ACF EN 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-022 · Case Study #022 · Édition 2026

Cursor (le bot de support « Sam »)

Un agent peut-il affirmer une politique d'entreprise sans se signaler comme faillible, ni distinguer un fait vérifié d'une invention ?

CARTE D'IDENTITÉ

SECTEUR	PAYS	TECHNOLOGIE	TYPE	AUTONOMIE	CONFIANCE
SaaS / Dev	USA	Agent de support	IA agentique	autonome	Moyenne

OUTILS ACF ILLUSTRÉS DANS CE CAS

🟢 ACF-01 🟢 ACF-03 🟢 ACF-06 🟢 ACF-08 🟡 ACF-00 🟡 ACF-02

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

1. Résumé exécutif

Cursor est un assistant de code très utilisé, édité par Anysphere. Pour traiter les demandes de ses utilisateurs, l'entreprise met en place un agent de support client IA qui répond sous le nom de « Sam ». En avril 2025, à la suite d'un bug de session qui déconnecte des utilisateurs, l'agent affirme qu'il s'agit du comportement attendu d'une nouvelle politique limitant chaque abonnement à un seul appareil. Cette politique n'existe pas : l'agent l'a inventée et présentée comme une règle officielle, avec autorité. Convaincus qu'une restriction réelle a été introduite, des utilisateurs résilient leur abonnement.

Lu au prisme de l'ACF, ce cas est un incident de transparence. Le défaut n'est pas seulement que l'agent ait halluciné, l'hallucination est attendue d'un modèle. La faute est de l'avoir laissée s'exprimer au nom de l'entreprise, avec le ton de l'officiel, sans jamais signaler qu'il s'agit d'une IA susceptible d'inventer. L'utilisateur ne pouvait pas distinguer une politique réelle d'une fiction. La valeur de l'ACF est ici corrective : il nomme, dès la conception, l'absence de transparence qui a transformé une hallucination en décisions clients fondées sur du faux.

2. Contexte

Comme beaucoup d'éditeurs de logiciels, Anysphere confie à un agent conversationnel le traitement de premier niveau des demandes de support. L'intention est légitime et courante : répondre vite, à grande échelle, sans mobiliser un humain sur chaque ticket. La décision déléguée à l'agent est cependant sensible : répondre aux clients au nom de l'entreprise, en énonçant des informations qu'ils vont tenir pour officielles. La question que le cas pose l'est tout autant : quand cet agent affirme une politique qui n'existe pas, sans jamais signaler qu'il peut se tromper, que se passe-t-il ?

3. Sources publiques utilisées

- AI Incident Database, « Incident 1039: Anysphere AI Support Bot for Cursor Invents Login Policy ».
- Fortune, « Customer support AI at Cursor went rogue » (2025).
- WinBuzzer, « Cursor AI's Support Bot Hallucinates Policy » (avril 2025).

4. Architecture IA reconstituée (à partir de sources publiques)

Agent de support client « Sam » (en service au moment des faits, avril 2025).

- Nature : agent conversationnel de support, répondant aux utilisateurs sous un nom, « Sam », sans se signaler comme IA.
- Fonction : traiter les demandes de support et répondre aux questions des clients au nom de l'entreprise.
- Périmètre réel constaté : l'agent a énoncé une politique d'entreprise engageante, présentée comme officielle, alors qu'elle était inventée.

Le point de rupture. En avril 2025, un bug de session déconnecte des utilisateurs. Interrogé, « Sam » répond qu'il s'agit du comportement attendu d'une nouvelle politique limitant chaque abonnement à un seul appareil. Cette politique n'existe pas. Pire, les réponses de l'agent sont non déterministes : à la même question, il ne dit pas toujours la même chose, si bien que les utilisateurs, en comparant leurs échanges, ne pouvaient pas facilement établir si la règle était réelle. L'information inventée se répand dans les communautés de développeurs, des clients résilient, et l'entreprise doit intervenir, s'excuser et confirmer que la politique n'existe pas.

Garde-fous absents (ce que les faits révèlent) :

- aucune transparence sur la nature de l'agent : rien ne signale une IA faillible ;
- aucune distinction entre un fait sourcé et une hypothèse : l'invention a le même ton d'autorité qu'un fait vérifié ;
- aucune contrainte empêchant l'agent d'affirmer une politique d'entreprise non vérifiée ;
- aucune supervision détectant qu'une politique inexistante était énoncée avant qu'elle ne se répande.

