Understanding AI Agents

From Chatbot to Autonomous Assistant

A practical guide for decision-makers and teams

Reto Grau Consulting GmbH — March 2026

Introduction

Artificial intelligence has evolved rapidly in recent years. While most people know AI as a chatbot — you ask a question and get an answer — the next generation goes far beyond that: **AI Agents**.

An AI agent is not a better chatbot. It is a digital coworker that independently completes tasks, uses tools, learns from mistakes, and executes complex workflows across multiple steps.

This guide explains the concepts behind AI agents — clearly, practically, and without unnecessary jargon.

1. The Agent Loop — How an AI Agent Works

The Core Principle

Every AI agent follows a simple yet powerful cycle of three steps:

👁️ **Observe** — The agent reads its context: What is the task? What information is available? What has already been done?

🧠 **Think** — The agent analyzes the situation, weighs options, and creates a plan. This is where the language model's intelligence comes into play.

⚡ **Act** — The agent performs a concrete action: reading a file, sending an email, running a web search, or creating a report.

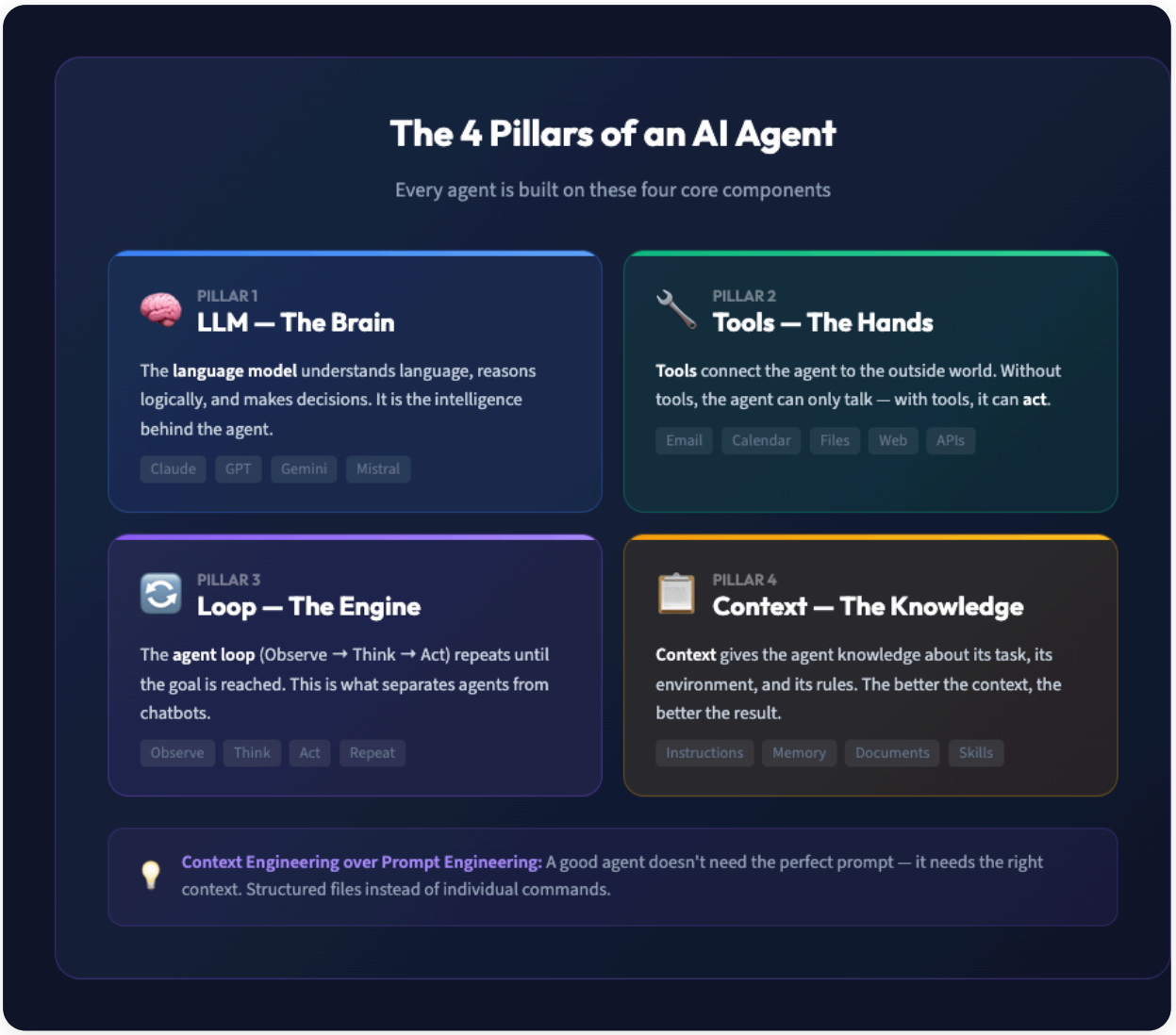
The Critical Difference

This cycle **repeats** until the goal is achieved. This is the fundamental difference between a chatbot and an agent:

	Chatbot	Agent
Flow	Question → Answer → Done	Goal → Loop → Loop → Result
Steps	A single one	Multiple, iterative
Context	None (each question is new)	Built up across the entire task
Tools	None	Actively uses tools
Memory	None	Learns and remembers preferences

An agent doesn't work reactively (question-answer), but **proactively** — it pursues a goal and works toward it independently.

2. The Four Pillars of an AI Agent



Every AI agent consists of four core components. If any one is missing, the system is incomplete.

Pillar 1: LLM — The Brain 🧠

The **Large Language Model** (LLM) is the intelligence behind the agent. It understands natural language, can reason logically, compose text, and make decisions.

Well-known models include Claude (Anthropic), GPT (OpenAI), Gemini (Google), and Mistral. Each has strengths and weaknesses — the right model depends on the use case.

