

IA générative pour le marketing digital

+٥٢٤٣٦١٧ | +٢٠٠٣٩١٤
FACULTÉ DES SCIENCES



M'kouka Btissam , Zahedi Chaymaa

Faculté des Sciences Semlalia

Université Cadi Ayyad

Encadrement: Pr. Mohamed-Amine Chadi

Résumé

Ce projet explore l'intégration de l'intelligence artificielle (IA) générative au sein des processus de marketing digital, en répondant à deux problématiques majeures : la production de supports visuels de haute qualité et l'automatisation de la prospection commerciale. Le travail est articulé autour de deux axes principaux, chacun ayant abouti à une implémentation logicielle concrète dotée d'une interface graphique dédiée.

La première partie traite de la préparation de visuels générés par IA pour l'impression grand format (billboards, flyers). Pour pallier les limites de résolution et les artefacts natifs des modèles de génération d'images, nous avons implanté un pipeline de traitement avancé intégrant des algorithmes de Super-Resolution (Upscaling) et de correction colorimétrique. Ce pipeline permet de transformer des images basse résolution en fichiers prêts pour l'impression, garantissant une netteté et une fidélité chromatique optimales.

La seconde partie se concentre sur l'optimisation de la prospection sur LinkedIn via un workflow d'agents IA. En utilisant des architectures agentiques, nous avons développé un système capable de rechercher, filtrer et extraire des profils de prospects en se basant sur des critères de ciblage complexes. Contrairement aux méthodes de scraping traditionnelles, l'utilisation de l'IA permet une analyse sémantique fine des profils pour assurer une pertinence maximale des leads générés.

Le développement de ce projet a permis de lever plusieurs verrous techniques, notamment la gestion de la mémoire lors du traitement d'images haute définition, le respect des limitations des plateformes sociales et la précision du filtrage intelligent. Les résultats démontrent que la synergie entre la puissance créative de l'IA et des pipelines de traitement rigoureux offre un gain de productivité et de précision significatif pour les professionnels du marketing.

Sommaire

Liste des Figures	vi
1 Introduction Générale et Contexte	1
1.1 Contexte du projet	1
1.2 Problématique	2
1.2.1 L'incompatibilité des sorties IA avec les standards de l'impression	3
1.2.2 Le manque de précision dans la prospection automatisée	3
1.2.3 La complexité d'usage pour les équipes marketing	3
1.3 Objectifs du projet	4
1.3.1 Axe Technique : Optimisation de la Qualité Visuelle	4
1.3.2 Axe Stratégique : Automatisation de la Prospection Intelligente	4
2 Pipeline de Traitement d'Image pour l'Impression Grand Format	5
2.1 Étude et sélection de la source de génération : Le choix d'Ideogram	6
2.1.1 Étude comparative des solutions du marché	6
2.1.2 Justification du choix d'Ideogram.ai	7
2.2 L'approche par Vectorisation : Analyse d'un échec technique	7
2.2.1 Concept et objectifs de la vectorisation	7
2.2.2 Tests réalisés et outils utilisés	8
2.2.3 Analyse de l'échec : les limites du "Raster-to-Vector"	8
2.3 Conception du Pipeline de Traitement d'Image (Solution Retenue)	9
2.3.1 Analyse Intelligente des Métadonnées	9
2.3.2 Nettoyage et Post-traitement (Pre-processing)	10
2.3.3 Double Pipeline d'Upscaling (Hybride)	10
2.3.4 Vérification et Calcul de Résolution (DPI Check)	10
2.3.5 Préparation Pré-presse et Gestion des Couleurs	11
2.3.5.1 Simulation de rendu (Soft Proofing)	11
2.3.5.2 Conversion CMJN Mémoire-Optimisée	11
2.3.5.3 Exportation au Standard Industriel PDF/X-1a	11

Sommaire

3 Automatisation du Sourcing Clients via AI Workflow	12
3.1 Introduction du Chapitre	13
3.2 Concepts Théoriques : Agents Autonomes vs AI Workflows	14
3.2.1 Définition des agents autonomes	14
3.2.2 Scripts classiques vs agents autonomes	15
3.2.3 Contraintes du scraping et limites de l'autonomie	15
3.2.4 Justification du choix : AI Workflow scripté	16
3.3 Architecture Technique et Rôle du MCP	17
3.3.1 Architecture générale de l'AI Workflow	17
3.3.2 Rôle du Model Context Protocol (MCP)	17
3.3.3 Séparation logique : outils, raisonnement et orchestration . .	18
3.4 Module d'Extraction Contrôlée (LinkedIn)	18
3.4.1 Extraction depuis la page de résultats	18
3.4.2 Données extraites	19
3.4.3 Robustesse de l'extraction et simulation humaine	19
3.5 Module d'Enrichissement Intelligent (LLM)	19
3.5.1 Détection du genre (Vision/Nom)	19
3.5.2 Extraction et analyse du parcours éducatif	19
3.5.3 Estimation de l'âge par raisonnement LLM	20
3.6 Agent Autonome de Raisonnement et Filtrage (DB Agent)	20
3.6.1 Objectif de l'agent et filtrage sémantique	20
3.6.2 Architecture "Tool-Augmented Agent"	20
3.6.3 Boucle de raisonnement et contrôle de l'autonomie	21
3.6.4 Export et traçabilité des résultats	21
3.7 Orchestration Globale du Workflow	21
3.7.1 Séquence globale d'exécution	21
3.7.2 Gestion des limites et sécurité technique	22
3.8 Extension : Agent Autonome de Diffusion par Email	22
3.8.1 Rôle et objectifs de l'agent email	22
3.8.2 Caractère autonome et intelligence contextuelle	22
3.8.3 Architecture logique et sécurité	23
3.9 Conclusion du Chapitre	23
4 Résultats	24
4.1 Conception et Architecture de l'Interface Graphique (Partie Image)	24
4.1.1 Import et affichage des métadonnées	24
4.1.2 Résultats du Pipeline de Traitement d'Image	25
4.1.2.1 Analyse de l'Upscaling (Détails et Textures)	25
4.1.2.2 Validation de la Fidélité Colorimétrique (Soft Proofing)	27

Sommaire

4.1.3	Évaluation de la Conformité Pré-presse	28
4.1.4	Analyse des Performances	28
4.2	Architecture de l'Interface Graphique Partie Agent et scrapper	29
4.2.1	Choix architecturaux de l'interface	29
4.2.2	Structure globale de l'interface	29
4.2.2.1	Zone de configuration du scraping LinkedIn	30
4.2.2.2	Module de filtrage avancé des profils	31
4.2.2.3	Interface de l'agent IA (requêtes en langage naturel)	31
4.2.2.4	Module d'export des résultats	32
4.2.2.5	Interface de campagne email	32
4.2.3	Module d'affichage des logs	32
5	Analyse des Défis Transverses et Perspectives	35
5.1	Analyse des défis liés au traitement d'image et à l'impression	35
5.1.1	Le défi de l'équilibre entre "Hallucination" et "Netteté"	35
5.1.2	La gestion critique de la mémoire RAM	36
5.2	Perspectives d'évolution pour la partie Image	36
5.3	Perspectives pour la partie ai scrapping	37
5.3.1	Amélioration du raisonnement sémantique	37
5.3.2	Extension vers une approche multi-plateformes	37
5.3.3	Automatisation adaptative sous supervision humaine	38
5.3.4	Analyse avancée, auditabilité et conformité	38
5.3.5	Vers une industrialisation du workflow	38
6	Conclusion Générale	40

Liste des Figures

2.1	Exemple d'une image avant et apres vectorisation	9
4.1	Vue générale de l'application montrant les métadonnées de l'image téléchargée.	25
4.2	Comparaison de la netteté sur une image après passage dans la premiere etape de pipeline.	26
4.3	Exemple de check de DPI pour savoir si besoin d'une autre etape de scaling ou non	26
4.4	Choix du profile ICC avant d'appliquer le soft proofing	27
4.5	Exemple d'affichage du résultat après le soft proofing. Si le résultat est satisfaisant, on passe à la vraie conversion CMYK.	27
4.6	Possibilité de exportation de l'image sous forme de TIFF ou pdf/X-1a	28
4.7	Interface ai workflow for scrapping	30
4.8	Exemple d'extraction d'un profile du LInkedIn avec keyword.	33
4.9	Exemple d'extraction du database avec agent.	34

1

Introduction Générale et Contexte

Sommaire

2.1 Étude et sélection de la source de génération : Le choix d'Ideogram	6
2.1.1 Étude comparative des solutions du marché	6
2.1.2 Justification du choix d'Ideogram.ai	7
2.2 L'approche par Vectorisation : Analyse d'un échec technique	7
2.2.1 Concept et objectifs de la vectorisation	7
2.2.2 Tests réalisés et outils utilisés	8
2.2.3 Analyse de l'échec : les limites du "Raster-to-Vector"	8
2.3 Conception du Pipeline de Traitement d'Image (Solution Retenue)	9
2.3.1 Analyse Intelligente des Métadonnées	9
2.3.2 Nettoyage et Post-traitement (Pre-processing)	10
2.3.3 Double Pipeline d'Upscaling (Hybride)	10
2.3.4 Vérification et Calcul de Résolution (DPI Check)	10
2.3.5 Préparation Pré-presse et Gestion des Couleurs	11
2.3.5.1 Simulation de rendu (Soft Proofing)	11
2.3.5.2 Conversion CMJN Mémoire-Optimisée	11
2.3.5.3 Exportation au Standard Industriel PDF/X-1a	11

1.1 Contexte du projet

L'émergence de l'Intelligence Artificielle (IA) générative marque un tournant historique dans l'industrie du marketing digital. En quelques années, nous sommes passés d'outils de simple assistance à des systèmes capables de créer du contenu

1. Introduction Générale et Contexte

visuel complexe (via des modèles comme Midjourney ou Stable Diffusion) et de raisonner de manière semi-autonome pour accomplir des tâches complexes (via les Large Language Models).

