

KIT PÉDAGOGIQUE ACF

La transparence

Un agent qui parle aux gens doit-il dire qu'il est une IA, et ne pas présenter une invention comme un fait ? Deux cas réels, un principe.

La question

Un agent répond au nom
d'une entreprise.

Sait-on qu'il est une IA faillible ?

- Cacher sa nature ou présenter une invention comme un fait, c'est tromper.
- Le référentiel **ACF** fait de la transparence une **règle**.

Le langage ACF : quatre objets

Humain responsable

Il répond des propos de l'agent. Déléguer ne décharge pas.

Agent

Il parle au nom de l'entreprise. Il peut inventer avec autorité.

Constitution

Se signaler comme IA, ne pas affirmer une politique non vérifiée.

Supervision

Vérifier que l'agent se signale et n'énonce pas de faits inventés.

Inventer avec **autorité**

**Une IA qui énonce une invention
avec le ton de l'officiel,
sans dire qu'elle est une IA,
trompe l'utilisateur.**

- Le problème n'est pas l'hallucination, c'est l'absence de transparence.
- L'utilisateur doit pouvoir distinguer un fait d'une hypothèse.

Cas 1 : Cursor (**le bot « Sam », incident**)

- Avril 2025 : un bug déconnecte des utilisateurs de Cursor (assistant de code).
- Le bot de support « Sam » **invente une politique** (« un seul appareil ») et l'affirme comme officielle.
- Réponses **non déterministes** ; la fausse règle se répand ; des clients **résilient**.
- L'entreprise s'excuse : la politique **n'existe pas**. Aucun signalement de faillibilité.

Sources publiques : AI Incident Database, Fortune, WinBuzzer (2025).

Cas 2 : AI Act européen (article 50, référence)

- L'UE impose des **obligations de transparence** pour les IA au contact du public.
- Un agent conversationnel doit **se signaler comme une machine**.
- Les contenus générés (texte, image, deepfakes) doivent être **marqués / divulgués**.
- La transparence devient une **obligation**, pas une option (en vigueur 2 août 2026).

Sources publiques : EU AI Act art. 50, Commission européenne.

Le contraste

Cursor (Sam)

Se tait sur sa nature, invente une politique avec autorité.
Confiance trompée.

AI Act art. 50

L'agent doit se signaler, les contenus IA sont marqués.
L'utilisateur sait à qui il parle.

Un agent qui parle aux gens, **opaque ou transparent.**

La règle d'or

Un agent qui s'adresse
à des personnes doit
se signaler et ne pas présenter
une **invention** comme un fait.

À vous de jouer

- Écrivez les règles de transparence et de non-affirmation (fiche **ACF-03**).
- Définissez le contrôle des politiques énoncées par l'agent (fiche **ACF-05**).
- Cartographiez qui répond des propos de l'agent (fiche **ACF-01**).
- Tracez les affirmations et les divulgations (fiche **ACF-08**).

La leçon

**La confiance ne vient pas
d'un agent qui a l'air sûr de lui.
Elle vient de la transparence.**