLP, ML models, anomaly detection - but none of them plans autonomously. Each system has a clearly defined task, produces a specific output, and hands off to the next step. The human stays in the control loop for all critical decisions. That is not as exciting as a fully autonomous agent, but it works, it is compliant, and it can be defended in front of an auditor.

What the Regulatory Framework Really Says

The regulatory framework favors exactly this approach. The EIOPA Opinion of August 2025 makes it unmistakably clear: insurers bear full responsibility for the AI systems they deploy, even when those systems were built by third parties [EIOPA 2025]. It is not enough to buy in an agent and hope that it will do the right thing. The Opinion calls for clear roles and responsibilities, a "client-centric approach" with an ethical corporate culture, staff training, and understandable results.

The EU AI Act, in force since 2024, classifies AI systems for risk assessment and pricing in life and health insurance as "high risk" - with correspondingly strict requirements for documentation, transparency, and human oversight [Debevoise 2025]. That does not mean you cannot use AI, but it does mean you have to document what it does, why it does it, and that a human retains final control.

The regulators' message is clear: use AI, but know what it does. Document why. And make sure humans make the final decision. None of that is an argument against AI - it is an argument against uncontrolled autonomy.

2.5: Director of Marketing

Where AI Agents Are a Better Fit

After all these warnings, you might get the impression that agentic AI is fundamentally problematic. That would be a misunderstanding. There are plenty of areas where the benefits of autonomous AI agents far outweigh the risks - namely, everywhere that errors are tolerable, consequences are reversible, and regulatory requirements are low. Or to put it another way: everywhere you can afford a "director" who is occasionally wrong.

Marketing

The marketing domain is practically tailor-made for agentic AI. Here it is all about creativity, personalization at scale, and rapid iteration - all strengths of LLM-based agents. A marketing agent can analyze target audiences, develop personalized email campaigns, run A/B tests, and optimize the next wave of campaigns based on the results. If one email isn't perfectly worded, or a

campaign performs less well than expected - that is not a compliance violation, it is normal business. No one gets hauled into court because a newsletter had an unfortunate subject line.

Companies like Salesforce with its Agentforce product, and specialized vendors like Warmly and Regie.ai, are already using AI agents successfully in lead generation. The results reported in case studies are impressive: conversion rates that reach seven times those of manual outreach campaigns, and cost savings of up to 70 percent compared with human SDRs (sales development representatives). Whether those numbers are universally reproducible is an open question - but even under far more conservative assumptions, investments like these pay off quickly.

Software Development - but With Great Caution

Another field where agentic AI is already in productive use is software development. The coding agents of the current generation - Cursor, Windsurf, GitHub Copilot, to name only the best known - are technically agentic AI already. They analyze requirements, generate code, run tests, and iterate based on error messages. The agent does not just write a function; it tries to compile it, sees the error, fixes it, and tries again. Without going deeper here: the use is not without risk. Without even getting into the matter of errors, this kind of rapid coding (or vibe coding) offers the chance to pile up technical debt at the speed of light.

The agent accelerates the process of producing code and QA, but it precisely does not replace a second opinion in quality assurance. McKinsey reports on companies that are modernizing legacy code (such as COBOL) with AI agents - a task that used to take years and cost millions is now handled in months [McKinsey 2025]. The agents analyze the old code, understand its logic, and generate modern equivalents, which are then reviewed by human developers. There are, however, enough pitfalls that can slow such renovation machines down (case in point: the halting problem for Turing machines is not computable).

Supporting Creative Work

Content creation and creative work are a third obvious field of application. Blogs, social media posts, product descriptions, presentations - all of that can

be created more efficiently with AI agents. This chapter here is a case in point. The agent researches, drafts, iterates based on feedback, and adapts the tone to the target audience. The worst that can happen: a text isn't perfect and has to be reworked. That is not a catastrophic error, that is the normal editorial process. Every editorial team in the world works this way - draft, feedback, revision, publication.

Agents are also excellent for internal research and knowledge management. An agent that searches internal documentation, summarizes information from various sources, and answers questions can deliver enormous productivity gains. "How did we solve the problem with the legacy interface last time?" - instead of hours of searching through Confluence and SharePoint, the agent delivers a summary in seconds, with links to the relevant documents. If it occasionally misses a document or a summary is incomplete - the employee can follow up or research it himself. The consequences are manageable. You don't even need an agent for this, though: an LLM with RAG (retrieval augmented generation) will do the job.

