

LE PETIT LIVRE DE L'IA · CHAPITRE 0.2 · FICHE DE RÉVISION, STYLE EXAMEN

Erreurs, inventions et biais : les limites à connaître

Fiabilité Une IA peut inventer une référence, un chiffre ou un arrêt avec un aplomb parfait. Elle reproduit aussi les biais des textes dont elle a appris. Connaître ces limites, c'est savoir où placer la relecture humaine.

1 L'essentiel à savoir par cœur

- Les **inventions** (on dit aussi « hallucinations ») : sources, articles de loi, chiffres qui n'existent pas.
- Les **biais** : stéréotypes de genre, d'âge ou d'origine repris des données d'apprentissage.
- L'**obsolescence** : montants 2026, barèmes, taux ; le modèle peut être resté à une année antérieure.
- La parade : **vérifier chaque fait** dans la source officielle, et ne jamais laisser l'IA décider d'une personne.

3 limites : inventions, biais, obsolescence	**1** source officielle par fait	**LPD 21** décision automatisée	**5** questions avant d'envoyer

2 La base légale, article par article

- L'exactitude des données est un principe de la LPD : celui qui traite des données doit s'assurer qu'elles sont exactes (LPD art. 6, al. 5).
- Une décision fondée exclusivement sur un traitement automatisé, qui a des effets juridiques pour la personne, doit lui être annoncée ; elle peut demander qu'une personne la revoie (LPD art. 21).
- Le travailleur exécute son travail avec soin (CO 321a) : utiliser une IA ne décharge pas de vérifier.
- L'employeur répond des actes de ses auxiliaires et de ses outils envers les tiers (CO 101).

3 Les règles, cas par cas

Situation	Règle	Article
Un article de loi cité par l'IA	Le retrouver sur Fedlex avant de l'utiliser	LPD 6 al. 5
Un montant ou un taux 2026	Contrôler dans la source officielle de l'année	Bonne pratique
Un arrêt du Tribunal fédéral cité	Vérifier son existence et son contenu sur bger.ch	Bonne pratique
Un texte sur une personne (évaluation, certificat)	Relire chaque jugement : biais possibles	CO 328, 330a
Une décision sur une candidature	Jamais uniquement automatisée ; informer et permettre un réexamen	LPD 21

4 La méthode, dans l'ordre

1. Vérifier chaque chiffre.
2. Retrouver chaque article.
3. Comparer les résultats par groupes.
4. Garder une personne pour décider.
5. Dater la source.

> **« Ce qui sonne juste n'est pas forcément juste. »**

6 Les pièges

Demander à l'IA de vérifier sa propre réponse : elle peut confirmer son erreur avec la même assurance.

Côté RH et paie : Organiser la relecture

- Définir qui relit quoi : un texte interne peut être relu rapidement ; tout ce qui sort de l'entreprise ou touche une personne passe par une relecture complète.
- Afficher près des écrans les cinq questions « avant d'envoyer ».
- Partager en équipe, sans chercher de coupable, les erreurs repérées : c'est la meilleure formation.

7 Le cas type du chapitre

Le courriel trop sûr de lui. Un assistant rédige pour un collaborateur une réponse : « Selon l'article 329f du CO, vous avez droit à 16 semaines de congé de paternité payé. »

a. Qu'est-ce qui cloche ?

b. Comment l'aurait-on repéré ?

c. Quelle règle interne en tirer ?

Corrigé en fin de fiche.

8 Le coin de l'expert

- Les biais ne se voient pas sur un cas isolé : il faut comparer les résultats par groupes (femmes et hommes, âges) sur un échantillon.
- Demander à l'outil ses sources ne suffit pas : il peut citer des sources inventées ; seule la consultation de la source compte.
- Pour les textes juridiques, préférer un outil qui cite des passages exacts de documents fournis par l'entreprise (lois, règlement interne).
- La « juste confiance » : le ton assuré d'un assistant inspire spontanément la crédibilité ; avec la routine et la pression du temps, le biais d'automatisation fait passer des erreurs pourtant visibles. Fixer le niveau de vérification selon l'usage et laisser le temps de vérifier.
- Quatre degrés de vérification, du plus léger au plus sûr : relecture critique, comparaison avec un second outil ou une source, contrôle documenté par échantillon, travail refait entièrement pour les sujets à fort enjeu.

10 Questions d'examen

1. Vrai ou faux ? Une IA peut citer un article de loi qui n'existe pas.
2. Vrai ou faux ? Demander à l'IA de se vérifier suffit.
3. Vrai ou faux ? Une décision purement automatisée sur une personne doit lui être annoncée.
4. Vrai ou faux ? Les biais d'une IA se repèrent facilement sur un seul cas.
5. Vrai ou faux ? Le travailleur reste tenu de travailler avec soin quand il utilise une IA.
6. Vrai ou faux ? Les montants 2026 donnés par une IA sont toujours à jour.
7. Cas express : L'IA cite un arrêt du Tribunal fédéral : on peut le reprendre tel quel ?
8. Résolvez le cas type de la section 7, « Le courriel trop sûr de lui », questions a à c.
9. Citez les étapes de la méthode, dans l'ordre, et la base légale du chapitre.

11 Corrigés

1. Vrai : On parle d'invention ou d'hallucination.

2. Faux : Elle peut confirmer son erreur ; il faut la source.

3. Vrai : LPD art. 21.

4. Faux : Il faut comparer des groupes sur un échantillon.

5. Vrai : CO 321a.

6. Faux : Le modèle peut dater d'une année antérieure.

7. Non. Il faut vérifier son existence et son contenu sur bger.ch.

8.

1. Le congé de l'autre parent est de deux semaines (CO 329g) ; l'article 329f concerne le congé de maternité (14 semaines). Le chiffre et l'article sont faux.

2. En vérifiant l'article sur Fedlex, comme le prévoit la liste « avant d'envoyer ».

3. Tout article ou montant cité dans un courriel sortant est vérifié dans la source ; voir les congés familiaux.

9. 1) Vérifier chaque chiffre. ; 2) Retrouver chaque article. ; 3) Comparer les résultats par groupes. ; 4) Garder une personne pour décider. ; 5) Dater la source. Base : LPD art. 6 al. 5, 21 ; CO 101, 321a, 328, 330a.