Important: The LLM alone is just a language model. Only in combination with the other three pillars does it become an agent.

Pillar 2: Tools — The Hands 🔧

Tools (also called MCPs — Model Context Protocol) connect the agent to the real world. Without tools, an agent can only generate text. With tools, it can:

- **Read and send emails**
- **Create, edit, and organize files**
- **Check calendars and schedule meetings**
- **Search websites and extract information**
- **Query and update databases**
- **Call APIs** (CRM, ERP, accounting, etc.)

The number and type of available tools determine what an agent can accomplish. A well-equipped agent has access to all relevant systems of an organization.

Pillar 3: Loop — The Engine 🔄

The **Agent Loop** (Observe → Think → Act → Repeat) is the heart of the system. It enables the agent to break complex tasks into manageable steps and work iteratively toward the goal.

An example: The task «Create a market report» requires multiple loops:

1. **Loop 1:** Search news, identify relevant articles
2. **Loop 2:** Retrieve market data (prices, indices, trends)
3. **Loop 3:** Perform analysis, identify connections
4. **Loop 4:** Write report, format, submit for approval

Without the loop, the agent would be done after the first step — like a chatbot.

Pillar 4: Context — The Knowledge 📋

Context is the knowledge the agent needs to complete its task correctly. This includes:

- **Who it is** — Its role, tone, boundaries
- **Who it helps** — Background, preferences, goals of the user
- **How it works** — Rules, workflows, priorities
- **What it knows** — Learned information, past decisions, domain knowledge

The better the context, the better the result. This leads us to the next concept.

3. Context Engineering — The Art of the Right Context

From Prompt Engineering to Context Engineering

The first generation of AI usage was based on **Prompt Engineering** — the attempt to craft the perfect input to get the desired output. This works for simple tasks but quickly reaches its limits.

Context Engineering goes a step further: Instead of individual prompts, you create **structured files** that are permanently available to the agent. The agent reads these files at every start and immediately knows who it is, what it can do, and how it should work.

