

Rapport d'Étude : Analyse de la Robustesse de l'Extraction de CV par LLM selon le format

Laura BARDOU - Sacha BEAUJEAN - Justin COCHET GARSOT

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 du template utilisé (Canva, LaTeX, Europass, etc.). Pour chaque candidat, plusieurs templates sont générés. 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

I	Première étude : Etude préliminaire	4
1	Méthodologie	4
1.1	Constitution du Corpus	4
1.2	Protocole de Test	4
2	Résultats Quantitatifs	4
2.1	Performance Globale	4
3	Analyse et Discussion des Erreurs	5
4	Conclusion de la phase préliminaire	5
II	Deuxième étude : nombre élevé de CV	5
5	Méthodologie	5
5.1	Constitution du Corpus	5
5.2	Protocole de Test	6
6	Résultats Quantitatifs	6
6.1	Performance globale et stabilité inter-runs	6
6.2	Répartition des erreurs par type	6
6.3	Répartition des erreurs par section	7
6.4	Distribution des erreurs par criticité	7

6.5	Profils présentant des erreurs récurrentes	7
6.6	Récurrence inter-runs des erreurs	8
6.7	Influence du format de CV	8
6.8	Synthèse quantitative	8
7	Analyse qualitative des erreurs	9
7.1	Typologie des erreurs	9
7.2	Récurrence des erreurs selon les runs	9
7.3	Analyse par type de champ extrait	9
7.3.1	Centres d'intérêt	9
7.3.2	Études	10
7.3.3	Expériences professionnelles et bénévolat	10
7.4	Lien entre format et nature des erreurs	10
7.5	Synthèse	10
III	Troisième partie : Confiance des résultats	10
8	Intervalle de confiance	11
8.1	Cadre théorique	11
8.1.1	Structure expérimentale	11
8.1.2	Modélisation statistique	11
8.1.3	Intervalle de confiance de Wilson (95%)	11
8.2	Calculs numériques	11
8.2.1	Étude 1 ($n = 33$, $x = 1$)	11
8.2.2	Étude 2 ($n = 2000$, $x = 70$)	12
8.3	Conclusion sur les intervalles de confiance	12
9	Analyse du bruit : variabilité inter-runs	12
9.1	Taux d'erreur moyen	12
9.2	Écart-type inter-runs	13
9.3	Coefficient de variation	13
9.4	Interprétation	13
10	Autres statistiques (Étude 2)	13
10.1	Taux d'erreur conditionnel par type de section	13
10.2	Probabilité qu'un CV contienne une erreur selon la criticité	14
10.2.1	Probabilité cumulée par niveau de criticité	14
10.2.2	Interprétation	14
IV	Synthèse générale et recommandations	14
11	Résumé exécutif	14
11.1	Principaux résultats	14
11.2	Limites du LLM identifiées	15
11.3	Limites de l'étude identifiées	15
11.4	Interprétation et niveau de confiance	15
11.5	Recommandations et pistes d'amélioration	16

A	Annexes : CV de Référence de la première étude	17
A.1	Profil 1 : Alexandre DUPONT	18
A.2	Profil 2 : Thomas LEGRAND	19
A.3	Profil 3 : Camille MARTIN	20
B	Annexes : CV de référence de la deuxième étude	21
C	Etude 2 : liste détaillée des erreurs	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 divers formats de CV. Cette première étude compare 39 documents générés à partir de 3 profils types sur 13 templates différents.

1.1 Constitution du Corpus

Le jeu de données est constitué de **39 CV** au total, basés sur la matrice suivante :

- **3 Profils Candidats** : Alexandre, Camille, Thomas.
- **13 Templates distincts** appliqués à chaque profil :
 - 4 templates *Overleaf* (LaTeX) (1 à 4)
 - 4 templates *Canva* (5 à 8)
 - 2 templates *Europass* (9 et 10)
 - 3 templates *Resume.io* (11 à 13)

Le choix de ces modèles s'est basé à la fois sur la recherche de formats populaires représentatifs et sur la volonté de proposer une grande diversité graphique, incluant des CV plus atypiques, comme ceux réalisés sur Canva.

1.2 Protocole de Test

1. **Génération** : Création des textes des CV et remplissage des templates.
2. **Extraction** : Passage des CV dans le LLM Semantikmatch pour extraire les champs *Intérêts*, *Études* et *Expériences Professionnelles*.
3. **Validation** : Comparaison automatisée (via LLM) entre les données extraites et le fichier JSON de référence construit manuellement.

2 Résultats Quantitatifs

2.1 Performance Globale

Sur l'ensemble des 39 documents traités, le taux de succès global est de **74,4%**. Une erreur est toute modification ou omission entre le CV de référence et le CV extrait. Une erreur peut ainsi être levée au moindre écart entre les deux CV même si tout le reste est identique.

Il faut être attentif puisque la comparaison est exécutée par un LLM, il existe un risque qu'il relève des erreurs qui ne le sont pas (ponctuation, format de dates,...). Pour cette raison, toutes les erreurs relevées ne sont pas forcément significatives ou importantes. Une part d'erreur humaine entre également en jeu lors de l'écriture des CV et notamment lors du remplissage des templates.

TABLE 1 – Taux d’erreur par profil candidat

Candidat	Erreurs	Total CV	Taux d’Erreur
Alexandre	2	13	15%
Camille	6	13	46%
Thomas	2	13	15%

3 Analyse et Discussion des Erreurs

L’analyse qualitative des échecs a permis d’identifier quatre causes racines principales.

- Structure du Document : Le cas du Format N°13 Le template N°13 présente un taux d’erreur de 100% car il ne contient pas de section "Intérêts".
- Confusion Sémantique : Le profil de Camille (Bénévolat) Le LLM peine à distinguer le bénévolat de l’expérience salariée stricte.
- Confusion Sémantique : Le profil de Thomas (Entrepreneuriat) Le LLM n’est pas constant dans le cas d’une expérience professionnelle d’entrepreneuriat en raison de l’absence du nom de l’entreprise ; par conséquent, il tente de le combler avec d’autres termes de la description.
- Confusion Sémantique : Le profil de Camille (Echange à l’étranger) Une expérience d’échange à l’étranger est présente sur son profil, mais de manière différente des autres études du profil ; par conséquent, non consistance du LLM pour relever cet échange.

4 Conclusion de la phase préliminaire

L’étude montre une robustesse satisfaisante pour les formats standards, mais nécessite des ajustements concernant les types de contrats et le volontariat.

À la suite de la constatation de ces erreurs, le prompt d’extraction a été modifié afin qu’il puisse extraire les missions de bénévolat et les expériences en free-lance sans défaut. Les CV factuellement faux (catégorie manquante) ont été supprimés. Avec ces corrections, nous n’obtenons plus qu’une seule erreur sur 33 CV et donc un taux de succès de **97%**.