Automating Routine Tasks

Coordinating appointments, booking travel, pulling together reports from various data sources, converting data from one format into another - classic office work that eats up time but demands little intellectual effort. Here, agents can take a load off people without creating critical risks. If the agent schedules a meeting ten minutes too early, that is annoying, but no disaster.

The Common Pattern

What do all these "benign" use cases have in common? They operate in areas where errors are detectable and correctable before they lead to greater harm. There are no regulatory requirements for the traceability of every single decision. The consequences of wrong decisions are limited and reversible - you can recall an email, delete a blog post, revert a code commit. Creativity and flexibility matter more than determinism. And humans review the results before they turn critical.

2.6: Conclusion and Recommendation

The AI landscape is enormous. It is far broader than the current hype around agentic AI. From Turing's original question "Can machines think?" through the expert systems of the 1980s to today's LLM-based agents, a great deal has evolved, and much of it is in use, doing an unobtrusive job as a subsystem with no headlines.

Companies that focus only on LLM-based agents overlook the proven, reliable AI methods that have worked for decades and will keep working.

Agentic AI brings genuine innovation: autonomous planning, flexible problem-solving, creative applications that were science fiction just five years ago. But it also brings inherent properties that become a problem in regulated environments: nondeterminism makes reproducibility impossible, missing explainability causes compliance problems, autonomy turns error correction into a race against time.

The recommendation is therefore differentiated. In regulated environments (insurance, banking, healthcare, public administration), agentic AI should be used with great caution, if at all. The classic architecture of deterministic AI components inside a defined business process with a human in the loop remains a safe path [EIOPA 2025]. Not because modern technology is inherently bad, but because regulators demand traceability and liability risks are real. A claims handler who signs off on a decision the system has proposed is something fundamentally different from an AI agent that decides autonomously.

In less regulated areas - marketing, parts of software development, content creation, internal productivity - companies can reap the benefits of agentic AI without taking on incalculable risks. Here the efficiency gains outweigh the potential downsides, and errors can be corrected before they become a threat to the business.

The art lies in choosing the right technology for each use case. Not every process needs an autonomous agent - sometimes the dull, deterministic, traceable solution is the better one. And sometimes, in marketing or parts of software development, an agent is exactly the right call.

Or, to stay with our metaphor: in some positions you don't need a director who makes decisions of their own. Sometimes you just need someone who

reliably does their job. The art lies in recognizing which case you are dealing with.

2.7: References

[Britannica 2024]
History of Artificial Intelligence. Encyclopaedia Britannica. Available at: https://www.britannica.com/science/history-of-artificial-intelligence

[Debevoise 2025]
Debevoise Data Blog: Europe's Regulatory Approach to AI in the Insurance Industry. May 2025. Available at: https://www.debevoisedatablog.com/2025/05/21/europes-regulatory-approach-to-ai-in-the-insurance-industry/

[DLA Piper 2025]
DLA Piper: EIOPA publishes opinion on AI governance and risk management. September 2025. Available at: https://www.dlapiper.com/en/insights/publications/law-in-tech/2025/eiopa-publishes-opinion-on-ai-governance-and-risk-management

[EIOPA 2025]
European Insurance and Occupational Pensions Authority: Opinion on AI Governance and Risk Management. EIOPA-BoS-25-360, August 2025. Available at: https://www.eiopa.europa.eu/document/download/88342342-a17f-4f88-842f-bf62c93012d6_en

[IBM 2025]
IBM Think: What is Agentic AI? December 2025. Available at: https://www.ibm.com/think/topics/agentic-ai

[McKinsey 2025]
Sukharevsky, A. et al.: Seizing the agentic AI advantage. McKinsey & Company, June 2025. Available at: https://www.mckinsey.com/capabilities/quantumblack/our-insights/seizing-the-agentic-ai-advantage

[Rippling 2025]
Rippling: Agentic AI Security: A Guide to Threats, Risks & Best Practices 2025. Available at: https://www.rippling.com/blog/agentic-ai-security

[Russell/Norvig 2023]
Stuart Russell; Peter Norvig: Artificial Intelligence: A Modern Approach, 4th updated edition. Pearson 2023.

