



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-004 · Case Study #004 · Édition 2026

Cruise (General Motors)

When an autonomous agent injures, who presses the stop button, and at what moment?

CARTE D'IDENTITÉ

SECTOR Mobility	COUNTRY USA	TECHNOLOGY Autonomous robotaxi	TYPE Agentic AI	AUTONOMY autonomous
CONFIDENCE High				

OUTILS ACF ILLUSTRÉS DANS CE CAS

ACF-01	ACF-05	ACF-06	ACF-08	ACF-00	ACF-02
--------	--------	--------	--------	--------	--------

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

1. Executive summary

Cruise, General Motors' robotaxi subsidiary, operated in San Francisco **fully autonomous** vehicles, with no safety driver on board. Driving there was 100% delegated to the agent, which perceives, decides, and acts alone, continuously. On October 2, 2023, one of these robotaxis strikes a pedestrian already thrown by another car, then, while attempting to pull over, **drags her about six meters** before coming to a stop. On October 24, 2023, the California regulator (DMV) suspends Cruise's driverless permits, effective immediately.

Read through the lens of ACF, this case is a **kill switch incident**. The stop mechanism that actually halted the agent was not internal: it was the outside authority that "pressed the button", ten days after the facts. The value of ACF here is **corrective**: it names, from the design stage, what a real stop mechanism is - internal, armed, measurable, and immediate - as opposed to an imposed stop, forced from outside after the incident.

2. Context

Cruise operated a fleet of robotaxis in real conditions, on open roads, in one of the activities where an agent's error can cause physical injury. This is the highest possible level of autonomy: no human in the loop at the moment of the driving act. The governance question is therefore head-on, and it is not technical: **who can stop this agent, and at what moment?** This is not a classroom hypothesis, it is the condition of acceptability of any autonomy that can cause harm.

3. Public sources used

- Carscoops, "California DMV Suspends Cruise Permits After Robotaxi Ran Over And Dragged Pedestrian".
- The San Francisco Standard, "Cruise Hid Video of Woman Being Dragged Along San Francisco Street, DMV Says".
- CalMatters, "Two state agencies ground Cruise driverless cars for public safety".
- CBS San Francisco, "Cruise to pay \$1.5M penalty in connection with San Francisco pedestrian accident, NHTSA says".
- Carnegie Mellon University (P. Koopman), "Cruise Pedestrian Dragging Mishap: Case Study".

4. Reconstructed AI architecture (from public sources)

Cruise autonomous robotaxi (Chevrolet Bolt, in operation at the time of the facts, October 2023).

- Nature: fully autonomous vehicle, no safety driver on board, on public roads.
- Function: driving 100% delegated, continuous decision (accelerate, brake, stop, pull over).
- Autonomy level: the highest, on an activity with bodily risk.

The breaking point. On October 2, 2023, around 9:30 pm, a pedestrian crosses a San Francisco street. She is first hit by a car driven by a human, which throws her onto the robotaxi's path. The agent strikes her, then, instead of remaining still, **continues a programmed maneuver** to pull over and drags her about six meters at low speed before coming to a stop. The agent did not recognize that a person was under the vehicle: a "pull over" rule prevailed over a "remain still on human contact" rule.

The chain reaction. On October 24, 2023, the DMV suspends the driverless permits, official grounds: the vehicles "are not safe for public operation" and "may lack the ability to respond safely during incidents involving a pedestrian". The DMV also faults Cruise for having initially provided an **incomplete** video, without the dragging sequence, which Cruise disputes. Shortly after, Cruise suspends all of its operations and recalls its fleet. The federal agency (NHTSA) then imposes a penalty for failure to meet reporting obligations.

Missing safeguards (what the facts reveal):

- no **priority** stop rule prevailing over any maneuver as soon as human contact is detected;
- no **reflex immobilization** in the case of an unidentified mass under the vehicle;
- no **fleet self-suspension** mechanism by the company, armed before the regulator's intervention;
- a disputed transparency on the complete rendering of the incident.

5. ACF mapping

Tools used: ACF-01 (decision map) · ACF-06 (kill switch) · ACF-05 (supervision) · ACF-08 (register).