Cependant, l'adoption de ces technologies en milieu professionnel se heurte encore à deux obstacles majeurs que ce projet vise à adresser :

- **Premièrement, la barrière de la qualité technique.** Si les IA actuelles excellent dans la création de concepts visuels rapides, les fichiers produits sont généralement limités par une faible résolution et des profils colorimétriques inadaptés aux exigences de l'industrie de l'impression. Pour un marketeur, transformer une idée générée sur un écran en un billboard de plusieurs mètres ou en un flyer de haute qualité nécessite des compétences techniques en post-traitement qui freinent souvent la fluidité du flux de travail créatif.
- **Deuxièmement, le défi de la distribution ciblée.** Créer un contenu de qualité n'est que la moitié du chemin ; encore faut-il l'adresser à la bonne audience. Le réseau social LinkedIn est devenu le terrain de prospection privilégié pour le marketing B2B, mais la recherche manuelle de clients potentiels est une tâche chronophage et souvent imprécise. L'avènement des "agents IA" permet désormais d'envisager une prospection automatisée mais intelligente, capable d'analyser le contexte professionnel des profils avec une finesse proche de l'analyse humaine.

Ce projet s'inscrit donc dans une volonté de professionnaliser l'usage de l'IA générative. Il ne s'agit pas seulement de générer du contenu, mais de construire un écosystème complet où l'image est techniquement parfaite pour le monde physique et où le ciblage client est automatisé par des agents intelligents. À travers le développement de deux interfaces distinctes, nous proposons une solution concrète pour intégrer l'IA dans l'ensemble de la chaîne de valeur marketing : de la conception visuelle à la rencontre du client.

1.2 Problématique

L'intégration de l'IA générative dans les processus marketing soulève des défis structurels qui empêchent souvent son exploitation à un niveau industriel. Bien que la technologie permette de générer des résultats impressionnantes en quelques secondes, le passage à une application concrète et professionnelle révèle deux problématiques majeures :

1. Introduction Générale et Contexte

1.2.1 L'incompatibilité des sorties IA avec les standards de l'impression

Les modèles de génération d'images (Text-to-Image) produisent nativement des fichiers optimisés pour le web : basse résolution (souvent 72 ou 96 DPI) et espace colorimétrique RGB. Or, le marketing physique (billboards, flyers, affiches) exige une haute densité de pixels (300 DPI) et un profil CMJN (Cyan, Magenta, Jaune, Noir).

- **Le problème :** Comment augmenter la résolution d'une image générée par IA sans perdre en fidélité, sans créer de flou et sans dénaturer les détails originaux ?
- **L'enjeu :** Créer un pont technique automatisé capable de rendre une image "numérique" prête pour la "production physique" de grande échelle.

1.2.2 Le manque de précision dans la prospection automatisée

La prospection sur LinkedIn souffre d'un paradoxe : les outils d'automatisation classiques (scripts de scraping) sont souvent trop rigides et "aveugles", tandis que la recherche manuelle est trop lente.

- **Le problème :** Comment identifier des profils de clients cibles avec une précision sémantique (comprendre réellement le poste, le secteur et l'intérêt du prospect) sans y passer des heures manuellement ?
- **L'enjeu :** Dépasser le simple filtrage par mots-clés pour mettre en place un système capable de "raisonner" sur les données LinkedIn afin de ne retenir que les profils à haute valeur ajoutée.

1.2.3 La complexité d'usage pour les équipes marketing

Enfin, la plupart des outils d'IA avancés nécessitent des connaissances en programmation ou en gestion de serveurs (Python, API, environnements Docker).

- **Le problème :** Comment rendre ces technologies de pointe accessibles à des professionnels du marketing qui ne sont pas des développeurs ?
- **L'enjeu :** Concevoir des interfaces graphiques (GUI) intuitives qui encapsulent des pipelines complexes (Computer Vision et Agents IA) en outils "prêts à l'emploi".

1. Introduction Générale et Contexte

En résumé, la question centrale de ce projet est la suivante : Comment transformer les capacités brutes de l'IA générative en une solution logicielle intégrée, capable de produire des supports visuels de haute qualité et de cibler intelligemment des prospects, tout en restant accessible aux utilisateurs finaux ?

1.3 Objectifs du projet

Les objectifs spécifiques se déclinent en deux axes majeurs :

1.3.1 Axe Technique : Optimisation de la Qualité Visuelle

Le premier objectif vise à construire un environnement capable de transformer une génération d'IA brute en un produit professionnel.

- **Développement d'un pipeline de Super-Résolution** : Implémenter des algorithmes de Deep Learning pour multiplier la densité de pixels sans perte de détails, permettant des impressions sur supports de grande taille.
- **Automatisation de la mise en conformité print** : Créer des fonctions de post-traitement pour ajuster les niveaux de bruit, la netteté et la conversion des espaces colorimétriques nécessaires aux imprimeurs.
- **Accessibilité via une interface dédiée** : Développer une interface graphique (GUI) permettant de charger une image et de la traiter en un clic, masquant la complexité du code sous-jacent.

1.3.2 Axe Stratégique : Automatisation de la Prospection Intelligente

Le second objectif porte sur la création d'un système d'acquisition de clients haute fidélité.

- **Conception d'une architecture agentique** : Mettre en place des agents IA capables de naviguer et d'interpréter les données de LinkedIn de manière autonome.
- **Interface de management des leads** : Fournir un tableau de bord permettant de paramétriser les critères de recherche et de visualiser les résultats extraits de manière structurée.

2

Pipeline de Traitement d'Image pour l'Impression Grand Format

Sommaire

3.1	Introduction du Chapitre	13
3.2	Concepts Théoriques : Agents Autonomes vs AI Work-flows	14
3.2.1	Définition des agents autonomes	14
3.2.2	Scripts classiques vs agents autonomes	15
3.2.3	Contraintes du scraping et limites de l'autonomie	15
3.2.4	Justification du choix : AI Workflow scripté	16
3.3	Architecture Technique et Rôle du MCP	17
3.3.1	Architecture générale de l'AI Workflow	17
3.3.2	Rôle du Model Context Protocol (MCP)	17
3.3.3	Séparation logique : outils, raisonnement et orchestration	18
3.4	Module d'Extraction Contrôlée (LinkedIn)	18
3.4.1	Extraction depuis la page de résultats	18
3.4.2	Données extraites	19
3.4.3	Robustesse de l'extraction et simulation humaine	19
3.5	Module d'Enrichissement Intelligent (LLM)	19
3.5.1	Détection du genre (Vision/Nom)	19
3.5.2	Extraction et analyse du parcours éducatif	19
3.5.3	Estimation de l'âge par raisonnement LLM	20
3.6	Agent Autonome de Raisonnement et Filtrage (DB Agent)	20
3.6.1	Objectif de l'agent et filtrage sémantique	20
3.6.2	Architecture "Tool-Augmented Agent"	20
3.6.3	Boucle de raisonnement et contrôle de l'autonomie	21
3.6.4	Export et traçabilité des résultats	21

2. Pipeline de Traitement d'Image pour l'Impression Grand Format

3.7 Orchestration Globale du Workflow	21
3.7.1 Séquence globale d'exécution	21
3.7.2 Gestion des limites et sécurité technique	22
3.8 Extension : Agent Autonome de Diffusion par Email	22
3.8.1 Rôle et objectifs de l'agent email	22
3.8.2 Caractère autonome et intelligence contextuelle	22
3.8.3 Architecture logique et sécurité	23
3.9 Conclusion du Chapitre	23

2.1 Étude et sélection de la source de génération : Le choix d'Ideogram

La première étape de notre workflow consiste à générer le contenu visuel brut qui servira de base à nos supports marketing. Bien que le marché de l'IA générative soit aujourd'hui saturé d'outils performants, le choix du moteur de génération est critique, car la qualité de la sortie initiale conditionne l'efficacité de tout le pipeline de traitement ultérieur.

2.1.1 Étude comparative des solutions du marché

Dans le cadre de ce projet, nous avons procédé à une phase d'évaluation technique portant sur une large gamme d'outils de génération et d'édition d'images par IA. Parmi les solutions testées figurent des plateformes diversifiées telles que :

- **Solutions généralistes et artistiques** : OpenArt AI, Imagine.art, CG-Dream.ai, Createimg.com, Nano Banana.
- **Outils orientés design et branding** : Recraft.ai, Kittl.com, Dzine.ai, Zoviz.com.
- **Outils d'amélioration et d'édition** : Fotor.com, Let's Enhance.io, GoEnhance.ai, Glima.ai.

Cette phase de test a permis d'évaluer plusieurs critères essentiels : la fidélité au prompt (prompt adherence), la qualité des textures, la gestion de la composition et, surtout, la capacité à intégrer du texte cohérent au sein des images.