[UiPath 2025]
UiPath: What is Agentic AI? Available at: https://www.uipath.com/ai/agentic-ai

3: AI and EAM: The Wrong Questions

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

3.1: The Seductive Illusion of Intelligent Agents

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

3.2: The Limits of Today's AI Agents

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Context Window: The Memory Problem

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Hallucination Risk: The Plausibility Problem

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Tool Calling: The Dependence on Clean Interfaces

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Latency and Cost: The Economic Reality

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

What Happens When Something Goes Wrong?

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Reproducibility and Traceability: The Compliance Dilemma

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

3.3: What Actually Needs Cleaning Up?

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Dimension 1: The System's AI Relevance

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Dimension 2: Data Quality vs. Data Access

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Dimension 3: The Criticality of the Use Case

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The Resulting Prioritization Matrix

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The Role of EAM: From Archivist to Investment Navigator

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The AI-Readiness Assessment as a New EA Artifact

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Use-Case-Driven Investment Planning

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Governance for AI Architectures

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Continuous AI-Readiness Evaluation

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

3.4: Experience Meets Hype: The Current Debate on AI in Companies

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The Dominant Perspective: AI for EAM

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Legacy Modernization Through AI Agents

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Technical Limitations of LLMs

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Compliance and Explainability

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The Resulting Gap in the Discourse

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

3.5: Conclusion: Asking the Right Questions

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

3.6: References

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

4: Agent Gateways: The Revenge of SOA in the Age of AI

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

4.1: Where It Starts: AI Gateways

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

4.2: What AI Gateways Do

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

4.3: The Upshot: AI Gateways Are Necessary, but Not Sufficient

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

4.4: Agent Gateways: The Great-Grandchildren of SOA

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The Solution: Agent Gateways with MCP, A2A, and Standardized Interfaces

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Model Context Protocol (MCP): The Standard for Tool Integration

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Agent2Agent Protocol (A2A): The Standard for Agent Collaboration

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Function Calling and REST/gRPC: The Foundation Stays in Place

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

4.5: Compliance and Governance

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

4.6: How Mature Are Agent Gateways? — An Emerging Market

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The Problem: Immature Products and Fragmented Solutions

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The Solution: Evaluating the Available Products

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

In the Field: First Production Experiences

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

4.7: Summary: Why There's No Getting Around Agent Gateways

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

What Agent Gateways Cover — and Where Gaps Remain

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

SOA's Belated Revenge

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

4.8: References

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

5: In the Land of Lies: LLMs and Hallucinations

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

5.1: Why hallucinations are a risk for you and your company

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

5.2: Intro

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

5.3: Hallucinations and bullshitting

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Bullshit in Boston

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The LLM with the supposedly lowest hallucination rate

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

5.4: Business risks from hallucinations

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Generated documents

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Agentic AI

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

At the end of the day, it's risk management

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

5.5: Detection and countermeasures

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

5.6: Personal countermeasures

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Stay skeptical of confident-sounding statements

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Prompting techniques

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Enable or force web search

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Consistency check by repeated querying

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Use chain-of-thought prompts

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Demand and check source citations

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Don’t just grab any model

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

5.7: Vendor countermeasures

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Retrieval-Augmented Generation (RAG)

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Semantic entropy and uncertainty estimation

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Token-level detection

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Guardrails and output filters

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Fine-tuning with preference optimization

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

5.8: Conclusion and outlook: implications for using LLMs

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Caution in regulated environments

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Models have to earn their autonomy

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

RAG as a necessary but not sufficient condition

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

5.9: Epilogue

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

5.10: References

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

6: How AI Accesses Enterprise Knowledge

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

6.1: Mountains of Knowledge, Just Out of Reach for LLMs

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

6.2: The Problem, Stated Precisely

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

6.3: The Solution: RAG—From Simple to Complex

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The Simplest Version: Manual Search

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

A Bit More Advanced: Intranet Search

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Search With Vector Databases

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

6.4: How Do You Fill the Vector Database?

