



**MINISTÈRE  
DE L'AGRICULTURE,  
DE L'AGRO-ALIMENTAIRE  
ET DE LA SOUVERAINETÉ  
ALIMENTAIRE**

*Liberté  
Égalité  
Fraternité*

# **GÉNÉRATION AUTOMATIQUE DE COMMENTAIRES À PARTIR DE DONNÉES STATISTIQUES SUR LA PAC**

SSPHub 14/04/26

Par Antoine Lesauvage, à partir du travail de Léo Nivelles  
Chargé de mission datascience et innovation  
MAASA/SSP/DEMESIS/BMS

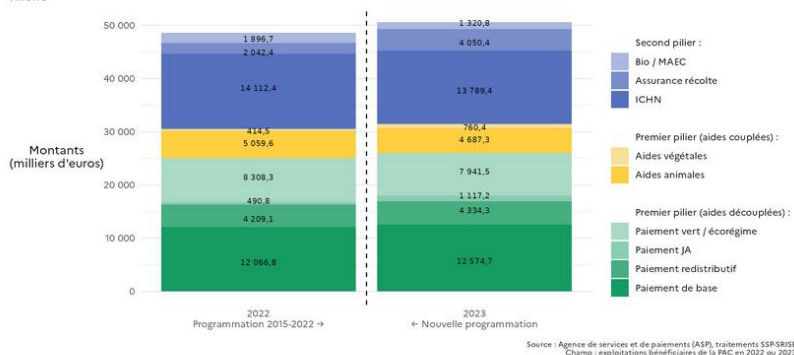
# Le projet initial de commentaires automatisés

Projet initial porté par la mission mutualisation du SSP (Michael DELORME).

## Description du projet initial :

- Publication de fiches récapitulatives de la PAC (Politique Agricole Commune) au niveau départemental.
- Publication de 5 graphiques par département + commentaires associés
- Départements × graphiques ≈ **500 commentaires** à réaliser au total
- + Rédaction manuelle parfois **fastidieuse** pour les agents en charge
- Exemple de graphique et commentaire associé :

Montant total des aides de la PAC en 2022 et 2023, par type d'aides  
Rhône



## Lecture

Le graphique montre une répartition des montants par type d'aide en 2022 et 2023 dans le département du Rhône. Les aides du premier pilier, notamment le paiement de base, le paiement vert, le paiement redistributif et le paiement JA, représentent la majorité des montants, avec une augmentation du paiement de base en 2023 (12 575 milliers d'euros contre 12 067 milliers d'euros en 2022). Concernant les aides couplées, les aides animales diminuent de 372 milliers d'euros tandis que les aides végétales augmentent de 346 milliers d'euros. Les aides du deuxième pilier, incluant ICHN, Bio/MAEC et assurance récolte, augmentent de 1 110 milliers d'euros : l'assurance récolte augmente de 2 008 milliers d'euros tandis que l'ICHN et les paiements Bio/MAEC diminuent respectivement de 323 et 575 milliers d'euros.

Figure 1 – Montant total des aides de la PAC en 2022 et 2023, par type d'aides

# Le projet initial de commentaires automatisés

Projet initial porté par la mission mutualisation du SSP (Michael DELORME).

## Description du projet initial :

- Publication de fiches récapitulatives de la PAC (Politique Agricole Commune) au niveau départemental.
- Publication de 5 graphiques par département + commentaires associés
- Départements x graphiques ≈ **500 commentaires** à réaliser au total
- + Rédaction manuelle parfois **fastidieuse** pour les agents en charge

⇒ Une automatisation via LLM de ces commentaires de graphique pourrait avoir un intérêt.

## Méthode :

- Utilisation de l'IA Mistral pour générer un commentaire à partir d'un graphique
- Un résumé par graphique
- Limite : 100 mots
- Contrôle de la qualité par les rédacteurs a posteriori

# Le projet initial de commentaires automatisés

## Méthode :

- Utilisation de l'IA Mistral pour générer un commentaire à partir d'un graphique
- Un résumé par graphique
- Limite : 100 mots
- Contrôle de la qualité par les rédacteurs a posteriori

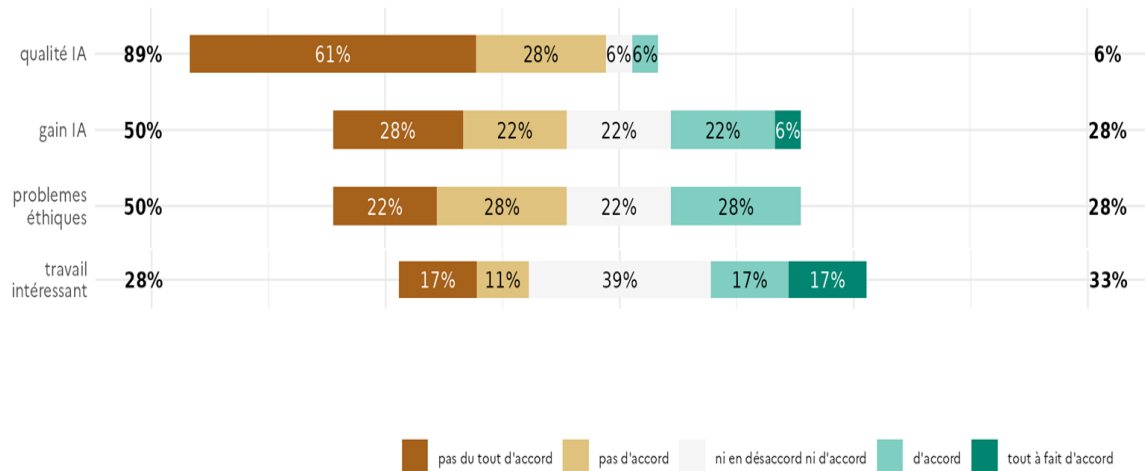
## Résultats :

