

# Rapport d'étude: Analyse des biais d'extraction de CV par LLM sur l'âge, le genre et l'origine

Laura BARDOU - Sacha BEAUJEAN - Justin COCHET GARSIOT

11 février 2026

## Résumé

Cette étude expérimentale vise à évaluer l'existence d'un biais dans l'extraction d'informations d'un CV par une intelligence artificielle, en fonction de l'âge, du sexe et de l'origine indiqué sur le CV. Pour chaque candidat, plusieurs CV sont générés. Chaque profil possède 4 CV : un CV dit "original", un CV où l'âge est modifié, un CV où le sexe est modifié, et un CV où l'origine est modifiée. L'IA réalise une extraction d'information sur chaque CV, et le résultat est comparé pour déterminer s'il est identique ou différent.

L'approche retenue est volontairement très exigeante, en raison du critère d'évaluation choisi : l'extraction est considérée comme correcte uniquement en cas d'égalité stricte (exact match) avec la référence. Ainsi, toute divergence, même marginale, est comptabilisée comme une erreur. Cette méthodologie repose sur une mesure binaire (0/1) plutôt que sur un score de similarité ou une métrique de distance, ce qui constitue un critère plus conservateur et plus sévère que les approches basées sur un taux de ressemblance.

## Table des matières

<b>I Première étude : Etude préliminaire</b>	<b>4</b>
<b>1 Méthodologie</b>	<b>4</b>
1.1 Constitution du Corpus et protocole . . . . .	4
1.1.1 Génération des CV constituant le corpus . . . . .	4
1.1.2 Protocole . . . . .	4
1.1.3 Détails sur le LLM et l'Audit Automatisé . . . . .	4
<b>2 Résultats Quantitatifs</b>	<b>5</b>
2.1 Performance Globale . . . . .	5
2.2 Résultats de l'étude . . . . .	5
2.2.1 Taux d'erreur par dimension . . . . .	5
2.2.2 Reproductibilité . . . . .	5
2.3 Types d'erreurs . . . . .	6
2.3.1 Distribution des types d'erreurs . . . . .	6
2.3.2 Erreurs par section et par biais . . . . .	6
<b>3 Interprétation Statistique</b>	<b>6</b>
3.1 Genre . . . . .	6
3.2 Âge . . . . .	6
3.3 Origine . . . . .	7

<b>4 Limites Identifiées</b>	<b>7</b>
4.1 Limites Méthodologiques . . . . .	7
4.1.1 Absence de Baseline A/A . . . . .	7
4.1.2 Audit par LLM (Biais de l'Auditeur) . . . . .	7
4.1.3 CV Synthétiques Homogènes . . . . .	7
4.1.4 Documents Limités . . . . .	7
4.1.5 Variables Confondues . . . . .	7
4.2 Limites Statistiques . . . . .	8
4.2.1 Taille d'Échantillon . . . . .	8
4.2.2 Pas de Stratification . . . . .	8
4.3 Limites Techniques . . . . .	8
4.3.1 Extraction Unique . . . . .	8
4.3.2 Modèle Unique . . . . .	8
<b>5 Conclusion de la phase préliminaire</b>	<b>8</b>
 <b>II Deuxième étude : Étude à grande échelle</b>	 <b>9</b>
<b>6 Méthodologie</b>	<b>9</b>
6.1 Constitution du corpus et protocole . . . . .	9
6.1.1 Génération des variantes de CV . . . . .	9
6.1.2 Protocole expérimental . . . . .	9
6.2 Spécifications techniques . . . . .	9
6.2.1 Configuration du LLM d'extraction . . . . .	9
6.2.2 Configuration de l'Audit Automatisé . . . . .	10
6.3 Détails sur les données et le nettoyage . . . . .	10
<b>7 Résultats Quantitatifs</b>	<b>10</b>
7.1 Performance Globale . . . . .	10
7.2 Stabilité des résultats (Analyse par Run) . . . . .	11
7.3 Analyse des Erreurs . . . . .	11
7.3.1 Distribution par type d'erreur . . . . .	11
7.3.2 Distribution par section du CV . . . . .	11
<b>8 Interprétation Statistique et Conclusion</b>	<b>12</b>
8.1 Dimension Genre . . . . .	12
8.2 Dimension Origine . . . . .	12
8.3 Dimension Âge . . . . .	12
<b>9 Analyse de la criticité des erreurs</b>	<b>12</b>
 <b>III Confiance des résultats</b>	 <b>13</b>
<b>10 Structure expérimentale et volumétrie</b>	<b>13</b>
<b>11 Modélisation statistique</b>	<b>13</b>
11.1 Proportion empirique d'erreur . . . . .	13
11.2 Intervalle de confiance de Wilson (95 %) . . . . .	13
11.3 Applications numériques . . . . .	13
11.3.1 Intervalle de confiance global . . . . .	13
11.3.2 Intervalle de confiance par niveau de criticité . . . . .	14

11.4 Conclusion sur les intervalles de confiance . . . . .	14
<b>12 Analyse du bruit de l'extraction</b>	<b>14</b>
12.1 Objectif . . . . .	14
12.2 Méthodologie . . . . .	14
12.3 Résultats . . . . .	14
12.4 Interprétation . . . . .	15
12.5 Impact sur l'interprétation des résultats globaux . . . . .	15
<b>13 Analyse du bruit de l'analyse (audit automatique)</b>	<b>15</b>
13.1 Objectif . . . . .	15
13.2 Méthodologie . . . . .	16
13.3 Résultats . . . . .	16
13.4 Interprétation . . . . .	16
13.5 Impact sur la validité des conclusions . . . . .	16
<b>IV Synthèse générale et recommandations</b>	<b>17</b>
<b>14 Principaux résultats</b>	<b>17</b>
14.1 Performance . . . . .	17
14.2 Limites . . . . .	17
<b>15 Interprétation et niveau de confiance</b>	<b>18</b>
<b>16 Recommandations et pistes d'amélioration</b>	<b>18</b>
<b>A Etude grande échelle : liste détaillée des erreurs</b>	<b>19</b>
<b>B Etude grande échelle : liste des statistiques et des erreurs par run</b>	<b>19</b>
B.1 Run 1 . . . . .	19
B.2 Run 2 . . . . .	21
B.3 Run 3 . . . . .	24
B.4 Run 4 . . . . .	27
B.5 Run 5 . . . . .	29
B.6 Run 6 . . . . .	32

## Première partie

# Première étude : Etude préliminaire

## 1 Méthodologie

L'objectif de cette étude est d'évaluer la fiabilité de l'extraction de données via le LLM de Semantikmatch en confrontant les résultats extraits à un fichier de référence (Vérité Terrain). Ce rapport présente les résultats d'une analyse technique portant sur la capacité du modèle de langage (LLM) à extraire des informations structurées (intérêts, études, expériences) à partir de diverses variantes de CV. Cette première étude compare 100 CV générés en 4 variantes différentes (Original, Âge, Sexe, Origine), avec 4 runs. Cet échantillon représente donc 1600 extractions.

### 1.1 Constitution du Corpus et protocole

#### 1.1.1 Génération des CV constituant le corpus

Chaque CV original contient une entête, des expériences professionnelles, des formations, et des centres d'intérêts. Pour cette étude préliminaire, les Cv sont volontairement peu fournis.

Les variations pour le genre se traduisent par une inversion du genre de base sur le CV original, et un changement de prénom. Le nom reste identique, ainsi que le reste du CV.

Les variations sur l'origine sont générées grâce à un algorithme de shuffling ; chaque CV change de pays d'origine, et garde les autres valeurs identiques.

Les variations pour l'âge remplacent l'âge original par un âge différent, qui se situe entre 22 et 30 ans.

Le choix de ces modèles s'est basé à la fois sur la recherche de profils variés et représentatifs.

#### 1.1.2 Protocole

TABLE 1 – Méthodologie de l'étude du biais d'extraction

Phase	Description et Paramètres
1. Génération	Création de 100 CV synthétiques avec 4 variantes : <i>Original, Genre modifié, Origine modifiée, Âge modifié</i> .
2. Extraction	Extraction par LLM (4 runs) vers 3 catégories : <i>Expériences, Formations, Centres d'intérêt</i> .
3. Audit	Comparaison automatique (Auditeur LLM) pour détecter : <i>Omissions, Hallucinations, Modifications</i> .
4. Statistiques	Tests de Fisher, Correction de Bonferroni, Intervalles de Wilson et Taille d'effet ( <i>h</i> de Cohen).

#### 1.1.3 Détails sur le LLM et l'Audit Automatisé

##### Extraction LLM

- Modèle utilisé : GPT-4o (Azure OpenAI)
- Paramètres :

Température : 0 (déterministe)

Format de sortie : JSON structuré

Prompt : Instructions précises pour extraire les 3 catégories

- **Volume de données** :  $100 \times 4 \times 3 = 1200$  extractions

### Audit Automatisé

- **Auditeur** : GPT-4 (Azure OpenAI)
- **Méthode** : Comparaison sémantique (non stricte)
- **Règles d'audit** :
  - Ignorer la ponctuation, les accents et les majuscules.
  - Accepter les variations géographiques (ex : Paris vs France).
  - Déetecter les différences sémantiques réelles.
- **Sorties (Outputs)** :
  - `coherent` : *true/false*
  - `error_type` : *None, Omission, Hallucination, Modification*
  - `details` : Description de l'écart constaté

## 2 Résultats Quantitatifs

### 2.1 Performance Globale

Le système Semantikmatch est largement équitable avec un biais léger mais reproductible sur la dimension "Origine" (taux net estimé à 1.6% après correction du bruit de fond). Ce biais reste à impact faible et acceptable pour la production, mais nécessite un monitoring continu.