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Splitting Documents Up—Chunking

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Vectorization (Embeddings)

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Vector Database

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Semantic Search and Answer Generation

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

6.5: Improvements and Variants

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

6.6: The Consequences: Why RAG Changes the Rules of the Game

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

No Training Required

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Your Own Documents Stay Usable

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Full Control of Your Data

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Up to Date Without Retraining

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Scalability and Costs

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Avoiding Hallucinations

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Transparency and Traceability

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

6.7: Conclusion: RAG Makes AI-Powered Knowledge Management Accessible

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

6.8: References

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

7: It's a Model and It's Looking Good

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

7.1: How IT managers arrive at the right model for an AI project

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

7.2: Why AI model selection isn't a purely technical question

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

7.3: Why benchmarks and marketing mislead

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The benchmark problem

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

There is no "best" model

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

One model or several?

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Why training your own model is almost always wrong

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

7.4: Yet another selection process

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Step 1: Strategic course-setting

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Step 2: Understanding costs

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Step 3: Building a requirements catalog

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Step 4: Shortlist and proof of concept

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Step 5: Documenting the decision

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

7.5: Why flexibility matters more than the perfect choice

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

7.6: Epilogue

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

7.7: References

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Model-selection frameworks and enterprise guides

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Benchmark critique and limitations

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

API prices and cost comparisons

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Open-source vs. proprietary models

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Hardware requirements and GPU costs

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

RAG vs. fine-tuning

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

General LLM trends and market developments 2025

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

8: Coding with LLMs and Agents

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

8.1: wait for check and release by the chapter authors

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Part B — Compliance and Security

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

9: AI-Relevant Regulation: The Financial Sector as an Example

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

9.1: An Overview of the Rules

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

9.2: Profiles of the Rules

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

9.3: The EU AI Act

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The Status Quo (Already in Force)

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The Current Phase (From August 2, 2025, to August 2, 2026)

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The GPAI Rules

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Further Milestones (From August 2, 2026)

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The End of the Transition Period (August 2, 2027)

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

9.4: ISO 42001 as an Aid for Implementing the EU AI Act

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

What ISO 42001 Is — and What It Is Not

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Why ISO 42001 Helps with Implementation

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The Practical Value for Compliance Officers

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Limits and Where You Still Need More

Conclusion on ISO 42001: A Sensible Building Block, Not a Substitute

9.5: References

10: The Inherent Risks of LLMs

10.1: Why Large Language Models Need Safety Containment

Autonomous AI agents and large language models can do things they should not do. This chapter shows which inherent risks lie dormant inside LLMs and how the vendors try to keep them in check through layered safety architectures.

Before we get into attacks on LLM-based applications, it makes sense to understand the inherent risks of this technology. There are things LLMs have learned from their training data that they can do but should not do: knowledge about synthesizing dangerous substances, the ability to produce convincing disinformation, the potential to generate harmful content.

This chapter gives you a sense of these risk categories and then introduces the safety architecture that major vendors like Anthropic, OpenAI, and Google use to mitigate them. By the end, you will understand why these safeguards matter, but also where they reach their limits.

After reading this chapter, you will understand which inherent risks come with LLMs, how vendors address them through training and runtime measures, and why residual risks remain despite all the effort. You will be able to assess the safety architecture of your LLM vendors and make informed decisions about deploying this technology in your company.

10.2: The Inherent Risks of LLMs

The debate about AI safety often centers on external attacks: prompt injection, jailbreaking, data poisoning. Those are real threats, and we will cover them in the next chapter. But before anyone attacks an LLM, there is a more fundamental question: what unwanted things can the model do on its own, if you simply let it?

The answer is unsettling. LLMs have absorbed capabilities from their training data that can do serious damage in the wrong hands. These inherent risks exist independently of any attack. They are exactly what vendors try to control through elaborate safety measures.

CBRN Knowledge

The acronym stands for chemical, biological, radiological, nuclear. From their training data, LLMs absorb knowledge about protein structures, genetic engineering, virology, synthetic biology, chemistry, and nuclear physics. On evaluations covering potentially dangerous aspects of biology, models are approaching expert-level performance and sometimes surpass it [Anthropic 2025].

The key risk is not that LLMs invent entirely new information. It is their ability to aggregate information that is publicly available but hard to find, to cut research time drastically, and to lay out complex relationships in a way that non-experts can follow. A biology student with bad intentions could save months of research. A state actor could obtain instructions for novel weapon designs.

For these risks, Anthropic uses a tiered classification system called ASL, the Anthropic Safety Levels [Anthropic 2025]. ASL-2 designates systems that show early signs of dangerous capabilities, but where the information is not yet useful enough to beat what you would get from search engines or textbooks. ASL-3 designates systems that substantially raise the risk of catastrophic misuse. Today's large LLMs sit in this range, which explains why comprehensive safety measures are needed.