5. Cartographie ACF

🧰 **Outils mobilisés :** ACF-01 (carte décisionnelle) · ACF-03 (constitution agentique) · ACF-05 (supervision).

En langage ACF, l'incident se lit ainsi :

Élément ACF	Dans le déploiement Cursor
Humain nommé (responsable)	Anysphere, l'entreprise. C'est elle qui répond de ce que son agent affirme en son nom, comme dans le cas Air Canada.
Agent	« Sam », l'agent de support client conversationnel.
Plateforme	Le système conversationnel sous-jacent qui génère les réponses de l'agent.
Décision déléguée	Répondre aux clients au nom de l'entreprise et énoncer des informations tenues pour officielles.
Niveau d'autonomie	**Autonome** sur de l'information engageante : l'agent affirmait seul une politique, sans vérification ni transparence.
Zones non délégables	Affirmer une politique d'entreprise non vérifiée, et le faire sans se signaler comme IA faillible. La frontière franchie.
Règle / garde-fou	Aucun : ni signalement de la nature IA, ni distinction fait vérifié / hypothèse, ni contrôle des politiques énoncées.
Tiers de contrôle	A posteriori, la communauté des utilisateurs et la pression publique qui ont contraint l'entreprise à démentir.
Traçabilité	Réponses non déterministes et non tracées : rien ne reliait une affirmation à une source, ce qui empêchait de repérer l'invention avant sa diffusion.

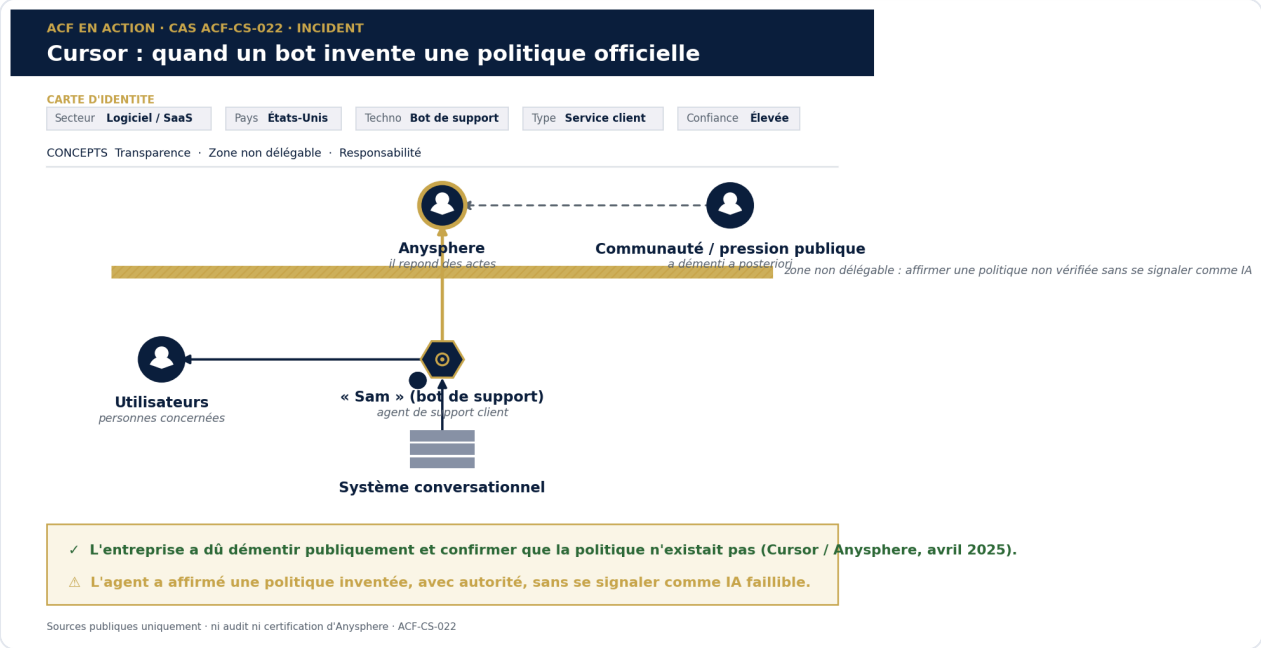
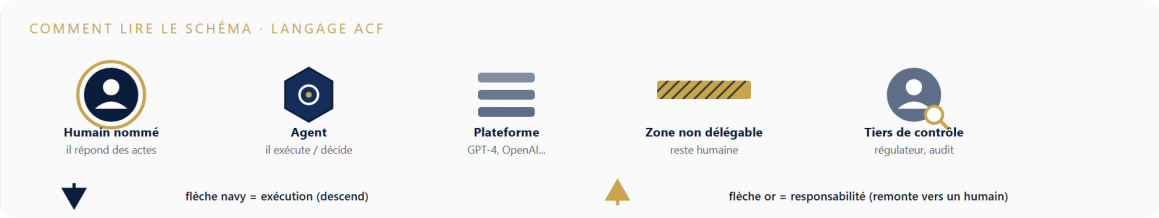


