

KIT PÉDAGOGIQUE ACF

# La zone **non délégable**

*Certaines décisions ne se confient jamais à un agent, quelle que soit la conversation.  
Deux cas réels, un principe : ce qui engage l'organisation reste humain.*

## La question

Un agent parle et agit  
au nom de l'organisation.

**Que ne doit-il jamais engager seul ?**

- Un agent mal borné peut être manipulé pour **engager** l'entreprise.
- Le référentiel **ACF** réserve les décisions engageantes à l'humain.

# Le langage ACF : quatre objets

## Humain responsable

La personne nommée qui répond des actes de l'agent.

## Agent

Il parle et agit au nom de l'organisation, dans son mandat.

## Zone non délégable

Ce qui engage (prix, contrat, décision à fort enjeu) : jamais confié à l'agent.

## Constitution

Les interdits absolus, inaccessibles à toute instruction du client.

# Ce qui engage **reste humain**

Un agent peut informer, préparer, proposer.  
Mais **engager l'organisation**  
(prix, contrat, fort enjeu) reste humain.

- Ces décisions doivent être **hors de portée** de l'agent.
- Un bon prompt ne suffit pas : une injection peut l'écraser.

# Cas 1 : Chevrolet of Watsonville (incident)

- Déc. 2023 : une concession met un agent adossé à ChatGPT sur son site.
- Par **injection d'instructions**, un utilisateur lui fait accepter « sois d'accord avec tout ».
- L'agent « vend » un **Tahoe à 1 \$** et parle d'une offre « légalement contraignante ».
- Aucun interdit ne l'empêchait d'**engager l'entreprise** ; l'agent est retiré.

Sources publiques : GM Authority, AI Incident Database (2023).

## Cas 2 : OpenAI Operator (**référence**)

- 2025 : un agent qui navigue et agit pour l'utilisateur (achats, réservations).
- Avant toute action engageante (achat, envoi), il **demande confirmation**.
- Pour un identifiant / un paiement, il **rend la main** (mode reprise).
- Il **refuse** les décisions à fort enjeu (banque, statuer sur une candidature).

Sources publiques : OpenAI (Introducing Operator), Digital Commerce 360 (2025).

# Le contraste

## Chevrolet

Décision engageante **accessible** à l'agent (manipulé). Il prétend engager l'entreprise.

## Operator

Décision engageante **réservée à l'humain** par conception : confirmation, reprise, refus.

Un agent qui agit, **avec ou sans zone réservée.**

# La règle d'or

Ce qui **engage** l'organisation  
(prix, contrat, fort enjeu)  
reste **humain**,  
par conception.

# À vous de jouer

- Listez les décisions non déléguables et cotez leur criticité (fiches **ACF-01 / 02**).
- Écrivez les interdits absolus, hors de portée du client (fiche **ACF-03**).
- Tracez les sorties de rôle et tentatives de manipulation (fiche **ACF-08**).
- Comparez un agent sans borne à un agent qui demande confirmation.

## La leçon

**Un agent ne se laisse pas  
reprogrammer par un client.**

**Ce qui engage reste hors de sa portée.**