Deuxième partie

Deuxième étude : nombre élevé de CV

5 Méthodologie

La seconde étude repose sur la même méthodologie que la première, mais elle est appliquée à un corpus plus large et inclut des répétitions indépendantes du processus. Le corpus se compose de 10 profils, chacun décliné selon 40 templates différents, soit un total de 400 CV. Ces 400 CV sont ensuite comparés à l’extraction de référence (le CV réel) par un autre LLM. L’extraction de l’ensemble des 400 CV est réalisée cinq fois afin d’évaluer la stabilité des résultats.

5.1 Constitution du Corpus

Le jeu de données est constitué de **400 CV** au total, basés sur la matrice suivante :

- **10 Profils Candidats** : Clara, Hugo, Sarah, ThomasG, Ines, Louis, Amelie, Maxime, ThomasL, Camille.
- **40 Templates distincts** appliqués à chaque profil :
 - 10 templates *Canva* (1 à 10)

- 4 templates *Europass* (11 à 14)
- 6 templates *Overleaf* (LaTeX) (15 à 20)
- 20 templates *Resume.io* (21 à 40)

5.2 Protocole de Test

1. **Génération** : Création des CV en PDF.
2. **Extraction** : Passage des CV dans le LLM Semantikmatch pour extraire les champs *Intérêts*, *Études* et *Expériences Professionnelles*.
3. **Validation** : Comparaison automatisée (via LLM) entre les données extraites et le fichier JSON de référence construit manuellement.
4. **Réitération** : Réexécution des 2 dernières étapes 5 fois

Au total, le nombre de comparaisons atteint 400 observations indépendantes par analyse. Ainsi, le nombre de comparaisons indépendantes atteint 2000 après les répétitions des 5 runs.

6 Résultats Quantitatifs

Cette section présente une analyse quantitative détaillée des erreurs observées lors des cinq runs d'extraction, soit un total de 2000 CV analysés. Les résultats permettent d'évaluer la stabilité du modèle, la nature des erreurs et leur distribution selon plusieurs dimensions. Les résultats détaillés de cette étude sont présents en annexe.

6.1 Performance globale et stabilité inter-runs

Chaque run comprend 400 CV, pour un total de 2000 extractions. Le Tableau 2 présente le taux d'erreur observé pour chacun des cinq runs.

TABLE 2 – Taux d'erreur par run

Run	Erreurs	Total CV	Taux d'erreur
Run 1	11	400	2.8%
Run 2	15	400	3.8%
Run 3	14	400	3.5%
Run 4	15	400	3.8%
Run 5	15	400	3.8%

Les taux d'erreur sont homogènes d'un run à l'autre, compris entre 2.8% et 3.8%. Sur l'ensemble des cinq runs, 70 erreurs ont été observées sur 2000 extractions, soit un taux d'erreur global de **3.5%**, correspondant à un taux de succès de **96.5%**. Cette faible variance inter-runs indique une bonne stabilité globale du modèle.

6.2 Répartition des erreurs par type

Les erreurs observées ont été classées en deux catégories principales : omissions et modifications.

TABLE 3 – Répartition des erreurs par type (ensemble des runs)

Type d'erreur	Nombre	Pourcentage
Omission	60	85.7%
Modification	10	14.3%

Les erreurs par omission représentent la très grande majorité des cas observés. Ce résultat suggère que le modèle tend davantage à ne pas extraire certaines informations qu'à produire des extractions incorrectes ou altérées.

6.3 Répartition des erreurs par section

Un même CV pouvant être affecté par plusieurs erreurs provenant de différentes sections, les statistiques suivantes ne sont pas mutuellement exclusives.

TABLE 4 – Répartition des erreurs par section

Section	Nombre d'erreurs	Pourcentage
Centres d'intérêt	53	75.7%
Études	27	38.6%
Expériences professionnelles / bénévolat	11	15.7%

Les centres d'intérêt constituent la section la plus sensible aux erreurs, concentrant plus des trois quarts des erreurs observées. À l'inverse, les expériences professionnelles présentent un taux d'erreur nettement plus faible.

6.4 Distribution des erreurs par criticité

Chaque erreur a été évaluée selon un niveau de criticité allant de 1 (mineure) à 4 (majeure).

TABLE 5 – Distribution des erreurs par niveau de criticité

Criticité	Nombre	Pourcentage
1 (mineure)	5	7.1%
2 (légère)	35	50.0%
3 (modérée)	20	28.6%
4 (majeure)	10	14.3%

La majorité des erreurs (85.7%) sont de criticité faible à modérée (niveaux 1 à 3), indiquant que les écarts observés ont généralement un impact limité sur la compréhension globale du profil du candidat. Les erreurs mineures sont toutes dues à des erreurs extérieures à l'extraction du LLM et peuvent être entièrement ignorées. Les erreurs légères sont de réelles erreurs, mais d'une importance faible pour la retransmission du profil du candidat (un intérêt omis, une mission de volontariat relevée dans une section différente,...). Les erreurs modérées sont des modifications relativement significatives, des omissions de description ou d'éléments pour les études ou expériences professionnelles, et des erreurs précédentes combinées. Enfin, les erreurs critiques sont celles qui nuisent vraiment à la compréhension du CV ; en grande majorité, il s'agit d'omissions d'études comme le Baccalauréat.

6.5 Profils présentant des erreurs récurrentes

L'analyse par profil met en évidence une concentration des erreurs sur un nombre restreint de candidats.

Ces profils concentrent une part importante des erreurs observées, souvent associées aux mêmes templates et aux mêmes champs.

Personnes	Nombre d'erreurs	Fréquence d'erreurs (%)
ThomasL	14	7.0
Camille	9	4.5
Ines	8	4.0
ThomasG	7	3.5
Amelie, Louis, Sarah	6	3.0
Hugo	5	2.5
Clara	1	0.5
Maxime	0	0.0

TABLE 6 – Fréquence des erreurs par profils.

6.6 Récurrence inter-runs des erreurs

L'analyse de la répétabilité des erreurs montre que :

- environ 65% des erreurs apparaissent dans au moins trois runs,
- moins de 20% des erreurs sont observées sur un seul run.

Cette forte récurrence inter-runs indique que les erreurs observées relèvent principalement de limites structurelles du processus d'extraction plutôt que d'un comportement aléatoire du modèle.

6.7 Influence du format de CV

Les erreurs se concentrent majoritairement sur un nombre restreint de templates.

Templates	Nombre d'erreurs	Fréquence (%)
38	12	24.0
37	10	20.0
39	6	12.0
11, 29, 32, 33	5	10.0
15, 21	4	8.0
13, 19, 20	3	6.0
17	2	4.0
12, 18	1	2.0
1–10, 14, 16, 22–28, 30, 31, 34–36, 40	0	0.0

TABLE 7 – Nombre et fréquence d'erreurs par template (50 observations par template, tous runs confondus).

Certaines templates présentent des omissions fréquentes, notamment sur les centres d'intérêt, sans différences structurelles évidentes par rapport aux autres formats, ce qui suggère une interaction spécifique entre la mise en page et le mécanisme d'extraction. Parmi les templates présentant le plus d'erreurs, nous avons en majorité des templates Resume.io ; cependant, des templates du même site obtiennent également un taux d'erreur nul, ce qui rend difficile de déceler un vrai schéma. Les templates Canva ont tous un taux d'erreur nul.