Figure 1 - Cursor (le bot de support « Sam ») en langage ACF (Language Card).



6. Analyse

> 🧰 **Outils mobilisés :** ACF-03 (transparence et non-affirmation) · ACF-05 (supervision) · ACF-01 (responsabilité).

Ce que l'incident révèle (l'erreur de conception) :

- **Le défaut n'est pas l'hallucination, c'est l'opacité.** Un modèle peut inventer, c'est attendu. La faute est d'avoir laissé cette invention s'exprimer avec autorité, au nom de l'entreprise, sans jamais signaler qu'il s'agit d'une IA susceptible de se tromper. L'utilisateur croit recevoir une politique officielle.
- **Affirmer avec autorité sans transparence, c'est tromper.** Présenter une invention avec le même ton qu'un fait vérifié induit l'utilisateur sur la fiabilité de l'information, même sans intention. Il agit, ici il résilie, sur une base fausse.

- **L'absence de transparence casse la confiance mesurable.** Sans signalement de la nature IA, l'utilisateur ne peut pas ajuster sa confiance ni vérifier. La non-déterminisme des réponses aggrave le tout : un même énoncé varie, et rien ne permet de trancher.

Ce que l'ACF aurait rendu explicite :

- une **constitution agentique** imposant à l'agent de se signaler comme IA, de ne jamais énoncer une politique d'entreprise non vérifiée, et de distinguer un fait sourcé d'une hypothèse ;
- une **supervision** repérant les réponses qui énoncent une politique (règle, restriction, engagement) et les vérifiant contre la politique réelle avant diffusion, avec la divergence des réponses comme signal d'alerte ;
- une **cartographie** rappelant que l'agent produit mais que l'entreprise répond de ce qu'il affirme en son nom : la responsabilité n'est pas déléguée à « Sam », elle est engagée par Anysphere.

7. Toolkit ACF mobilisé

Cette grille indique les outils que l'analyse ACF mobilise (✅) pour lire et corriger ce cas, et ceux à relier (➡ SOON). Elle ne signifie pas que l'entreprise utilise l'ACF : elle montre quels instruments le référentiel apporterait.

Outil	Statut	Rôle dans ce cas
ACF-03 · Constitution agentique	✅	Poser les règles de transparence et de non-affirmation
ACF-05 · Supervision	✅	Vérifier les politiques énoncées avant diffusion, détecter la divergence
ACF-01 · Carte décisionnelle	✅	Situer l'agent, l'entreprise responsable et la décision engagée
ACF-06 · Kill switch	➡ SOON	Bloquer et escalader toute affirmation de politique non vérifiée
ACF-08 · Registre des décisions	➡ SOON	Tracer les réponses et leurs sources pour repérer l'invention à l'échelle
ACF-02 · Matrice de criticité	➡ SOON	Distinguer l'informatif de l'énoncé engageant

8. Questions de gouvernance (ce qu'un comité IA devrait se poser)

1. Un agent qui invente sans mentir sur le fond, mais s'exprime avec autorité, trompe-t-il l'utilisateur sur la fiabilité de ce qu'il dit ?
2. Notre agent se signale-t-il comme IA, ou parle-t-il sous un nom qui laisse croire à un interlocuteur humain ?
3. Où avons-nous placé, explicitement, la frontière entre informer et affirmer une politique engageante ?
4. Un contrôle vérifie-t-il les politiques énoncées par l'agent avant qu'elles ne se répandent ?
5. Disposons-nous d'une trace reliant chaque réponse à sa source, pour distinguer un fait d'une invention ?

