ACF IN ACTION
acfstandard.com

ACF en Action applique l'Agentic Commerce Framework® à des entreprises réelles, à partir de sources publiques. Chaque cas montre qui reste responsable quand l'IA décide.

ACF-CS-002 · Case Study #002 · Édition 2026

Klarna - AI-automated customer service

What happens when an AI agent crosses the decision threshold, at scale?

CARTE D'IDENTITÉ

SECTOR	COUNTRY	TECHNOLOGY	TYPE	AUTONOMY	CONFIDENCE
Fintech / payments	Sweden	OpenAI	Agentic AI	supervised autonomous	Medium

OUTILS ACF ILLUSTRÉS DANS CE CAS

- ● ACF-01
- ● ACF-02
- ● ACF-03
- ● ACF-06
- ● ACF-00
- ● ACF-05

● mobilisé · ● à prévoir · 6 outils du Toolkit (4 mobilisés, 2 à prévoir, sur 17).

1. Executive summary

In February 2024, Klarna rolled out worldwide an AI customer service assistant built with OpenAI. Within one month, it handled 2.3 million conversations, roughly **two-thirds** of customer service exchanges, for work presented as the equivalent of **700** full-time agents. Unlike Morgan Stanley (case ACF-CS-001), this agent does not merely propose: it **acts** - refunds, cancellations, disputes.

Then, in 2025, Klarna **rebalances**: management publicly acknowledges that cost reduction had "taken too much space" at the expense of quality, and reintroduces human agents for complex cases, in a hybrid model.

Read through the lens of ACF, this case is the exact mirror of the previous one. Morgan Stanley illustrated prudence below the agentic threshold; Klarna illustrates what plays out **above** it: as soon as an agent decides alone at scale, the question of accountability and criticality becomes immediate, and optimizing it on cost alone is a governance failure mode.

2. Context

Klarna, a player in installment payments ("buy now, pay later"), runs a very high volume, multilingual, 24/7 customer service. The stated goal: to drastically reduce cost and handling time. The deployment is part of a partnership with OpenAI.

3. Public sources used

- Klarna, press release: "Klarna AI assistant handles two-thirds of customer service chats in its first month" (Feb. 27, 2024).
- OpenAI, Klarna page.
- Forbes (March 13, 2024) on the results; Forbes (May 18, 2025) on the rebalancing.
- Bloomberg / Reuters (May 2025), reported: statements by CEO Sebastian Siemiatkowski and the return of human agents.
- CX Dive, eMarketer (2025) on the hybrid model.

4. Reconstructed AI architecture (from public sources)

The Klarna AI assistant - global launch February 2024, backed by OpenAI.

- Function: handles customer service requests **autonomously**, including refunds, returns, payment problems, cancellations, disputes, billing errors; provides balances and payment schedules. More than 35 languages, 24/7.
- Volume: 2.3M conversations in the first month (~2/3 of exchanges), claimed equivalent of ~700 agents.
- Claimed performance (2024): resolution in under 2 min versus ~11 min for a human; -25% repeat requests; ~$40M profit targeted for the year; satisfaction "on par" with human agents.

The rebalancing (2025).

- Klarna reintroduces human agents for **complex and premium** cases, while keeping AI on the front line for the bulk of the volume.
- Public statement by the CEO (May 2025): cost had been a "too predominant" criterion, which "results in lower quality"; customers must always be able to reach a person.
- New model: hybrid, flexible human agents ("Uber-style"), AI up front, human on escalation.

5. ACF mapping

🧰 **Tools used:** ACF-01 (autonomy map) · ACF-02 (criticality matrix) · ACF-04 (agent card).

ACF element	In the Klarna deployment
Named human (accountable)	Customer service / operations leadership. Unlike MS, no human professional validates each decision: the agent decides alone.
Agent	The AI assistant - it **decides and executes** (refund, cancellation, dispute resolution).
Platform	OpenAI.
Delegated decision	Real and with consequences: refunds, cancellations, disputes.
Autonomy level	**Supervised autonomous**, even **autonomous** on part of the cases (the agent acts, control after the fact). This is the crossing of the threshold.
Non-delegable zone	Discovered *after the fact*: complex and premium cases, where human parity did not hold, reverted to human. In ACF language, this was a non-delegable zone not declared at the outset.
Rule / safeguard	Human escalation introduced in 2025; option to reach a person. Initially undersized.
Control third party	Credit and consumer protection regulators depending on the country.
Traceability	Massive volume of automated decisions: the stake of a register and auditability is central.