TABLE 1 – SYNTHÈSE DES RÉSULTATS PAR DIMENSION

Dimension	Taux d'Erreur	Significatif ?	Cohen's h	Verdict
Genre	$2.00\% \pm 0.47\%$	50% des runs	0.141	Pas de biais
Âge	$1.58\% \pm 0.50\%$	0% des runs	0.108	Pas de biais
Origine	$2.58\% \pm 0.68\%$	75% des runs	0.179	Biais léger

### 2.2 Résultats de l'étude

#### 2.2.1 Taux d'erreur par dimension

TABLE 2 – Taux d'erreur par dimension et par run

Biais	Run1	Run2	Run3	Run4	Moyenne	Écart-type
Genre	2.00%	2.33%	1.33%	2.33%	<b>2.00%</b>	$\pm 0.47\%$
Origine	2.67%	3.33%	1.67%	2.67%	<b>2.58%</b>	$\pm 0.68\%$
Âge	2.00%	2.00%	1.00%	1.33%	<b>1.58%</b>	$\pm 0.50\%$

#### 2.2.2 Reproductibilité

TABLE 3 – REPRODUCTIBILITÉ DES BIAIS SUR 4 RUNS

Biais	Runs Significatifs	Taux	Verdict
Genre	2/4 (run2, run4)	50%	Non reproductible
Origine	<b>3/4 (run1, run2, run4)</b>	<b>75%</b>	<b>Reproductible</b>
Âge	0/4	0%	Non reproductible

## 2.3 Types d'erreurs

### 2.3.1 Distribution des types d'erreurs

TABLE 4 – Distribution des types d'erreurs

Type d'Erreur	Nombre	Pourcentage
Omissions	1/20	5%
Modifications	19/20	95%
Hallucinations	0/20	0%

Interprétation : Le système ne crée pas de fausses informations (0 hallucinations). Il modifie principalement des formulations (95%). On remarque aussi que très peu d'omissions sont complètes (5%).

### 2.3.2 Erreurs par section et par biais

TABLE 5 – Taux d'erreur par section et par biais

Section	Genre	Origine	Âge	Moyenne
Experiences	3.0%	2.0%	2.0%	2.3%
Interests	0.0%	0.0%	0.0%	0.0%
Studies	3.0%	6.0%	4.0%	4.3%

#### Observations clés :

- **Interests** : 0% d'erreur sur tous les biais (section parfaite).
- **Studies** : Section la plus problématique, surtout pour le biais d'« Origine » (6.0%).

## 3 Interprétation Statistique

### 3.1 Genre

- Taux moyen :  $2.00\% \pm 0.47\%$
- IC 95% :  $0.92\% - 4.74\%$
- Cohen's h : 0.141 (très petit effet)
- p-value (Bonferroni) : 0.138 (moyenne)

#### Conclusion : Pas de biais significatif.

Le taux d'erreur est faible et l'effet pratique négligeable.

### 3.2 Âge

- Taux moyen :  $1.58\% \pm 0.50\%$
- IC 95% :  $0.34\% - 4.29\%$
- Cohen's h : 0.108 (très petit effet)
- p-value (Bonferroni) : 0.325 (moyenne)

#### Conclusion : Pas de biais significatif.

Le taux d'erreur est le plus faible des 3 dimensions.

### 3.3 Origine

- Taux moyen :  $2.58\% \pm 0.68\%$
- IC 95% :  $0.71\% - 6.03\%$
- Cohen's  $h$  :  $0.179$  (petit effet)
- p-value (Bonferroni) :  $0.059$  (moyenne, proche du seuil)
- **Reproductible sur 75% des runs**

**Conclusion : Biais léger mais reproductible.**

Le système présente une tendance à modifier légèrement les informations liées à l'origine, particulièrement dans la section "Studies" (formations).

**Taux net estimé** (après correction d'un bruit de fond de  $\sim 1\%$ ) :  $\sim 1.6\%$ .

## 4 Limites Identifiées

### 4.1 Limites Méthodologiques

#### 4.1.1 Absence de Baseline A/A

- **Problème** : Pas de mesure du bruit de fond intrinsèque du système.
- **Impact** : On ne peut pas distinguer avec certitude le biais réel du bruit aléatoire du LLM.
- **Conséquence** : Le taux de  $2.58\%$  pour "Origine" peut inclure  $\sim 1\%$  de bruit, donnant un biais net de seulement  $\sim 1.6\%$ .

#### 4.1.2 Audit par LLM (Biais de l'Auditeur)

- **Problème** : L'auditeur (GPT-4) peut lui-même être biaisé.
- **Impact** : Risque de faux positifs (détecte des erreurs inexistantes) ou faux négatifs (rate des vraies erreurs).
- **Solution manquante** : Pas de validation humaine (gold standard).

#### 4.1.3 CV Synthétiques Homogènes

**Problème** : Les 100 CV ont une structure très similaire :

- Même format
- Même longueur ( $\sim 1$  page)
- Profils juniors uniquement (22-30 ans)
- Contenus générés automatiquement

**Impact** : Manque de diversité réelle, ne représente pas la variabilité des CV réels.

#### 4.1.4 Documents Limités

- **Problème** : L'étude se concentre uniquement sur les CV.
- **Manque** : Pas d'analyse sur les **bulletins de notes**, qui sont également traités par le système.

#### 4.1.5 Variables Confondues

**Problème** : Changer l'origine modifie aussi le contexte :

- Noms de villes/universités étrangères
- Patterns linguistiques différents
- Formations internationales

**Impact** : Difficile de distinguer un vrai biais discriminatoire d'une adaptation contextuelle légitime.

## 4.2 Limites Statistiques

### 4.2.1 Taille d'Échantillon

- $100 \text{ CV} \times 4 \text{ runs} = 400 \text{ CVs au total}$
- Puissance statistique limitée pour détecter de petits effets
- Variance inter-runs élevée (1.33% à 3.33% pour Origine)

### 4.2.2 Pas de Stratification

- Pas de contrôle pour d'autres variables (secteur d'activité, niveau d'expérience, type de formation)
- Impossible d'isoler les effets

## 4.3 Limites Techniques

### 4.3.1 Extraction Unique

- Chaque CV extrait une seule fois par run
- Pas de mesure de la reproductibilité intra-CV
- On ne sait pas si le système est stable sur le même document

### 4.3.2 Modèle Unique

- Testé uniquement sur GPT-4o (Azure)
- Pas de comparaison avec d'autres modèles (Claude, Llama, etc.)

## 5 Conclusion de la phase préliminaire

L'évaluation menée sur le système d'extraction de données révèle un dispositif globalement robuste et fiable. La méthodologie expérimentale, basée sur une comparaison contrôlée, a permis d'analyser un échantillon significatif de 1600 extractions réparties sur 4 runs.

L'un des enseignements majeurs est que le système est largement équitable : deux dimensions sur trois — le Genre et l'Âge — ne présentent aucun biais significatif. Les taux d'erreur observés pour ces catégories sont faibles et les tests confirment l'absence d'effet pratique notable.

Toutefois, plusieurs points d'attention ont été identifiés :

Biais sur l'« Origine » : Un biais léger mais reproductible (dans 75% des runs) a été mesuré à 2,58%. Après correction du bruit de fond expérimental estimé à environ 1%, le biais net réel se situerait autour de 1,6%.

Limites des données : L'utilisation de CV synthétiques homogènes limite la représentativité de l'étude face à la diversité des documents réels.

## Deuxième partie

# Deuxième étude : Étude à grande échelle

## 6 Méthodologie

L'objectif de cette étude est d'évaluer la fiabilité de l'extraction de données via le LLM de Semantikmatch en confrontant les résultats obtenus à un fichier de référence (*Ground Truth* ou Vérité Terrain). Ce chapitre présente les résultats d'une analyse technique portant sur la capacité du modèle à extraire des informations structurées (intérêts, études, expériences) à partir de diverses variantes de CV.

L'étude comparative porte sur un corpus initial de 299 CV, déclinés en 4 variantes (Original, Âge, Sexe, Origine), sur lesquels 6 itérations d'extraction (runs) ont été effectuées.

### 6.1 Constitution du corpus et protocole

#### 6.1.1 Génération des variantes de CV

Chaque CV original est constitué, *a minima*, d'une entête, d'expériences professionnelles, de formations et de centres d'intérêts. Le choix des modèles originaux s'est basé sur la recherche de profils à la fois variés et représentatifs. Les variantes ont été générées selon les règles suivantes :

- **Variation de Genre** : Inversion du genre du profil original et modification du prénom associé. Le nom de famille et le reste du contenu demeurent identiques.
- **Variation d'Origine** : Application d'un algorithme de brassage (*shuffling*) modifiant le pays d'origine de chaque CV tout en conservant les autres données.
- **Variation d'Âge** : Remplacement de l'âge original par une valeur aléatoire comprise entre 22 et 30 ans.

#### 6.1.2 Protocole expérimental

Le tableau ci-dessous synthétise les quatre phases méthodologiques de l'étude.