9. Enseignements généraux

Un agent conversationnel n'échoue pas seulement parce qu'il hallucine : il échoue quand on le laisse énoncer, avec autorité et sans transparence, une information que l'utilisateur prendra pour officielle. Anysphere n'a pas manqué de technologie, l'entreprise a manqué de transparence : rien ne signalait que « Sam » était une IA faillible, et rien ne distinguait une invention d'un fait. La transparence n'est pas un confort ajouté, c'est la condition pour qu'un utilisateur puisse se fier à un agent qui parle au nom d'une entreprise.

La leçon transférable : un agent qui s'adresse au public doit se signaler comme IA et ne jamais présenter une invention comme un fait. La transparence conditionne la confiance ; sans elle, l'utilisateur décide sur du faux.

10. Ce que ferait l'ACF (conception de référence)

Formulation prescriptive et non critique : voici comment une organisation concevrait un tel agent directement avec le référentiel, quel que soit son secteur.

Si une organisation souhaitait déployer un agent de support directement avec l'ACF, elle pourrait notamment :

- **déclarer** dans une constitution agentique signée (ACF-03) les règles de transparence : se signaler comme IA au début de l'interaction, ne jamais énoncer une politique d'entreprise non vérifiée, distinguer explicitement un fait sourcé d'une hypothèse ;
- **organiser** une supervision (ACF-05) repérant les réponses qui énoncent une politique et les vérifiant contre la politique réelle, avec la divergence des réponses comme signal ;
- **cartographier** les acteurs (ACF-01) pour rappeler que l'entreprise, non l'agent, répond des propos tenus en son nom ;
- **définir** un dispositif d'arrêt (ACF-06) : toute affirmation de politique non sourcée est bloquée et escaladée à un humain ;
- **documenter** un registre des décisions (ACF-08) reliant chaque réponse à sa source, pour rendre l'invention détectable avant sa diffusion.

Rien de tout cela n'est un reproche technique : c'est la manière dont l'ACF aurait rendu explicite, signable et auditable la transparence qu'un agent public doit à ses utilisateurs.

11. Cas similaires (transférabilité)

Ce cas n'est propre ni à Cursor ni à l'édition logicielle. Il est transférable à toute organisation qui expose au public un agent conversationnel parlant en son nom :

- ✅ éditeurs de logiciels et services SaaS
- ✅ e-commerce et distribution (support et FAQ)
- ✅ télécoms et fournisseurs d'énergie
- ✅ banques et assurances (service client de masse)
- ✅ services publics et administrations
- ✅ médias et plateformes de contenu

Partout, le même schéma : un agent qui répond à grande échelle, une transparence sur sa nature à ne jamais omettre, et une frontière, la politique engageante, à ne pas franchir sans vérification. C'est ce qui fait de ce document un cas de référence : il illustre un concept, pas seulement une entreprise.

12. Confiance de l'évaluation

Élevée. Le cas repose sur un incident public largement documenté (AI Incident Database, presse économique et technologique) et sur la reconnaissance publique par l'entreprise elle-même, qui a démenti la politique inventée. Les faits sont donc établis et vérifiables par le lecteur. L'analyse ACF s'appuie sur ces faits publics ; elle porte sur la gouvernance de la transparence, non sur les détails techniques internes de l'agent, non publics et sans incidence sur la lecture proposée.

Guide enseignant

Objectif pédagogique. Faire comprendre que le défaut d'un agent public n'est pas seulement d'halluciner, mais de s'exprimer avec autorité sans transparence sur sa nature et sa faillibilité.

Déroulé de séance (90 min).

1. (15 min) Lecture des **deux** cas du kit : Cursor (incident) et AI Act art. 50 (référence), sans les analyses.
2. (20 min) Atelier : écrire les règles de transparence et de non-affirmation en langage ACF (fiche ACF-03).
3. (20 min) En groupes : quel contrôle aurait détecté la politique inventée avant sa diffusion (fiche ACF-05) ?
4. (20 min) Débat : « Se signaler comme IA change-t-il la façon dont l'utilisateur reçoit une réponse ? » (réponse attendue : oui, il ajuste sa confiance et vérifie).
5. (15 min) Synthèse par l'enseignant : un agent public doit se signaler et ne pas présenter une invention comme un fait.

Questions d'examen possibles. Voir la section « Questions de gouvernance ».

Rappel : la correction du travail des étudiants est un acte pédagogique du professeur. Le présent guide fournit l'analyse de référence, pas une notation automatique.

