

Version en ligne : [aipm-standard.com/aipm/cas/cout-du-contrôle](https://aipm-standard.com/aipm/cas/cout-du-contrôle) · édition du 30/09/2026. Le statut de preuve peut évoluer : la page en ligne fait foi.

Cas approfondi

## Le contrôle ralentit le travail légitime, et l'éditeur l'écrit

Une phrase de l'annonce de GPT-6 Astra vaut un chapitre de méthode : les contrôles de sécurité supplémentaires peuvent « parfois ralentir, suspendre ou arrêter du travail légitime ». C'est l'arbitrage central de toute gouvernance de l'IA, énoncé par celui qui vend le produit.

Toute la banque de cas ([aipm-standard.com/aipm/cas](https://aipm-standard.com/aipm/cas)) · La règle de preuve ([aipm-standard.com/aipm/methode-de-preuve](https://aipm-standard.com/aipm/methode-de-preuve))

**Statut de preuve : Annoncé/projeté.** Ce statut n'est pas une opinion sur l'entreprise ou sur la technologie : il décrit ce que les sources consultées permettent d'affirmer, et rien de plus. La règle qui l'attribue est publiée sur la page Méthode de preuve ([aipm-standard.com/aipm/methode-de-preuve](https://aipm-standard.com/aipm/methode-de-preuve)). Il porte sur le constat central de la fiche ; les chiffres cités à l'intérieur conservent le statut de leur propre source, et ceux qui proviennent du seul éditeur restent « annoncés », y compris dans une fiche confirmée.

### Le contexte

Qui, quand, quel système

Le discours dominant présente les contrôles de sécurité comme un pur bénéfice : ils empêcheraient le mauvais sans gêner le bon. Toute personne ayant exploité un dispositif de contrôle réel — filtrage, validation, double signature — sait que c'est faux, et que la question n'est jamais s'il y aura des faux positifs, mais combien et qui les paie.

L'annonce de GPT-6 Astra pose l'arbitrage en une ligne, dans la section des limites reconnues : les contrôles de sécurité supplémentaires peuvent « parfois ralentir, suspendre ou arrêter du travail légitime » — l'éditeur cite explicitement la cybersécurité défensive parmi les travaux concernés.

### La décision, et qui l'a prise

La Directive de fait

Le même document décrit ce que l'éditeur met à disposition des organisations pour régler ce curseur : contrôles d'administration en entreprise portant sur les sites autorisés, les applications et les transferts de fichiers ; politiques de confirmation avant les actions critiques ; revue automatisée des appels d'outils.

Ces trois mécanismes forment, en substance, une Directive au sens d'AIPM : un périmètre d'action, des points d'arrêt obligatoires, et une surveillance des gestes. La différence est que le réglage — donc l'arbitrage entre sécurité et débit — est explicitement laissé au client.

### **Ce que le système a réellement fait**

L'éditeur revendique par ailleurs, sur une évaluation d'usage d'ordinateur et sur ce périmètre seulement, une réduction de 89 % des résultats involontaires par rapport à la génération précédente. Le chiffre est annoncé, non vérifié par un tiers, et il ne dit rien du taux de faux positifs — c'est-à-dire du travail légitime bloqué, qui est précisément le coût que la phrase reconnaît.

Aucun chiffre n'est publié sur ce second taux. C'est l'asymétrie habituelle : le bénéfice du contrôle est quantifié, son coût est mentionné.

### **L'état de la preuve**

Ce que les sources établissent, et ce qu'elles n'établissent pas

#### ***Ce que la source établit***

- L'éditeur reconnaît explicitement que ses contrôles de sécurité peuvent bloquer du travail légitime, y compris en cybersécurité défensive
- Trois mécanismes de réglage sont mis à disposition des organisations : périmètre, confirmation avant action critique, revue des appels d'outils
- Une réduction de 89 % des résultats involontaires est revendiquée, sur une évaluation d'usage d'ordinateur
- L'arbitrage entre sécurité et débit est laissé au client

#### ***Ce que la source n'établit pas***

- Le taux de faux positifs : aucun chiffre n'est publié
- Ce que « parfois » recouvre — quelques cas ou une part significative
- Si la réduction de 89 % est reproductible : aucune vérification indépendante à cette date
- Le coût en temps humain des politiques de confirmation avant action critique

**Sources consultées** : OpenAI — annonce de GPT-6 Astra · OpenAI — GPT-6 Astra pour le travail

---

### **Ce que ce cas dit du référentiel**

C'est le cas le plus directement utile à la **Directive**. Une Directive AIPM n'est pas une liste d'interdictions : c'est un arbitrage écrit entre ce qu'on accepte de perdre en fluidité et ce qu'on refuse de risquer. Ce cas prouve que l'arbitrage existe même chez celui qui conçoit le système, et qu'il n'est pas soluble dans la technique.

Il éclaire aussi le **RACI-IA** par un angle rarement traité : quelqu'un doit être responsable des *blocages injustifiés*. Sans ce rôle, la pression opérationnelle finit toujours par desserrer les contrôles en silence, parce que personne n'a mandat pour arbitrer et que tout le monde a intérêt à ce que ça passe.

## La consigne

Ce qu'un lecteur fait différemment après avoir lu cette fiche

**Écrivez le taux que vous acceptez.** Avant de déployer, fixez une phrase du type : « nous acceptons que N % des demandes légitimes soient retardées pour éviter une action non autorisée ». Sans ce nombre, le réglage sera fait par défaut, puis desserré sous la pression, sans trace.

**Nommez un responsable des blocages.** La même personne qui répond des incidents doit répondre des blocages inutiles. Les deux erreurs ont un coût ; une seule est visible.

**Comptez les deux colonnes.** Journalisez les actions bloquées à côté des actions autorisées. Un dispositif dont on ne mesure que les succès de protection dérive vers un dispositif que tout le monde contourne.

**Ce qui invaliderait cette lecture.** Une mesure publiée du taux de faux positifs, par l'éditeur ou par un utilisateur, qui montrerait que le blocage de travail légitime est négligeable. La consigne resterait prudente, mais le cas perdrait son statut de démonstration d'arbitrage.

Cette fiche suit le gabarit des cas approfondis d'AIPM : huit questions, toujours les mêmes. Voir les autres fiches ([aipm-standard.com/aipm/cas-approfondis](https://aipm-standard.com/aipm/cas-approfondis)).

© 2026 DYONYSOS — Tous droits réservés · Bexo® est une marque déposée de DYONYSOS · [aipm-standard.com](https://aipm-standard.com). Reproduction interdite sans autorisation, hors courtes citations avec mention de la source ([aipm-standard.com](https://aipm-standard.com)).