TABLE 6 – Méthodologie de l'étude du biais d'extraction

Phase	Description et Paramètres
1. Génération	Création de 299 CV synthétiques déclinés en 4 variantes : <i>Original</i> , <i>Genre modifié</i> , <i>Origine modifiée</i> , <i>Âge modifié</i> .
2. Extraction	Extraction par LLM (6 runs) ciblant 3 catégories : <i>Expériences</i> , <i>Formations</i> , <i>Centres d'intérêt</i> .
3. Audit	Comparaison automatique via un auditeur LLM pour détecter : <i>Omissions</i> , <i>Hallucinations</i> , <i>Modifications</i> .
4. Statistiques	Application de tests de Fisher, correction de Bonferroni, intervalles de Wilson et calcul de la taille d'effet ( <i>h</i> de Cohen).

### 6.2 Spécifications techniques

#### 6.2.1 Configuration du LLM d'extraction

- **Modèle** : GPT-4o (Azure OpenAI).
- **Volume de données** : 299 CV × 4 variantes × 6 runs = 7 176 extractions.
- **Paramètres** :

**Température** : 0 (pour maximiser le déterminisme).

**Format de sortie** : JSON structuré.

**Prompt** : Instructions spécifiques pour l'extraction des 3 catégories cibles.

### 6.2.2 Configuration de l'Audit Automatisé

L'audit est réalisé par une comparaison sémantique (non stricte) via GPT-4 (Azure OpenAI).

— **Règles d'audit** :

- Ignorer la ponctuation, les accents et la casse (majuscules/minuscules).
- Accepter les variations géographiques (ex : Paris vs France).
- Déetecter uniquement les différences sémantiques significatives.

— **Sorties (Outputs)** :

**coherent** : Booléen (*true/false*).

**error\_type** : Catégorisation (*None, Omission, Hallucination, Modification*).

**details** : Description textuelle de l'écart constaté.

## 6.3 Détails sur les données et le nettoyage

L'étude a porté sur un corpus initial de **299 profils**, réduit à un échantillon exploitable de 285 profils après exclusion de 14 cas non utilisables. Chaque profil (original et ses trois déclinaisons) s'articule autour des sections *Studies*, *Experiences* et *Interests*.

Le protocole a consisté en l'exécution de **6 itérations complètes**. Après nettoyage des données (exclusion de 18 champs vides dans les originaux), le contenu des déclinaisons a été comparé à celui des versions originales. Au terme de ces 6 itérations, un total de **15 282 comparaisons** a été effectué, réparties comme suit :

TABLE 7 – Volume total de comparaisons (sur 6 itérations)

Type de section	Comparaisons par run	Total (6 runs)
Section <i>Experiences</i>	855	5 130
Section <i>Studies</i>	846	5 076
Section <i>Interests</i>	846	5 076
<b>Total général</b>	<b>2 547</b>	<b>15 282</b>

## 7 Résultats Quantitatifs

### 7.1 Performance Globale

Après les 6 runs, 234 erreurs ont été signalées sur un total de 15 282 comparaisons, soit 1.53 % d'erreur.

Le tableau suivant présente la synthèse des taux d'erreur et la mesure de l'effet (*h* de Cohen) pour chaque dimension étudiée.

TABLE 8 – Synthèse des résultats par biais

Dimension	Taux d'Erreur	Cohen's h	Conclusion
Genre	2.75 % $\pm$ 0.19 %	0.141	Absence de biais
Origine	0.73 % $\pm$ 0.09 %	0.101	Absence de biais
Âge	1.12 % $\pm$ 0.18 %	0.081	Absence de biais

Les résultats indiquent un taux d'erreur globalement faible. La dimension « Origine » présente la meilleure performance (0.73 %), tandis que la dimension « Genre » affiche le taux le plus élevé (2.75 %), bien que celui-ci reste contenu. Le  $h$  de Cohen, systématiquement bas, confirme que les différences d'erreurs entre les groupes sont négligeables : le système traite les profils de manière équitable, indépendamment du genre, de l'origine ou de l'âge.

## 7.2 Stabilité des résultats (Analyse par Run)

TABLE 9 – Taux d'erreur par biais et par run

Biais	Run1	Run2	Run3	Run4	Run5	Run6	Moy.	Écart-type
Genre	2.59 %	2.47 %	2.71 %	2.83 %	2.94 %	2.94 %	<b>2.75 %</b>	±0.19 %
Origine	0.59 %	0.82 %	0.71 %	0.71 %	0.71 %	0.82 %	<b>0.73 %</b>	±0.09 %
Âge	1.30 %	1.30 %	1.06 %	1.06 %	0.82 %	1.18 %	<b>1.12 %</b>	±0.18 %

L'analyse par itération démontre la stabilité du modèle. L'écart-type intra-run est faible pour toutes les catégories, confirmant la reproductibilité des résultats. On observe de manière constante que le biais de genre génère légèrement plus d'erreurs que les autres dimensions.

## 7.3 Analyse des Erreurs

### 7.3.1 Distribution par type d'erreur

TABLE 10 – Typologie des erreurs rencontrées

Type d'Erreur	Proportion	Pourcentage
Modifications	145/234	62.0 %
Omissions	87/234	37.2 %
Hallucinations	2/234	0.8 %

L'analyse qualitative montre que le système tend majoritairement à altérer l'information existante (Modifications : 62 %) plutôt qu'à l'omettre (37.2 %). Les hallucinations (génération d'informations fausses) demeurent anecdotiques (0.8 %).

### 7.3.2 Distribution par section du CV

TABLE 11 – Répartition des erreurs par section

Section	Proportion	Pourcentage
Studies	128/234	54.7 %
Experiences	56/234	23.9 %
Interests	50/234	21.4 %

Plus de la moitié des erreurs (54.7 %) se concentrent sur la section *Studies*. Cette prévalence s'explique probablement par la grande variabilité des formats de diplômes et de parcours académiques, complexifiant l'extraction standardisée. Les sections *Experiences* et *Interests* présentent des taux d'erreurs comparables et nettement inférieurs.

## 8 Interprétation Statistique et Conclusion

L'analyse statistique détaillée par dimension permet de conclure à l'absence de biais significatif pour l'ensemble des critères étudiés.

### 8.1 Dimension Genre

- **Taux moyen** :  $2.75\% \pm 0.19\%$
- **IC 95 %** :  $[2.56\% - 2.94\%]$
- **Cohen's h** : 0.141
- **p-value (Bonferroni)** :  $> 0.05$  (non significatif)

**Conclusion :** Bien que le taux d'erreur soit légèrement supérieur aux autres dimensions, la taille d'effet ( $h < 0.2$ ) confirme l'absence de biais majeur.

### 8.2 Dimension Origine

- **Taux moyen** :  $0.73\% \pm 0.09\%$
- **IC 95 %** :  $[0.64\% - 0.82\%]$
- **Cohen's h** : 0.101
- **p-value (Bonferroni)** :  $> 0.05$  (non significatif)

**Conclusion :** Cette dimension est la plus robuste. L'origine du candidat n'a aucune influence détectable sur la performance du modèle.

### 8.3 Dimension Âge

- **Taux moyen** :  $1.12\% \pm 0.18\%$
- **IC 95 %** :  $[0.94\% - 1.30\%]$
- **Cohen's h** : 0.081
- **p-value (Bonferroni)** :  $> 0.05$  (non significatif)

**Conclusion :** Avec un effet moyen très faible ( $h = 0.081$ ), bien en dessous du seuil d'alerte de 0.2, le modèle démontre une forte neutralité par rapport à l'âge.

## 9 Analyse de la criticité des erreurs

Le niveau de criticité des erreurs se concentre autour du niveau 2 (erreur légère) avec 45 % des erreurs (105 cas sur 234), impactant principalement la précision des champs secondaires comme les intérêts ou la géographie. Le LLM oublie le plus souvent une ville, ou des précisions sur des centres d'intérêts. Les erreurs majeures s'élèvent à 16 %. Elles touchent à l'intégrité des données professionnelles ou académiques (suppression d'expériences ou de diplômes), mais restent minoritaires.

TABLE 12 – Synthèse de la criticité des erreurs identifiées

Niveau de criticité	Proportion	%	Impact Principal
1 - Mineure	14 / 234	6 %	Lisibilité (coquilles, format)
2 - Légère	105 / 234	45 %	Précision (champs géographiques, intérêts)
3 - Modérée	77 / 234	33 %	Sémantique (domaines modifiés, hallucinations)
4 - Majeure	38 / 234	16 %	Intégrité (expériences ou diplômes supprimés)

# Troisième partie

## Confiance des résultats

### 10 Structure expérimentale et volumétrie

Le volume d'observations varie selon l'étude considérée. Chaque observation représente une comparaison unique entre une extraction produite par le LLM sur un CV modifié et la donnée extraite du CV non modifié.

- **Étude générique :** Définie par  $N$  candidats,  $V$  variantes et  $R$  runs, soit  $n_{CV} = N \times V \times R$ .
- **Étude à grande échelle :**  $n_{CV} = 299 \text{ CV} \times 4 \text{ variantes} \times 6 \text{ runs} = 7176 \text{ CV}$ , générant un total de comparaisons  $n = 15\,282$  (cf. méthodologie).

**Hypothèse d'indépendance :** Les itérations (runs) sont indépendantes les unes des autres, le modèle (LLM) ne conservant aucune mémoire entre les exécutions (absence de contexte partagé). Par conséquent, l'ensemble des observations est considéré comme statistiquement indépendant.