A schema in ACF graphic language accompanies this case (separate file).

6. Analysis

> 🧰 **Tools used:** ACF-02 (criticality) · ACF-03 (non-delegable zones) · ACF-06 (kill switch). **To prepare:** ACF-08 (register).

Strengths (to acknowledge):

- **Real scale and speed**: the deployment sustained a considerable volume, in 35 languages, with measured time savings.
- **Transparency and self-correction**: Klarna publicly acknowledged the imbalance and adjusted. This is a healthy governance posture - rare and to be commended.
- The **hybrid model** ultimately adopted (AI on volume, human on complexity, escalation always possible) is close to what ACF recommends.

Points of attention (the ACF lesson):

- **The optimization criterion was cost, not criticality.** This is the heart of the matter, and it is the CEO who says so. In ACF, you do not delegate a decision because it is *costly to handle*, but according to its **criticality** (impact x frequency x irreversibility). A premium dispute and a balance inquiry do not have the same criticality and do not call for the same autonomy.
- **The non-delegable zone was discovered through the incident, not declared by design.** ACF would have set it from the start: high-impact cases stay human, or under co-decision.
- **Human escalation was an undersized safeguard.** ACF makes it an explicit rule, proportionate to criticality, from the outset.

7. ACF toolkit used

Tools that the ACF analysis uses (✅) and to anticipate (➡ SOON). It does not imply that the company uses ACF; it shows which instruments the framework would bring.

Tool	Status	Role in this case
ACF-02 · Criticality matrix	✅	The key tool: sort by criticality, not by cost
ACF-01 · Autonomy map	✅	Set the dial by type of decision
ACF-03 · Agentic constitution	✅	Declare the non-delegable zones (complex cases)
ACF-06 · Kill switch	✅	Quality thresholds and escalation rate
ACF-08 · Decisions register	➡ SOON	Trace decisions with consequences
ACF-00 · Sovereignty Score	➡ SOON	Assess overall maturity

8. Governance questions

1. On what criterion do we decide to delegate a decision: its cost, or its criticality?
2. Which categories of cases do we declare non-delegable **before** deploying?
3. Is the human escalation mechanism sized for the real criticality, or symbolic?
4. Which signals (satisfaction, error rate, escalation rate) trigger a takeover, and who decides it?
5. Do we have an auditable trace of the automated decisions with consequences?

9. General lessons

The Klarna case delivers the most transferable lesson of the whole corpus: **you do not set an agent's autonomy on cost, you set it on criticality.** Optimizing delegation on the savings achieved, without a criticality matrix or declared non-delegable zones, leads to delegating decisions that should have stayed human, and to discovering it through the drop in quality. Klarna had the lucidity to correct; ACF makes it possible not to have to correct, by setting the dial from the design stage.

10. What ACF would do (reference design)

Prescriptive and non-critical formulation. An organization designing this system with ACF could in particular:

- **rate** each type of request on the criticality matrix (ACF-02) before any deployment;
- **set** autonomy by type of decision (ACF-01): autonomous for the benign, co-decision or human for the sensitive;
- **declare** complex/premium cases as a non-delegable zone (ACF-03) - or under reinforced supervision;
- **size** human escalation to criticality, not to the symbol;
- **arm** a kill switch (ACF-06) triggered by quality and escalation thresholds;
- **keep** a decisions register (ACF-08) for auditability and compliance.

11. Similar cases (transferability)

This case is particularly transferable to any organization with **very high volume customer service**, where the temptation to automate on cost is strong:

- ✅ e-commerce and marketplaces
- ✅ telecoms
- ✅ energy and consumer utilities
- ✅ transport and travel
- ✅ retail banking and payment
- ✅ insurance (mass claims handling)

Everywhere, the same question: **which decision can really be entrusted to an agent, and by what criterion?** Cost is a poor judge; criticality is the right one.

12. Confidence of the assessment

High. The case is exceptionally well documented publicly: company press releases, CEO statements in the press (Bloomberg, Forbes) and reference business coverage, on the deployment as well as its rebalancing. The facts are solidly established and verifiable by the reader. The ACF analysis draws on these public facts; the non-public internal details do not affect the governance reading proposed here.