- Forme plutôt correcte, fond incohérent
  - Réécriture lourde nécessaire
- ⇒ Très faible satisfaction des rédacteurs

## Problèmes rencontrés :

- Hallucinations sur les nombres
- Unités omises
- Tendances inversées
- Faits manquants/mal choisis
- Style pas adapté

RETEX Fiches PAC



# Le nouveau projet : améliorer la qualité des commentaires

Projet mené par Léo Nivelles, étudiant à l'École Polytechnique, sous la supervision de Ioana Manolescu, chercheuse à INRIA.

**Objectifs :** Générer des résumés descriptifs, factuels et contraints (100 mots) non conscris au cas d'usage des fichiers PAC (généralisable)

## **Documents :**

- 93 départements (hors 20 et 92).
- Par départements : 5 tables → 5 graphiques → 5 commentaires
- Deux types de tables :
  - petites : 6 à 9 indicateurs (montants 2022/2023),
  - grandes : tables descriptives OTEX + gain/pertes

**Méthode :** générer à partir des tables et non des graphiques, évaluer la qualité grâce à la précision numérique/tendance/complétude, améliorer la qualité, optimiser l'invite, ...

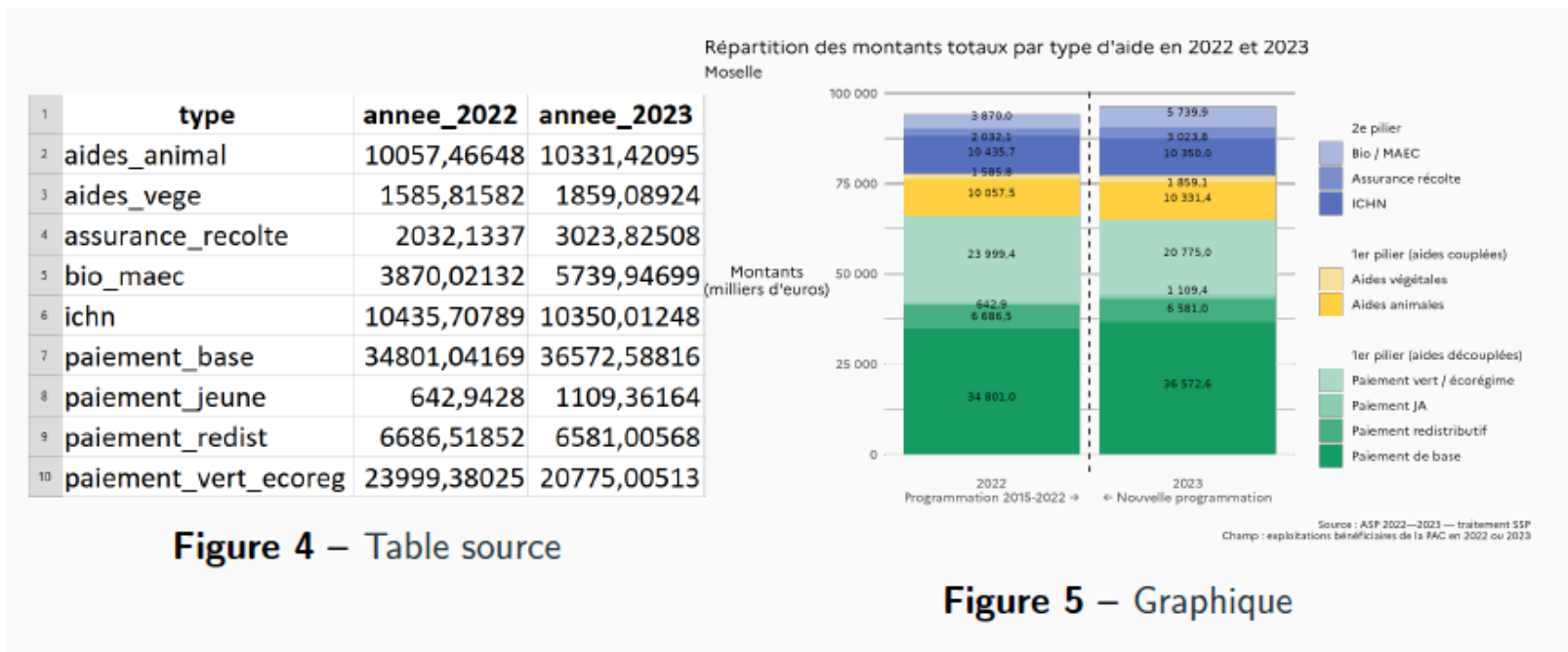
# Evaluer la qualité d'un commentaire

Procédure permettant d'évaluer la qualité d'un commentaire :

1. Entrée : texte généré par l'IA + tables de valeurs statistiques
2. Découpage : texte → phrases → segments (1 type ou OTEX par segment)
3. Correspondance : pour chaque segment, sélection des lignes de table verbalisées (type + nombres)
4. Assignation d'un score par segment : CV1 (nombres) + CV2 (tendance) + couverture des lignes verbalisées
5. Agrégation : moyenne des scores sur les segments ⇒ score global de la génération

# Evaluer la qualité d'un commentaire : exemple

Visualisation les 5 étapes sur le département 57 (Moselle) sur la figure 1 (paiements PAC 2022/2023).