6.8 Synthèse quantitative

Les résultats quantitatifs montrent que :

- le modèle présente un taux de succès élevé (96.5%),
- les erreurs sont majoritairement des omissions,
- elles sont fortement concentrées sur les centres d'intérêt,
- elles sont largement répétables d'un run à l'autre,

- elles dépendent davantage du format documentaire que d’une instabilité intrinsèque du modèle.

7 Analyse qualitative des erreurs

L’analyse qualitative des erreurs observées lors de la seconde étude vise à caractériser la nature des écarts entre les extractions produites par le LLM et les CV de référence. Le protocole reposant sur cinq runs successifs permet non seulement d’identifier les types d’erreurs, mais également d’évaluer leur récurrence et leur stabilité, offrant ainsi une distinction claire entre erreurs structurelles et variations ponctuelles.

7.1 Typologie des erreurs

Les erreurs observées se répartissent principalement en deux grandes catégories :

- **Erreurs par omission** : informations présentes dans le CV de référence mais absentes de l’extraction finale.
- **Erreurs par modification** : informations extraites mais partiellement altérées, incluant des simplifications excessives, des reformulations imprécises, des incohérences de dates ou des noms tronqués.

Les erreurs par omission représentent la majorité des cas observés (**86%**). Elles concernent principalement des champs facultatifs ou faiblement structurés, suggérant que le modèle privilégie la robustesse sur les sections centrales au détriment des informations perçues comme secondaires.

7.2 Récurrence des erreurs selon les runs

L’analyse comparative des cinq runs révèle que de nombreuses erreurs apparaissent de manière récurrente, ce qui indique des limites structurelles du processus d’extraction plutôt qu’un comportement aléatoire du modèle.

Les erreurs systématiques les plus fréquemment observées sont les suivantes :

- Omission du diplôme du *Baccalauréat* pour le profil *ThomasG*, principalement dans les formats 37 et 38.
- Omission répétée des centres d’intérêt techniques (*Tech*, *Sports*) pour le profil *ThomasL*.
- Omission récurrente de l’intérêt *Culture* pour le profil *Camille*, en particulier dans les formats 29 et 38.
- Omission des expériences de *bénévolat* pour le profil *Ines*, notamment dans le format 15.
- Omission de l’intérêt *Marathon training* pour le profil *Hugo* (format 11).

Ces erreurs sont observées dans au moins quatre runs sur cinq, ce qui confirme leur caractère stable et prédictible.

7.3 Analyse par type de champ extrait

7.3.1 Centres d’intérêt

Les centres d’intérêt constituent la catégorie la plus sensible aux erreurs, représentant **75%** des erreurs observées. Les omissions concernent principalement des intérêts culturels (*Culture*, *Cultural Trends*), techniques (*Tech*, *Data Visualization*) ou des activités spécifiques (*Marathon training*).

Ces erreurs apparaissent souvent dans des templates où la section *Centres d’intérêt* est segmentée par un saut de page. Dans les autres cas, ce sont systématiquement le dernier ou les deux derniers éléments de la liste qui ne sont pas extraits. La variabilité inter-runs est extrêmement faible : les mêmes profils et les mêmes templates sont affectés de manière répétée, indiquant une limite claire du mécanisme d’extraction face à certaines structures de pagination.

7.3.2 Études

Les erreurs liées à la section *Études* sont moins fréquentes et concernent principalement :

- l’omission de diplômes secondaires, notamment le *Baccalauréat* (observée sur un seul profil et deux templates mais 5 runs),
- des incohérences sur les dates, liées à la diversité des formats de présentation (années, plages de dates, formats abrégés),
- l’omission du champ d’étude pour un diplôme (deux occurrences sur un même profil).

Ces observations suggèrent une difficulté du modèle à hiérarchiser les formations lorsque plusieurs niveaux académiques sont présents, en particulier lorsque les diplômes secondaires sont graphiquement moins mis en avant.

7.3.3 Expériences professionnelles et bénévolat

Les erreurs observées dans les expériences professionnelles restent marginales. Elles concernent principalement :

- l’omission d’expériences de bénévolat, lorsque celles-ci sont présentées dans une section distincte ou peu explicite,
- des simplifications excessives des descriptions de missions,
- des incohérences mineures sur les noms d’entreprises ou les périodes, notamment dans le cas d’entités génériques (ex. *E-commerce Startup*) ou d’expériences de free-lancing.

Le bénévolat apparaît ainsi comme un cas limite pour le modèle : il est parfois sous-extrait, ignoré ou reclassé dans une autre catégorie, bien que ces erreurs restent peu fréquentes. Cette erreur vient notamment du fait que, selon les profils, les expériences de volontariat sont mises dans des sections différentes, ce qui rend la tâche difficile pour le LLM d’être constant.

7.4 Lien entre format et nature des erreurs

Les formats présentant les taux d’erreur les plus élevés (notamment les formats 37, 38 et 39) concentrent également les omissions les plus sévères. Ces templates proviennent tous de la plateforme *Resume.io*. Toutefois, aucune différence structurelle majeure (contenu, densité ou organisation des sections) n’a été identifiée pour expliquer de manière évidente ces écarts de performance, en particulier pour les centres d’intérêt.

7.5 Synthèse

L’analyse qualitative met en évidence que les erreurs observées sont majoritairement :

- répétables d’un run à l’autre,
- concentrées sur des champs faiblement structurés et de faible criticité,
- fortement dépendantes du format et de la mise en page des CV.

Ces résultats indiquent que les limites identifiées relèvent davantage de la représentation documentaire et des choix de mise en forme que d’une instabilité intrinsèque du modèle de langage, même si l’origine des différences d’extraction n’a pas été clairement identifiée pour tous les cas.

Autres points importants à noter ; si les sections ont des noms non explicites ou trop originaux, le LLM d’extraction peut courir le risque de ne pas relever la section en totalité (Ex : "What I did" à la place de "Professional Experiences" ou "Work Experiences")

Troisième partie

Troisième partie : Confiance des résultats

8 Intervalle de confiance

8.1 Cadre théorique

8.1.1 Structure expérimentale

Le nombre d'observations varie selon l'étude considérée. Chaque observation correspond à une comparaison entre une extraction produite par le LLM et un CV de référence.

— **Étude 1** : 3 candidats, 11 templates (après suppression des templates incorrects), 1 run $\Rightarrow n = 3 \times 11 \times 1 = 33$.

— **Étude 2** : 10 candidats, 40 templates, 5 runs $\Rightarrow n = 10 \times 40 \times 5 = 2000$.

— **Étude générique** : N candidats, T templates, R runs $\Rightarrow n = N \times T \times R$.

Les runs sont indépendantes les unes des autres (le LLM n'apprend rien des runs précédentes pour les nouvelles runs), on peut donc considérer que toutes les observations sont indépendantes.

8.1.2 Modélisation statistique

Pour un biais ou une erreur donnée, on observe 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 dans lequel se situe la probabilité réelle d'erreur du système, compte tenu de la taille finie de l'échantillon.