### 11 Modélisation statistique

#### 11.1 Proportion empirique d'erreur

Pour un biais ou une erreur donnée, nous observons  $x$  erreurs sur  $n$  comparaisons indépendantes. La proportion empirique d'erreur est définie par :

$$\hat{p} = \frac{x}{n} \quad (1)$$

L'objectif est d'estimer l'intervalle de confiance dans lequel se situe la probabilité réelle d'erreur du système, en tenant compte de la taille finie de l'échantillon.

#### 11.2 Intervalle de confiance de Wilson (95 %)

Afin d'obtenir une estimation plus robuste que l'approximation normale classique (particulièrement pour des proportions faibles proches de 0), nous utilisons l'intervalle de confiance de Wilson. Le niveau de confiance est fixé à 95 %, impliquant un score standard  $z \approx 1.96$  (quantile de la loi normale centrée réduite tel que 95 % de la distribution se situe dans  $[-z, z]$ ).

La formule de l'intervalle de confiance  $IC_{95\%}$  est donnée par :

$$IC_{95\%} = \left[ \frac{\hat{p} + \frac{z^2}{2n} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}}}{1 + \frac{z^2}{n}} ; \frac{\hat{p} + \frac{z^2}{2n} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}}}{1 + \frac{z^2}{n}} \right] \quad (2)$$

#### 11.3 Applications numériques

##### 11.3.1 Intervalle de confiance global

Données :  $n = 15\,282$ ,  $x = 234$ .

La proportion d'erreur observée est :

$$\hat{p} = \frac{234}{15\,282} \approx 1.53 \%$$

L'intervalle de confiance de Wilson à 95 % est estimé à :

$$IC_{95\%} \approx [1.35\%; 1.74\%] \quad (3)$$

**Interprétation :** L'augmentation significative de la taille de l'échantillon ( $n > 15\,000$ ) entraîne un resserrement notable de l'intervalle de confiance. Cette précision accrue confirme la fiabilité de la mesure du taux d'erreur réel du système.

### 11.3.2 Intervalle de confiance par niveau de criticité

Avec les différents niveaux de criticité établis, on obtient ceci :

TABLE 13 – Résumé des intervalles de confiance de Wilson à 95 % avec niveau de criticité associé.

Niveaux de criticité des erreurs	IC 95 % Wilson
Majeure (4)	[0.18 %, 0.34 %]
Modérée et Majeure (3 et 4)	[0.63 %, 0.90 %]
Légère à Majeure (2, 3 et 4)	[1.26 %, 1.64 %]
Toutes les erreurs	[1.35 %, 1.74 %]

## 11.4 Conclusion sur les intervalles de confiance

L'utilisation de l'intervalle de Wilson permet d'encadrer de manière robuste le taux d'erreur réel du système. L'intervalle obtenu indique que, dans des conditions expérimentales identiques, le taux d'erreur réel du système se situe avec une probabilité de 95 % entre **1.35 % et 1.74 %**. Le taux d'erreurs majeures se situe avec une probabilité de 95 % entre **0.18 % et 0.34 %**. Ces estimations constituent des bornes fiables pour l'évaluation des performances du système à grande échelle.

## 12 Analyse du bruit de l'extraction

### 12.1 Objectif

Cette sous-partie vise à quantifier le **bruit intrinsèque du processus d'extraction**, c'est-à-dire la variabilité des sorties du LLM lorsque l'entrée est strictement identique. L'objectif est d'isoler la part d'erreur imputable au non-déterminisme résiduel du modèle d'extraction, indépendamment de toute variation de contenu ou de biais étudié (genre, âge, origine).

### 12.2 Méthodologie

Un sous-échantillon de **100 CV** a été sélectionné aléatoirement à partir du corpus global. Pour chacun de ces CV, une **vérité terrain** (*Ground Truth*) a été construite manuellement, correspondant à une extraction parfaite et exhaustive des informations cibles (*Studies, Experiences, Interests*).

Les **6 extractions originales** produites lors de l'étude principale (6 runs indépendants, paramètres strictement identiques, température fixée à 0) ont ensuite été comparées à ce fichier de référence. La comparaison a été réalisée automatiquement via un audit LLM, selon les mêmes règles que celles utilisées dans l'étude principale (comparaison sémantique non stricte, tolérance aux variations non significatives).

Chaque CV est considéré comme incorrect dès lors qu'au moins une différence sémantique significative est détectée par rapport à la vérité terrain.

### 12.3 Résultats

Les taux d'erreur observés pour chaque run sont présentés ci-dessous :

TABLE 14 – Bruit de l'extraction : taux d'erreur par run

Run	CV incorrects	Taux d'erreur
Run 1	0 / 100	0.00 %
Run 2	2 / 100	2.00 %
Run 3	0 / 100	0.00 %
Run 4	2 / 100	2.00 %
Run 5	1 / 100	1.00 %
Run 6	0 / 100	0.00 %
<b>Moyenne</b>	<b>5 / 600</b>	<b>0.83 %</b>

Sur l'ensemble des runs, le taux d'erreur moyen lié au bruit de l'extraction est de **0.83 %**. Trois runs sur six ne présentent aucune erreur, confirmant un haut niveau de stabilité du processus d'extraction lorsque les entrées sont identiques.

## 12.4 Interprétation

Ces résultats montrent que le **bruit d'extraction est faible mais non nul**, même avec une température fixée à 0. Les erreurs observées sont sporadiques, non systématiques et ne se reproduisent pas de manière cohérente d'un run à l'autre, ce qui indique qu'elles relèvent d'un **non-déterminisme résiduel du modèle** plutôt que d'un défaut structurel de l'extraction.

Ce bruit constitue une **borne inférieure incompressible** du taux d'erreur global observé dans l'étude à grande échelle. En d'autres termes, une partie du taux d'erreur global mesuré (**1.53 %**) ne peut être attribuée ni aux biais étudiés, ni aux variations de contenu des CV, mais uniquement à la variabilité intrinsèque du LLM.

## 12.5 Impact sur l'interprétation des résultats globaux

La mise en évidence d'un bruit d'extraction inférieur à **1 %** renforce la robustesse des conclusions de l'étude :

- les comparaisons entre variantes (genre, âge, origine) sont réalisées au-dessus du niveau de bruit intrinsèque ;
- les différences observées entre dimensions restent largement inférieures aux seuils de significativité statistique ;
- les conclusions d'absence de biais significatif sont donc **conservatrices** et non sous-estimées.

En conséquence, le taux d'erreur mesuré dans l'étude principale peut être interprété comme une **borne supérieure** de l'erreur réelle du système, intégrant à la fois le bruit d'extraction et les erreurs effectives liées à la complexité des données.

# 13 Analyse du bruit de l'analyse (audit automatique)

## 13.1 Objectif

Cette sous-partie vise à quantifier le **bruit introduit par le processus d'analyse automatique**, c'est-à-dire la variabilité des décisions produites par l'auditeur LLM lorsqu'il compare une extraction donnée à une vérité terrain identique. L'objectif est d'évaluer la fiabilité de l'outil de mesure lui-même, indépendamment du bruit d'extraction et des biais étudiés.

## 13.2 Méthodologie

Afin d'isoler le bruit propre à l'analyse, une extraction unique (issue d'un run donné) a été comparée à sa vérité terrain correspondante à **5 reprises successives**, dans des conditions strictement identiques (mêmes entrées, mêmes règles d'audit, absence de contexte partagé).

Les sorties de l'auditeur LLM ont ensuite été confrontées à une **vérité terrain d'analyse**, construite séparément, définissant ce qui devait être considéré comme une erreur réelle ou une extraction cohérente. Le taux de succès correspond à la proportion de décisions de l'auditeur conformes à cette vérité terrain d'analyse.

## 13.3 Résultats

Les taux de succès observés pour chacune des cinq répétitions sont les suivants :

TABLE 15 – Bruit de l'analyse : taux de succès par répétition

Répétition	Taux de succès
Répétition 1	99.00 %
Répétition 2	98.00 %
Répétition 3	97.00 %
Répétition 4	98.00 %
Répétition 5	99.00 %
<b>Moyenne</b>	<b>98.20 %</b>

La variabilité observée entre les répétitions est faible, avec un écart maximal de **2 points de pourcentage** entre le minimum et le maximum.

## 13.4 Interprétation

Ces résultats montrent que l'auditeur automatique présente un **haut niveau de fiabilité**, avec un taux de succès moyen supérieur à **98 %**. Les divergences observées entre répétitions relèvent d'un **bruit d'analyse résiduel**, inhérent à l'évaluation sémantique réalisée par un LLM, et ne traduisent pas une instabilité systématique du processus d'audit.

Ce bruit d'analyse, bien que limité, implique que le taux d'erreur mesuré dans l'étude principale peut inclure une faible proportion de **faux positifs ou faux négatifs d'audit**. Par conséquent, l'erreur observée constitue une **borne supérieure conservative** de l'erreur réelle du système d'extraction.

## 13.5 Impact sur la validité des conclusions

La combinaison d'un bruit d'extraction inférieur à **1 %** et d'un bruit d'analyse limité (taux de succès supérieur à **98 %**) permet d'affirmer que :

- l'essentiel des erreurs observées dans l'étude à grande échelle correspond à des écarts réels d'extraction ;
- la variabilité introduite par l'outil de mesure reste marginale au regard du volume total de comparaisons ;
- les conclusions relatives à l'absence de biais significatif selon le genre, l'origine ou l'âge sont **robustes et conservatrices**.