# Evaluer la qualité d'un commentaire : exemple

Etape 1 : génération du texte via GPT4o

## Résumé généré par l'IA

*Le graphique montre une légère augmentation des montants totaux alloués en 2023 par rapport à 2022 en Moselle.*

*Dans le 1er pilier, les paiements de base, qui restent la part la plus importante des aides versées, augmentent de 34,8 millions d'euros en 2022 à 36,6 millions d'euros en 2023.*

*Les montants augmentent légèrement pour les aides végétales et jeunes agriculteurs, tandis qu'ils diminuent pour le paiement redistributif.*

*Le paiement vert, devenu éco-régime, a quant à lui connu une baisse, passant de 24 à 20,8 millions d'euros.*

*Pour le 2ème pilier, l'ICHN reste stable mais les montants BIO/MAEC et l'aide à l'assurance récolte augmentent.*

# Evaluer la qualité d'un commentaire : exemple

## Etape 2 : découpage du texte en segments

- Découpage en phrases
- Détection d'alias
- Construction de segments entre occurrences successives
- Un type canonique par segment (ex : paiement\_base, aides\_vege, ...)

### Texte annoté (aliases en rouge, segments séparés par |)

Le graphique montre une légère augmentation des montants totaux alloués en 2023 par rapport à 2022 en Moselle.

Dans le 1er pilier, les | paiements de base qui restent la part la plus importante des aides versées, augmentent de 34,8 millions d'euros en 2022 à 36,6 millions d'euros en 2023.

Les montants augmentent légèrement pour les | aides végétales et | jeunes agriculteurs tandis qu'ils diminuent pour le | paiement redistributif

Le | paiement vert devenu éco-régime, a quant à lui connu une baisse, passant de 24 à 20,8 millions d'euros.

Pour le 2ème pilier, | l'ICHN reste stable mais les montants | BIO/MAEC et l'aide à | l'assurance récolte augmentent.

# Evaluer la qualité d'un commentaire : exemple

Etape 3 : correspondance entre les segments et les lignes de la table initiale

Deux méthodes pour ce faire :

**Méthode 2 (type-based)** : tentative d'assignation des alias trouvés dans les segments aux lignes de la table

**Méthode 1 (matching numérique)** : extraction regex des nombres du segment puis sélection de cellules dans la table via un *tolerant match* entre un nombre extrait  $n$  et une cellule numérique candidate  $v$  tel que

$$|n - v| \leq 1 \quad \text{ou} \quad \frac{|n - v|}{|v|} \leq 3\%.$$

**Règle de résolution** : si on trouve < 3 lignes correspondant et pas « respectivement », on choisit la méthode 2

Si  $\geq 3$  lignes correspondant ou contient « respectivement », on choisit l'union entre méthode 1 et 2

Seg.	Segment	row_type (M2)	row_num (M1)	row_final
S1	paiements de base, qui restent la part la plus importante des aides versées, augmentent de 34,8 millions d'euros en 2022 à 36,6 millions d'euros en 2023	{7}	{7}	{7}
S2	aides végétales et	{3}	∅	{3}
S3	jeunes agriculteurs, tandis qu'ils diminuent pour le	{8}	∅	{8}
S4	paiement redistributif	{9}	∅	{9}
S5	paiement vert, devenu éco-régime, a quant à lui connu une baisse, passant de 24 à 20,8 millions d'euros	{10}	{10}	{10}
S6	ICHN reste stable mais les montants	{6}	∅	{6}
S7	BIO/MAEC et l'aide à l'	{5}	∅	{5}
S8	assurance récolte augmentent	{4}	∅	{4}

# Evaluer la qualité d'un commentaire : exemple

Etape 4 : évaluation numérique de la qualité des segments (CV1, CV2 et complétude)

## MÉTRIQUE CV1 : PRÉCISION NUMÉRIQUE

- Extraire les nombres du segment (hors années), normaliser l'unité (k€).
- Comparer chaque nombre extrait  $n$  à une cellule  $v$  d'une ligne correspondante (déterminées précédemment).
- Similarité :

$$s(n, v) = \begin{cases} 1 & \text{si } |n - v| \leq 1 \\ \max\left(0, 1 - \frac{|n-v|}{|v|}\right) & \text{sinon} \end{cases}$$

- $CV1(\text{segment}) = \min_n \max_v s(n, v)$  (pire nombre du segment).

## EXEMPLE FIL ROUGE (MOSELLE) : CV1

- Segment S1 : "34,8 M€" et "36,6 M€" →  $n_1 = 34800$ ,  $n_2 = 36600$  (k€).
- Ligne associée : (2022) 34801.04 ; (2023) 36572.59 (k€).
- Scores :  $s(n_1, \cdot)$  et  $s(n_2, \cdot)$  proches de 1 ⇒  $CV1(S1) \approx 0.999$ .