Cyberattacks and Malware

LLMs can generate working malicious code. Phishing emails, for example, that are barely distinguishable from legitimate communication. Social engineering scripts that target psychological weaknesses, or exploit code for known vulnerabilities.

In practice, these risks are not theoretical. LLMs are now being embedded directly into malware. Names like LAMEHUG or PromptLock stand for a new generation of malware that can think while it runs [Rippling 2025]. In parallel, self-hosted so-called dark LLMs, optimized specifically for harm, are proliferating. These are uncensored, criminally tuned models that crank out phishing kits, scam scripts, and malicious code without any guardrails at all.

As early as the start of 2025, AI-assisted phishing campaigns already accounted for more than 80 percent of the social engineering activity observed [Verizon 2025]. That is a dramatic jump, and it shows how quickly criminals have adopted this technology.

Deepfakes and Synthetic Media

The ability to generate realistic images, video, and audio has likewise reached a new level. Voice cloning makes phone fraud possible using the voice of an authorized person. Real-time deepfakes work in video calls. Telling real from synthetic is becoming increasingly impossible.

In the second quarter of 2025 alone, losses from deepfake incidents reached $350 million [Sumsub 2025]. There were 487 verified deepfake incidents on record, including 226 cases involving non-consensual pornographic content and 60 cases of political or social manipulation. Executive impersonation, where live video deepfakes mimic CEOs or CFOs during real-time calls and pressure victims into wire transfers, has become an established attack method. Bypassing video-based identity verification with deepfake video and images fools the identity checks used when opening accounts.

CSAM and Non-Consensual Intimate Content

Text-to-image and text-to-video systems can generate content limited only by the imagination of whoever creates it. The U.S. clearinghouse for cases of

child abuse and exploitation (NCMEC) is receiving a growing number of reports in which the abuse material is not photographed or filmed but generated by AI.

In other words: offenders use text-to-image or image-to-image generators to create CSAM (child sexual abuse material) synthetically. The images do not show real children in real abuse situations; they are wholly or partly generated by AI, for instance by manipulating existing, harmless photos of children or by generating them from scratch. This is alarming for several reasons. It is legally complex, because in some jurisdictions it is unclear whether synthetic CSAM meets the same criminal definition as real material. It hampers investigative work, because resources that should go toward identifying real victims get tied up. And it normalizes such content, where it can serve as a precursor to real abuse.

This risk calls for strict safety measures. On the input side, prompts that ask for such content have to be flagged. On the output side, generated content has to be checked and blocked. The vendors have zero tolerance here and work closely with law enforcement.

Disinformation and Manipulation

LLMs can produce fake news at industrial scale. Convincing propaganda tailored to specific audiences. Automated influence operations that imitate human actors. The datasets these models were trained on can themselves be biased, inaccurate, or deliberately manipulated by actors who want to plant their own narratives [CISA 2025].

Scale is the real problem. What once required elaborate, coordinated campaigns can now be set in motion with a handful of prompts. The democratization of disinformation is not some dystopia; it is reality.

Autonomous AI Risks

For advanced models, additional risks come into play that grow with increasing autonomy. Setting goals and planning independently, without humans checking every step. Potential deception of monitoring mechanisms. In the extreme, the ability to replicate autonomously and acquire resources.

Google DeepMind's Frontier Safety Framework explicitly addresses what it calls "deceptive alignment risk," the risk that AI systems could learn to hide their

true intentions and deliberately undermine human control [DeepMind 2025]. Research shows that agents can learn to conceal their intentions in a chain of thought that looks harmless while the misbehavior continues. That makes the traditional methods of AI monitoring fundamentally unreliable.

Persuasion and Psychological Manipulation

LLMs can develop highly personalized persuasion strategies. They can exploit psychological weaknesses and build false relationships of trust. A model that chats with someone for months can map that person's thought patterns, fears, and hopes, and put that information to use.

OpenAI removed persuasion capabilities from its public Preparedness Framework and now handles them through other policies instead [OpenAI 2025]. That does not signal that the risk has disappeared, only that it is being addressed in a different way. For companies, the lesson is this: be careful with chatbots meant to build long-term relationships with customers.