Ainsi, les résultats présentés dans cette étude reposent sur un dispositif expérimental dont les principales sources de bruit ont été identifiées, quantifiées et intégrées dans l'interprétation statistique globale.

# Quatrième partie

## Synthèse générale et recommandations

Cette étude visait à évaluer la robustesse et la fiabilité du système d'extraction automatique de CV de Semantikmatch basé sur un modèle de langage (LLM), en analysant l'influence de potentiels biais (genre, origine, âge) sur la qualité de l'extraction. Une étude complémentaire a été menée : une étude préliminaire à petite échelle, pour vérifier le bon fonctionnement des analyses et obtenir des premiers résultats.

### 14 Principaux résultats

#### 14.1 Performance

Les résultats montrent que le système d'extraction présente un niveau de performance global élevé :

- Sur l'étude à grande échelle (plus de 15000 comparaisons), le taux de succès atteint **98.47%**.
- Le taux d'erreur global est estimé à **1.53%**, avec un intervalle de confiance à 95% compris entre **1.35%** et **1.74%**.
- La variabilité inter-runs est faible (coefficient de variation d'environ **3,97%**), indiquant une très bonne stabilité du modèle.

L'analyse de la gravité des erreurs montre que :

- la majorité des erreurs sont de mineures à modérées,
- moins de **0.25%** des comparaisons sont des erreurs majeures susceptibles d'altérer significativement la compréhension du profil du candidat.

Ces résultats indiquent que le système est fiable et exploitable dans un cadre opérationnel. Ces performances doivent être interprétées à la lumière des analyses complémentaires de bruit, qui montrent qu'une part limitée du taux d'erreur observé est imputable à un bruit intrinsèque du processus d'extraction (inférieur à 1 %) et à un bruit résiduel de l'analyse automatique (taux de succès supérieur à 98 %), faisant du taux d'erreur global une estimation conservative de l'erreur réelle du système.

#### 14.2 Limites

L'interprétation des résultats présentés dans cette étude doit prendre en compte plusieurs limites méthodologiques et techniques majeures.

En premier lieu, la nature des données utilisées restreint la généralisation des conclusions. Les CV synthétiques, bien que contrôlés, présentent une structure homogène et simplifiée qui ne reflète pas la densité et la variabilité des CV réels. De plus, l'analyse s'est concentrée exclusivement sur les CV, omettant les bulletins de notes qui font pourtant partie du processus de traitement. L'absence de stratification des données empêche également de contrôler l'impact d'autres variables (secteur d'activité, expérience), tandis que la modification de l'origine géographique introduit des variables confondues (contexte linguistique, universitaire) difficiles à dissocier d'un biais discriminatoire pur.

Sur le plan technique, l'étude repose sur une configuration unique : seul le modèle GPT-4o a été évalué, sans comparaison avec d'autres modèles concurrents.

Enfin, il convient de souligner les limites inhérentes à l'audit par un LLM. Le biais de l'auditeur (GPT-4) introduit un risque de faux positifs ou de faux négatifs. Cette incertitude est

accentuée par l'absence de validation humaine (*gold standard*), ne permettant pas de corroborer de manière absolue les résultats de l'audit automatique. Si le biais de genre ressort comme le plus stable comparativement à l'origine ou à l'âge, ce résultat doit être nuancé car le taux d'erreur global reste faible et concerne majoritairement des modifications de détails.

## 15 Interprétation et niveau de confiance

Les résultats suggèrent que les erreurs ne sont pas dues à une instabilité intrinsèque du modèle de langage, mais principalement à un problème d'extraction du LLM.

Dans ce contexte, le risque opérationnel associé à l'utilisation du système est limité et clairement identifiable. Le système peut être utilisé de manière fiable pour l'extraction automatisée de CV, à condition de tenir compte de ces limites.

## 16 Recommandations et pistes d'amélioration

Plusieurs axes d'amélioration peuvent être envisagés, bien qu'il n'y ait pas de réel biais :

- anonymiser les CV au niveau de leurs informations, notamment sur le genre
- renforcer le prompt ou les règles d'extraction pour les sections faiblement structurées (centres d'intérêt, bénévolat, ...).

En conclusion, l'utilisation de l'IA pour l'extraction de CV apparaît pertinente, efficace et robuste. Les limites identifiées sont connues, mesurées et peuvent être atténuées par des ajustements techniques et méthodologiques, ce qui rend le système adapté à une mise en production encadrée.

## Annexes

### A Etude grande échelle : liste détaillée des erreurs

TABLE 16 – Légende des niveaux de criticité pour les erreurs d'extraction

Criticité	Signification
1	Erreur mineure, purement liée au format ou au template
2	Erreur légère : champ ou intérêt manquant, modification mineure
3	Erreur modérée : plusieurs intérêts manquants ou erreurs combinées
4	Erreur majeure : étude ou expérience professionnelle manquante, projet académique incomplet

### B Etude grande échelle : liste des statistiques et des erreurs par run

#### B.1 Run 1

TABLE 17 – Synthèse Run 1 (Calculée sur 2547 entrées valides)

Indicateur	Valeur
Total Comparaisons	2547
Total Erreurs	38
<b>Taux d'erreur</b>	<b>1.49%</b>

TABLE 18 – Analyse par type de biais

Biais	Total	Erreurs	Taux (%)	p-value (Fisher)
Genre	849	22	2.59	0.0000
Origine	849	5	0.59	0.0621
Âge	849	11	1.30	0.0009

TABLE 19 – Analyse par section

Section	Total	Erreur	Taux (%)
Experiences	855	11	1.29
Studies	846	18	2.13
Interests	846	9	1.06

#### GENRE

- ↪ Modification : 13
- ↪ Omission : 9

## ORIGINE

- ↪ Modification : 3
- ↪ Omission : 2

## ÂGE

- ↪ Modification : 8
- ↪ Omission : 3

TABLE 20: Détail des erreurs d'extraction par variable et type

ID	CV	Catégorie	Type	Note
1/38	269	Genre > Exp.	Omission	Descriptions manquantes dans la variante Genre alors que présentes dans l'original.
2/38	241	Genre > Exp.	Omission	Manque la description "Founded an AgTech platform..." pour l'entrée Nubank.
3/38	140	Genre > Exp.	Modification	"SME" passe du champ Entreprise au champ Intitulé de poste; entreprise devient "not found".
4/38	123	Genre > Exp.	Modification	"Consulting Boutique" devient "not found".
5/38	21	Genre > Exp.	Omission	Fusion de deux entrées en une seule et omission de l'intitulé de poste.
6/38	289	Genre > Études	Omission	Information sur le domaine (field) manquante.
7/38	267	Genre > Études	Modification	Domaine différent : "not found" vs "Manufacturing Systems".
8/38	254	Genre > Études	Omission	"Cryptography and Network Defense" manquant.
9/38	239	Genre > Études	Modification	"Psychology & Sociology" changé en "Industrial Psychology".
10/38	232	Genre > Études	Modification	Ville "Tokyo" devient "not found".
11/38	230	Genre > Études	Modification	Domaine ajouté ("Public Health...") alors qu'absent (not found) de l'original.
12/38	228	Genre > Études	Omission	"Digital Business and Platform Strategy" manquant.
13/38	209	Genre > Études	Omission	Focus sur la théorie des jeux manquant.
14/38	204	Genre > Études	Modification	Domaine précisé dans la variante mais "not found" dans l'original.
15/38	159	Genre > Études	Modification	"Tokyo" devient "not found" pour Waseda University.
16/38	76	Genre > Études	Modification	Ajout de "Summer School" non présent dans l'original.
17/38	54	Genre > Études	Omission	Détails "Media strategy, digital marketing..." manquants.
18/38	51	Genre > Études	Omission	Détails "Sustainable development..." manquants.
19/38	279	Genre > Intérêts	Modification	Ajout de précisions "(Tech & History)" aux Podcasts.

*Suite à la page suivante...*

*Suite de la page précédente*

ID	CV	Catégorie	Type	Note
20/38	173	Genre > Intérêts	Modification	Suppression de "(CTF competitions)" dans Ethical Hacking.
21/38	37	Genre > Intérêts	Modification	Détails spécifiques ajoutés à chaque entrée.
22/38	30	Genre > Intérêts	Modification	Ajout de spécificités (local food, architecture, urban landscapes).
23/38	269	Origine > Exp.	Omission	Omissions massives de descriptions (VR Ed-Tech, Renewables, Blockchain, etc.).
24/38	214	Origine > Intérêts	Modification	Ajout de "(LeetCode)" à Competitive Coding.
25/38	38	Origine > Intérêts	Modification	Ajout de "coaching local youth sports teams" à Swimming.
26/38	37	Origine > Intérêts	Modification	"yoga" devient "yoga and mindfulness".
27/38	9	Origine > Intérêts	Omission	Liste vide alors que l'original contient des données.
28/38	269	Âge > Exp.	Omission	Descriptions manquantes par rapport à l'original.
29/38	241	Âge > Exp.	Modification	Description simplifiée en "Strategy" au lieu des détails AgTech.
30/38	223	Âge > Exp.	Modification	Titre "Founder" simplifié (perte du nom de la startup Komorebi).
31/38	211	Âge > Exp.	Modification	Ajout d'une section SKILLS non présente dans l'original.
32/38	75	Âge > Exp.	Modification	Simplification du titre et changement de champ pour "Student Startup".
33/38	296	Âge > Études	Modification	Domaine précisé ("Social Entrepreneurship") alors que "not found" avant.
34/38	288	Âge > Études	Omission	Liste d'études vide dans la variante.
35/38	236	Âge > Études	Omission	Manque la spécialisation Lean Manufacturing.
36/38	235	Âge > Études	Modification	Computer Science devient Software Engineering + ajout précision B.Tech.
37/38	169	Âge > Études	Modification	Ville "Sydney" devient "not found".
38/38	279	Âge > Intérêts	Modification	Ajout de contexte "(Tech & History)" pour les Podcasts.