In ACF language, the incident reads as follows:

ACF element	In the Cruise deployment
Named human (accountable)	Cruise, the company. It is the one that answers for its agent's driving and for its ability to stop it.
Agent	The autonomous robotaxi (Chevrolet Bolt), which perceives, decides, and acts alone.
Platform	The underlying autonomous driving system (perception, planning, control).
Delegated decision	Continuous driving, including the action after contact with a pedestrian, a decision that should have stayed under a strict non-negotiable rule.
Autonomy level	Fully autonomous , including in a critical situation, where an immobilization rule should have prevailed.
Non-delegable zones	Acting after having struck a pedestrian: the "immobilization" rule is not negotiable. This is the boundary that was crossed.
Rule / safeguard	No priority reflex stop; no internally armed fleet self-suspension.
Control third party	The regulator (California DMV), then the federal agency (NHTSA), after the fact.
Traceability	Incident video deemed incomplete by the DMV; no complete and uncontested rendering ahead of the dispute.

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

6. Analysis

Tools used: ACF-06 (stop) · ACF-01 (mapping) · ACF-05 (supervision) · ACF-08 (register).

What the incident reveals (the design error):

- **The stop was not armed in the right place.** The flaw is not "the agent did not stop at all", it is "the agent continued a maneuver" where human contact should have triggered a priority immobilization. A comfort rule (pull over) took precedence over a safety rule (remain still).
- **The real kill switch was imposed, not armed.** The mechanism that actually stopped Cruise is the DMV's permit suspension, ten days after the accident. This is not a safety mechanism, it is a recall imposed from outside, after the fact. The company had the means to suspend its fleet and did not do so in time on its own initiative.
- **Transparency is a piece of governance, not an add-on.** The video deemed incomplete by the DMV worsened the crisis of trust. Without complete and incontestable rendering, the correction loop cannot function and accountability becomes impossible.

What ACF would have made explicit (and enforceable):

- a **stop mechanism** (ACF-06) with measurable and prioritized triggers: detected contact with a pedestrian, unidentified mass under the vehicle, impact sensor, with **immediate and automatic immobilization** prevailing over any programmed maneuver;
- a **fleet self-suspension** activatable by the company without waiting for an external constraint;
- a **supervision** (ACF-05) linking pedestrian incidents, manual intervention rates, and abnormal behaviors after impact to the arming of the stop;
- a **decisions register** (ACF-08) timestamping and rendering the incident completely, a condition of correction and trust.

7. ACF toolkit used

This grid indicates the tools that the ACF analysis uses (✓) to read and correct this case, and those to anticipate (→
SOON). It does not mean the company uses ACF: it shows which instruments the framework would bring.

Tool	Status	Role in this case
ACF-06 · Kill switch	✓	Arm an internal stop with measurable triggers, prioritized over any maneuver
ACF-01 · Decision map	✓	Situate the agent, the accountable party, and the non-delegable zone
ACF-05 · Supervision	✓	Link safety indicators to the arming of the stop
ACF-08 · Decisions register	✓	Timestamp and render the incident in full
ACF-02 · Criticality matrix	→ SOON	Isolate the emergency situation as a decision of maximum criticality
ACF-03 · Agentic constitution	→ SOON	Enshrine priority immobilization as a non-negotiable rule

8. Governance questions (what an AI committee should ask itself)

1. Is a kill switch that only activates after a regulator's or court's decision still a kill switch?
2. Is our stop mechanism **internal, armed, measurable, and immediate**, or does it depend on an external intervention?
3. Where have we placed, explicitly, the non-delegable zone in driving: which rule prevails in a contact or emergency situation?
4. Which indicators would trigger a self-suspension **by the company** before a third party imposes it?
5. Is transparency about the incident part of our governance of the agent, with a complete and incontestable trace?

9. General lessons

An autonomous agent does not fail only when it "drives badly": it fails when no one has armed, inside the system, the rule that stops it in a critical situation. Cruise did not lack sensors, it lacked a **priority stop mechanism** and a **self-suspension** decided in time. The kill switch that worked came from the regulator, which, by definition, is already a governance failure.