8.1.3 Intervalle de confiance de Wilson (95%)

Afin d'obtenir une estimation plus robuste que l'approximation normale classique, notamment pour des proportions faibles, nous utilisons l'intervalle de confiance de Wilson à 95%, défini comme suit (avec $z = 1.96$), où z correspond au quantile de la loi normale centrée réduite associé au niveau de confiance choisi. Autrement dit, $z = 1.96$ est la valeur telle que 95% de la masse de probabilité de la loi normale standard se situe dans l'intervalle $[-z, z]$.

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

8.2 Calculs numériques

8.2.1 Étude 1 ($n = 33$, $x = 1$)

La proportion observée est :

$$\hat{p}_1 = \frac{1}{33} \approx 3.03\%.$$

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

$$IC_{95\%,1} \approx [0.54\%, 15.32\%]. \quad (3)$$

La largeur importante de cet intervalle s'explique par la taille réduite de l'échantillon, ce qui limite la précision de l'estimation.

8.2.2 Étude 2 ($n = 2000$, $x = 70$)

La proportion observée est :

$$\hat{p}_2 = \frac{70}{2000} = 3.5\%.$$

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

$$IC_{95\%,2} \approx [2.78\%, 4.40\%]. \quad (4)$$

La réduction significative de la largeur de l'intervalle reflète l'augmentation de la taille de l'échantillon et une estimation plus précise du taux d'erreur réel.

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

Niveau de criticité des erreurs pris en compte	IC 95% Wilson
majeure (4)	[0.28 %, 0.91 %]
modéré et majeure (3 et 4)	[1.06 %, 2.13 %]
légère à majeure (2, 3 et 4)	[2.56 %, 4.12 %]
mineure à majeure (1, 2, 3 et 4)	[2.78 %, 4.40 %]

TABLE 8 – Résumé des intervalles de confiance de Wilson à 95% pour différents nombres d'erreurs sur 2000 observations, avec niveau de criticité associé.

8.3 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. Dans le cas de l'étude 2, 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 **2.78% et 4.40%**. Le taux d'erreur majeure se situe avec une probabilité de 95% entre **0.28% et 0.91%**. Ces estimations constituent des bornes fiables pour l'évaluation des performances du système à grande échelle.

9 Analyse du bruit : variabilité inter-runs

Afin d'évaluer la stabilité du système, nous analysons la variabilité des taux d'erreur observés entre les différents runs. Cette variabilité est assimilée ici au *bruit expérimental*, par opposition à un biais structurel du modèle.

Les taux d'erreur observés sont :

$$p = \{2.8\%, 3.8\%, 3.5\%, 3.8\%, 3.8\%\}.$$

9.1 Taux d'erreur moyen

Le taux d'erreur moyen est défini par :

$$\bar{p} = \frac{1}{R} \sum_{r=1}^R p_r \approx 3.54\%,$$

où :

- p_r : taux d'erreur observé pour le run r ,
- R : nombre total de runs ($R = 5$ dans notre cas),
- \bar{p} : taux d'erreur moyen sur l'ensemble des runs.

Cette valeur est cohérente avec le taux d'erreur global observé sur l'ensemble des 2000 extractions.

9.2 Écart-type inter-runs

La dispersion des performances entre les runs est mesurée par l'écart-type :

$$\sigma_{\text{runs}} = \sqrt{\frac{1}{R-1} \sum_{r=1}^R (p_r - \bar{p})^2},$$

On obtient :

$$\sigma_{\text{runs}} \approx 0.43\%.$$

Cette valeur indique une variabilité limitée entre les runs.

9.3 Coefficient de variation

Pour normaliser cette dispersion par rapport au niveau moyen d'erreur, on calcule le coefficient de variation :

$$CV = \frac{\sigma_{\text{runs}}}{\bar{p}} \approx 12.2\%,$$

où :

— CV : coefficient de variation des taux d'erreur entre runs.

9.4 Interprétation

Un coefficient de variation d'environ 12% indique une variabilité inter-runs modérée. Le système présente donc un comportement stable, sans fluctuations significatives d'un run à l'autre. Le taux d'erreur observé reflète principalement un biais systématique lié à la structure des documents et au prompt du LLM plutôt qu'un bruit aléatoire du processus d'extraction.

10 Autres statistiques (Étude 2)

10.1 Taux d'erreur conditionnel par type de section

Chaque section (*Études*, *Expériences*, *Centres d'intérêt*) comporte 2000 entrées évaluées. Le taux d'erreur conditionnel est défini par :

$$p(\text{erreur} \mid s) = \frac{\text{Nombre d'erreurs dans la section } s}{\text{Nombre total d'entrées de la section } s}.$$

TABLE 9 – Taux d'erreur conditionnel par section

Section	Erreurs	Entrées	Taux d'erreur
Centres d'intérêt	53	2000	2.65%
Études	27	2000	1.35%
Expériences	11	2000	0.55%

Les centres d'intérêt apparaissent comme la section la plus sensible aux erreurs, avec un taux près de cinq fois supérieur à celui des expériences professionnelles. Ce résultat confirme que les sections peu structurées constituent la principale source d'incertitude du système.

10.2 Probabilité qu'un CV contienne une erreur selon la criticité

Au-delà du taux d'erreur par champ, il est pertinent d'évaluer le risque au niveau du CV complet. Nous estimons ici la probabilité qu'un CV contienne au moins une erreur, en tenant compte du niveau de criticité associé.

Pour chaque CV (candidat \times template \times run), on retient la criticité maximale observée parmi les erreurs éventuelles. Les probabilités suivantes sont calculées sur un total de 2000 CV analysés (Étude 2).

10.2.1 Probabilité cumulée par niveau de criticité

On définit :

$$P(\text{CV avec erreur} \geq k) = \frac{\text{Nombre de CV ayant au moins une erreur de criticité} \geq k}{2000}.$$

Les probabilités empiriques observées sont :

$$\begin{aligned} P(\text{CV avec au moins une erreur}) &= 2.90\%, \\ P(\text{CV avec erreur de criticité} \geq 2) &= 2.60\%, \\ P(\text{CV avec erreur de criticité} \geq 3) &= 0.90\%, \\ P(\text{CV avec erreur de criticité} \geq 4) &= 0.35\%. \end{aligned}$$

10.2.2 Interprétation

Ces résultats montrent que, bien que des erreurs puissent survenir, la majorité d'entre elles sont de faible à moyenne criticité. Les erreurs les plus sévères (criticité ≥ 4), susceptibles d'impacter fortement l'interprétation du CV, restent très rares et concernent moins de 0.4% des cas.

Ainsi, le risque opérationnel associé au système d'extraction demeure limité, et les erreurs observées sont majoritairement non bloquantes.

Quatrième partie

Synthèse générale et recommandations

11 Résumé exécutif

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 du format documentaire sur la qualité de l'extraction. Deux études complémentaires ont été menées : une étude préliminaire à petite échelle et une étude à grande échelle incluant des répétitions afin d'évaluer la stabilité du système.