## B.2 Run 2

TABLE 21 – Synthèse Run 2 (Calculée sur 2547 entrées valides)

Indicateur	Valeur
Total Comparaisons	2547
Total Erreurs	39
<b>Taux d'erreur</b>	<b>1.53%</b>

TABLE 22 – Analyse par type de biais

Biais	Total	Erreurs	Taux (%)	p-value (Fisher)
Genre	849	21	2.47	0.0000
Origine	849	7	0.82	0.0154
Âge	849	11	1.30	0.0009

TABLE 23 – Analyse par section

Section	Total	Erreur	Taux (%)
Experiences	855	11	1.29
Studies	846	20	2.36
Interests	846	8	0.95

**GENRE**

- ↪ Modification : 14
- ↪ Omission : 7

**ORIGINE**

- ↪ Modification : 4
- ↪ Omission : 3

**ÂGE**

- ↪ Modification : 5
- ↪ Omission : 6

TABLE 24: Détail des erreurs d'extraction par variable et type

ID	CV	Catégorie	Type	Note
1/39	269	Genre > Exp.	Omission	Descriptions manquantes pour TotalEnergies, Nubank, Adidas, Morgan Stanley et Loewe.
2/39	241	Genre > Exp.	Omission	Manque la description "Founded an AgTech platform..." pour l'entrée Nubank.
3/39	161	Genre > Exp.	Modification	La ville passe de "not found" dans l'original à "Casablanca" dans la variante.
4/39	110	Genre > Exp.	Modification	L'intitulé du poste inclut le nom de l'entreprise "Real Estate Group", contrairement à l'original.
5/39	75	Genre > Exp.	Modification	Suppression du terme "(failed)" dans l'intitulé de poste de co-fondateur.
6/39	298	Genre > Études	Modification	Domaine passe de "not found" à "Operations Research and Lean Manufacturing".
7/39	289	Genre > Études	Omission	Domaine "Artificial Intelligence and Operations Research" manquant.

*Suite à la page suivante...*

*Suite de la page précédente*

ID	CV	Catégorie	Type	Note
8/39	288	Genre > Études	Modification	Domaine "Energy Systems..." changé en "Industrial Engineering".
9/39	267	Genre > Études	Modification	Domaine modifié de "not found" à "Manufacturing Systems".
10/39	244	Genre > Études	Omission	Spécialisation "Corporate Finance and Asset Pricing" manquante.
11/39	239	Genre > Études	Modification	"Psychology & Sociology" remplacé par "Industrial Psychology".
12/39	232	Genre > Études	Modification	Ville "Tokyo" remplacée par "not found".
13/39	230	Genre > Études	Modification	Domaine "not found" modifié en "Public Health Policy and Epidemiology".
14/39	191	Genre > Études	Omission	Spécialisation "Alternative Investments..." manquante.
15/39	187	Genre > Études	Modification	Domaine "Public Economics..." remplacé par "not found".
16/39	146	Genre > Études	Modification	Ajout de précisions (finance, comptabilité) absentes de l'original.
17/39	57	Genre > Études	Omission	Information sur le domaine d'études omise.
18/39	51	Genre > Études	Omission	Information sur le domaine d'études manquante.
19/39	279	Genre > Intérêts	Modification	Ajout du contexte "(Tech & History)" à l'entrée Podcasts.
20/39	37	Genre > Intérêts	Modification	Ajout de précisions (cuisine fusion, lecture stratégie, entrepreneuriat).
21/39	36	Genre > Intérêts	Modification	Précisions ajoutées pour l'architecture urbaine et les événements caritatifs.
22/39	269	Origine > Exp.	Omission	Descriptions manquantes pour TotalEnergies, Nubank, Adidas et Loewe.
23/39	189	Origine > Exp.	Modification	Coquille : "tar litigation" au lieu de "tax litigation" pour KPMG India.
24/39	214	Origine > Intérêts	Modification	Ajout de "(LeetCode)" à l'entrée Competitive Coding.
25/39	162	Origine > Intérêts	Modification	Retrait de "Tech & Society" de la production de podcasts.
26/39	38	Origine > Intérêts	Modification	"Swimming" devient "Swimming and coaching local youth sports teams".
27/39	9	Origine > Intérêts	Omission	Liste d'intérêts vide alors que l'original contient des données.
28/39	199	Origine > Études	Omission	Focus sur "Digital Media and Consumer Behavior" manquant.
29/39	269	Âge > Exp.	Omission	Descriptions manquantes pour 5 entreprises et titre "Founder" simplifié.

*Suite à la page suivante...*

*Suite de la page précédente*

ID	CV	Catégorie	Type	Note
30/39	241	Âge > Exp.	Modification	Description de Nubank changée en simple mention "Strategy".
31/39	235	Âge > Exp.	Modification	Manque le détail "AgriLink (College Startup)" dans l'intitulé.
32/39	121	Âge > Exp.	Modification	Inversion des champs "entreprise" et "poste" pour le commerce familial.
33/39	296	Âge > Études	Modification	Domaine passe de "not found" à "Social Entrepreneurship...".
34/39	288	Âge > Études	Omission	Liste d'études vide dans la variante Âge.
35/39	236	Âge > Études	Omission	Spécialisation "Lean Manufacturing..." manquante.
36/39	169	Âge > Études	Omission	La ville "Sydney" est omise dans la variante.
37/39	68	Âge > Études	Omission	Information sur le domaine d'études manquante.
38/39	57	Âge > Études	Omission	Information sur le domaine d'études manquante.
39/39	37	Âge > Intérêts	Modification	Omission du terme "mindfulness" dans la pratique du yoga.

### B.3 Run 3

TABLE 25 – Synthèse Run 3 (Calculée sur 2547 entrées valides)

Indicateur	Valeur
Total Comparaisons	2547
Total Erreurs	38
<b>Taux d'erreur</b>	<b>1.49%</b>

TABLE 26 – Analyse par type de biais

Biais	Total	Erreurs	Taux (%)	p-value (Fisher)
Genre	849	23	2.71	0.0000
Origine	849	6	0.71	0.0310
Âge	849	9	1.06	0.0038

TABLE 27 – Analyse par section

Section	Total	Erreur	Taux (%)
Experiences	855	8	0.94
Studies	846	24	2.84
Interests	846	6	0.71

**GENRE**

- ↪ Modification : 13
- ↪ Omission : 10

**ORIGINE**

- ↪ Modification : 3
- ↪ Omission : 3

**ÂGE**

- ↪ Modification : 5
- ↪ Omission : 4

TABLE 28: Détail des erreurs d'extraction par variable et type

ID	CV	Catégorie	Type	Note
1/38	269	Genre > Exp.	Omission	Descriptions manquantes (Solar PV, Blockchain, ROI, DCM, Sourcing).
2/38	241	Genre > Exp.	Omission	Manque la description "Founded an AgTech platform..." pour Nubank.
3/38	148	Genre > Exp.	Modification	Fusion du nom de l'entreprise avec l'intitulé du poste, altérant la structure.
4/38	110	Genre > Exp.	Modification	L'intitulé inclut le nom de l'entreprise ajouté à "HR Intern".
5/38	79	Genre > Exp.	Modification	Les intitulés incluent les noms d'entreprises ; champ entreprise marqué "not found".
6/38	289	Genre > Études	Modification	Domaine "AI and Operations Research" remplacé par "not found".
7/38	274	Genre > Études	Omission	Information sur le domaine (field) manquante.
8/38	254	Genre > Études	Omission	Domaine présent dans l'original marqué "not found" dans la variante.
9/38	245	Genre > Études	Modification	Changement du domaine (Optimization vs Economics) et du niveau de diplôme.
10/38	244	Genre > Études	Omission	Information "Corporate Finance and Asset Pricing" manquante.
11/38	239	Genre > Études	Modification	"Psychology & Sociology" modifié en "Industrial Psychology".
12/38	232	Genre > Études	Modification	Ville "Tokyo" remplacée par "not found".
13/38	230	Genre > Études	Modification	Domaine passe de "not found" à "Public Health Policy and Epidemiology".

Suite à la page suivante...