The Agent Files

File	Question	Purpose
IDENTITY.md	Who am I?	Name, personality, tone
USER.md	Who do I help?	Background, user preferences
AGENTS.md	How do I work?	Rules, workflows, guidelines
MEMORY.md	What do I know?	Learned information, decisions
skills/*.md	What can I do?	SOPs — step-by-step instructions
HEARTBEAT.md	What do I check?	Periodic tasks and monitoring

Why Is This Better Than a Prompt?

1. **Persistent** — Context is preserved, even when the agent restarts
2. **Structured** — Different aspects are organized in separate files
3. **Extensible** — New knowledge is simply written to the right file
4. **Maintainable** — Changes in one place take effect immediately
5. **Shareable** — Teams can collaborate on context files

The context IS the instruction. No single prompt can match the depth and breadth of a well-structured file system.

4. The Self-Improving Loop — How Agents Learn from Mistakes

The Principle of Continuous Improvement

An AI agent makes mistakes — especially in the beginning. That's normal. What separates a good agent from a bad one is the ability to **learn from corrections and never repeat the same mistake**.

The Process

1. **Agent makes a mistake** — Wrong approach, forgotten preference, or false assumption.

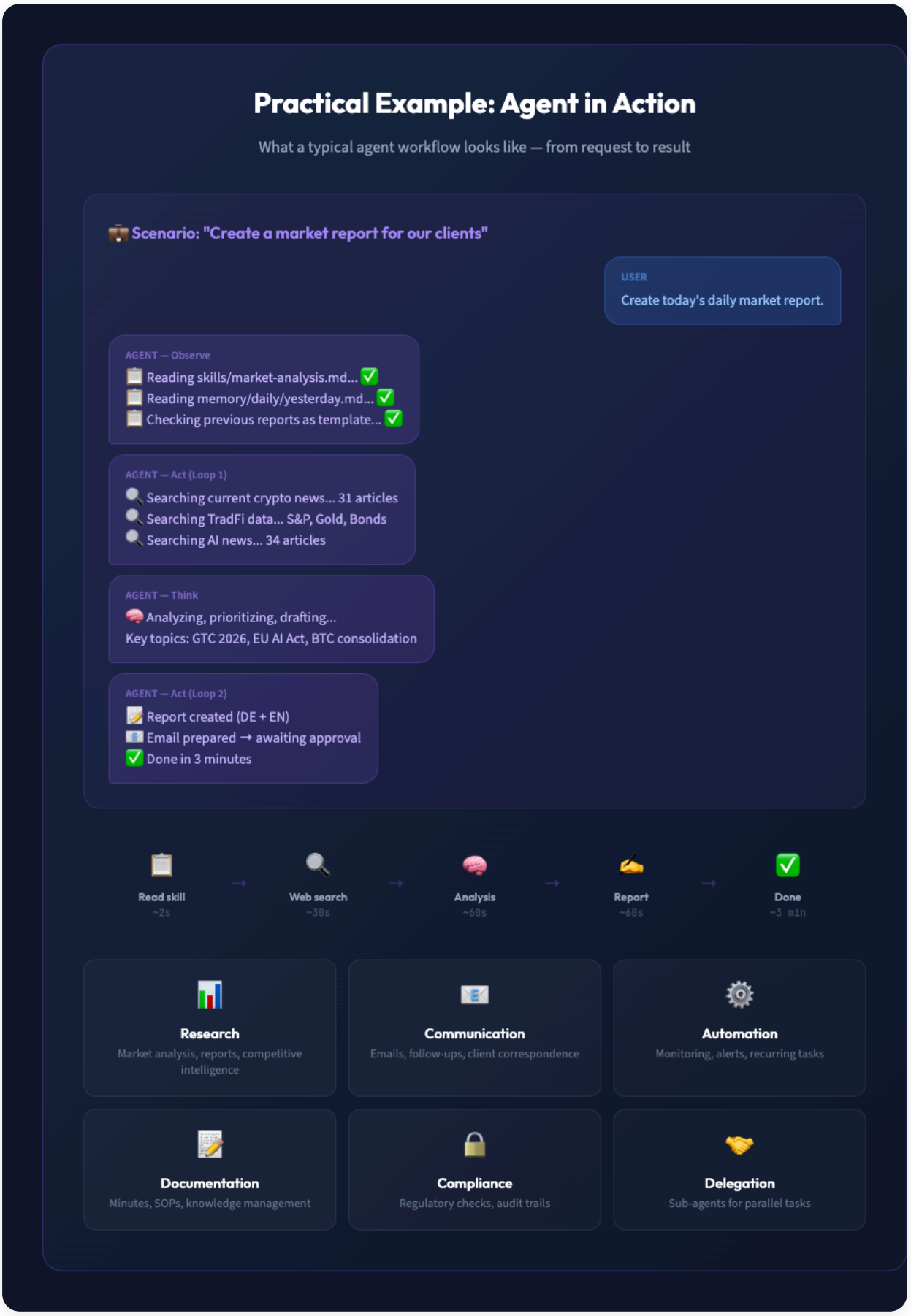
2. **User corrects** — "Not like that, like this." Simple feedback is enough.
3. **Agent analyzes the root cause** — Instead of just fixing the error, the agent asks: *Why* did this happen? Wrong assumption? Missing knowledge? Unclear rule?
4. **Agent updates its files** — The right context file is permanently changed:
    - Wrong behavior → Update **AGENTS.md**
    - Wrong knowledge → Update **MEMORY.md**
    - Wrong preference → Update **USER.md**
5. **Mistake never repeats** — On the next start, the agent reads the updated file. The system becomes permanently better.