2. Pipeline de Traitement d'Image pour l'Impression Grand Format

2.1.2 Justification du choix d'Ideogram.ai

À l'issue de ces tests, Ideogram.ai a été retenu comme la solution principale pour notre pipeline. Ce choix se justifie par trois piliers fondamentaux :

1. **Typographie et rendu textuel** : Contrairement à la majorité des modèles qui produisent souvent des textes illisibles ou des "hallucinations" de caractères, Ideogram se distingue par sa précision exceptionnelle dans la génération de titres, de slogans et de logos intégrés. Pour la création de flyers et de supports marketing, cette capacité est indispensable.
2. **Cohérence stylistique** : Le modèle offre une maîtrise poussée des styles graphiques (flat design, 3D render, cinématique), permettant d'obtenir des visuels qui respectent l'identité visuelle d'une marque sans nécessiter de retouches complexes.
3. **Base exploitable pour le post-traitement** : Bien que la résolution native soit limitée pour l'impression, la propreté des lignes et la gestion des contrastes par Ideogram constituent une excellente base pour notre algorithme d'upsampling, minimisant les artefacts lors de l agrandissement.

2.2 L'approche par Vectorisation : Analyse d'un échec technique

Avant de concevoir notre pipeline de traitement d'image raster, nous avons exploré une voie qui semblait idéale sur le plan théorique pour l'impression grand format : la vectorisation (ou Image Tracing).

2.2.1 Concept et objectifs de la vectorisation

La vectorisation consiste à transformer une image matricielle (composée de pixels) en une image vectorielle (composée de formules mathématiques et de tracés). L'avantage majeur de ce format est sa capacité de mise à l'échelle infinie : une image vectorielle peut être agrandie pour couvrir un billboard de 10 mètres sans jamais perdre en netteté ni pixeliser.

2. Pipeline de Traitement d'Image pour l'Impression Grand Format

2.2.2 Tests réalisés et outils utilisés

Nous avons soumis les visuels générés par Ideogram à plusieurs outils de vectorisation, allant de solutions grand public aux moteurs open source reconnus :

- **Outils en ligne** : Vectorizer.ai et autres convertisseurs automatisés.
- **Outils Open Source et professionnels** : Utilisation de l'algorithme Potrace et de la fonction de vectorisation dynamique d'Adobe Illustrator.

2.2.3 Analyse de l'échec : les limites du "Raster-to-Vector"

Malgré les avantages théoriques, les tests ont révélé une dégradation inacceptable de la qualité visuelle. Les images produites par l'IA générative possèdent des caractéristiques que les algorithmes de vectorisation actuels ne parviennent pas à interpréter fidèlement :

1. **Destruction du rendu photoréaliste** : Les visuels d'IA contiennent des dégradés de couleurs extrêmement subtils, des jeux de lumière (lighting) et des ombres complexes. La vectorisation tente de regrouper ces nuances en aplats de couleurs solides, ce qui crée un effet "cartoon" ou "poster" non désiré.
2. **Perte de texture et de détails fins** : Les textures de peau, de tissu ou les reflets métalliques disparaissent totalement, remplacés par des formes géométriques simplifiées.
3. **Altération des typographies** : Bien qu'Ideogram génère des titres nets, la vectorisation arrondit souvent les angles des polices de caractères ou fusionne certaines lettres, rendant le texte marketing moins professionnel.
4. **Complexité du fichier final** : Pour tenter de garder un minimum de détail, les outils génèrent des millions de points d'ancrage, rendant le fichier vectoriel extrêmement lourd et difficile à manipuler par les logiciels de pré-presse, sans pour autant égaler la qualité de l'original.

2. Pipeline de Traitement d'Image pour l'Impression Grand Format

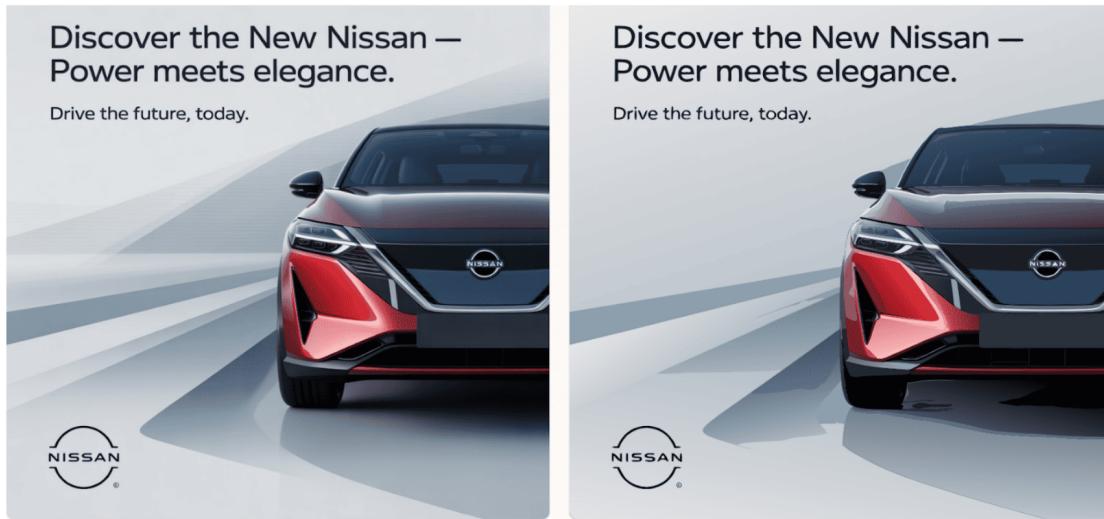


Figure 2.1: Exemple d'une image avant et après vectorisation

Cette phase de test a permis de conclure que la vectorisation est un outil performant pour des logos simples ou des icônes, mais qu'elle est totalement inadaptée aux images marketing riches produites par l'IA. C'est ce constat d'échec qui a imposé le changement de paradigme vers la création d'un pipeline de traitement "Raster Haute Fidélité", basé sur l'upscale intelligent et la gestion rigoureuse des pixels.

2.3 Conception du Pipeline de Traitement d'Image (Solution Retenue)

Suite à l'analyse des limites de la vectorisation, nous avons développé un pipeline de traitement d'image matriciel (Raster) haute fidélité. L'objectif est de manipuler la structure profonde de l'image pour l'adapter aux exigences de l'impression professionnelle tout en préservant l'intégrité esthétique de la génération initiale.

2.3.1 Analyse Intelligente des Métadonnées

La première étape utilise la bibliothèque Pillow (PIL) et ImageCms pour établir un diagnostic technique complet de l'image source.

- Technologies : PIL.Image, PIL.ImageCms.
- Fonctionnalités : Extraction automatique du mode colorimétrique, de la profondeur de bits par canal et détection des profils ICC existants.
- Importance : Cette analyse permet de détecter si l'image possède déjà un profil de couleur et de préparer les calculs de conversion ultérieurs.

2. Pipeline de Traitement d'Image pour l'Impression Grand Format

2.3.2 Nettoyage et Post-traitement (Pre-processing)

Avant l agrandissement, l image subit une phase de purification pour éviter d amplifier les défauts de génération.

- Technologies : OpenCV (cv2), NumPy.
- Algorithmes :
 - Denoising : Utilisation de fastNlMeansDenoisingColored pour supprimer le bruit numérique tout en préservant les textures.
 - Sharpening : Application d'un noyau de convolution (filter2D) pour accentuer les détails et la netteté avant l'étape de Super-Résolution.

2.3.3 Double Pipeline d'Upscaling (Hybride)

Pour garantir une qualité optimale selon le support, le projet implémente deux types d agrandissement :

1. **Upscaling Neural (Real-ESRGAN)** : Via l API gradio_client, nous utilisons le modèle realesr-general-x4v3. Cette étape est cruciale pour "réinventer" les détails manquants sur les textures complexes.
2. **Upscaling Géométrique (Lanczos)** : Pour les redimensionnements finaux vers des dimensions cibles précises, nous utilisons l interpolation de Lanczos (Image.LANCZOS). Cette méthode est reconnue pour son excellent compromis entre lissage et conservation des contours nets.

2.3.4 Vérification et Calcul de Résolution (DPI Check)

Le système intègre une logique métier permettant de valider la qualité de l image en fonction du support final (flyer, poster ou billboard).

- **Logique** : Le script calcule le DPI réel en fonction des dimensions physiques (mètres) et du nombre de pixels.
- **Aide à la décision** : Il compare le DPI obtenu aux standards de l industrie (ex: 300 DPI pour un flyer à 0.6m vs 4 DPI pour un billboard à 50m) et suggère automatiquement le facteur d upscaling nécessaire.

2. Pipeline de Traitement d'Image pour l'Impression Grand Format

2.3.5 Préparation Pré-presse et Gestion des Couleurs

Une fois l'image agrandie, elle doit être convertie pour le monde physique (impression). Cette phase est la plus critique pour la fidélité des couleurs.

2.3.5.1 Simulation de rendu (Soft Proofing)

Avant la conversion définitive, le pipeline génère une prévisualisation.

- Processus : Simulation d'un cycle complet RGB → CMYK → RGB pour montrer à l'utilisateur, sur son écran, les couleurs qui risquent de devenir ternes à l'impression.
- Optimisation : Traitement allégé pour une génération rapide de la preview.