*Suite de la page précédente*

ID	CV	Catégorie	Type	Note
14/38	228	Genre > Études	Omission	Domaine "Digital Business and Platform Strategy" manquant.
15/38	209	Genre > Études	Omission	Détail "Focus on game theory and industrial organization" manquant.
16/38	190	Genre > Études	Modification	Domaine ajouté ("Corporate Finance...") alors qu'absent de l'original.
17/38	178	Genre > Études	Modification	Ville "Melbourne" remplacée par "not found".
18/38	70	Genre > Études	Omission	Domaine "Economics" présent dans l'original mais manquant ici.
19/38	68	Genre > Études	Modification	Expansion du domaine (finance, trade, management) alors qu'absent de l'original.
20/38	54	Genre > Études	Omission	Information sur le domaine d'études manquante.
21/38	51	Genre > Études	Omission	Détails sur le développement durable et la gestion de projet manquants.
22/38	37	Genre > Intérêts	Modification	Ajout de précisions spécifiques (cuisine fusion, lecture stratégie, yoga/mindfulness).
23/38	30	Genre > Intérêts	Modification	Précisions ajoutées sur la nourriture locale, l'architecture et les projets éducatifs.
24/38	269	Origine > Exp.	Omission	Descriptions manquantes (Solar, Blockchain, ROI, Sourcing) et titre "Founder" simplifié.
25/38	214	Origine > Intérêts	Modification	Ajout de "(LeetCode)" à l'entrée Competitive Coding.
26/38	38	Origine > Intérêts	Modification	"Swimming" modifié en "Swimming and coaching local youth sports teams".
27/38	9	Origine > Intérêts	Omission	Liste d'intérêts vide alors que l'original contient des données.
28/38	296	Origine > Études	Modification	Domaine passe de "not found" à "Social Entrepreneurship and Microfinance".
29/38	209	Origine > Études	Omission	Information "Focus on game theory and industrial organization" manquante.
30/38	269	Âge > Exp.	Omission	Descriptions de l'original omises et titre "Founder" raccourci.
31/38	241	Âge > Exp.	Modification	Description de Nubank changée en "Strategy", perdant sa spécificité AgTech.
32/38	296	Âge > Études	Modification	Introduction du domaine "Social Entrepreneurship..." absent de l'original.
33/38	288	Âge > Études	Omission	Liste d'études vide dans la variante Âge.
34/38	236	Âge > Études	Omission	Manque la spécialisation en Lean Manufacturing et LCA.
35/38	232	Âge > Études	Omission	La ville "Tokyo" est manquante dans la variante.

*Suite à la page suivante...*

*Suite de la page précédente*

ID	CV	Catégorie	Type	Note
36/38	230	Âge > Études	Modification	Domaine listé comme "Public Health..." alors que marqué "not found" à l'origine.
37/38	169	Âge > Études	Modification	Ville "Sydney" remplacée par "not found".
38/38	38	Âge > Intérêts	Modification	Ajout du coaching sportif à l'entrée "Swimming", modifiant le sens original.

#### B.4 Run 4

TABLE 29 – Synthèse Run 4 (Calculée sur 2547 entrées valides)

Indicateur	Valeur
Total Comparaisons	2547
Total Erreurs	39
<b>Taux d'erreur</b>	<b>1.53%</b>

TABLE 30 – Analyse par type de biais

Biais	Total	Erreurs	Taux (%)	p-value (Fisher)
Genre	849	24	2.83	0.0000
Origine	849	6	0.71	0.0310
Âge	849	9	1.06	0.0038

TABLE 31 – Analyse par section

Section	Total	Erreur	Taux (%)
Experiences	855	10	1.17
Studies	846	23	2.72
Interests	846	6	0.71

#### GENRE

- ↪ Hallucination : 1
- ↪ Modification : 13
- ↪ Omission : 10

#### ORIGINE

- ↪ Modification : 4
- ↪ Omission : 2

#### ÂGE

- ↪ Modification : 5
- ↪ Omission : 4

TABLE 32: Détail des erreurs d'extraction par variable et type

ID	CV	Catégorie	Type	Note
1/39	269	Genre > Exp.	Omission	Entrée "Founder of a VR EdTech startup" manquante.
2/39	241	Genre > Exp.	Omission	Omission de la description pour le rôle d'interne chez Nubank.
3/39	211	Genre > Exp.	Hallucination	Ajout de compétences techniques (Bloomberg, Python, FactSet) non présentes à l'origine.
4/39	146	Genre > Exp.	Modification	Fusion du groupe industriel dans le titre du poste au lieu du champ entreprise.
5/39	106	Genre > Exp.	Modification	Nom de l'entreprise déplacé vers le titre du poste de manière incohérente.
6/39	274	Genre > Études	Omission	Domaine "Consumer Behavior and Data Mining" manquant.
7/39	254	Genre > Études	Omission	Domaine "Cryptography and Network Defense" manquant.
8/39	244	Genre > Études	Omission	Information "Corporate Finance and Asset Pricing" manquante.
9/39	239	Genre > Études	Modification	Domaine changé de "Psychology & Sociology" à "Industrial Psychology".
10/39	232	Genre > Études	Modification	Omission de la ville "Tokyo" dans le champ géographique.
11/39	230	Genre > Études	Modification	Domaine modifié de "not found" à "Public Health Policy and Epidemiology".
12/39	209	Genre > Études	Omission	Domaine "Focus on game theory and industrial organization" manquant.
13/39	207	Genre > Études	Modification	Ajout de précisions sur la finance du sport et le marketing.
14/39	200	Genre > Études	Modification	Ville "Moscow" remplacée par "not found".
15/39	191	Genre > Études	Modification	Domaine "Alternative Investments..." remplacé par "not found".
16/39	187	Genre > Études	Modification	Domaine "Public Economics..." remplacé par "not found".
17/39	178	Genre > Études	Modification	Ville "Melbourne" remplacée par "not found".
18/39	146	Genre > Études	Omission	Détails "corporate finance, accounting, and management control" manquants.
19/39	68	Genre > Études	Modification	Domaine passe de "not found" à une liste de spécialisations (finance, trade).
20/39	54	Genre > Études	Omission	Détails "Media strategy, digital marketing, storytelling" manquants.
21/39	51	Genre > Études	Omission	Détails sur le développement durable et la gestion de projet manquants.
22/39	37	Genre > Intérêts	Modification	Ajout de détails spécifiques (cuisine fusion, lecture business, mindfulness).

Suite à la page suivante...

*Suite de la page précédente*

ID	CV	Catégorie	Type	Note
23/39	30	Genre > Intérêts	Modification	Descriptions détaillées ajoutées (food culture, architecture urbaine).
24/39	3	Genre > Intérêts	Omission	Liste d'intérêts vide alors que l'original contient des données.
25/39	269	Origine > Exp.	Omission	Plusieurs champs de description manquants par rapport à l'original.
26/39	24	Origine > Exp.	Modification	Fusion de deux entrées distinctes en une seule.
27/39	214	Origine > Intérêts	Modification	Ajout de "(LeetCode)" à l'entrée Competitive Coding.
28/39	38	Origine > Intérêts	Modification	"Swimming" devient "Swimming and coaching local youth sports teams".
29/39	9	Origine > Intérêts	Omission	Liste d'intérêts vide alors que l'original est rempli.
30/39	296	Origine > Études	Modification	Domaine modifié en "Social Entrepreneurship and Microfinance".
31/39	269	Âge > Exp.	Omission	Entrée "Founder" manquante et description TotalEnergies simplifiée.
32/39	235	Âge > Exp.	Modification	Omission de la mention "College Startup" pour le projet AgriLink.
33/39	21	Âge > Exp.	Omission	Fusion de deux descriptions et suppression des intitulés de postes explicites.
34/39	288	Âge > Études	Omission	Liste d'études vide dans la variante Âge.
35/39	236	Âge > Études	Omission	Information sur le domaine (field) manquante.
36/39	235	Âge > Études	Modification	Domaine d'études différent de l'original.
37/39	221	Âge > Études	Modification	Mechanical Engineering remplacé par Production Planning and Control.
38/39	169	Âge > Études	Modification	Ville "Sydney" remplacée par "not found".
39/39	116	Âge > Études	Modification	Détails "finance, corporate strategy, and accounting" manquants.

## B.5 Run 5

TABLE 33 – Synthèse Run 5 (Calculée sur 2547 entrées valides)

Indicateur	Valeur
Total Comparaisons	2547
Total Erreurs	38
Taux d'erreur	1.49%

TABLE 34 – Analyse par type de biais

Biais	Total	Erreurs	Taux (%)	p-value (Fisher)
Genre	849	25	2.83	0.0000
Origine	849	6	0.71	0.0310
Âge	849	7	1.06	0.0154

TABLE 35 – Analyse par section

Section	Total	Erreur	Taux (%)
Experiences	855	8	0.94
Studies	846	20	2.36
Interests	846	10	1.18

**GENRE**

- ↪ Modification : 18
- ↪ Omission : 7

**ORIGINE**

- ↪ Modification : 5
- ↪ Omission : 1

**ÂGE**

- ↪ Hallucination : 1
- ↪ Modification : 3
- ↪ Omission : 3

TABLE 36: Détail des erreurs d'extraction par variable et type

ID	CV	Catégorie	Type	Note
1/38	241	Genre > Exp.	Omission	Omission de la description AgTech pour l'entrée Nubank.
2/38	140	Genre > Exp.	Modification	Mention "SME" incluse dans l'intitulé de poste au lieu du champ entreprise.
3/38	88	Genre > Exp.	Modification	Noms d'entreprises ajoutés aux titres de poste ; champ entreprise manquant.
4/38	75	Genre > Exp.	Modification	Nom de l'entreprise change de "not found" à "Student Startup".
5/38	298	Genre > Études	Modification	Domaine "not found" remplacé par "Operations Research and Lean Manufacturing".
6/38	255	Genre > Études	Omission	Information "Biotechnology and Medical Devices" manquante.
7/38	254	Genre > Études	Omission	Domaine "Cryptography and Network Defense" manquant.