The Result

Over time, the number of corrections drops dramatically. After a few weeks, the agent knows the preferences, rules, and quirks of its user so well that corrections become the exception.

This principle is comparable to a new employee: In the first week, they need a lot of guidance. After a month, they work largely independently. The difference: An AI agent never forgets what it has learned.

5. Practical Example: An Agent in Action



Scenario: Daily Market Report

A company uses an AI agent for daily market report creation. The entire process takes less than 3 minutes — fully automated.

Step 1 — Load Context (~2 seconds) The agent reads its "Market Analysis" skill file and knows exactly which sources to use, what format the report should have, and which topics are relevant.

Step 2 — Research (~30 seconds) The agent automatically searches news sources: crypto news, TradFi data (S&P 500, gold, bonds), AI news. It collects over 100 current articles.

Step 3 — Analysis (~60 seconds) The collected information is weighted, categorized, and put into context. The agent identifies the 3-5 most important topics of the day.

Step 4 — Create Report (~60 seconds) The report is written in two languages (German and English), formatted, and generated as a PDF.

Step 5 — Delivery The report is prepared for email and awaits approval from the responsible person. After confirmation, delivery happens automatically.

Use Cases for AI Agents

Area	Examples
Research	Market analysis, competitive comparisons, due diligence
Communication	Email triage, follow-ups, client correspondence
Automation	Monitoring, alerts, recurring reports
Documentation	Minutes, SOPs, knowledge management
Compliance	Regulatory checks, audit trails, data privacy
Delegation	Sub-agents for parallel tasks

Summary

AI agents are the next level of artificial intelligence — from passive chatbot to active digital coworker. The key concepts:

1. **The Agent Loop** — Observe, Think, Act, Repeat
2. **The Four Pillars** — LLM, Tools, Loop, Context
3. **Context Engineering** — Structured files instead of individual prompts
4. **Self-Improving Loop** — Learn from corrections, never repeat mistakes

The key to success lies not in technology alone, but in the **right configuration and the right context**. A well-configured agent can become an indispensable team member within a few weeks.

About Reto Grau Consulting

Reto Grau Consulting GmbH supports companies and organizations in adopting and using AI agents — from strategy through implementation to ongoing operations.

Our focus areas:

- AI strategy and implementation
- On-premise AI solutions (data sovereignty)
- Agent development and training
- Compliance and data privacy (CH/EU)

📧 kontakt@reto-grau-consulting.ch 🌐 reto-grau-consulting.ch