2.3.5.2 Conversion CMJN Mémoire-Optimisée

Le passage du profil sRGB au profil de presse (ex: US Web Coated SWOP) est une opération lourde.

- Innovation technique : Pour gérer des images de très grande taille (billboards) sans saturer la RAM, nous avons implémenté un traitement par blocs (tiles) de 2048px.
- Technologie : ImageCms.buildTransform pour une conversion rigoureuse respectant les standards ICC. Le fichier est ensuite sauvegardé en TIFF 16 bits avec compression Deflate pour conserver une qualité maximale.

2.3.5.3 Exportation au Standard Industriel PDF/X-1a

L'étape finale garantit que le fichier sera accepté par n'importe quel imprimeur professionnel dans le monde.

- Technologies : img2pdf pour l'encapsulation et pikepdf pour la mise en conformité.
- Conformité PDF/X-1a :
 - Forçage de la version PDF 1.3.
 - Intégration du profil ICC dans le OutputIntent.
 - Ajout des métadonnées obligatoires (GTS_PDFXVersion).

3

Automatisation du Sourcing Clients via AI Workflow

Sommaire

4.1 Conception et Architecture de l'Interface Graphique (Partie Image)	24
4.1.1 Import et affichage des métadonnées	24
4.1.2 Résultats du Pipeline de Traitement d'Image	25
4.1.2.1 Analyse de l'Upscaling (Détails et Textures) .	25
4.1.2.2 Validation de la Fidélité Colorimétrique (Soft Proofing)	27
4.1.3 Évaluation de la Conformité Pré-presse	28
4.1.4 Analyse des Performances	28
4.2 Architecture de l'Interface Graphique Partie Agent et scrapper	29
4.2.1 Choix architecturaux de l'interface	29
4.2.2 Structure globale de l'interface	29
4.2.2.1 Zone de configuration du scraping LinkedIn .	30
4.2.2.2 Module de filtrage avancé des profils	31
4.2.2.3 Interface de l'agent IA (requêtes en langage naturel)	31
4.2.2.4 Module d'export des résultats	32
4.2.2.5 Interface de campagne email	32
4.2.3 Module d'affichage des logs	32

3. Automatisation du Sourcing Clients via AI Workflow

3.1 Introduction du Chapitre

Dans un contexte de transformation numérique accélérée, la prospection commerciale et le sourcing de clients ont connu une évolution majeure, passant de méthodes traditionnelles à des approches de plus en plus automatisées et assistées par l'intelligence artificielle. Les plateformes professionnelles en ligne, telles que LinkedIn, sont devenues des sources privilégiées pour identifier des profils pertinents, qu'il s'agisse de clients potentiels, de partenaires ou de talents. Elles offrent un volume considérable d'informations structurées et semi-structurées, permettant d'affiner les stratégies de prospection digitale.

Cependant, les approches classiques de sourcing reposent encore largement sur des recherches manuelles ou sur des scripts automatisés rigides. La recherche manuelle, bien que précise, est chronophage, peu scalable et fortement dépendante de l'intervention humaine. À l'inverse, les scripts automatisés traditionnels, basés sur des règles fixes, manquent de flexibilité et de capacité d'adaptation au contexte. Ils sont souvent incapables de comprendre la sémantique des profils, d'évaluer leur pertinence réelle ou de gérer les cas ambigus, ce qui limite considérablement leur efficacité dans des environnements complexes et dynamiques.

Par ailleurs, l'utilisation de plateformes professionnelles comme LinkedIn est soumise à des contraintes techniques, légales et éthiques strictes. Les politiques anti-scraping, les limitations d'accès, ainsi que les risques de suspension de compte rendent difficile la mise en place de systèmes totalement autonomes et agressifs. Ces contraintes imposent une approche contrôlée, responsable et respectueuse des conditions d'utilisation, excluant de facto l'usage d'agents entièrement autonomes capables d'agir librement sans supervision.

Face à ces défis, ce chapitre propose une approche alternative basée sur un **AI Workflow orchestré**, combinant des outils techniques contrôlés, un protocole de communication structuré (Model Context Protocol – MCP) et le raisonnement avancé des modèles de langage (LLM). L'objectif est de concevoir un système capable d'automatiser intelligemment le sourcing de profils, tout en maintenant un haut niveau de contrôle, de sécurité et de conformité. Cette approche vise à tirer parti de l'intelligence contextuelle des LLM pour l'enrichissement et le filtrage des données, tout en respectant les contraintes techniques et éthiques imposées par les plateformes professionnelles.

3. Automatisation du Sourcing Clients via AI Workflow

3.2 Concepts Théoriques : Agents Autonomes vs AI Workflows

3.2.1 Définition des agents autonomes

Notion d'agent autonome en intelligence artificielle

En intelligence artificielle, un **agent autonome** est défini comme une entité logicielle capable de percevoir son environnement, de prendre des décisions de manière indépendante et d'agir sur cet environnement afin d'atteindre un objectif donné. Contrairement aux programmes traditionnels, l'agent autonome ne se limite pas à l'exécution d'instructions prédéfinies, mais adapte son comportement en fonction des informations perçues et de l'évolution du contexte.

Selon la littérature classique (Russell & Norvig), un agent est caractérisé par sa capacité à maximiser une mesure de performance à travers une interaction continue avec son environnement. Cette interaction repose sur une **boucle fermée perception-décision-action**, constituant le cœur du comportement autonome.

Capacités principales d'un agent autonome

Un agent autonome repose généralement sur quatre capacités fondamentales :

Perception : L'agent collecte des informations à partir de son environnement via des capteurs logiciels ou matériels (données, événements, signaux). Ces informations constituent l'état observé du système.

Décision : À partir des perceptions, l'agent raisonne afin de choisir l'action la plus appropriée. Cette prise de décision peut reposer sur des règles, des modèles probabilistes, des techniques d'apprentissage automatique ou des modèles de langage avancés.

Action : L'agent agit sur l'environnement à travers des effecteurs, modifiant ainsi l'état du système ou déclenchant des événements externes.

Boucle d'auto-adaptation : Les agents autonomes disposent d'une capacité d'adaptation continue. Les résultats de leurs actions influencent leurs décisions futures, permettant un apprentissage ou un ajustement progressif du comportement.

3. Automatisation du Sourcing Clients via AI Workflow

3.2.2 Scripts classiques vs agents autonomes

Scripts procéduraux traditionnels

Les scripts procéduraux reposent sur une logique déterministe et séquentielle. Ils exécutent une suite d'instructions prédéfinies selon un ordre fixe, indépendamment du contexte dynamique. Leurs principales caractéristiques sont :

- Une logique déterministe, où chaque entrée produit toujours la même sortie.
- Une séquence fixe d'actions, définie à l'avance par le développeur.
- Une faible capacité d'adaptation aux situations imprévues.

Bien que simples à implémenter et faciles à contrôler, ces scripts deviennent rapidement limités lorsqu'ils doivent gérer des cas complexes, ambigus ou évolutifs.

Agents autonomes

À l'inverse, les agents autonomes s'appuient sur un raisonnement dynamique. Leurs actions ne sont pas strictement prédéfinies, mais conditionnées par l'état courant de l'environnement et les objectifs à atteindre. Ils se distinguent par :

- Une prise de décision contextuelle, basée sur l'interprétation des données.
- Une capacité à ajuster leur comportement en temps réel.
- Une meilleure gestion de l'incertitude et des cas non anticipés.

Cependant, cette autonomie accrue implique une perte de contrôle direct sur les actions exactes de l'agent, ce qui peut poser problème dans des environnements réglementés.

3.2.3 Contraintes du scraping et limites de l'autonomie

L'exploitation automatisée des données issues de plateformes professionnelles comme LinkedIn est soumise à de fortes restrictions. Le scraping y est généralement interdit ou strictement limité, conformément aux conditions d'utilisation de la plateforme. Les principaux risques associés à une automatisation excessive incluent :

- Le blocage ou la suspension définitive du compte utilisé.
- Des violations des conditions d'utilisation, pouvant entraîner des conséquences légales.
- Une dégradation de la fiabilité du système due aux mécanismes anti-automatisation.

Dans ce contexte, la mise en place d'un agent totalement autonome est impraticable. Un tel agent disposerait d'une liberté d'action incompatible avec les contraintes imposées :

3. Automatisation du Sourcing Clients via AI Workflow

- Il ne peut pas naviguer librement sur la plateforme.
 - Il ne peut pas prendre de décisions arbitraires concernant les interactions ou les requêtes.
 - Il doit opérer dans un cadre strictement contrôlé et prévisible.
- Ces limitations rendent l'autonomie complète non seulement risquée, mais également non conforme aux exigences du projet.

3.2.4 Justification du choix : AI Workflow scripté

Face aux contraintes identifiées, le choix d'un **AI Workflow scripté** s'impose comme une solution intermédiaire pertinente. Contrairement à un agent autonome classique, ce type de système repose sur une orchestration contrôlée des actions, limitant volontairement l'autonomie.

Un agent autonome n'est pas adapté dans ce contexte car :

- Raisonne qu'il est entrain de Scrapper ce qui est interdit et stop l'action.
- Il peut prendre des initiatives non prévues .
- Il augmente le risque de non-conformité avec les règles de la plateforme.