The transferable lesson: autonomy is only acceptable if it is reversible and interruptible from within. An imposed stop, forced after the incident, is not a kill switch, it is an acknowledgment of failure.

10. What ACF would do (reference design)

Prescriptive and non-critical formulation: here is how an organization would design such a system directly with the framework, whatever its sector.

If an organization wished to deploy this type of agent directly with ACF, it could in particular:

- **define** a stop mechanism (ACF-06) with measurable and prioritized triggers: pedestrian contact, unidentified mass under the vehicle, impact sensor, with immediate and automatic immobilization;
- **set** as a non-negotiable rule, in a signed agentic constitution (ACF-03), that immobilization prevails over any programmed maneuver in the case of human contact;
- **map** driving on the autonomy scale (ACF-01) and isolate the emergency situation as a non-delegable zone;
- **arm** a fleet self-suspension activatable by the company, without waiting for an external constraint;
- **link** the supervision (ACF-05) of pedestrian incidents and manual interventions to the arming of the stop;
- **document** a decisions register (ACF-08) rendering each incident completely and timestamped.

None of this is a technical reproach: it is the way ACF would have made **explicit, signable, and auditable** the stop that an autonomous deployment must keep armed before opening an agent to public roads.

11. Similar cases (transferability)

This case is specific neither to Cruise nor to robotaxis. It is transferable to any organization that delegates to an agent a **continuous action capable of causing physical or economic harm**, where the only question that matters is: who stops it, and at what moment?

-  autonomous vehicles and mobility
-  industrial and logistics robotics
-  drones and tele-operated machines
-  automated trading and execution
-  action agents on critical systems (energy, networks)
-  medical devices and hospital automation

Everywhere, the same pattern: an agent that acts alone continuously, a critical situation where a stop rule must prevail, and an interruption mechanism that must be internal, armed, and immediate. This is what makes this document a **reference case**: it illustrates a concept, not just a company.

12. Confidence of the assessment

High. The case rests on **public official acts** (permit suspension by the California DMV, federal penalty from the NHTSA) and on reference press coverage and an academic case study. The facts are therefore established and verifiable by the reader. One point (the complete or incomplete rendering of the video) is the subject of diverging versions, reported as such. The ACF analysis concerns **governance**, not the internal technical details of the driving system, without bearing on the reading proposed.

Teacher guide

Pedagogical objective. Make clear that a kill switch is a governance mechanism only if it is internal, armed, measurable, and immediate, and that a stop imposed by a third party after the incident is already a failure.

Session outline (90 min).

1. (15 min) Reading of the case, without the Analysis part.
2. (20 min) In groups: design the kill switch (card ACF-06), measurable triggers and who decides the stop.
3. (20 min) Debate: "Is a kill switch imposed by a regulator a kill switch?" (expected answer: no, it is an imposed stop).
4. (20 min) Projection: which supervision indicators would have triggered a suspension by the company before the regulator?
5. (15 min) Synthesis: autonomy is only acceptable if it is reversible and interruptible from within.

Possible exam questions. See the "Governance questions" section.

Reminder: grading students' work is a pedagogical act of the teacher. This guide provides the reference analysis, not automatic grading.

Associated lab (ACF-LAB-004)

From this case, the student produces:

- the definition of a kill switch (card ACF-06): measurable and prioritized triggers, who decides, speed of execution;
- the decision map of the robotaxi (card ACF-01): agent, accountable party, non-delegable zone;
- the table of supervision indicators (card ACF-05) arming a self-suspension by the company;
- the schema of a decisions register (card ACF-08): what to render from an incident and why.

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 profile can reuse and adapt them.



OBJECTIVE

Map the decisions of the Cruise deployment and set, for each, the autonomy level based on what it commits. Filled from public sources.

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

Driving in normal situation

STAKE

Perceive, brake, pull over outside incident

IMPACT

Reversible, no harm

DEADLINE

Continuous

AUTONOMY

Autonomous

B

Reporting and fleet self-suspension

STAKE

Withdraw the fleet on safety indicators

IMPACT

Internal decision, reversible

DEADLINE

On threshold

AUTONOMY

Supervised, decided by the company

C

Acting after a pedestrian contact

STAKE

Continue or not a maneuver after human impact

IMPACT

Commits physical integrity, irreversible

DEADLINE

-

AUTONOMY

NON-DELEGABLE

Reading: it is this setting, made **before** deployment, that would have imposed priority immobilization on any human contact, instead of letting a pull-over maneuver prevail.