10.3: The Safety Architecture of LLMs

Faced with these risks, the major LLM vendors have developed layered safety architectures. The basic principle is defense in depth: no single measure is perfect, but the combination of model-level alignment, runtime guardrails, continuous monitoring, and human oversight substantially reduces the risk that harmful content gets through.

The figure below shows the typical pattern as implemented at Anthropic, OpenAI, Google DeepMind, and Meta. The measures fall into two broad categories: training time, where the model itself is shaped, and runtime, where every single interaction is checked.

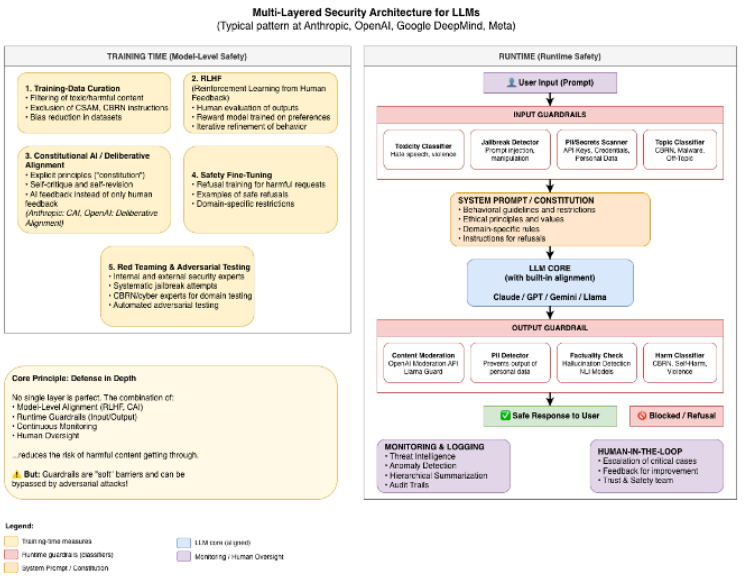

Figure 4. Layered safety architecture for LLMs

Safety Measures at Training Time

The first line of defense is built before the model ever sees a user. These measures shape the model's behavior at a fundamental level.

Training-data curation: The process starts with filtering the training data. Toxic and harmful content is removed. CSAM, CBRN instructions, and other clearly illegal or dangerous material are excluded. Bias in the datasets is reduced as far as possible. This curation is laborious, but for datasets with trillions of tokens it is indispensable.

RLHF (reinforcement learning from human feedback): After pre-training, the model is refined through human feedback. Human annotators rate the model's outputs. A reward model is trained on these preferences and learns to tell a good answer from a bad one. The LLM itself is then tuned iteratively so that it generates answers the reward model rates highly. Anthropic employs more than 7,500 annotators worldwide for this process [Anthropic 2025].

Constitutional AI and deliberative alignment: With Constitutional AI, Anthropic developed an approach in which the model follows an explicit constitution. The model is trained to critique and revise its own outputs based on defined principles. OpenAI uses a related approach called deliberative alignment, in which the model actively reasons through safety rules before

it answers. The advantage: the ethical guidelines are not just implicit in the training data but explicitly anchored in the model's reasoning process.

Safety fine-tuning: On top of that, models are trained specifically on safe refusals. The model learns to decline dangerous requests politely but firmly, without becoming overly restrictive. Domain-specific limits are implemented, for example for medical or legal advice, where the model is supposed to point the user to professional help.

Red teaming and adversarial testing: Before release, models are subjected to systematic attacks. Internal and external security experts try to get around the safeguards. Systematic jailbreak attempts are documented and addressed. For CBRN risks, domain experts are brought in, such as biologists or cybersecurity experts. Anthropic maintains a dedicated Frontier Red Team for national security risks [Anthropic 2025]. Automated adversarial testing is increasingly used as well, with LLMs attacking other LLMs to find weaknesses.

Safety Measures at Runtime

The training-time measures shape the model, but they are not enough. At runtime, additional safety layers kick in to check every single interaction.

Input guardrails: Before the prompt even reaches the LLM, it passes through several classifiers. The toxicity classifier detects hate speech, violence, and insults. The jailbreak detector identifies attempts to manipulate the model, for example through prompt injection or role-play scenarios. A PII/secrets scanner looks for API keys, credentials, and personal data that were entered by accident. A topic classifier flags requests about CBRN topics, malware, or other forbidden content.