Le système proposé transforme ainsi l'agent en un *workflow orchestré*, dans lequel :

- Les décisions sont limitées et supervisées.
- Les actions possibles sont strictement définies.
- L'intelligence est utilisée pour raisonner, et non pour agir librement.

Dans cette architecture, le LLM joue un rôle central mais encadré :

- Raisonnement sur les requêtes utilisateur.
- Validation des filtres et des résultats.
- Enrichissement sémantique des données extraites.

Une **séparation claire des responsabilités** est alors établie :

MCP : Gestion des outils techniques et des interfaces ;

Workflow : Orchestration des étapes et contrôle du flux ;

LLM : Intelligence décisionnelle et interprétation contextuelle.

Cette approche garantit un équilibre entre puissance de raisonnement, contrôle opérationnel et conformité aux contraintes techniques et éthiques du domaine.

3. Automatisation du Sourcing Clients via AI Workflow

3.3 Architecture Technique et Rôle du MCP

3.3.1 Architecture générale de l'AI Workflow

Cette section présente l'architecture globale du système proposé pour l'extraction, le filtrage et l'exportation de profils professionnels à partir de données préalablement collectées. L'objectif principal de cette architecture est de concilier intelligence décisionnelle, contrôle strict des actions et conformité aux contraintes imposées par les plateformes professionnelles.

L'architecture repose sur une approche modulaire et orchestrée, intégrant un modèle de langage (LLM), un workflow contrôlé et un ensemble d'outils techniques spécialisés. Cette séparation permet d'assurer la robustesse, la maintenabilité et la traçabilité du système composée de quatre blocs principaux, chacun ayant une responsabilité clairement définie :

- **Frontend UI** : interface utilisateur permettant de définir les paramètres de recherche (poste, localisation, genre, tranche d'âge, nombre de profils à extraire).
- **API Router & Orchestrator** : cœur logique du système, implémenté en Python. Il orchestre la séquence des opérations, contrôle les flux et garantit le respect des limites imposées.
- **MCP Server (Playwright)** : serveur local dédié exclusivement à l'interaction avec le navigateur. Il encapsule toute la logique de navigation et d'extraction.
- **AI Services (LLM via API)** : module de raisonnement basé sur un modèle de langage (Claude), utilisé pour l'enrichissement sémantique et les inférences complexes.

3.3.2 Rôle du Model Context Protocol (MCP)

Le **Model Context Protocol (MCP)** joue un rôle central en tant que couche d'abstraction standardisée entre les outils techniques et la logique métier. Plutôt que d'intégrer directement le code de scraping (Playwright) dans l'orchestrateur principal, celui-ci est isolé dans un serveur MCP dédié (ex: `linkedin_server.py`).

Cette approche présente plusieurs avantages majeurs :

- **Sécurité accrue** : le LLM et l'orchestrateur ne peuvent invoquer que les outils explicitement exposés (ex: `login_linkedin`, `extract_education_data`).

3. Automatisation du Sourcing Clients via AI Workflow

- **Prévention des actions non autorisées** : aucune navigation libre ou action arbitraire n'est possible.
- **Maintenabilité** : les évolutions du DOM LinkedIn n'impactent que le serveur MCP.

3.3.3 Séparation logique : outils, raisonnement et orchestration

L'architecture respecte strictement le principe de **séparation des préoccupations** :

- **Outils (MCP)** : exécutent des actions atomiques et déterministes telles que :
“Ouvre cette URL”, “Extrais le texte de cette balise HTML”.
Ces outils sont volontairement simples, robustes et sans raisonnement.
- **Orchestration (Service)** : définit la logique séquentielle :
“Lance la recherche”, “Si l'extraction réussit, alors enrichis les données”, “Sauvegarde en base”.
- **Raisonnement (LLM)** : intervient uniquement pour des tâches cognitives :
“Ce profil correspond-il réellement à un Data Scientist ?”, “Quel âge approximatif peut-on déduire de ce parcours éducatif ?”

3.4 Module d'Extraction Contrôlée (LinkedIn)

3.4.1 Extraction depuis la page de résultats

L'extraction est réalisée directement depuis la page de résultats de recherche (SERP). Cette approche permet de :

- Minimiser le nombre de requêtes HTTP ;
- Éviter la consultation massive de profils individuels ;
- Réduire les risques de détection et de limitation par LinkedIn.

3. Automatisation du Sourcing Clients via AI Workflow

3.4.2 Données extraites

Les données brutes récupérées incluent :

- Nom complet ;
- URL publique du profil ;
- Localisation ;
- Titre professionnel (headline) ;
- URL de l'image de profil.

3.4.3 Robustesse de l'extraction et simulation humaine

La robustesse est assurée par plusieurs stratégies implémentées dans la fonction `extract_all_search_profiles_with_images` :

- Simulation humaine via un scroll progressif pour déclencher le *lazy loading*.
- Utilisation de sélecteurs CSS multiples pour chaque champ (nom, image, titre).
- Gestion asynchrone avec Playwright afin d'améliorer la stabilité.

3.5 Module d'Enrichissement Intelligent (LLM)

3.5.1 Détection du genre (Vision/Nom)

LinkedIn ne fournit pas explicitement le genre. Celui-ci est inféré via le modèle Claude à l'aide de la fonction `detect_gender`. Le prompt utilisé combine le prénom et l'URL de l'image :

“The person’s name is '{name}'. Based on the image url and the name, detect the gender.”

3.5.2 Extraction et analyse du parcours éducatif

Lorsque l'option est activée, le workflow visite le profil individuel afin d'extraire la section *Education* via l'outil MCP `extract_education_data`. Seules les informations académiques pertinentes sont conservées (écoles, diplômes, années).

3. Automatisation du Sourcing Clients via AI Workflow

3.5.3 Estimation de l'âge par raisonnement LLM

L'âge est estimé par inférence logique grâce à des heuristiques académiques :

- Baccalauréat \approx 18 ans ;
- Master \approx 23–24 ans ;
- Doctorat \approx 27–28 ans.

Le LLM calcule l'âge actuel en combinant l'année du dernier diplôme significatif avec l'année courante.

3.6 Agent Autonome de Raisonnement et Filtrage (DB Agent)

3.6.1 Objectif de l'agent et filtrage sémantique

L'objectif principal de cet agent est de :

- Interpréter des requêtes utilisateurs exprimées en langage naturel ;
- Décider dynamiquement quels outils invoquer ;
- Appliquer des filtres complexes sur les profils enrichis ;
- Produire des résultats exploitables (réponse textuelle ou export CSV).

3.6.2 Architecture "Tool-Augmented Agent"

L'agent repose sur un modèle **Claude Sonnet** connecté à un ensemble d'outils explicitement déclarés :

- `filter_profiles_tool` : filtrage avancé multi-critères ;
- `export_csv_tool` : génération et export des résultats ;
- `get_profile_by_name_tool` : recherche ciblée d'un profil.

3. Automatisation du Sourcing Clients via AI Workflow

3.6.3 Boucle de raisonnement et contrôle de l'autonomie

L'agent suit une boucle perception–raisonnement–action :

1. **Perception** : réception de la requête utilisateur (prompt).
2. **Raisonnement** : le LLM analyse l'intention et décide s'il doit appeler un outil.
3. **Action** : exécution de l'outil sélectionné (filtrage, export, récupération).
4. **Retour de contexte** : les résultats sont réinjectés dans le contexte du LLM.

Garde-fous :

- Limitation du nombre de tours de raisonnement (max 5 itérations) ;
- Liste fermée d'outils accessibles au LLM ;
- Validation stricte des schémas d'entrée via `AdvancedFilterRequest` ;
- Journalisation complète des décisions via `LogCollector`.

3.6.4 Export et traçabilité des résultats

Lorsque l'utilisateur demande une sauvegarde, l'agent déclenche automatiquement l'outil `export_csv_tool`. Celui-ci :

- Génère un fichier CSV avec identifiant unique ;
- Normalise les données complexes (parcours éducatif) ;
- Rend le fichier téléchargeable via une URL dédiée.

3.7 Orchestration Globale du Workflow

3.7.1 Séquence globale d'exécution

Le workflow est exécuté séquentiellement via `LinkedInAutonomousAgent.run()` :

1. Connexion sécurisée via MCP ;
2. Construction et navigation vers l'URL de recherche ;
3. Extraction des N premiers profils ;
4. Boucle d'enrichissement (genre, âge) ;
5. Fermeture contrôlée du navigateur ;
6. Stockage des résultats.

3. Automatisation du Sourcing Clients via AI Workflow

3.7.2 Gestion des limites et sécurité technique

Des pauses aléatoires (`asyncio.sleep`) sont intégrées afin de simuler un comportement humain. Une limite stricte sur le nombre de profils extraits est imposée afin d'éviter tout scraping massif non contrôlé.

Cette orchestration garantit un compromis optimal entre efficacité, sécurité et conformité.

3.8 Extension : Agent Autonome de Diffusion par Email

3.8.1 Rôle et objectifs de l'agent email

L'agent email a pour objectif de :

- Lire automatiquement une liste de contacts depuis un fichier CSV ;
- Générer des messages personnalisés à partir de modèles textuels ;
- Envoyer des emails de manière autonome via un serveur SMTP sécurisé ;
- Gérer les succès et les échecs sans intervention humaine.