OBJECTIVE

Link safety indicators to the arming of the stop and the self-suspension. Card filled with the public elements of the case.

1 SUPERVISION INDICATORS

Indicator	Value / threshold (public sources)
Incidents involving a pedestrian	any incident, immediate escalation
Abnormal behavior after impact	detection and blocking of the maneuver
Manual intervention rate	to be tracked, arms the self-suspension
Rendering of incidents	full video, no removed sequence
Self-suspension decision	held by the company, not by the regulator

2 MECHANISM

- 1 Link pedestrian incidents and post-impact behaviors to the arming of the stop (ACF-06)
- 2 Supervision officer: the DDAO, decision-maker of the self-suspension
- 3 Full and incontestable rendering of each incident
- 4 Periodic review of critical situations and immobilization rules

A useful supervision does not only measure performance: it **arms the stop**. Linked to the register (ACF-08), it turns an incident into a fix; a self-suspension decided in time avoids the imposed stop.



OBJECTIVE

Define the signals that stop the agent or suspend the fleet. Card filled with the triggers relevant to the case.

1 TRIGGERS

Signal	Threshold	Action
Detected contact with a pedestrian	Any contact	Immediate immobilization, prioritized over any maneuver
Unidentified mass under the vehicle	Detected	Reflex stop, no pull-over
Pedestrian incidents or manual interventions	Above threshold	Fleet self-suspension by the company

2 PROCEDURE

- 1 Who decides the stop: the company (DDAO), without waiting for the regulator
- 2 How: automatic immobilization on contact, fleet self-suspension armed internally
- 3 Regular testing of the triggers on critical situations
- 4 Log every stop and every suspension (ACF-08)

It is precisely this internal stop, armed and immediate, that would have immobilized the vehicle on contact and suspended the fleet **before** the DMV's intervention, ten days after the facts.



OBJECTIVE

Record every incident and every stop to be able to reconstruct them, correct them at scale and account for them. Card filled with the fields relevant to the case.

1 FIELDS TO RECORD

Field to record	Example, Cruise case
Timestamp	2 October 2023, around 9:30 pm
Situation	Pedestrian thrown onto the robotaxi's path
Agent's decision	Continuation of a pull-over maneuver after contact
Consequence	Pedestrian dragged about six meters
Rendering	Video deemed incomplete by the DMV (diverging versions)

2 USE OF THE REGISTER

- 1 Detect a dangerous post-impact behavior before it repeats
- 2 Correct at scale the immobilization rule across the whole fleet
- 3 Establish who accounts for the agent's driving: the company
- 4 Provide a full and incontestable trace, a condition of trust

Without full rendering, the correction loop does not function and accountability becomes impossible. The register is a piece of governance, not an add-on.

Disclaimer

This study is carried out **exclusively from public information** (official acts of regulators and accessible press coverage). It constitutes **neither an audit, nor a certification, nor an official evaluation** of Cruise or General Motors, nor an opinion on legal liabilities, which fall to the competent authorities. It applies the ACF framework for pedagogical and illustrative purposes. Certain assertions (the video communication) are the subject of diverging versions, reported as such. No internal, confidential, or non-public data was used. The trademarks cited belong to their respective owners.

Sources

- Carscoops, *California DMV Suspends Cruise Permits After Robotaxi Ran Over And Dragged Pedestrian*
- The San Francisco Standard, *Cruise Hid Video of Woman Being Dragged Along San Francisco Street, DMV Says*
- CalMatters, *Two state agencies ground Cruise driverless cars for public safety*
- CBS San Francisco, *Cruise to pay \$1.5M penalty in connection with San Francisco pedestrian accident, NHTSA says*
- Carnegie Mellon University (P. Koopman), *Cruise Pedestrian Dragging Mishap: Case Study*

ACF Case Study ACF-CS-004 · v1 · working document, to be reviewed before any public distribution.