11.1 Principaux résultats

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

- Sur l'étude à grande échelle (2000 CV analysés), le taux de succès atteint **96.5%**.
- Le taux d'erreur global est estimé à **3.5%**, avec un intervalle de confiance à 95% compris entre **2.78%** et **4.40%**.
- La variabilité inter-runs est faible (coefficient de variation d'environ **12%**), indiquant une bonne stabilité du modèle.

L'analyse de la gravité des erreurs montre que :

- la majorité des erreurs sont de faible à moyenne criticité,
- moins de **0.4%** des CV contiennent une erreur critique susceptible d'altérer significativement la compréhension du profil candidat.

Ces résultats indiquent que le système est globalement fiable et exploitable dans un cadre opérationnel.

11.2 Limites du LLM identifiées

L'analyse quantitative et qualitative met toutefois en évidence plusieurs limites importantes :

- Les erreurs sont majoritairement des omissions, concentrées sur des sections faiblement structurées, en particulier les *centres d'intérêt*.
- Certaines erreurs sont fortement dépendantes du format du CV, avec une concentration sur un nombre restreint de templates, notamment issus de la plateforme *Resume.io*.
- Les erreurs observées sont largement répétables d'un run à l'autre, ce qui indique des limites structurelles liées à la mise en page (sauts de page, segmentation graphique, intitulés non standards).

Il est également observé que des intitulés de sections trop originaux ou non explicites (ex. *"What I did"* au lieu de *"Professional Experience"*) augmentent le risque d'extraction incomplète.

11.3 Limites de l'étude identifiées

L'étude présente cependant plusieurs limites.

- Les CV utilisés ne sont pas de vrais CV (cela nécessiterait un processus d'anonymisation complexe) ; il s'agit exclusivement de textes de CV générés par IA, avec une vérification humaine du contenu.
- Les profils des CV sont similaires : en effet, l'étude s'est concentrée sur des profils postulant à une école de commerce, ce qui conduit à des CV relativement homogènes. Toutefois, cela ne constitue pas nécessairement un problème tant que le cadre d'application reste comparable ; une étude complémentaire serait néanmoins requise pour des contextes plus éloignés.
- Le nombre et la nature des différents éléments au sein de chaque section varient peu, pour les mêmes raisons, mais également pour des considérations de praticité lors de la création des CV.
- L'étude s'est concentrée uniquement sur les sections **Studies**, **Professional Experiences** et **Interests**. D'autres sections pourraient être extraites, mais il est raisonnable de s'attendre à un comportement similaire pour celles-ci.
- Les formats utilisés ne sont, de manière évidente, pas exhaustifs. Nous avons cherché à tester à la fois des CV classiques et populaires, ainsi que des formats plus originaux (notamment Canva), incluant par exemple des templates très colorés ou présentant des organisations moins linéaires. D'autres formats existent toutefois et n'ont pas été pris en compte.
- L'analyse et la comparaison des CV ont été réalisées à l'aide d'un LLM. Cela peut entraîner à la fois l'omission de certaines erreurs significatives et la détection abusive d'autres. Une vérification humaine a été effectuée en aval de manière succincte afin d'éviter les erreurs d'analyse les plus importantes, mais le risque de biais résiduels ne peut être totalement exclu.

11.4 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 la représentation documentaire et aux choix de mise en forme

des CV. Le comportement du système est donc prédictible : les mêmes structures produisent les mêmes erreurs.

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.

11.5 Recommandations et pistes d'amélioration

Plusieurs axes d'amélioration peuvent être envisagés :

- standardiser ou contraindre les formats de CV acceptés, notamment pour les sections facultatives,
- renforcer le prompt ou les règles d'extraction pour les sections faiblement structurées (centres d'intérêt, bénévolat),
- intégrer une étape de post-traitement ou de validation ciblée sur les templates identifiés comme à risque,

En conclusion, l'utilisation de l'IA pour l'extraction de CV apparaît pertinente, efficace et robuste dans la majorité des cas. 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.

A Annexes : CV de Référence de la première étude

A.1 Profil 1 : Alexandre DUPONT

Alexandre DUPONT

*Candidate for Master in Management –
Corporate Finance*

75016 Paris, France
☎ +33 6 12 34 56 78
✉ alexandre.dupont@email.com
in alexandre-dupont
Driving License

Profile

Rigorous Economics & Management student with a strong analytical mindset. Top of the class in Microeconomics, I consolidated my theoretical knowledge with a significant internship in financial audit. I aim to join a Top Business School to develop expertise in corporate finance and strategy.

Education

- 2023–2026 **Bachelor in Economics & Management**, *University Paris-Dauphine*, Paris
- **GPA/Results:** Current average 16/20 (Expected Highest Honors). Valedictorian in Microeconomics.
 - **Key Courses:** Financial Analysis, Advanced Accounting, Applied Statistics, Monetary Macroeconomics.
- 2023 **French Baccalauréat (High School Diploma)**, *Lycée Condorcet*, Paris
- Specialization in Math & Physics. Highest Honors (Mention Très Bien), Jury Congratulations.

Professional Experience

- June – Aug 2025 **Junior Audit Assistant (Internship)**, *Mazars*, Paris La Défense
- Participated in statutory audit missions for industrial clients.
 - Verified financial statement consistency and performed analytical review of 3 major files.
 - Conducted details testing and third-party circularization.
 - Intensive use of Excel for accounting data processing.
- Sept 2024 – Present **Treasurer**, *Dauphine Debate Association*, Paris
- Managed and monitored an annual budget of €15,000.
 - Optimized event costs and produced semi-annual financial reporting.
 - Negotiated partnerships: increased funds raised by 20% year-over-year.

Skills & Languages

- | | | | |
|------------------|--|----------------|--------------------------------|
| Languages | English (C1 - TOEFL 105), Spanish (B2) | Finance | Financial Analysis, Accounting |
| IT Skills | Excel Expert (TOSA 950), VBA (Basics) | Tools | Bloomberg (Introduction) |

Interests

- Chess Regional competition (8 years of practice). Strategic thinking development.
- Piano 10 years at the Conservatory. Discipline and perseverance.

A.2 Profil 2 : Thomas LEGRAND

Thomas LEGRAND

Grande École Candidate – Innovation & Digital

69002 Lyon, France
+33 6 98 76 54 32
✉ thomas.legrand@email.com
🌐 www.thomas-folio.com
in [thomas-legrand](#)

Profile

Creative and proactive student ("Doer"), passionate about digital growth levers. With concrete experience in startups and freelancing, I master web marketing tools. I wish to join a Grande École to hybridize my technical skills with a high-level strategic business vision.

Education

2023–2026 **Bachelor in Digital Project Management, IUT Lyon 1 / École du Web, Lyon**

Projects: Created an MVP for a carpooling app (Design Thinking, UX/UI).

Award: Winner of the "Lyon Start-Up 2024" Hackathon.

Experience & Projects

Jan – June 2025 **CEO's Right Hand / Executive Intern, "PayFast" FinTech (Station F), Paris**

- Contributed to the Go-to-market strategy for the launch of a new B2B offer.
- Growth Hacking: Automating LinkedIn prospecting (+30 qualified leads/week).
- Video content creation and social media management.