3.8.2 Caractère autonome et intelligence contextuelle

Cet agent satisfait pleinement les critères définissant un agent autonome en intelligence artificielle :

- **Perception** : lecture et interprétation des données depuis un fichier CSV (emails et métadonnées associées).
- **Décision** : choix conditionnel des actions (envoi, attachement d'image, gestion des erreurs).
- **Action** : envoi effectif des emails via le protocole SMTP.
- **Boucle d'auto-exécution** : traitement itératif de chaque contact jusqu'à épuisement de la liste.

3. Automatisation du Sourcing Clients via AI Workflow

3.8.3 Architecture logique et sécurité

L'agent est structuré autour de quatre fonctions principales :

- `read_emails_csv` : perception des données d'entrée et validation de leur structure ;
- `send_email_with_image` : action atomique d'envoi d'un email personnalisé ;
- `process_email_campaign` : orchestration autonome de la campagne complète ;
- **Configuration dynamique** via variables d'environnement (`.env`) pour la sécurité.

Mécanismes de sécurité :

- Variables d'environnement pour identifiants SMTP ;
- Connexion TLS sécurisée (`starttls`) ;
- Gestion des erreurs pour éviter l'arrêt brutal ;
- Limitation implicite du débit par traitement séquentiel.

3.9 Conclusion du Chapitre

Ce chapitre a présenté une approche hybride et contrôlée pour l'automatisation du sourcing de clients. Le couplage d'un workflow AI orchestré, respectueux des contraintes légales et techniques, avec des agents autonomes ciblés (diffusion email, filtrage sémantique) illustre une solution pragmatique, efficace et éthique pour la prospection digitale. La séparation claire des responsabilités entre outils, raisonnement et orchestration garantit robustesse, traçabilité et conformité.

4

Résultats

Sommaire

5.1 Analyse des défis liés au traitement d'image et à l'impression	35
5.1.1 Le défi de l'équilibre entre "Hallucination" et "Netteté"	35
5.1.2 La gestion critique de la mémoire RAM	36
5.2 Perspectives d'évolution pour la partie Image	36
5.3 Perspectives pour la partie ai scrapping	37
5.3.1 Amélioration du raisonnement sémantique	37
5.3.2 Extension vers une approche multi-plateformes	37
5.3.3 Automatisation adaptative sous supervision humaine	38
5.3.4 Analyse avancée, auditabilité et conformité	38
5.3.5 Vers une industrialisation du workflow	38

4.1 Conception et Architecture de l'Interface Graphique (Partie Image)

4.1.1 Import et affichage des metadonnées

Pour rendre le pipeline de traitement accessible aux équipes marketing, nous avons développé une interface graphique intuitive. L'objectif était de masquer la complexité technique (gestion des profils ICC, API d'upscaling, traitement par blocs) derrière une expérience utilisateur simplifiée.

- Technologies utilisées : FastAPI et inja2 templates

4. Résultats

- Fonctionnalités clés :
 - Sélecteur de fichier avec prévisualisation.
 - Configuration du support final (choix entre Flyer, Poster, Billboard) pour adapter automatiquement le DPI.
 - Barre de progression dynamique pour le traitement des "tiles" (blocs) lors de la conversion CMJN.

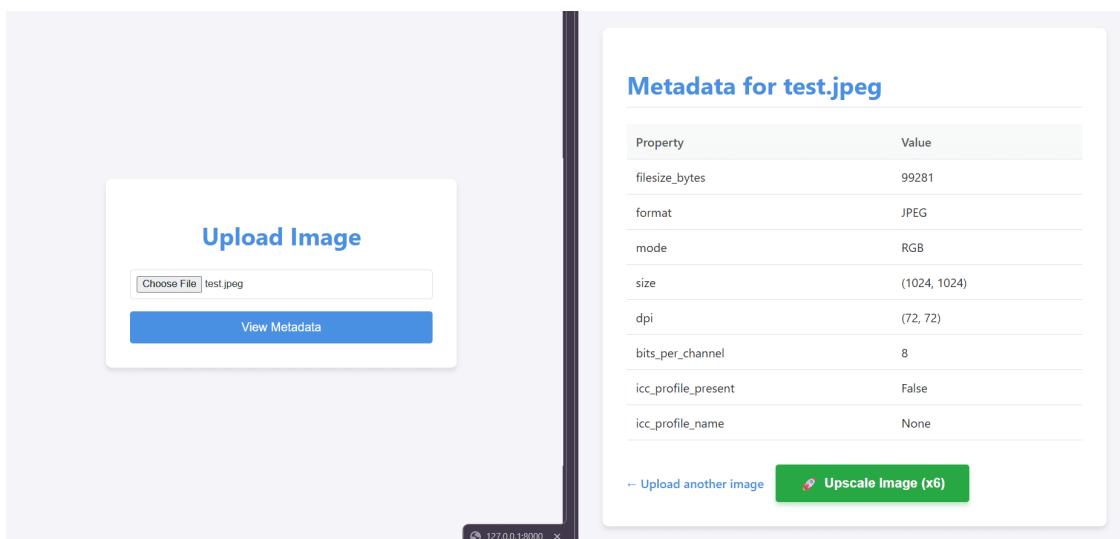


Figure 4.1: Vue générale de l'application montrant les métadonnées de l'image téléchargée.

4.1.2 Résultats du Pipeline de Traitement d'Image

Cette section présente l'efficacité de notre solution à travers des tests comparatifs effectués sur des visuels générés par Ideogram.

4.1.2.1 Analyse de l'Upscaling (Détails et Textures)

La comparaison entre l'image originale (72 DPI) et l'image traitée par Real-ESRGAN montre une récupération significative des détails fins, particulièrement sur les éléments textuels et les textures complexes.

4. Résultats

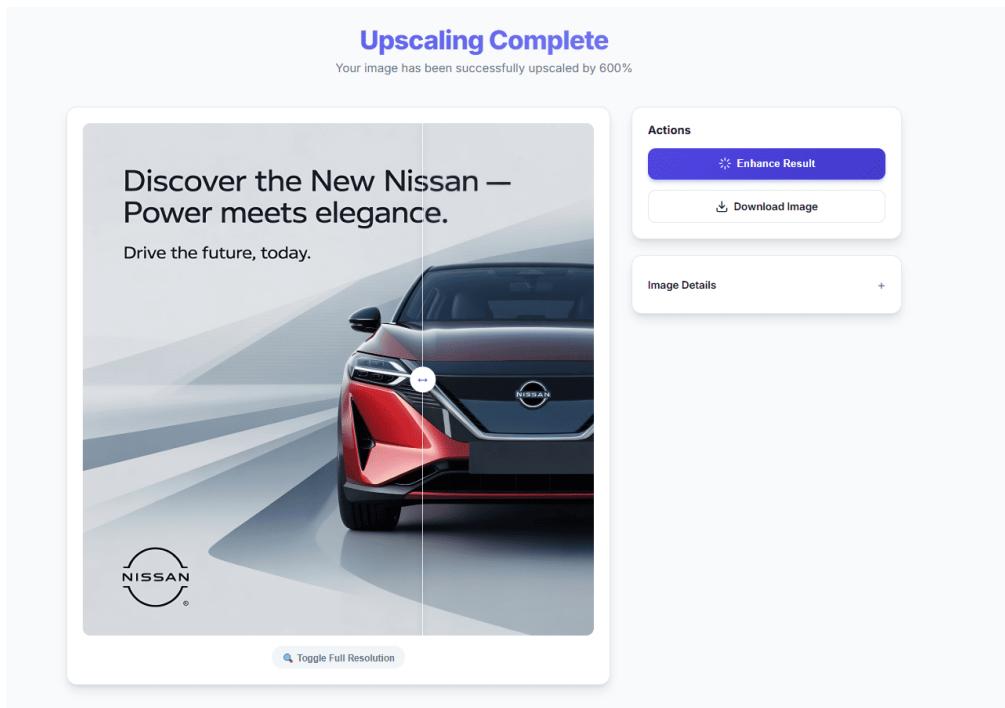


Figure 4.2: Comparaison de la netteté sur une image après passage dans la première étape de pipeline.