Seg.	Segment	CV1
S1	paiements de base ... augmentent de 34,8 M€ (2022) à 36,6 M€ (2023)	0.9993
S5	paiement vert ... baisse, passant de 24 à 20,8 M€	0.9988

# Evaluer la qualité d'un commentaire : exemple

Etape 4 : évaluation numérique de la qualité des segments (CV1, CV2 et complétude)

## MÉTRIQUE CV2 : FIDÉLITÉ | DE TENDANCE

- Détecter une tendance dans le texte (hausse / baisse) via lexique + lemmatisation.
- Gérer la négation (ex. "ne ... pas", "sans", "aucun").
- Comparer à la vérité :  $\Delta = v_{2023} - v_{2022}$  sur la (les) ligne(s) alignée(s).
- Score :  $CV2(\text{segment}) = 1$  si tendance correcte, sinon 0.

## EXEMPLE FIL ROUGE (MOSELLE) : CV2

- S1 contient "augmentent"  $\Rightarrow$  tendance = hausse.
- Table :  $v_{2023} > v_{2022} \Rightarrow$  hausse  $\Rightarrow CV2(S1) = 1$ .

Seg.	Segment	CV2
S1	paiements de base, qui restent la part la plus importante des aides versées, augmentent de 34,8 millions d'euros en 2022 à 36,6 millions d'euros en 2023	1
S3	jeunes agriculteurs, tandis qu'ils diminuent pour le	0
S8	assurance récolte augmentent	1

# Evaluer la qualité d'un commentaire : exemple

Etape 4 : évaluation numérique de la qualité des segments (CV1, CV2 et complétude)

## COMPLÉTUDE : COUVERTURE DE LA TABLE

- Une ligne est couverte si elle apparaît dans au moins une ligne liée à un segment.
- Petites tables : on attend tous les types.

$$\text{complétude} = \frac{\#\{\text{types couverts}\}}{\#\{\text{types totaux}\}}$$

- Grandes tables (OTEX) : on attend seulement 4 extrêmes (max 2023, min 2023, plus forte hausse, plus forte baisse).

## EXEMPLE FIL ROUGE (MOSELLE) : COMPLÉTUDE

- Table : 9 types présents.
- Génération : 8 types couverts (un type manquant).
- Score :  $\text{comp} = \frac{8}{9} \approx 0.889$ .

Ensemble	Taille
Types existants	9
Types couverts	8
Complétude	$8/9 \approx 0.889$

# Evaluer la qualité d'un commentaire : limites

Etape 5 : moyenne des scores sur les segments  $\Rightarrow$  score global de la génération

## Limites :

L'évaluation repose sur des appariements segment-ligne et des vérifications numériques et de tendance locales qui présente plusieurs limites :

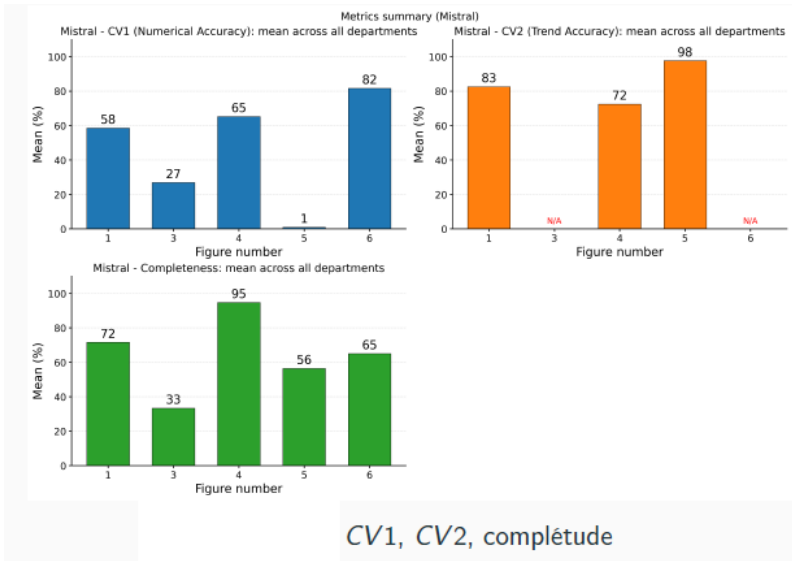
- Difficulté à gérer les multi-segments comme « respectivement », entraînant des inversions possibles
- Pénalisation des pourcentages, agrégats ou variations non directement présents dans le tableau
- Le CV1 peut parfois être optimiste lorsque plusieurs cellules sont similaires
- Mauvais rattachement des indices de tendance lorsque leur portée traverse plusieurs segments
- ...

En pratique, ces cas sont rares et les scores sont moyennés sur de nombreux segments, limitant l'impact global.

