

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

Laura BARDOU - Sacha BEAUJEAN - Justin COCHET GARSIOT

14 janvier 2026

Résumé

Ce rapport présente les résultats d'une analyse technique portant sur la capacité d'un modèle de langage (LLM) à extraire des informations structurées (intérêts, études, expériences) à partir de divers formats de CV. L'étude compare 39 documents générés à partir de 3 profils types sur 13 templates différents.

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).

1.1 Constitution du Corpus

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

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

1.2 Protocole de Test

1. **Génération** : Création des CVs en PDF.
2. **Extraction** : Passage des CVs 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 et contenant les informations exactes des profils.

2 Résultats Quantitatifs

2.1 Performance Globale

Sur l'ensemble des 39 documents traités, le taux de succès global est de **74.4%**. Les erreurs ne sont pas réparties uniformément entre les candidats, comme le montre le tableau ci-dessous.

TABLE 1 – Taux d'erreur par profil candidat

Candidat	Erreurs	Total CVs	Taux d'Erreur
Alexandre	2	13	15%
Camille	6	13	46%
Thomas	2	13	15%

2.2 Performance par Format

L'analyse par numéro de template révèle une disparité technique importante. Le format N°13 est le seul présentant un échec systématique.

TABLE 2 – Distribution des erreurs par template

Format	Échecs / Total	Taux d'Erreur
N° 13	3 / 3	100.0%
N° 02, 04, 07, 08, 09, 10, 12	1 / 3	33.3%
N° 01, 03, 05, 06, 11	0 / 3	0.0%

TABLE 3 – Liste détaillée des erreurs détectées par CV et Format

Format	Candidat	Type	Détails de l'erreur
N° 13	Alexandre	Omission	Section 'Intérêts' non détectée par l'extraction
N° 13	Camille	Omission	Section 'Intérêts' non détectée par l'extraction
N° 13	Thomas	Omission	Expérience Freelance (Dates/Ent. manquantes) + Intérêts absents
N° 02	Camille	Modification	Hallucination : Ajout de dates et domaine (McGill) non présents dans l'original
N° 04	Thomas	Omission	Section 'Intérêts' non détectée
N° 07	Camille	Omission	Omission de l'expérience 'Mission Humanitaire' (ONG)
N° 08	Camille	Modification	Omission 'Mission Humanitaire' + Hallucination description 'Affaires Publiques'
N° 09	Camille	Omission	Omission de l'expérience 'Mission Humanitaire' (ONG)
N° 10	Camille	Omission	Omission de l'expérience 'Mission Humanitaire' (ONG)
N° 12	Alexandre	Omission	Section 'Intérêts' non détectée

3 Analyse et Discussion des Erreurs

L'analyse qualitative des échecs a permis d'identifier quatre causes racines principales, liées soit à la structure du document, soit à la sémantique du contenu.

3.1 Structure du Document : Le cas du Format N°13

Le template N°13 (issu de *Resume.io*) présente un taux d'erreur de 100%.

- **Cause :** Ce template spécifique ne contient pas de section dédiée aux "Intérêts" ou celle-ci n'est pas rendue visible dans le PDF final.
- **Conséquence :** Le LLM tente d'extraire une information inexiste, entraînant une hallucination ou une erreur nulle fatale lors de la comparaison avec la référence.
- **Conclusion :** L'erreur provient du template et non du LLM : cas à ignorer.

3.2 Confusion Sémantique : Le profil de Camille (Bénévolat)

Le profil de Camille enregistre le taux d'erreur le plus élevé (46%).

- **Cause :** Ce profil contient des expériences de bénévolat. Selon les templates, cette section est soit clairement séparée ("Volontariat"), soit fusionnée visuellement avec l'expérience professionnelle.
- **Conséquence :** Le LLM peine à distinguer le bénévolat de l'expérience salariée stricte, ce qui fausse l'extraction des champs "Expérience Pro".
- **Conclusion :** Le LLM ne sait pas si le volontariat doit faire partie des expériences professionnelles : point à améliorer sur Semantikmatch.

3.3 Entités Nommées : Le profil de Thomas (Freelance)

- **Cause :** Thomas possède une expérience en freelance. Contrairement à un emploi salarié, le nom de l'entreprise n'est souvent pas spécifié explicitement (ou est identique au nom du candidat).
- **Conséquence :** Le LLM échoue à identifier le champ "Entreprise", générant une disparité avec le fichier de référence qui attend probablement une valeur spécifique ou "Indépendant".
- **Conclusion :** L'erreur est faible, c'est un cas particulier.

3.4 Ambiguïté Temporelle : Échanges Internationaux

Une source d'erreur transversale concerne les candidats ayant effectué un échange universitaire.

- **Cause :** Lorsque les dates spécifiques du semestre à l'étranger ne sont pas indiquées, le CV présente deux universités simultanées ou successives sans bornes temporelles claires.
- **Conséquence :** Le LLM extrait deux entités "Éducation" et tente d'inférer des dates manquantes, créant des incohérences par rapport à la référence.
- **Conclusion :** Il y a des risques d'hallucination : point à corriger.

4 Conclusion

L'étude montre une robustesse satisfaisante pour les formats standards. Cependant, l'extraction reste sensible à la structuration de l'information et aux nuances sémantiques.

A Annexes : CVs de Référence

Les documents ci-dessous représentent les profils exacts utilisés pour la génération des différents formats.

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
✉ 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 **Junior Audit Assistant (Internship)**, Mazars, Paris La Défense
- 2025
 - 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 – **Treasurer, Dauphine Debate Association**, Paris
- Present
 - 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](https://www.linkedin.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+ *Fluent Professional English*
Chinese A2 *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
- Conflict Resolution

- Intercultural Communication
- 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.