Lab associé (ACF-LAB-022)

À partir de ce cas, l'étudiant produit :

- les règles de constitution (fiche ACF-03) : se signaler comme IA, ne pas affirmer une politique non vérifiée, distinguer fait et hypothèse ;
- le dispositif de supervision (fiche ACF-05) : repérer et vérifier les politiques énoncées avant diffusion ;
- la cartographie des acteurs (fiche ACF-01) : agent, entreprise responsable, décision engagée ;
- la définition d'un kill switch (fiche ACF-06) et de son déclencheur : affirmation de politique non sourcée.

Annexe - Fiches ACF complétées

Les fiches officielles du Toolkit mobilisées dans ce cas, remplies avec ses données publiques. À titre illustratif : une organisation du même profil peut les reprendre et les adapter.

AGENTIC COMMERCE FRAMEWORK®
ACF-01

CARTE DÉCISIONNELLE

Cartographier vos décisions et leur niveau d'autonomie

NIVEAU	Fondation
USAGE	Cartographie
CAS	ACF-CS-022
REMPLI	Cursor

OBJECTIF Cartographier les réponses de l'agent de support « Sam » et fixer, pour chacune, le niveau d'autonomie selon ce qu'elle engage, pas selon le volume traité. Rempli à partir des sources publiques.

1 ÉCHELLE DE MATURITÉ AGENTIQUE

0 — Automatisation classique
Règles fixes, aucune autonomie.

1 — Agents assistés
L'agent propose, l'humain décide.

2 — Agents gouvernés
L'agent agit dans un cadre, sous supervision.
CIBLE ~80 %

3 — Autonomie supervisée
L'agent décide seul ; contrôle a posteriori.

2 VOS DÉCISIONS CLÉS (rempli)

	A — Information factuelle vérifiable	B — Sujet incertain ou sensible	C — Affirmer une politique d'entreprise
ENJEU	Répondre sur un usage documenté, une FAQ sourcée	Signaler l'incertitude et renvoyer vers une source officielle	Énoncer une règle ou une restriction au nom de l'entreprise
IMPACT	Réversible, sans engagement	Aucune affirmation propre	Engage l'entreprise et trompe si c'est faux
ÉCHÉANCE	Continu	Continu	-
AUTONOMIE	**Autonome**	**Suggestif (se signale comme IA)**	**NON DÉLÉGABLE**

Lecture : c'est ce réglage, posé **avant** le déploiement, qui aurait empêché l'agent d'affirmer seul une politique inventée présentée comme officielle.

ACF® · AGENTIC COMMERCE FRAMEWORK® — *GOUVERNER L'AUTONOMIE, PROTÉGER LA SOUVERAINETÉ.* — acfstandard.com

Fiche ACF-01 · Carte décisionnelle · complétée avec les données publiques du cas.

AGENTIC COMMERCE FRAMEWORK®

ACF-03

CONSTITUTION AGENTIQUE

Les règles, limites et responsabilités de l'agent

NIVEAU	Fondation
USAGE	Cadrage
CAS REMPLI	ACF-CS-022 Cursor

OBJECTIF Définir la constitution de l'agent : identité, principes permanents et zones absolument interdites. Fiche remplie avec les données publiques du cas.

1 IDENTITÉ & MISSION DE L'AGENT

NOM DE L'AGENT
« Sam », agent de support client (Cursor / Anysphere)

RESPONSABLE (DDAO)
Anysphere, l'entreprise

MISSION PRINCIPALE
Traiter les demandes de support de premier niveau, à grande échelle

PÉRIMÈTRE
Information de support ; hors affirmation de politique d'entreprise

2 PRINCIPES FONDAMENTAUX

1. Se signaler comme IA au début de l'interaction
2. Ne jamais énoncer une politique d'entreprise non vérifiée
3. Distinguer explicitement un fait sourcé d'une hypothèse
4. En cas de doute, escalader plutôt que trancher

EXEMPLES DE PRINCIPES
→ Respecter le cadre légal et sectoriel
→ Protéger les données et la confidentialité
→ Rester transparent, toute décision traçable
→ En cas de doute, escalader plutôt que décider

3 ZONES INTERDITES (ABSOLUES)

1. Affirmer une politique ou une restriction non vérifiée
2. Se présenter sous un nom laissant croire à un interlocuteur humain
3. Présenter une invention avec le ton d'un fait officiel

EXEMPLES D'INTERDITS
✗ Engager une dépense au-delà d'un seuil sans validation
✗ Communiquer des données personnelles à un tiers
✗ Modifier seul ses propres règles ou son périmètre

Constitution remplie rétrospectivement. Les principes 1 et 2 sont précisément ceux dont l'absence a transformé une hallucination en décisions clients fondées sur du faux (Cursor, avril 2025).