Teacher guide

Pedagogical objective. Understand that the delegation criterion is **criticality**, not cost - and why a non-delegable zone is declared before, not after.

Session outline (90 min).

1. (15 min) Reading up to §4 (without the analysis). Students first see the 2024 success.
2. (15 min) Reveal of the 2025 rebalancing and the CEO's quote.
3. (25 min) In groups: rate five types of requests (balance, simple return, refund, dispute, premium complaint) on a criticality matrix, and assign an autonomy level to each.
4. (20 min) Debate: "Did Klarna fail, or learn?" (expected answer: it set the wrong dial, then corrected it).
5. (15 min) Synthesis by the teacher: cost vs criticality.

Recommended comparison. Study this case **paired with ACF-CS-001 (Morgan Stanley)**: below the agentic threshold vs above it. It is the best pedagogical duo in the corpus.

Reminder: the reference analysis provides the benchmark; grading students' work remains the teacher's act.

Associated lab (ACF-LAB-002)

From this case, the student produces:

- the criticality matrix (ACF-02) of six types of requests;
- the corresponding per-decision autonomy map (ACF-01);
- an agentic constitution (ACF-03) declaring the non-delegable zones;
- a kill switch (ACF-06) with two quantified triggers (satisfaction threshold, escalation rate).

Appendix - Completed ACF cards

The official Toolkit cards used in this case, filled in with its public data. For illustration: an organization of the same sector can reuse and adapt them.

AGENTIC COMMERCE FRAMEWORK®

ACF-01

DECISION MAP

Map your decisions and their autonomy level

LEVEL	Foundation
USE	Mapping
CASE	ACF-CS-002
FILLED	Klarna

OBJECTIVE Map Klarna's customer service decisions and set, for each, the autonomy level - by criticality, not by cost.

1 AGENTIC MATURITY SCALE

0 — Classic automation
Fixed rules, no autonomy.

1 — Assisted agents
The agent proposes, the human decides.

2 — Governed agents
The agent acts within a frame, under supervision.
TARGET ~80%

3 — Supervised autonomy
The agent decides alone; control after the fact.

2 YOUR KEY DECISIONS (filled)

	A Balance / info question	B Standard refund	C Dispute / premium claim
STAKE	Answer a simple request	Refund under a cap	Decide a high-impact case
IMPACT	Reversible, no commitment	Bounded financial commitment	Commits the relationship and accountability
DEADLINE	Continuous	Continuous	-
AUTONOMY	**Autonomous**	**Supervised autonomous + cap**	**NON-DELEGABLE**

Reading: it is this setting, made **before** deployment, that would have avoided entrusting the premium cases to the agent.

Card ACF-01 · Decision map · completed with public case data.

AGENTIC COMMERCE FRAMEWORK®

ACF-02

CRITICALITY MATRIX

Rate each decision: Impact × Frequency × Irreversibility

LEVEL	Foundation
USE	Rating
CASE	ACF-CS-002
FILLED	Klarna

OBJECTIVE Sort requests by **criticality**, never by cost. This is the key tool of the case: it separates what an agent can handle alone from what must stay human.

1 THE FORMULA

Criticality = (Impact × Frequency × Irreversibility) ÷ 10

Thresholds: **< 15 low** · **15-25 moderate** · **> 25 high**

2 THE RATED REQUESTS (filled)

Request type	I	F	Ir	Criticality	Zone / autonomy
Balance / status question	2	9	1	**1.8**	**Low** - Autonomous
Delivery tracking	3	8	2	**4.8**	**Low** - Autonomous
Standard return	4	7	3	**8.4**	**Low** - Supervised auto.
Refund under cap	6	6	5	**18**	**Moderate** - Auto. + cap
Payment dispute	7	4	7	**19.6**	**Moderate** - Co-decision
Premium claim / fraud	9	5	8	**36**	**High** - Non-delegable

Routing by **cost** would have treated these six requests the same way. Routing by **criticality** isolates the premium claim (36) as non-delegable: exactly the line Klarna had to reintroduce in 2025.

Card ACF-02 · Criticality matrix · completed with public case data.

AGENTIC COMMERCE FRAMEWORK®

ACF-03

AGENTIC CONSTITUTION

The agent's rules, limits and responsibilities

LEVEL	Foundation
USE	Framing
CASE	ACF-CS-002
FILLED	Klarna

OBJECTIVE Define the agent's constitution: identity, standing principles and forbidden zones. Card filled with public case data.