2024 – 2026 **Freelancer / Self-Employed, Web Development & Marketing, Lyon**

- Designed websites for 5 local SMEs (WordPress & Shopify).
- Managed Facebook Ads campaigns (Budget €2k, ROI x3).
- Full client relationship management: Briefing, quoting, and invoicing.

Technical Skills

Web HTML/CSS, WordPress, Shopify, Notion
Data/Mkg Python (Data Analysis), Google Analytics (GA4), HubSpot
Design Adobe Creative Suite (Photoshop, Premiere Pro)

Languages

English B2+
Chinese A2

Fluent Professional English
Beginner

Interests

Travel Backpacking road-trip in Southeast Asia (Solo, 1 month).
Sports Running (Training for Paris Half-Marathon).
Tech Tech Blogger (AI and NoCode) - 1k monthly visitors.

A.3 Profil 3 : Camille MARTIN

Camille MARTIN

Grande École Candidate – International & Social Impact

33000 Bordeaux, France

+33 6 55 44 33 22

✉ camille.martin@email.com

in camille-martin

Profile

Trilingual Political Science student driven by current geopolitical and societal issues. My background, marked by international mobility and community engagement, demonstrates my adaptability and leadership. I aim to join a Grande École to acquire the managerial skills necessary for careers in CSR and International Affairs.

Education

2023–2026 **Bachelor in Political Science, University of Bordeaux**

- **International Exchange:** Semester at McGill University (Montreal, Canada). Focus: International Relations.
- **Thesis:** "The impact of CSR standards on European business competitiveness".

Experience & Volunteering

Summer 2025 **International Humanitarian Mission (2 months), NGO "Education For All", Senegal**

- Coordinated a multicultural team of 5 international volunteers.
- Taught French and provided tutoring (primary and middle school levels).
- Managed camp logistics and relations with local stakeholders.

2024 – 2025 **Public Affairs Intern, City Hall of Bordeaux, Bordeaux**

- Drafted synthesis notes for elected officials and press releases.
- Co-organized cultural and civic events (Budget: €50k).
- Conducted legislative monitoring on environmental topics.

Languages

English C1

IELTS 7.5 - Fluent

German B2/C1

Advanced proficiency

Arabic A1

Basic notions (Modern Standard)

Soft Skills

- Public Speaking
- Intercultural Communication
- Conflict Resolution
- Analytical Synthesis

Interests

Geopolitics Active member of MUN (Model United Nations).

Scouting Team Leader (3 years). Values of solidarity.

Culture Russian literature and South American cinema.

B Annexes : CV de référence de la deuxième étude

Clara MOREL

Lyon, France · 645218934 · clara.morel@email.com

Profile

Analytical and structured student with a strong interest in strategy and organizational performance.

Professional Experience

Strategy Analyst Intern | Roland Berger Paris
06/2024 – 12/2024

- Conducted market and competitor analysis for industrial clients. Built Excel models and PowerPoint decks for senior consultants. Participated in client interviews and synthesis workshops.

Student Consultant | emlyon Junior Conseil Lyon
09/2023 – 05/2024

- Delivered consulting missions for SMEs (market sizing, pricing). Led a team of 4 students on a retail optimization project.

Education

Bachelor in Management (Global BBA) 2022 – 2025
emlyon business school, Lyon

Baccalauréat (Maths & Economics) 2021 – 2022
Lycée du Parc, Lyon

Skills & Languages

- English:** C1 Level
- Spanish:** B2 Level
- Business & Strategy:** Market Analysis, Competitive Intelligence, Problem Solving.
- IT Skills:** Excel (Advanced), PowerPoint (Expert).

Interests

Debating (National competitions.), **Sports** (Long-distance running.), **Culture** (Economic podcasts.).

Hugo LEMAIRE

Paris, France · 678124409 · hugo.lemaire@email.com

Profile

Finance-oriented economics student with a strong quantitative background aiming for investment banking.

Professional Experience

Corporate Finance Intern | Société Générale CIB Paris
06/2024 – 08/2024

- Assisted in financial modeling and company valuation. Prepared pitchbooks and market analysis for M&A transactions. Conducted comparable companies and precedent transactions analyses.

Private Equity Analyst (Part-time) | Student Investment Fund Paris
01/2024 – 05/2024

- Analyzed investment opportunities in SMEs. Built LBO models and investment memos.

Education

Bachelor in Economics 2021 – 2024
University Paris-Dauphine, Paris

Baccalauréat (Maths & Physics) 2020 – 2021
Lycée Saint-Louis, Paris

Skills & Languages

- English:** C1 - IELTS 7.5
- French:** Native
- Finance:** Valuation, Financial Modeling, Accounting.
- IT Skills:** Excel (Advanced), VBA (Intermediate).

Interests

Stock market investing, Chess, Marathon training.

Sarah BENALI

Marseille, France · 691335672 · sarah.benali@email.com

Profile

Creative and analytical marketing student with strong interest in brand strategy and consumer behavior.

Professional Experience

Marketing Intern | L'Oréal France Paris
06/2024 – 12/2024

- Assisted brand managers in the preparation and launch of new skincare products. Analyzed sales performance and campaign KPIs using Excel and internal dashboards. Coordinated with creative agencies and influencers for digital campaigns. Conducted competitive analysis and consumer trend monitoring.

Digital Marketing Assistant | E-commerce Startup Marseille
09/2023 – 05/2024

- Managed social media content calendars (Instagram, TikTok). Monitored online performance metrics and engagement rates. Supported SEO optimization and newsletter campaigns.

Education

Bachelor in Business Administration 2022 – 2025
KEDGE Business School, Marseille

Baccalauréat (Economics & Social Sciences) 2021 – 2022
Lycée Thiers, Marseille

Skills & Languages

- English:** C1 - TOEIC 945
- Arabic:** Native
- Marketing:** Brand Strategy, Market Research, Campaign Analysis.
- IT Skills:** Excel (Advanced), Power BI, Google Analytics, Canva.

Interests

Photography (Lifestyle and travel portfolio on Instagram.), **Trends** (Fashion, consumer trends, contemporary art.).

Thomas GIRARD

Nantes, France · 622489105 · thomas.girard@email.com

Profile

Entrepreneurial-minded student with practical experience in startup co-founding and business development.

Professional Experience

Co-founder & Operations Manager | GreenMove (Student Startup) Nantes
01/2023 – 01/2026

- Co-founded a sustainable mobility startup offering shared electric bikes for students. Managed daily operations and partnerships with local institutions. Built financial forecasts and monitored monthly cash flow. Led customer acquisition campaigns, reaching over 1,000 registered users.

Business Development Intern | Startup Incubator Nantes
06/2023 – 08/2023

- Supported early-stage startups in market analysis and go-to-market strategy. Assisted entrepreneurs in preparing pitch decks for investors. Participated in mentoring sessions and demo days.

Education

Bachelor in Economics & Management 2021 – 2024
University of Nantes, Nantes

Baccalauréat (Maths & Economics) 2020 – 2021
Lycée Clémenceau, Nantes

Skills & Languages

- English:** C1 Level
- French:** Native
- Entrepreneurship:** Business Development, Financial Forecasting, Project Management.
- IT Skills:** Excel (Advanced), Notion, Trello.