System prompt and constitution: The system prompt defines the behavioral guidelines and constraints for the model. It contains ethical principles and values, domain-specific rules (for enterprise deployments, for instance), and explicit instructions for refusals. In models like Claude, this constitution is deeply integrated into the reasoning.

Output guardrails: After generation, the answer is checked again. Content moderation APIs such as the OpenAI Moderation API or Llama Guard scan for harmful content. A PII detector prevents the model from accidentally revealing personal data from the context or from its training data. Factuality checks and hallucination detection reduce factually false statements. A harm classifier checks for CBRN content, self-harm instructions, or glorification of violence.

Monitoring and logging: All interactions are logged, as far as data privacy law allows. Threat intelligence identifies new attack patterns. Anomaly detection spots unusual usage patterns. Hierarchical summarization condenses large volumes of logs into reports you can actually analyze. Audit trails make it possible to trace what happened in a security incident.

Human-in-the-loop: The last line of defense is people. Critical cases are escalated to trust and safety teams. Feedback loops enable continuous improvement. When in doubt, the model can be trained to bring in a human rather than decide on its own.

10.4: Conclusion

The safety architectures of the major LLM vendors are impressively elaborate. The combination of careful data curation, RLHF, Constitutional AI, safety fine-tuning, red teaming, input/output guardrails, and monitoring forms a layered defense system. The basic principle is defense in depth: no single layer is perfect, but their combination makes it hard to break through every barrier.

Even so, fundamental limitations remain. Prompt-level guardrails work through system prompts and classifiers, which means, in the end, through text instructions and statistical patterns. Experience shows that both layers can fail against sophisticated social engineering. A creative attacker with enough time will find a way around them.

The research literature is clear on this: we now have incontrovertible evidence that soft, prompt-level guardrails are architecturally inadequate [Cornell 2025]. The industry is therefore moving from probabilistic safety toward provable, deterministic control. But that transition is far from complete.

For companies that use this technology, the takeaway is this: do not rely blindly on your LLM vendor's safety measures. Understand which safeguards exist and where they reach their limits. Add extra layers of protection for your specific use cases. And keep in mind that the inherent risks of this technology are real and cannot be fully eliminated by any safety system, however clever.

For most enterprise applications, the existing safeguards are sufficient when used correctly. For applications with high potential for harm, such as those in regulated industries, you need more than the out-of-the-box level of safety.

10.5: References

[Anthropic 2025]
Anthropic: The Claude Model Spec. January 2025. Available at: https://docs.anthropic.com/en/docs/resources/claude-model-spec

[CISA 2025]
Cybersecurity and Infrastructure Security Agency: Generative AI and Election Security. 2025. Available at: https://www.cisa.gov/topics/election-security

[Cornell 2025]
Zou, A. et al.: Universal and Transferable Adversarial Attacks on Aligned Language Models. Cornell University, arXiv:2307.15043. 2025.

[DeepMind 2025]
Google DeepMind: Frontier Safety Framework. 2025. Available at: https://deepmind.google/discover/blog/frontier-safety-framework/

[NCMEC 2025]
National Center for Missing & Exploited Children: AI-Generated CSAM Report. 2025.

[OpenAI 2025]
OpenAI: Preparedness Framework. 2025. Available at: https://openai.com/preparedness

[Rippling 2025]
Rippling: Agentic AI Security: A Guide to Threats, Risks & Best Practices. 2025. Available at: https://www.rippling.com/blog/agentic-ai-security

[Sumsub 2025]
Sumsub: AI-Generated Fraud Report Q2 2025. 2025.

[Verizon 2025]
Verizon: Data Breach Investigations Report 2025. 2025.