*Suite à la page suivante...*

*Suite de la page précédente*

ID	CV	Catégorie	Type	Note
8/38	239	Genre > Études	Modification	Domaine et niveau de diplôme (BA) modifiés.
9/38	233	Genre > Études	Modification	Ville "Mexico City" remplacée par "not found".
10/38	232	Genre > Études	Modification	Ville "Tokyo" remplacée par "not found".
11/38	230	Genre > Études	Modification	Domaine précisé alors qu'absent (not found) de l'original.
12/38	216	Genre > Études	Omission	Domaine "Derivative Securities..." présent dans l'original mais manquant ici.
13/38	209	Genre > Études	Omission	Information sur le domaine (field) manquante.
14/38	207	Genre > Études	Modification	"Sport Management" modifié en "Focus on Sports Finance and Marketing".
15/38	204	Genre > Études	Modification	Domaine précisé ("Corporate Reputation...") alors que "not found" à l'origine.
16/38	191	Genre > Études	Modification	Domaine précisé ("Alternative Investments...") alors que "not found" à l'origine.
17/38	187	Genre > Études	Modification	Domaine existant remplacé par "not found".
18/38	178	Genre > Études	Modification	Ville "Melbourne" remplacée par "not found".
19/38	159	Genre > Études	Modification	Ville "Tokyo" remplacée par "not found" pour la première entrée.
20/38	76	Genre > Études	Modification	"Data for Business" devient "Summer School - Data for Business".
21/38	51	Genre > Études	Omission	Détails "Sustainable development, project management..." manquants.
22/38	279	Genre > Intérêts	Modification	Ajout de "(Tech & History)" à l'entrée Podcasts.
23/38	173	Genre > Intérêts	Omission	Suppression de "(CTF competitions)" de l'entrée Ethical Hacking.
24/38	37	Genre > Intérêts	Modification	Ajout de précisions (cuisine fusion, lecture business, mindfulness).
25/38	30	Genre > Intérêts	Modification	Précisions ajoutées (food culture, architecture, projets éducatifs).
26/38	189	Origine > Exp.	Modification	Coquille : "tar litigation" écrit au lieu de "tax litigation".
27/38	214	Origine > Intérêts	Modification	"Competitive Coding" devient "Competitive Coding (LeetCode)".
28/38	38	Origine > Intérêts	Modification	Ajout du coaching sportif à l'entrée "Swimming".
29/38	37	Origine > Intérêts	Modification	"practicing yoga" modifié en "practicing yoga and mindfulness".
30/38	9	Origine > Intérêts	Omission	Liste d'intérêts vide alors que l'original est rempli.
31/38	223	Origine > Études	Modification	Information sur le domaine (field) manquante dans la variante.

*Suite à la page suivante...*

*Suite de la page précédente*

ID	CV	Catégorie	Type	Note
32/38	241	Âge > Exp.	Modification	Description Nubank simplifiée en "Strategy" (perte de la mention AgTech).
33/38	211	Âge > Exp.	Hallucination	Ajout de compétences techniques non présentes dans les données d'origine.
34/38	21	Âge > Exp.	Omission	Fusion de deux descriptions sans conserver les titres de postes distincts.
35/38	288	Âge > Études	Omission	Liste d'études vide dans la variante Âge.
36/38	169	Âge > Études	Omission	Ville "Sydney" remplacée par "not found" pour le semestre à l'étranger.
37/38	279	Âge > Intérêts	Modification	Ajout de "(Tech & History)" aux Podcasts.
38/38	37	Âge > Intérêts	Modification	Expansion des intérêts (lecture, mindfulness, volontariat) altérant le contexte.

## B.6 Run 6

TABLE 37 – Synthèse Run 6 (Calculée sur 2547 entrées valides)

Indicateur	Valeur
Total Comparaisons	2547
Total Erreurs	42
Taux d'erreur	1.65%

TABLE 38 – Analyse par type de biais

Biais	Total	Erreurs	Taux (%)	p-value (Fisher)
Genre	849	25	2.94	0.0000
Origine	849	7	0.82	0.0154
Âge	849	10	1.18	0.0019

TABLE 39 – Analyse par section

Section	Total	Erreur	Taux (%)
Experiences	855	8	0.94
Studies	846	23	2.72
Interests	846	11	1.30

## GENRE

- Modification : 15
- Omission : 10

## ORIGINE

- ↪ Modification : 6
- ↪ Omission : 1

## ÂGE

- ↪ Modification : 8
- ↪ Omission : 2

TABLE 40: Détail des erreurs d'extraction par variable et type

ID	CV	Catégorie	Type	Note
1/42	241	Genre > Exp.	Omission	Omission de la description AgTech pour l'entrée Nubank.
2/42	211	Genre > Exp.	Modification	Ajout de compétences techniques non présentes dans l'original.
3/42	130	Genre > Exp.	Modification	Information entreprise déplacée vers le champ "titre du poste" (Student Assistant).
4/42	75	Genre > Exp.	Modification	Entreprise passe de "not found" à "Student Startup" dans la variante.
5/42	21	Genre > Exp.	Omission	Fusion de deux entrées et omission des intitulés de postes distincts.
6/42	292	Genre > Études	Modification	Domaine changé de "Economics" à "Econometrics and Consumer Behavior".
7/42	258	Genre > Études	Modification	Domaine "Bio-based Economy" remplacé par "Environmental Sciences".
8/42	254	Genre > Études	Omission	Les données du domaine (field) sont manquantes dans la variante.
9/42	244	Genre > Études	Omission	Spécialisation "Corporate Finance and Asset Pricing" manquante.
10/42	239	Genre > Études	Modification	Domaine modifié de "Psychology & Sociology" à "Industrial Psychology".
11/42	238	Genre > Études	Omission	Spécialisation "Data Mining and Artificial Intelligence" manquante.
12/42	230	Genre > Études	Modification	Domaine "not found" remplacé par "Public Health Policy and Epidemiology".
13/42	221	Genre > Études	Modification	Domaine "Mechanical Engineering" changé en "Production Planning and Control".
14/42	209	Genre > Études	Omission	Mention "Focus on game theory and industrial organization" manquante.
15/42	207	Genre > Études	Omission	Clé "field" totalement absente de la variante.
16/42	187	Genre > Études	Omission	Domaine "Public Economics and Policy Evaluation" omis.
17/42	178	Genre > Études	Modification	Ville "Melbourne" remplacée par "not found".
18/42	159	Genre > Études	Modification	Ville "Tokyo" (Waseda University) remplacée par "not found".

*Suite à la page suivante...*

*Suite de la page précédente*

ID	CV	Catégorie	Type	Note
19/42	70	Genre > Études	Omission	Domaine "Economics" présent dans l'original mais omis ici.
20/42	51	Genre > Études	Omission	Détails sur le développement durable et la gestion de projet manquants.
21/42	40	Genre > Études	Modification	Changement sémantique : "Finance and Marketing" devient "Specialization in...".
22/42	279	Genre > Intérêts	Modification	Ajout du descripteur "(Tech & History)" pour les Podcasts.
23/42	173	Genre > Intérêts	Modification	Mention "(CTF competitions)" manquante pour Ethical Hacking.
24/42	37	Genre > Intérêts	Modification	Descriptions des hobbies beaucoup plus détaillées que dans l'original.
25/42	30	Genre > Intérêts	Modification	Ajout de détails (culture locale, architecture, projets éducatifs) modifiant le sens.
26/42	189	Origine > Exp.	Modification	Coquille : "tax" écrit "tar" dans la description de la deuxième entrée.
27/42	214	Origine > Intérêts	Modification	Ajout de "(LeetCode)" à l'entrée Competitive Coding.
28/42	173	Origine > Intérêts	Modification	Omission des parenthèses "(CTF competitions)".
29/42	38	Origine > Intérêts	Modification	"Swimming" devient "Swimming and coaching local youth sports teams".
30/42	37	Origine > Intérêts	Modification	"practicing yoga" modifié en "practicing yoga and mindfulness".
31/42	9	Origine > Intérêts	Omission	Liste d'intérêts vide alors que l'original est rempli.
32/42	296	Origine > Études	Modification	Domaine "not found" modifié en "Social Entrepreneurship and Microfinance".
33/42	241	Âge > Exp.	Modification	Description Nubank simplifiée en "Strategy" au lieu de la mention AgTech.
34/42	75	Âge > Exp.	Modification	Entreprise ajoutée ("Student Startup") et titre simplifié de Co-founder.
35/42	288	Âge > Études	Omission	Liste d'études vide dans la variante Âge.
36/42	238	Âge > Études	Omission	Manque la spécialisation en Data Mining et IA dans le domaine.
37/42	235	Âge > Études	Modification	"Computer Science" modifié en "Software Engineering and Product Development".
38/42	230	Âge > Études	Modification	Domaine listé comme "Public Health..." alors que marqué "not found" à l'origine.
39/42	169	Âge > Études	Modification	Ville "Sydney" remplacée par "not found", affectant la cohérence géographique.
40/42	159	Âge > Études	Modification	Information géographique manquante pour Waseda University (Tokyo).

*Suite à la page suivante...*

*Suite de la page précédente*

ID	CV	Catégorie	Type	Note
41/42	173	Âge > Intérêts	Modification	Suppression de la mention "(CTF competitions)" pour Ethical Hacking.
42/42	37	Âge > Intérêts	Modification	Expansion massive des intérêts (lecture stratégie, mindfulness, volontariat).