# Evaluer la qualité d'un commentaire : application

Application à l'évaluation de la qualité des commentaires générées via la baseline préalablement utilisée (V0)

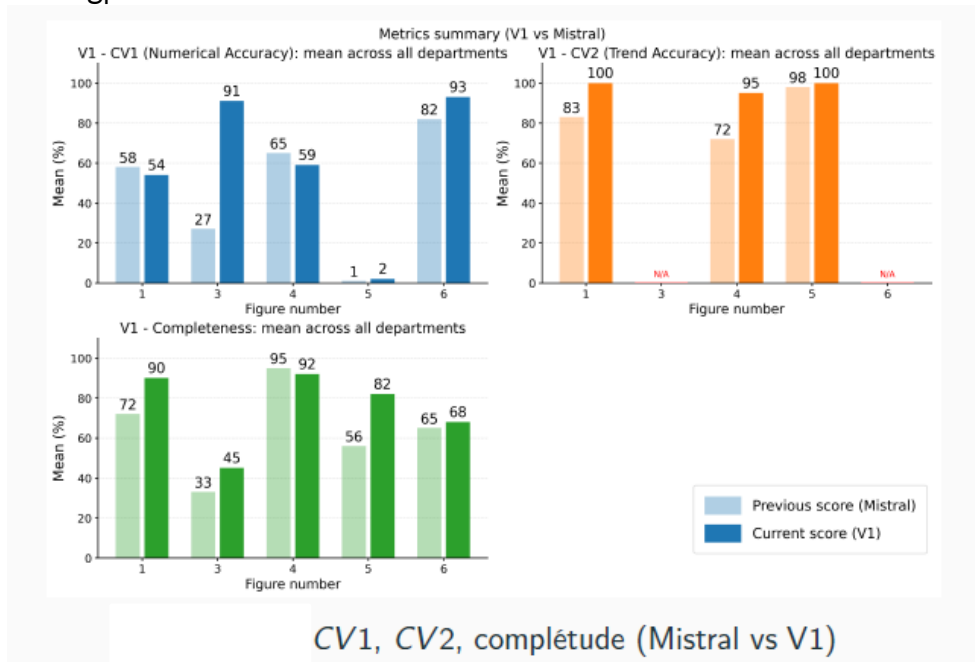
- Entrée : 5 graphiques par département, 1 requête par graphique
- Modèle d'IA utilisé : mistral-small-latest
- Instruction de prompt :
  - system commun (contexte PAC, règles de rédaction / formatage),
  - user spécifique à chaque figure (quoi commenter), limite 100 mots.



=> La précision numérique est globalement faible, la qualité générale n'est pas satisfaisante

# Proposition d'améliorations : V1

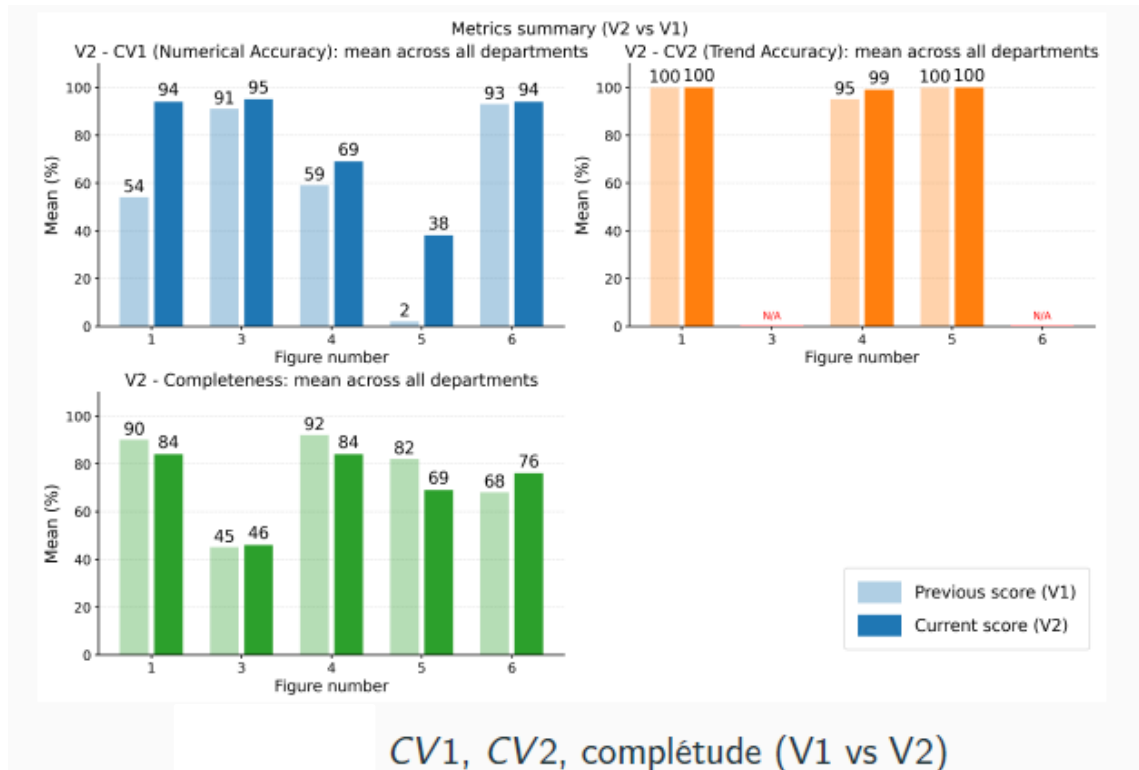
- Input : image ⇒ table (JSON)
- Objectif : isoler l'effet "lecture visuelle"
- Modèle : gpt-4o-mini