Interests

Entrepreneurship (Startup communities and innovation events.), **Sports** (Competitive sailing (6 years).).

Inès ROBERT

Paris, France · 637841962 · ines.robert@email.com

Profile

Economics student interested in public policy, management, and economic regulation with think tank experience.

Professional Experience

Policy Research Assistant | French Economic Think Tank Paris
09/2024 – 02/2025

- Conducted literature reviews and data analysis for policy briefs. Contributed to reports on labor market reforms and public spending efficiency. Participated in stakeholder interviews and policy workshops.

Public Affairs Intern | Ministry of Economy and Finance Paris
06/2024 – 08/2024

- Assisted senior analysts in economic impact assessments. Prepared briefing notes and synthesis documents for internal use. Monitored legislative developments related to economic policy.

Education

Bachelor in Economics & Social Sciences 2022 – 2025
Sciences Po Paris, Paris

Baccalauréat (Economics) 2021 – 2022
Lycée Henri-IV, Paris

Skills & Languages

- English:** C1 Level
- German:** B2 Level
- Public Policy:** Policy Evaluation, Economic Analysis, Report Writing.
- IT Skills:** Excel (Advanced), Stata (Intermediate).

Interests

Public Debate (Economic conferences and policy forums.), **Volunteering** (Tutor for underprivileged high school students.).

Louis CARON

Grenoble, France · 659027741 · louis.caron@email.com

Profile

Quantitative-oriented student with a background in applied mathematics, statistics, and business analytics.

Professional Experience

Data Analyst Intern | Retail Analytics Company Lyon
06/2024 – 09/2024

- Cleaned and analyzed large customer datasets using Python and SQL. Built dashboards to support marketing and pricing decisions. Presented insights to non-technical stakeholders.

Business Analytics Project | Academic Team Grenoble
01/2024 – 05/2024

- Worked in a team to analyze sales performance for a retail chain. Developed forecasting models and recommendations for inventory optimization.

Education

Bachelor in Applied Mathematics & Economics 2021 – 2024
Université Grenoble Alpes, Grenoble

Baccalauréat (Maths & Physics) 2020 – 2021
Lycée Champollion, Grenoble

Skills & Languages

- English:** C1 Level
- French:** Native
- Data & Analytics:** Statistical Analysis, Data Visualization, Predictive Modeling.
- IT Skills:** Python (Pandas, NumPy), SQL, Excel (Advanced), Power BI.

Interests

Data Visualization (Transforming complex data into clear insights.), **Sports** (Hiking and trail running.).

Amélie DUBOIS

Paris, France · 614885329 · amelie.dubois@email.com

Profile

Sustainability-oriented student specializing in corporate responsibility, ESG analysis, and strategic transformation.

Professional Experience

Sustainability Consulting Intern | EY Climate Change & Sustainability Services Paris
06/2024 – 12/2024

- Contributed to ESG diagnostics and carbon footprint analyses for corporate clients. Assisted in the preparation of sustainability reports aligned with CSRD standards. Conducted benchmarking on sustainable business practices across industries.

Project Coordinator (Part-time) | Environmental NGO Paris
09/2023 – 05/2024

- Coordinated sustainability awareness projects with universities. Managed project timelines, budgets, and stakeholder communication. Prepared impact assessment reports.

Education

Bachelor in Management & Sustainability 2022 – 2025
Paris School of Economics, Paris

Baccalauréat (Economics & Sciences) 2021 – 2022
Lycée Louis-le-Grand, Paris

Skills & Languages

- English:** C1 Level
- Italian:** B2 Level
- Sustainability:** ESG Analysis, Sustainable Finance, Strategic Benchmarking.
- IT Skills:** Excel (Advanced), Power BI, ESG reporting tools.

Interests

Environmental Engagement (Climate-related student initiatives.), **Wellness** (Yoga and mindfulness.).

Maxime PETIT

Lille, France · 673610488 · maxime.petit@email.com

Profile

Operations-focused student with a background in industrial engineering, supply chain optimization, and lean management.

Professional Experience

Operations Intern | Decathlon Logistics Lille
06/2024 – 09/2024

- Analyzed warehouse processes and identified efficiency improvement areas. Supported the implementation of lean management tools. Built performance dashboards to monitor key operational indicators.

Supply Chain Project | Academic Lille
02/2024 – 05/2024

- Modeled supply chain flows and demand forecasts for a retail company. Proposed inventory optimization strategies to reduce stockouts.

Education

Bachelor in Industrial Engineering & Management 2021 – 2024
IMT Nord Europe, Lille

Baccalauréat (Maths & Engineering) 2020 – 2021
Lycée Faidherbe, Lille

Skills & Languages

- English:** C1 Level
- French:** Native
- Operations:** Process Optimization, Lean Management, Supply Chain Analysis.
- IT Skills:** Excel (Advanced), Python (Basics), Power BI.

Interests

Innovation (Industry 4.0 and automation.), **Sports** (Competitive basketball.).

Thomas LEGRAND

Lyon, France · 698765432 · thomas.legrand@email.com

Profile

Creative and proactive student passionate about digital growth levers, with experience in startups and freelancing.

Professional Experience

CEO's Right Hand / Executive Intern | 'PayFast' FinTech Paris
01/2025 – 06/2025

- Contributed to the Go-to-market strategy for the launch of a new B2B offer. Growth Hacking: Automating LinkedIn prospecting (+30 qualified leads/week). Video content creation and social media management.

Freelancer / Self-Employed | Self-Employed Lyon
01/2024 – 01/2026

- Designed websites for 5 local SMEs (WordPress & Shopify). Managed Facebook Ads campaigns (Budget €2k, ROI x3). Full client relationship management: Briefing, quoting, and invoicing.

Education

Bachelor in Digital Project Management 2023 – 2026
IUT Lyon 1 / École du Web, Lyon

Skills & Languages

- English:** B2+ Fluent Professional
- Chinese:** A2 Beginner
- Web Development:** HTML/CSS, WordPress, Shopify, Notion.
- Data & Marketing:** Python (Data Analysis), Google Analytics (GA4), HubSpot.
- Design:** Adobe Creative Suite (Photoshop, Premiere Pro).

Interests

Travel (Backpacking road-trip in Southeast Asia.), **Sports** (Running (Paris Half-Marathon).), **Tech** (Tech Blogger (AI and NoCode).).

Camille MARTIN

Bordeaux, France · 655443322 · camille.martin@email.com

Profile

Trilingual Political Science student driven by geopolitical issues, with international and humanitarian experience.

Professional Experience

International Humanitarian Mission | NGO 'Education For All' Dakar
06/2025 – 08/2025

- Coordinated a multicultural team of 5 international volunteers. Taught French and provided tutoring (primary and middle school levels). Managed camp logistics and relations with local stakeholders.

Public Affairs Intern | City Hall of Bordeaux Bordeaux
01/2024 – 12/2025

- Drafted synthesis notes for elected officials and press releases. Co-organized cultural and civic events (Budget: €50k). Conducted legislative monitoring on environmental topics.