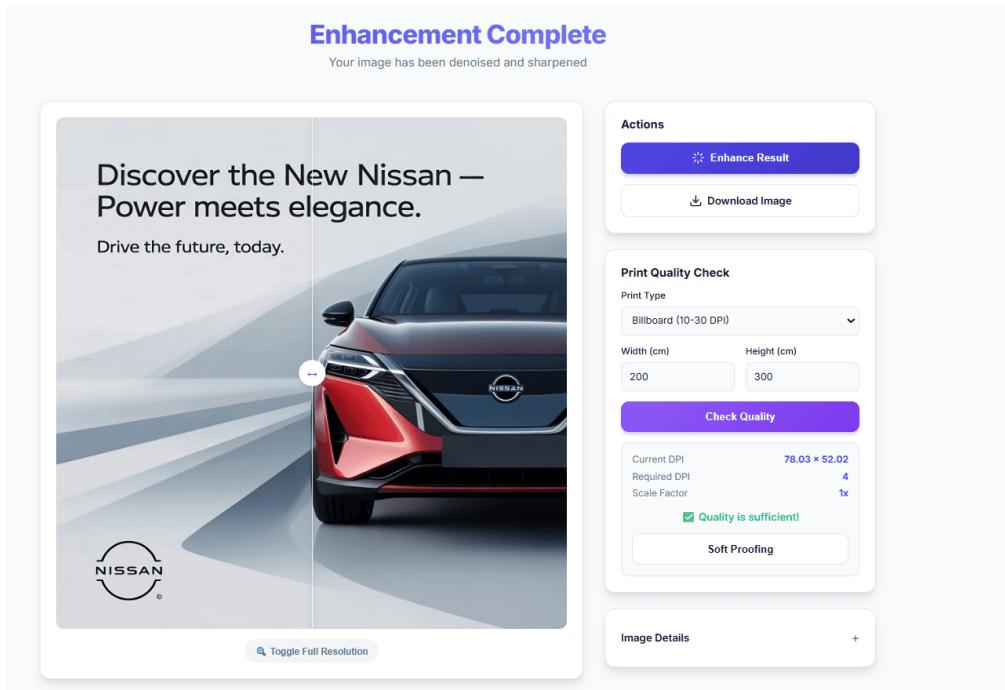


Figure 4.3: Exemple de check de DPI pour savoir si besoin d'une autre étape de scaling ou non

4. Résultats

4.1.2.2 Validation de la Fidélité Colorimétrique (Soft Proofing)

Le module de Soft Proofing permet de valider le rendu final avant l'exportation. Nous observons une transition fluide vers l'espace CMJN, avec une gestion maîtrisée de la saturation pour éviter les mauvaises surprises lors de l'impression physique.

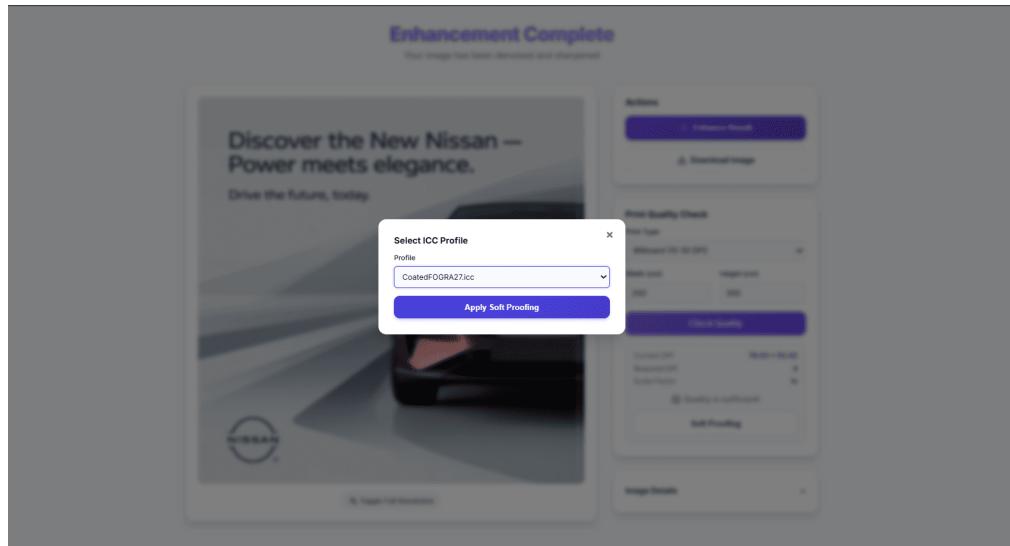


Figure 4.4: Choix du profile ICC avant d'appliquer le soft proofing

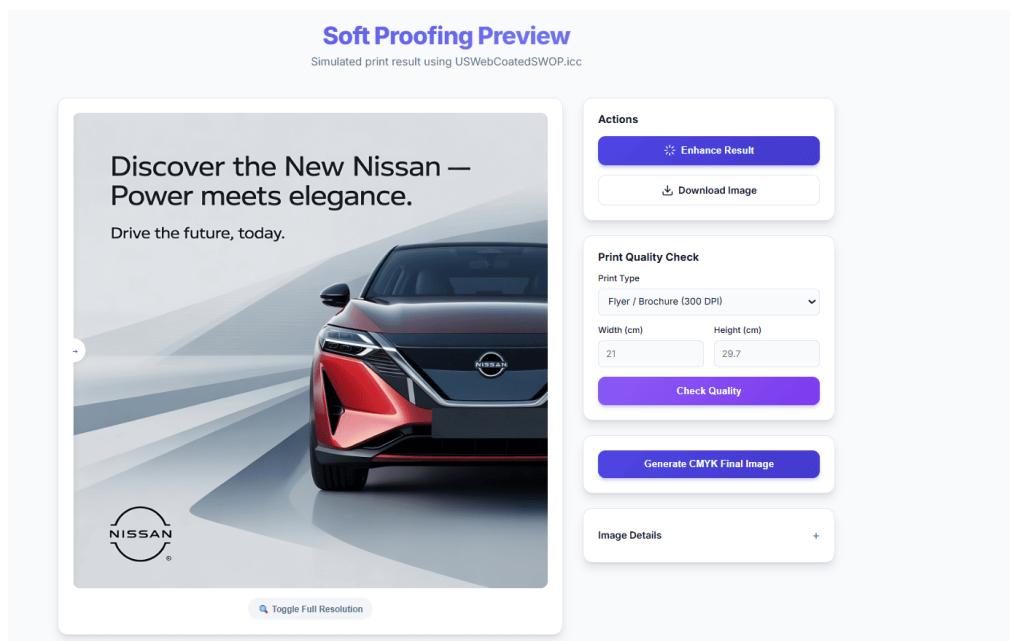


Figure 4.5: Exemple d'affichage du résultat après le soft proofing. Si le résultat est satisfaisant, on passe à la vraie conversion CMYK.

4. Résultats

4.1.3 Évaluation de la Conformité Pré-presse

L'export final a été testé pour vérifier sa compatibilité avec les standards industriels.

- Format TIFF : Validation de la profondeur 16 bits et de l'intégration correcte du profil ICC
- Format PDF/X-1a : Utilisation d'un outil de vérification (type Adobe Acrobat Preflight) pour confirmer la présence de l'OutputIntent et l'aplatissement des transparences.

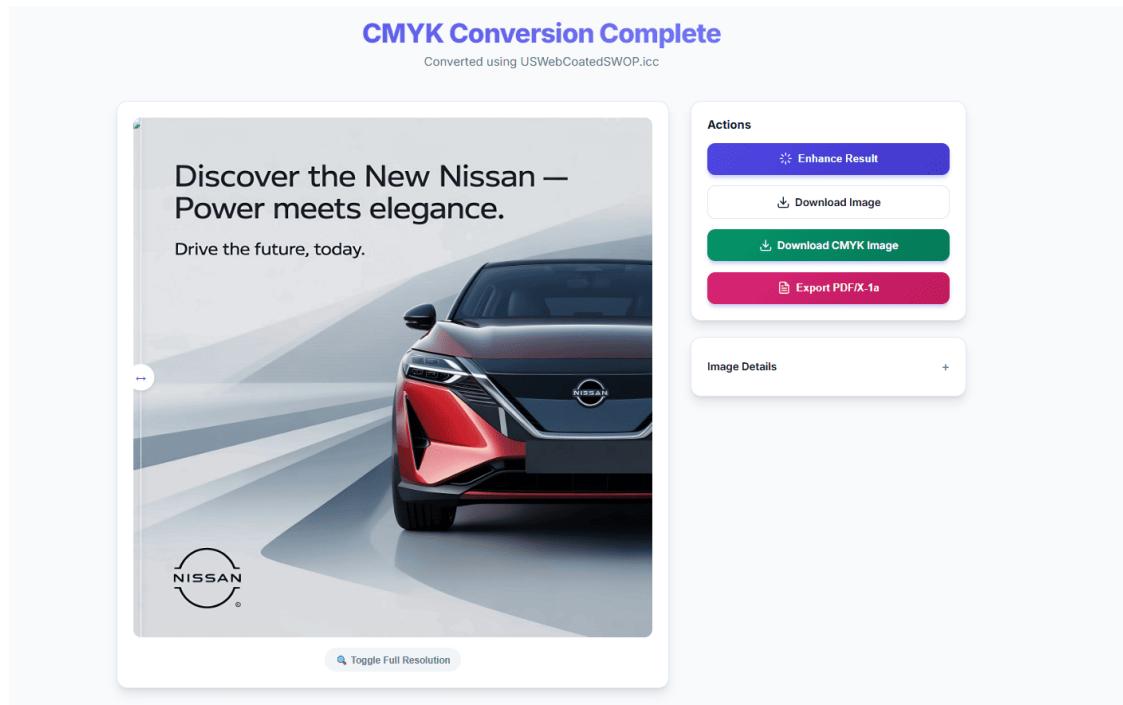


Figure 4.6: Possibilité de exportation de l'image sous forme de TIFF ou pdf/X-1a

4.1.4 Analyse des Performances

Le traitement par blocs (tiling) a permis de traiter des images de très grande taille sans dépassement de la mémoire vive (RAM). Pour une image destinée à un poster de 1 mètre, le temps moyen de traitement global (Upscaling + Conversion + Export) est de 10 minutes au maximum, ce qui reste largement inférieur à un traitement manuel sous Photoshop.