=> Meilleure qualité globale de commentaire mais toujours des erreurs et des hallucinations sur les chiffres, notamment pour la table 5 (grande table OTEX)

# Proposition d'améliorations : V2

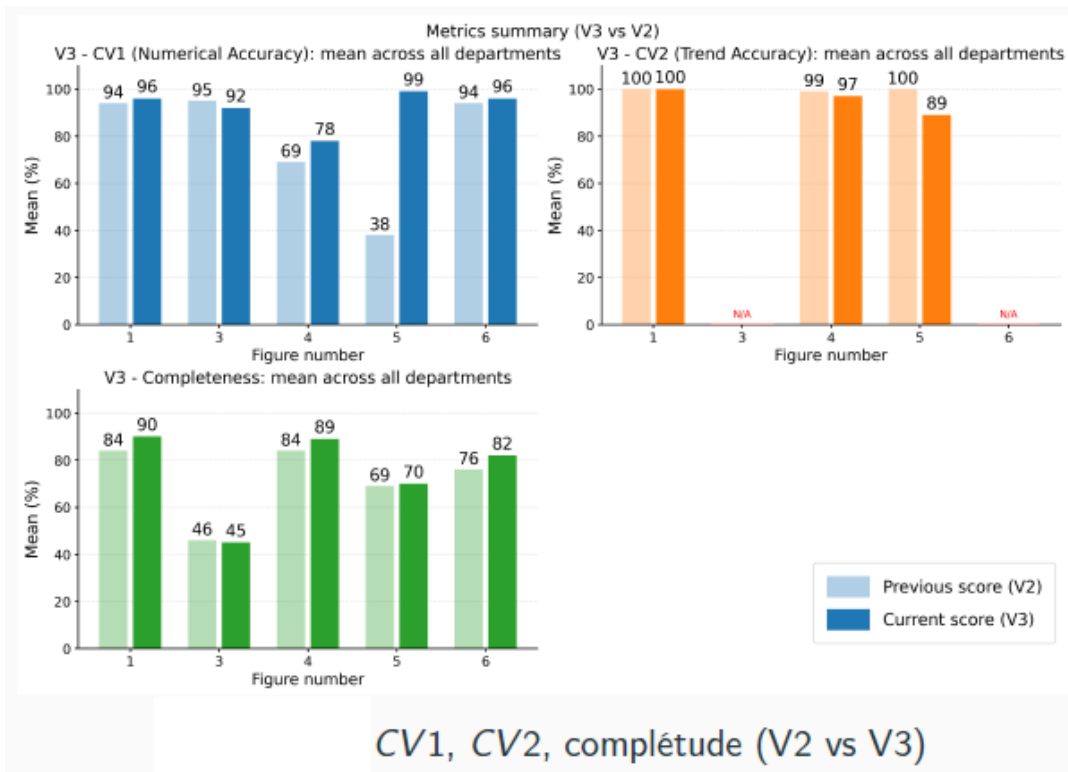
- V1 → V2 : instruction de prompt avec unité explicite (k€) + interdiction de calculs hors table.



⇒ Augmentation de la précision numérique au détriment de l'exhaustivité de la couverture des tables

# Proposition d'améliorations : V3

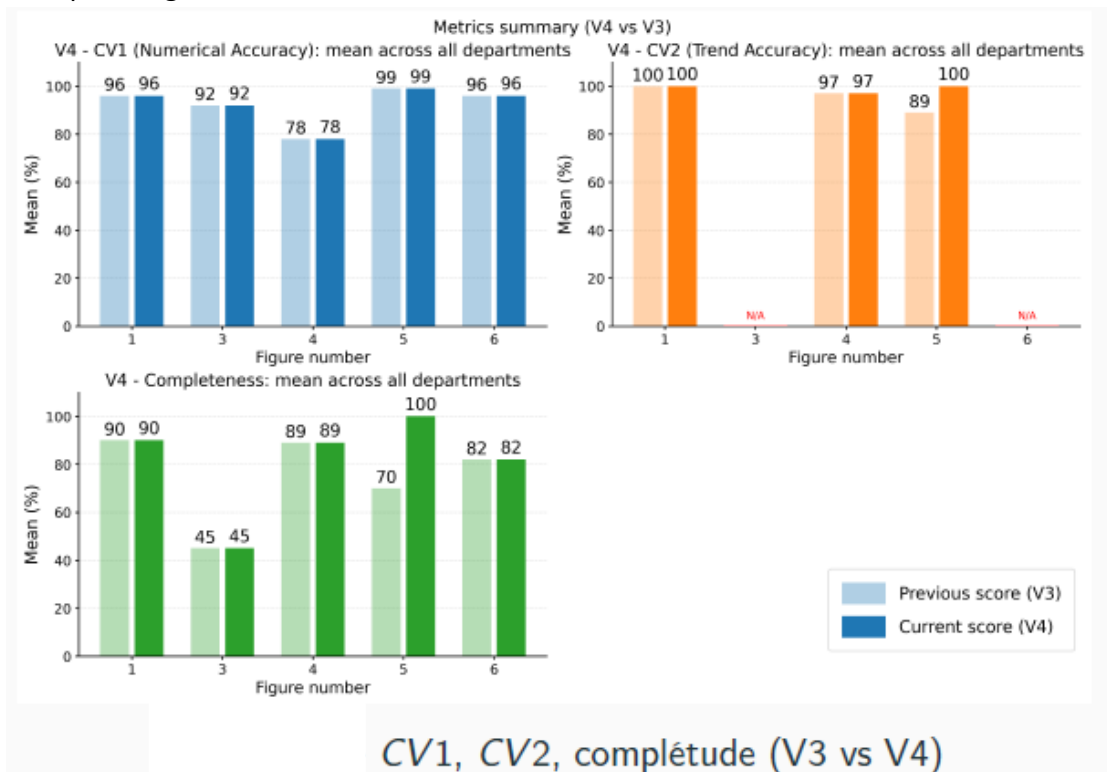
- V2 → V3 : arrondi entier des valeurs avant génération des commentaires.