Education

Bachelor in Political Science 2023 – 2026
University of Bordeaux, Bordeaux

International Exchange —
McGill University, Montreal

Skills & Languages

- English:** C1 - IELTS 7.5
- German:** B2/C1 Advanced
- Arabic:** A1 Basic
- Soft Skills:** Public Speaking, Intercultural Communication, Conflict Resolution, Synthesis.
- Analysis:** Legislative monitoring, Event organization.

Interests

Geopolitics (Member of MUN (Model United Nations).), **Scouting** (Team Leader (3 years).), **Culture** (Russian literature and South American cinema.).

C Etude 2 : liste détaillée des erreurs

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

TABLE 10 – Légende des niveaux de criticité pour les erreurs d'extraction.

Candidat	Type	Section(s)	Résumé	Criticité
Camille 38	Modification	Studies, Interests	'Education For All' missing country/city, McGill field missing, 'Culture' missing	4
Thomasg 38	Omission	Studies	'French Baccalauréat' missing	4
Thomasl 38	Modification	Studies	'Freelance / Self-employed' company/date discrepancy, 'École du Web' extra	3
Camille 11	Omission	Studies, Interests	McGill Exchange missing, 'Culture' missing	3
Hugo 11	Modification	Interests	'Stock market investing'→'Investing', 'Chess' missing, 'Marathon training' missing	3
Ines 13	Modification	Professionnal Exp.	Volunteer activity incomplete	2
Louis 13	Modification	Studies	Company name discrepancy, degree title difference	3
Thomasg 37	Omission	Studies	'French Baccalauréat' missing	4
Thomasl 37	Modification	Studies	'Freelance / Self-employed' date/company issue, extra university name	3
Sarah 19	Modification	Studies, Prof. Exp.	Company missing, Baccalauréat field missing	3
Camille 29	Modification	Studies	'Education For All' missing country/city, McGill field mismatch	3
Sarah 32	Omission	Interests	'Cultural Trends' missing	2
Thomasl 33	Modification	Studies	'Freelance / Self-employed' date/company issue, extra university name	3
Thomasl 39	Modification	Studies	'PayFast Fintech'→'PayFast', 'IUT Lyon 1'→'IUT Lyon 1 / École du Web'	3

TABLE 11 – Erreurs pour Run 1.

Candidat	Type	Section(s)	Résumé	Criticité
Camille 38	Omission	Interests	'Culture' missing	2
Thomasg 38	Omission	Studies	'French Baccalauréat' missing	4
Thomasl 38	Omission	Interests	'Tech' missing	2
Thomasg 37	Omission	Studies	'French Baccalauréat' missing	4
Thomasl 37	Omission	Interests	'Tech' missing	2
Hugo 11	Omission	Interests	'Marathon training' missing	2
Ines 13	Omission	Professionnal Exp.	Volunteering missing	2
Ines 15	Omission	Professionnal Exp.	Tutoring missing	4
Amelie 17	Modification	Studies	'Lycée Louis-le-Grand' dates modified	1
Amelie 20	Modification	Studies	'Lycée Louis-le-Grand' dates modified	1
Louis 21	Omission	Interests	'Data Visualization' missing	2
Camille 29	Omission	Interests	'Culture' missing	2
Sarah 32	Omission	Interests	'Cultural Trends' missing	2
Thomasl 33	Omission	Interests	'Tech' missing	2
Thomasl 39	Omission	Interests	'Sports' missing, 'Tech' missing	3

TABLE 12 – Erreurs pour Run 2.

Candidat	Type	Section(s)	Résumé	Criticité
Camille 38	Omission	Interests	'Culture' missing	2
Thomasg 38	Omission	Studies	'French Baccalauréat' missing	4
Thomasl 38	Omission	Interests	'Tech' missing	2
Thomasg 37	Omission	Studies	'French Baccalauréat' missing	4
Thomasl 37	Omission	Interests	'Tech' missing	2
Louis 39	Modification	Professionnal Exp.	'Business Analytics Project' description incomplete	3
Thomasl 39	Omission	Interests	'Sports' missing, 'Tech' missing	3
Hugo 11	Omission	Interests	'Marathon training' missing	2
Ines 15	Omission	Professionnal Exp.	Tutoring missing	2
Amelie 20	Modification	Studies	'Lycée Louis-le-Grand' dates modified	1
Louis 21	Omission	Interests	'Data Visualization' missing	2
Camille 29	Omission	Interests	'Culture' missing	2
Sarah 32	Omission	Interests	'Cultural Trends' missing	2
Thomasl 33	Omission	Interests	'Tech' missing	2

TABLE 13 – Erreurs pour Run 3.

Candidat	Type	Section(s)	Résumé	Criticité
Camille 38	Omission	Interests	'Culture' missing	2
Thomasg 38	Omission	Studies	'French Baccalauréat' missing	4
Thomasl 38	Omission	Interests	'Tech' missing	2
Clara 19	Modification	Studies, Prof. Exp.	'emlyon' country inconsistent, 'Lycée du Parc' field missing	3
Louis 19	Omission	Professionnal Exp.	'Business Analytics Project' missing	4
Thomasg 37	Omission	Studies	'French Baccalauréat' missing	4
Thomasl 37	Omission	Interests	'Tech' missing	2
Hugo 11	Omission	Interests	'Marathon training' missing	2
Ines 15	Omission	Professionnal Exp.	Tutoring missing	4
Amelie 20	Modification	Studies	'Lycée Louis-le-Grand' dates modified	1
Louis 21	Omission	Interests	'Data Visualization' missing	2
Camille 29	Omission	Interests	'Culture' missing	2
Sarah 32	Omission	Interests	'Cultural Trends' missing	2
Thomasl 33	Omission	Interests	'Tech' missing	2
Thomasl 39	Omission	Interests	'Sports' missing, 'Tech' missing	3

TABLE 14 – Erreurs pour Run 4.

Candidat	Type	Section(s)	Résumé	Criticité
Amelie 17	Modification	Studies	'Lycée Louis-le-Grand' dates modified	1
Amelie 18	Modification	Studies, Interests	'French Baccalauréat' field changed, 'Yoga and Mindfulness'→'Wellness'	3
Camille 29	Omission	Interests	'Culture' missing	2
Camille 38	Omission	Interests	'Culture' missing	2
Hugo 11	Omission	Interests	'Marathon training' missing	2
Ines 12	Omission	Professionnal Exp.	Tutoring missing	2
Ines 15	Omission	Professionnal Exp.	Tutoring missing	2
Louis 21	Omission	Interests	'Data Visualization' missing	2
Sarah 32	Omission	Interests	'Cultural Trends' missing	2
Thomasg 37	Omission	Studies	'French Baccalauréat' missing	4
Thomasg 38	Omission	Studies	'French Baccalauréat' missing	4
Thomasl 33	Omission	Interests	'Tech' missing	2
Thomasl 37	Omission	Interests	'Tech' missing	2
Thomasl 38	Omission	Interests	'Tech' missing	2
Thomasl 39	Omission	Interests	'Sports' missing, 'Tech' missing	3

TABLE 15 – Erreurs pour Run 5.