Fiche mémo · Leçon 33 · 29 min

Extraire les données des factures et des documents reçus

Une facture reçue en PDF se transforme en ligne de tableau en trois temps : extraire le texte, faire remplir par l'IA un schéma JSON strict, puis laisser du code recalculer les totaux, vérifier le SIREN et chercher les doublons. La comptable valide. Les factures électroniques des plateformes agréées, elles, arrivent déjà structurées.

Objectifs

- Situer ce que la réforme de la facture électronique change pour vos pièces reçues
- Écrire le schéma JSON et la consigne qui extraient une facture sans inventer de valeur
- Programmer les contrôles automatiques qui arrêtent une pièce douteuse avant la saisie

À retenir

- Depuis le 1er septembre 2026, toute entreprise reçoit des factures électroniques par une plateforme agréée ; les PME émettent à partir du 1er septembre 2027.
- Une facture électronique arrive déjà en données structurées : l'extraction par IA sert aux PDF, aux notes de frais et aux documents.
- L'IA recopie dans un schéma JSON strict : montants en nombres, dates en AAAA-MM-JJ, null si absent, un booléen « incertain ».
- Le code recalcule les totaux, vérifie le SIREN et cherche les doublons ; une seule alerte suffit à poser « à vérifier ».
- Payer, changer un IBAN et supprimer une pièce restent hors du flux, au rouge.

La facture électronique en France

1er sept. 2026 — Réception pour toutes

Toute entreprise doit pouvoir recevoir des factures électroniques par une plateforme agréée.

1er sept. 2026 — Émission, grandes et ETI

Les grandes entreprises et les ETI émettent leurs factures sous forme électronique.

1er sept. 2027 — Émission, PME et micro

Les PME et les micro-entreprises émettent à leur tour, avec la transmission de certaines données à l'administration.

Calendrier publié par l'administration. Entre les deux dates, des PDF continueront d'arriver de vos fournisseurs PME.

La demande à recopier

Demande à recopier La consigne système de l'extraction

```
Tu extrais les données d'une pièce d'achat reçue par [nom de l'entreprise], SIREN [notre
SIREN]. Tu reçois le texte brut du PDF.
Règles :
1. Recopie, ne calcule pas. Si un total n'est pas écrit sur la pièce, mets null : un programme
le recalculera.
2. Montants : des nombres avec un point décimal, sans espace ni symbole (1 234,56 € devient
1234.56). Un avoir garde des montants positifs et type_piece vaut avoir.
3. Dates au format AAAA-MM-JJ. Une date écrite 03/11/2026 est le 3 novembre 2026.
4. Le fournisseur est l'émetteur de la pièce, jamais [nom de l'entreprise].
5. Une ligne de TVA par taux présent (20, 10, 5.5, 2.1).
6. incertain = true si le texte est illisible par endroits, si plusieurs factures semblent
présentes, ou si tu as hésité sur un montant.
N'invente aucune valeur absente du texte.
```

Note : Remplacez les crochets par votre raison sociale et votre SIREN : sans eux, l'IA confond parfois l'émetteur et le destinataire.

Les erreurs fréquentes

L'erreur, et pourquoi elle coûte	La correction
Demander à l'IA de calculer le TTC ou le total de TVA quand la pièce ne l'écrit pas, puis enregistrer son résultat. Un modèle de langage prédit des chiffres comme il prédit des mots. Un total faux de quelques euros passe inaperçu, et la ligne fausse part chez le comptable avec la même allure qu'une ligne juste.	La consigne dit « recopie, ne calcule pas » et met null aux totaux absents. Le calcul et la vérification se font dans un nœud Code ou une condition du flux.
Laisser les montants en texte dans le schéma (« 1 234,56 € »). Le tableur ne les additionne pas, le tri est faux, et les séparateurs varient d'un fournisseur à l'autre : 1.234 peut vouloir dire mille deux cent trente-quatre ou un virgule deux.	Type number dans le schéma, règle de conversion écrite dans la consigne (1 234,56 € devient 1234.56), et contrôle des totaux qui échoue si la conversion s'est mal faite.