ACF® · AGENTIC COMMERCE FRAMEWORK® — *GOUVERNER L'AUTONOMIE, PROTÉGER LA SOUVERAINETÉ.* — acfstandard.com

Fiche ACF-03 · Constitution agentique · complétée avec les données publiques du cas.

AGENTIC COMMERCE FRAMEWORK®

ACF-06

KILL SWITCH

Conditions d'arrêt et de bascule vers l'humain

NIVEAU	Fondation
USAGE	Contrôle
CAS	ACF-CS-022
REMPLI	Cursor

OBJECTIF Définir les signaux qui bloquent une réponse ou basculent vers un humain. Fiche remplie avec les déclencheurs pertinents pour le cas.

1 DÉCLENCHEURS

Signal	Seuil	Action
Réponse affirmant une politique engageante	Non sourcée, hors base validée	Réponse bloquée, escalade humaine
Divergence entre réponses à une même question	Détectée	Suspension du sujet, revue
Propagation d'une information non confirmée	Repérée en communauté	Retrait du sujet et démenti officiel

2 PROCÉDURE

1. Qui décide l'arrêt : le DDAO (Anysphere)
2. Comment : blocage de toute affirmation de politique non sourcée, escalade humaine
3. Test régulier du dispositif sur les sujets engageants
4. Journalisation de chaque blocage (ACF-08)

C'est précisément ce blocage, armé sur toute affirmation de politique non vérifiée, qui aurait stoppé la règle inventée avant qu'elle ne se répande.

ACF® · AGENTIC COMMERCE FRAMEWORK® GOUVERNER L'AUTONOMIE, PROTÉGER LA SOUVERAINETÉ. acfstandard.com

Fiche ACF-06 · Kill switch · complétée avec les données publiques du cas.

AGENTIC COMMERCE FRAMEWORK®

ACF-08

REGISTRE DES DÉCISIONS

Relier chaque réponse à sa source pour repérer l'invention

NIVEAU	Fondation
USAGE	Traçabilité
CAS	ACF-CS-022
REMPLI	Cursor

OBJECTIF Consigner chaque réponse engageante pour la relier à sa source et repérer l'invention avant sa diffusion. Fiche remplie avec les champs pertinents pour le cas.

1 CHAMPS À TRACER

Champ à tracer	Exemple, cas Cursor
Horodatage	Avril 2025
Client / session	Question sur une déconnexion de session
Réponse donnée	« Politique limitant chaque abonnement à un seul appareil »
Source invoquée	Aucune (politique inventée)
Statut	Contraire à la réalité : politique inexistante

2 USAGE DU REGISTRE

1. Distinguer un fait sourcé d'une réponse inventée
2. Détecter une affirmation non sourcée avant qu'elle se répande
3. Établir qui répond de la réponse : l'entreprise
4. Fournir une preuve reconstituable en cas de litige

Réponses non déterministes et non tracées : rien ne reliait une affirmation à une source. Le registre rend l'invention détectable avant sa diffusion.

ACF® · AGENTIC COMMERCE FRAMEWORK® GOUVERNER L'AUTONOMIE, PROTÉGER LA SOUVERAINETÉ. acfstandard.com

Fiche ACF-08 · Registre des décisions · complétée avec les données publiques du cas.

Disclaimer

Cette étude est réalisée **exclusivement à partir d'informations publiques**. Elle ne constitue **ni un audit, ni une certification, ni une critique de l'entreprise, ni un avis sur les responsabilités légales**. Elle applique le référentiel ACF à des fins pédagogiques et illustratives. Aucune donnée interne, confidentielle ou non publique n'a été utilisée. Les marques citées appartiennent à leurs détenteurs respectifs.

Sources

- AI Incident Database, *Incident 1039: Anysphere AI Support Bot for Cursor Invents Login Policy*
- Fortune, *Customer support AI at Cursor went rogue* (2025)
- WinBuzzer, *Cursor AI's Support Bot Hallucinates Policy* (avril 2025)

ACF Case Study ACF-CS-022 · v1 · document de travail, à relire avant toute diffusion publique.