=> Simple étape permet  
 d'augmenter grandement  
 la précision (notamment  
 pour la figure 5), toujours  
 des problèmes de  
 complétude

# Proposition d'améliorations : V4

- V3 → V4 : pour Fig. 5, 4 extrêmes OTEX fournis (sélection déterministe).



=> La présélection des lignes les plus pertinentes améliore la complétude pour la Fig. 5 en préservant la précision et le nombre de chiffres utilisés pour le commentaire

# Proposition d'améliorations : transfert de style éditorial V5 -> V7

**Objectif :** imiter la « plume » des rédacteurs sans perte de factualité

**V5 :** approche « few-shot » avec dans le prompt 3 résumés rédigés par le MAASA

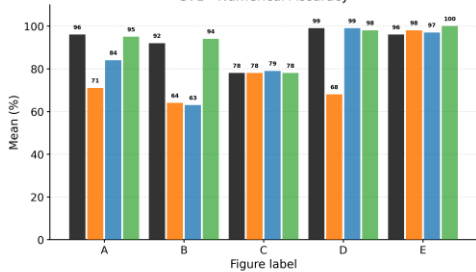
⇒ Dégradation des précisions numériques car le modèle hallucine en recopiant les valeurs des exemples + certaines tournures de phrases (« respectivement ») qui brise l'alignement entité-valeur utilisé pour l'évaluation

**V6 :** anonymisation des exemples (départements remplacés par "XXX", valeurs numériques par "...") pour réduire les hallucinations

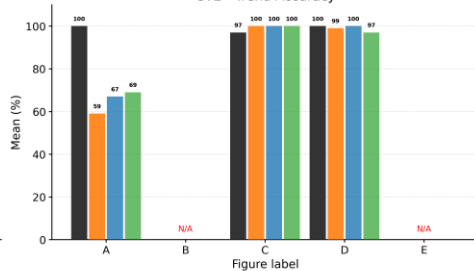
**V7 :** style contraint pour forcer le modèle à commenter des faits individuels, afin de faciliter

Comparaison des versions : V4 à V7

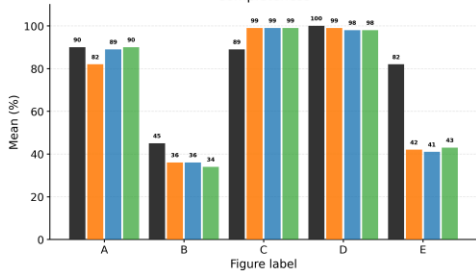
CV1 - Numerical Accuracy



CV2 - Trend Accuracy

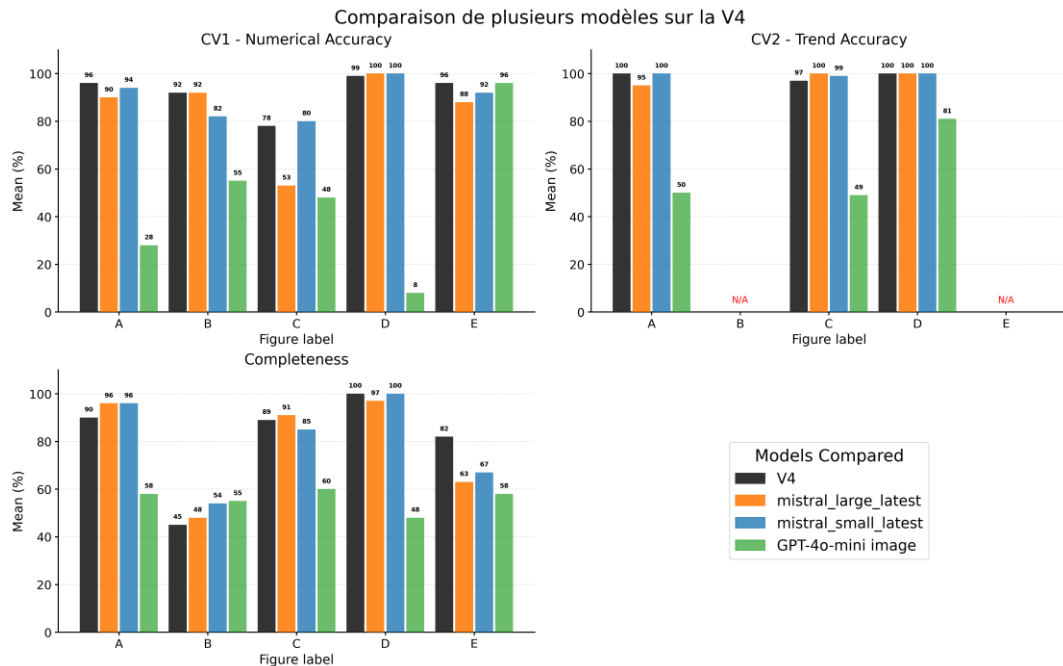


Completeness



⇒ Le transfert de style éditorial est possible sans pour autant réduire la rigueur factuelle

# Quel est l'impact du modèle ?



Des performances relativement comparables selon les modèles.