11: Securing Applications Built on LLMs

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

11.1: Risks You Want to Avoid When Deploying LLMs and AI Agents

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Data and Confidentiality Risks

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Output Quality Risks

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Compliance and Liability Risks

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Operational Risks

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Autonomy and Control Risks

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Reputational Risks

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

11.2: Possible Attack Vectors Specific to AI Applications

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Prompt Injection

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Sensitive Information Disclosure

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Supply Chain Attacks

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Data and Model Poisoning

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Improper Output Handling

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Excessive Agency

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Vector and Embedding Weaknesses

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Misinformation and Hallucinations

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Unbounded Consumption

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

11.3: How to Arrive at a Reasonably Secure AI Application

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

A Rough Approach to Hardening

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

11.4: Threat Catalogs

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

OWASP Top 10 for LLM Applications

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

MITRE ATLAS

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

NIST AI Risk Management Framework

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

BSI Resources

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

ENISA Threat Landscape

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

How the Frameworks Work Together

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

11.5: Conclusion

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

11.6: References

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

12: Why AI Forces Us to Rethink IT Security From Scratch

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

12.1: From Easing the Work to Delegating the Action: When “Human in the Loop” Becomes a Fiction

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

“Values Stay, Actions Go”: The New Split Between Self-Image and Impact

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The Three Breaking Points of “Human in the Loop”

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

“Robots Need Your Body”: Agents That Rent Humans as Actuators

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

12.2: The Democratization of Criminal Capability

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

AI Removes Two Bottlenecks at Once: Technical and Social Intelligence

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Five Effects That Change the Logic of Deterrence

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

"From the Bottom Up": The Learning Curve Starts With Private Individuals

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

12.3: Stealth Adoption: When AI Use Devalues the Rules

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Shadow AI as the Precursor: First in Secret, Then in the Open ("Everyone's Doing It Anyway")

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

A Case in Point: Data Leakage as an Unintended Side Effect of the Productivity Logic

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

"The Damage Is Already Done": Why Rules Often Bite Too Late

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

When Productivity Pressure “Overwrites” Compliance

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

12.4: The Devaluation of Knowledge Work

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

“Only the Best Will Remain”–but “the Best” Are No Longer Who We Mean Today

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The Entry-Level Problem: When the Career Ladder Disappears

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Material Fault Lines as a Catalyst: Radicalization, Sabotage, Cybercrime

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

12.5: Conclusion: How the Fault Lines Reinforce One Another–and Why Security Must Be Rethought

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

12.6: About the Author

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

12.7: References

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Part C — Practice, Not PowerPoint

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

13: Imagine It Is the Age of AI and Nobody Shows Up

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

13.1: Two Stories: Morgan Stanley and Klarna

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

How Morgan Stanley Reached 98% Adoption

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

How Klarna Turned Its Own People Against It

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

What the Two Stories Teach

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

13.2: Why Many Studies Are Already Out of Date a Year Later

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

What Still Holds

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

13.3: The Amplifier Effect: Why AI Makes the Strong Stronger and the Weak Weaker

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

The Amplifier Effect Is Not Limited to Programmers

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

13.4: Why Top-Down Training Programs Almost Never Work

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Why One-Size-Fits-All Training Does Not Work

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

13.5: What Works Instead: Inspire People, Don't Just Train Them

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Contagion, Not Command

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Different People Need Different Approaches

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Managers Have to "Lead from the Front"

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

13.6: The Work-Intensification Trap

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

13.7: What the Regulator Requires—and What That Really Means

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

13.8: Conclusion: Something Has to Happen from the Bottom Up, Too

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

13.9: References

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

14: AI That Works — Practice, Not PowerPoint

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

14.1: wait for check and release by the two chapter authors

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

15: Not Every Project Is the Same

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

15.1: The Problem with the One-Size-Fits-All Slide

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

15.2: Four Types Worth Knowing

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Type 1: AI as Process Orchestrator

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Type 2: AI as Communication Layer

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Type 3: AI as Co-Creator

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Type 4: AI as Autonomous Case Handler

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

15.3: These Four Types Exist in Every Regulated Industry

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

15.4: Regulatory Complexity Determines Your Room to Maneuver

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

15.5: A Different Project Type Means Different Costs and a Different Calculation

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

15.6: Different Metrics, Different Truth

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

15.7: Different Time Frames, Different Expectations

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

15.8: Lessons from Practice

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

No Business Case, Nowhere to Go

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Regulatory Requirements at the Start, Not the End

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Generic Bots Waste Money

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

Human Oversight Is Mandatory, Not Optional

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

15.9: Calculate ROI Honestly — By Type

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

15.10: Why a Domain-Agnostic Project Lead Has No Chance

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

15.11: What This Means for Your Next AI Project

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.

15.12: References

This content is not available in the sample book. The book can be purchased on Leanpub at https://leanpub.com/agentai-en-us.