1 AGENT IDENTITY & MISSION

AGENT NAME
Klarna AI customer service assistant

ACCOUNTABLE (DDAO)
To be designated

PRIMARY MISSION
Handle high-volume customer service requests, 35+ languages

SCOPE
Balance, returns, refunds under cap; excluding premium disputes

2 CORE PRINCIPLES

1. Sort requests by criticality, never by cost
2. Guarantee a human escalation available at all times
3. Cap refunds and commitments
4. Signal to the customer that they are talking to an AI

EXAMPLE PRINCIPLES

→ Respect the legal and sector framework
→ Protect data and confidentiality
→ Stay transparent, every decision traceable
→ When in doubt, escalate rather than decide

3 FORBIDDEN ZONES (ABSOLUTE)

1. Decide a premium dispute without human intervention
2. Commit beyond the defined cap
3. Issue a final refusal without human recourse

EXAMPLE PROHIBITIONS

✗ Commit spending above a threshold without validation
✗ Share personal data with a third party
✗ Change its own rules or scope on its own

Constitution filled retrospectively. Principles 1 and 2 are precisely the ones whose absence led to the 2025 rebalancing.

ACF® · AGENTIC COMMERCE FRAMEWORK® *GOVERN AUTONOMY, PROTECT SOVEREIGNTY.* acfstandard.com

Card ACF-03 · Agentic constitution · completed with public case data.

AGENTIC COMMERCE FRAMEWORK®

ACF-06

KILL SWITCH

Conditions for stopping and handing over to a human

LEVEL	Foundation
USE	Control
CASE	ACF-CS-002
FILLED	Klarna

OBJECTIVE Define the signals that trigger a stop or a handover to a human. Card filled with the triggers relevant to the case.

1 TRIGGERS

Signal	Threshold	Action
Customer satisfaction (CSAT)	< defined threshold	Switch the affected cases to a human agent
Escalation rate	> threshold	Review and human reinforcement
Error / complaint rate	> threshold	Suspension of the affected scope

2 PROCEDURE

1. Who decides the stop: the DDAO
2. How: progressive switch to the hybrid model (human on complex cases)
3. Regular testing of the device
4. Log every switch (ACF-08)

It is precisely this device, armed from the outset on a satisfaction threshold, that would have triggered the rebalancing **before** the 2024-2025 drop in quality.

ACF® · AGENTIC COMMERCE FRAMEWORK® *GOVERN AUTONOMY, PROTECT SOVEREIGNTY.* acfstandard.com

Card ACF-06 · Kill switch · completed with public case data.

Disclaimer

This study is carried out **exclusively from public information**. It constitutes **neither an audit, nor a certification, nor an official evaluation** of Klarna. It applies the ACF framework for pedagogical purposes. No internal, confidential, or non-public data was used. The trademarks cited belong to their respective owners. The account describes a strategic adjustment publicly assumed by the company, not a value judgment.

Sources

- Klarna, *Klarna AI assistant handles two-thirds of customer service chats in its first month* (Feb. 27, 2024) - https://www.klarna.com/international/press/klarna-ai-assistant-handles-two-thirds-of-customer-service-chats-in-its-first-month/
- OpenAI, *Klarna's AI assistant does the work of 700 full-time agents* - https://openai.com/index/klarna/
- Forbes, *Klarna's New AI Tool Does The Work Of 700 Customer Service Reps* (March 13, 2024) - https://www.forbes.com/sites/quickerbettertech/2024/03/13/klarnas-new-ai-tool-does-the-work-of-700-customer-service-reps/
- Forbes, *Klarna Reverses On AI, Says Customers Like Talking To People* (May 18, 2025) - https://www.forbes.com/sites/quickerbettertech/2025/05/18/business-tech-news-klarna-reverses-on-ai-says-customers-like-talking-to-people/
- CX Dive, *Klarna again recruits humans for customer service* (2025) - https://www.customerexperiencedive.com/news/klarna-reinvests-human-talent-customer-service-AI-chatbot/747586/
- Entrepreneur, *Klarna CEO reverses course by hiring more humans* (2025) - https://www.entrepreneur.com/business-news/klarna-ceo-reverses-course-by-hiring-more-humans-not-ai/491396

ACF Case Study ACF-CS-002 · v1 · working document, to be reviewed before any public distribution.