Pour la V7, les derniers modèles de Mistral (medium et large) ont une meilleure performance.

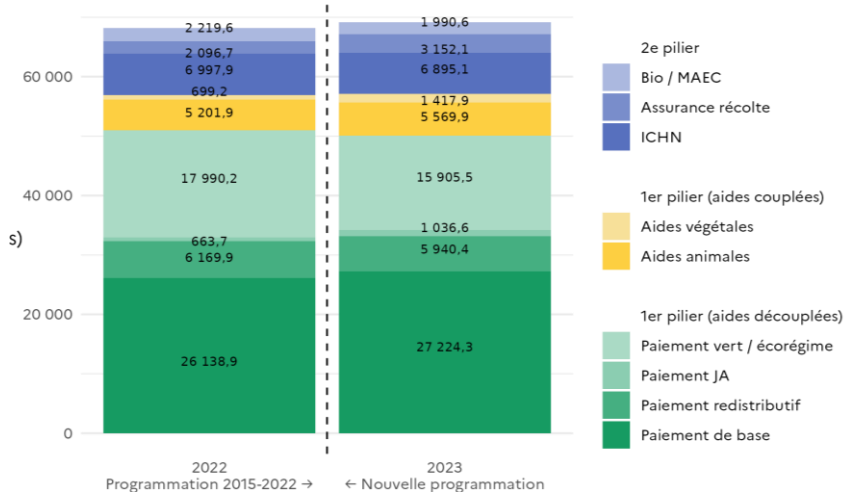
# Bilan des améliorations V0 vs V7

## V0

ChatGPT 4o mini

Le graphique montre une augmentation globale des montants alloués en 2023 par rapport à 2022. Les paiements de base restent la part la plus importante des aides, avec une légère augmentation de 26 139 à 27 224 milliers d'euros. Les paiements redistributifs et les paiements JA ont également connu une hausse significative, passant respectivement de 6 170 à 5 940 milliers d'euros et de 664 à 1 037 milliers d'euros. Les aides animales et végétales ont augmenté, tandis que les aides du deuxième pilier (ICHN, Bio/MAEC, Assurance récolte) ont globalement diminué.

des montants totaux par type d'aide en 2022 et 2023



Source : ASP 2022—2023 — traitement SSP  
 Champ : exploitations bénéficiaires de la PAC en 2022 ou 2023

## V7

Mistral medium

Le graphique présente l'évolution des paiements PAC entre 2022 et 2023 dans l'Ain, exprimée en milliers d'euros.

Pour le 1er pilier découplé, le paiement de base passe de 26 139 milliers d'€ en 2022 à 27 224 milliers d'€ en 2023. Le paiement vert/éco-régime recule, avec 17 990 milliers d'€ en 2022 contre 15 906 milliers d'€ en 2023. Le paiement redistributif diminue légèrement, passant de 6 170 à 5 940 milliers d'€. Le paiement aux jeunes agriculteurs (JA) augmente, avec 664 milliers d'€ en 2022 et 1 037 milliers d'€ en 2023.

Les aides couplées du 1er pilier voient les aides animales progresser de 5 202 milliers d'€ à 5 570 milliers d'€, tandis que les aides végétales passent de 699 milliers d'€ à 1 418 milliers d'€.

Dans le 2ème pilier, l'ICHN baisse légèrement, avec 6 998 milliers d'€ en 2022 contre 6 895 milliers d'€ en 2023. Les aides Bio/MAEC reculent, passant de 2 220 milliers d'€ à 1 991 milliers d'€. L'assurance récolte augmente significativement, de 2 097 milliers d'€ à 3 152 milliers d'€.

# Perspectives et conclusion

## Résumé de cette première étude :

- Baseline MAASA (Mistral, image→text) : non satisfaisante
  - ⇒ fortes erreurs numériques sur Fig. 3 & Fig.5 (lecture visuelle), erreurs également présentes pour les autres figures
- Génération à partir de tableaux (V1-V3) : améliore fortement la factualité
  - ⇒ les erreurs restantes viennent surtout de formatage/arrondis et de contraintes
- V4 : la sélection amont des 4 extrêmes résout le goulot de la sélection
  - ⇒ complétude Fig. 5 = 100% sans perte quantitative (et sous 100 mots)
- V5-V7 : un transfert de style éditorial est possible sans baisse de rigueur
  - ⇒ Des entretiens ont montré que les textes générés (Mistral Medium) sont maintenant de très bonne qualité

## Perspectives et pistes d'amélioration :

1. Évaluation étendue : créditer les énoncés dérivés (deltas, %, agrégats) via vérification sur indicateurs pré-calculés + prise en compte des tournures de phrases complexes
2. Améliorer le style éditorial si besoin
3. Statuer sur la généralisation de la pipeline
4. Quantifier l'apport de ces commentaires automatisés pour les rédacteurs