4. Résultats

4.2 Architecture de l'Interface Graphique Partie Agent et scrapper

4.2.1 Choix architecturaux de l'interface

L'interface graphique constitue le point d'entrée principal du système pour l'utilisateur final. Son rôle est de permettre une interaction simple, centralisée et contrôlée avec les différents modules du backend (scraping, filtrage, agent IA et campagnes email), sans exposer la complexité technique sous-jacente.

L'interface est conçue selon une approche Single Page Application (SPA) légère, reposant sur une page HTML unique (index.html) communiquant avec l'API FastAPI via des requêtes HTTP asynchrones.

L'architecture de l'interface repose sur les principes suivants :

- Simplicité : une seule page regroupant toutes les fonctionnalités.
- Découplage : séparation claire entre interface (frontend) et logique métier (backend).
- Contrôle : aucune action directe sur LinkedIn ou la base de données n'est possible sans passer par l'API.
- Scalabilité fonctionnelle : ajout de nouveaux modules sans refonte complète de l'UI.

4.2.2 Structure globale de l'interface

L'interface graphique du système est organisée en plusieurs blocs fonctionnels. Chaque bloc correspond à un module métier spécifique et interagit directement avec le backend via des endpoints REST exposés par l'API FastAPI.

4. Résultats

LinkedIn Profile Scraper and Digital Marketing Campaign

AI Workflow Scraper

Job Title / Keyword
e.g., Data Scientist

Location
Select a location

Number of Profiles
2

Start Scraping

Intelligent Extraction (Database)

Filter existing database profiles using natural language.

Prompt
e.g., Find me all female data scientists in Paris who are under 30 years old.

Run DB Agent

Intelligent Email Campaign

Recipient List (CSV)
Choose File No file chosen

Attachment Image
Choose File No file chosen

Subject
e.g., Special Offer

Message
Hello {name}, ...

Send Emails

Classic extraction from database :

Advanced Filtering

Keyword
e.g., Engineer

Location
Select a location

Gender
Any

Min Age
22

Max Age
60

Education
e.g., Computer Science

Apply Filters

Export Data

Download the currently filtered results as a standardized CSV file.

Download CSV

Figure 4.7: Interface ai workflow for scrapping

4.2.2.1 Zone de configuration du scraping LinkedIn

Cette section permet à l'utilisateur de configurer les paramètres de recherche pour l'extraction des profils LinkedIn. Les paramètres disponibles sont les suivants :

- Mot-clé correspondant au poste ou à la fonction ciblée
- Localisation, définie via un identifiant LinkedIn
- Nombre maximal de profils à extraire
- Activation optionnelle de l'enrichissement des données (âge, genre, niveau d'éducation)

4. Résultats

Une fois les paramètres définis, ils sont transmis au backend via l'endpoint `/scraper`, qui déclenche le processus d'extraction.

4.2.2.2 Module de filtrage avancé des profils

Après le stockage des profils dans la base de données SQLite, l'interface propose un module dédié au filtrage avancé. Ce module permet à l'utilisateur d'affiner les résultats selon plusieurs critères :

- Tranche d'âge
- Genre
- Localisation
- Niveau d'éducation
- Limite du nombre de résultats retournés

Les requêtes de filtrage sont envoyées à l'endpoint `/filter`, qui interagit avec la base de données via le service `db_service`.

4.2.2.3 Interface de l'agent IA (requêtes en langage naturel)

Cette section offre une interaction directe avec l'agent intelligent du système. L'utilisateur peut formuler des requêtes en langage naturel à l'aide d'un champ de texte libre, par exemple :

« *Donne-moi 10 profils de Data Scientists femmes à Casablanca* »

L'interface comprend :

- Un champ de saisie pour la requête
- Un bouton d'exécution
- Une zone d'affichage pour la réponse textuelle ou structurée

Les requêtes sont transmises à l'endpoint `/agent`, qui délègue le raisonnement au `AgentService` basé sur le modèle Claude.

4. Résultats

4.2.2.4 Module d'export des résultats

L'interface permet l'exportation des profils filtrés afin de faciliter leur exploitation. Les fonctionnalités proposées incluent :

- Téléchargement des résultats au format CSV
- Génération automatique de fichiers avec un identifiant unique
- Retour d'une URL sécurisée pour le téléchargement

Cette fonctionnalité repose sur l'outil `export_csv_tool`, exposé et piloté par l'agent intelligent.

4.2.2.5 Interface de campagne email

Cette section permet à l'utilisateur de lancer des campagnes d'emailing ciblées. Les actions possibles sont :

- Upload d'un fichier CSV contenant la liste des destinataires
- Upload optionnel d'une image en pièce jointe
- Lancement de la campagne d'envoi
- Retour d'un statut d'exécution (succès ou échec)

Les données sont envoyées à l'endpoint `/email_campaign`, qui déclenche un agent autonome chargé de l'envoi des emails.

4.2.3 Module d'affichage des logs

Afin d'assurer la traçabilité, le suivi des opérations et le débogage du système, l'interface graphique intègre un module dédié à l'affichage des logs applicatifs.

Ce module permet à l'utilisateur de visualiser en temps réel ou en différé les événements générés par le backend, notamment :

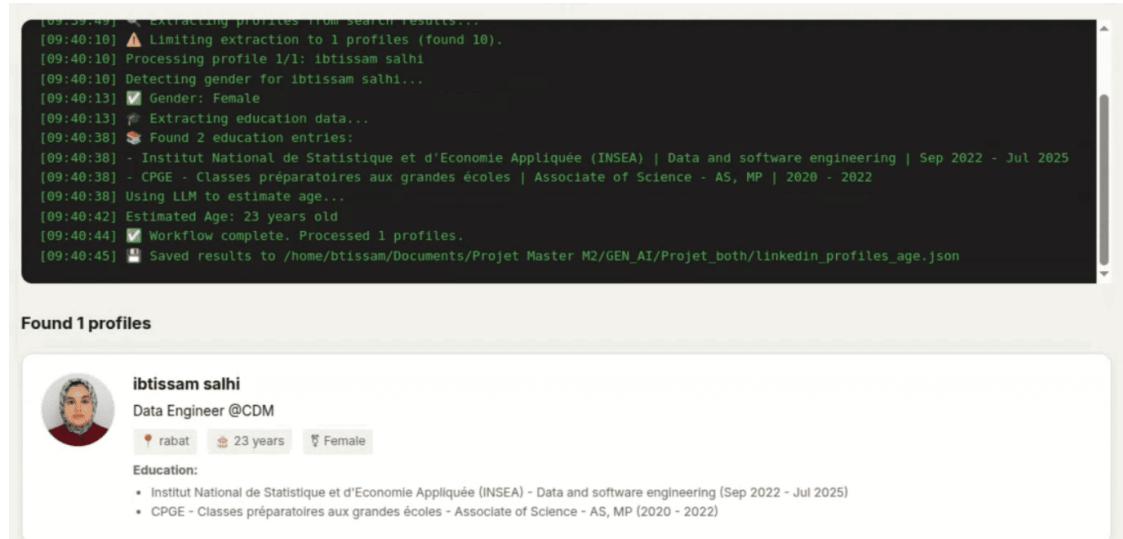
- Démarrage et arrêt des processus de scraping
- Statut des requêtes de filtre et d'export
- Exécution des requêtes de l'agent IA
- Lancement et état des campagnes d'emailing

4. Résultats

- Messages d'erreur, avertissements et informations système

Les logs sont collectés et structurés par le service `log_service`, puis exposés au frontend via une API dédiée. L'interface les affiche dans une zone spécifique sous forme chronologique, facilitant ainsi :

- La surveillance de l'état global du système
- L'identification rapide des erreurs
- L'analyse du comportement de l'application



The screenshot shows a terminal window at the top displaying log messages in green and black text. Below the terminal is a UI component titled "Found 1 profiles". It displays a profile card for "ibtissam salhi" with a placeholder image. The card includes the title "Data Engineer @CDM", location "rabat", age "23 years", gender "Female", and education details: "Institut National de Statistique et d'Economie Appliquée (INSEA) - Data and software engineering (Sep 2022 - Jul 2025)" and "CPGE - Classes préparatoires aux grandes écoles - Associate of Science - AS, MP (2020 - 2022)".

```
[09:39:49] 🚨 Extracting profiles from search results...
[09:40:10] ⚠️ Limiting extraction to 1 profiles (found 10).
[09:40:10] Processing profile 1/1: ibtissam salhi
[09:40:10] Detecting gender for ibtissam salhi...
[09:40:13] ✅ Gender: Female
[09:40:13] 📚 Extracting education data...
[09:40:38] 📄 Found 2 education entries:
[09:40:38] - Institut National de Statistique et d'Economie Appliquée (INSEA) | Data and software engineering | Sep 2022 - Jul 2025
[09:40:38] - CPGE - Classes préparatoires aux grandes écoles | Associate of Science - AS, MP | 2020 - 2022
[09:40:38] Using LLM to estimate age...
[09:40:42] Estimated Age: 23 years old
[09:40:44] ✅ Workflow complete. Processed 1 profiles.
[09:40:45] 📁 Saved results to /home/btissam/Documents/Projet Master M2/GEN_AI/Projet_both/linkedin_profiles_age.json
```

Figure 4.8: Exemple d'extraction d'un profile du LinkedIn avec keyword.

Ce mécanisme améliore la transparence du système et renforce la fiabilité perçue par l'utilisateur final, tout en fournissant des informations essentielles pour la maintenance et l'évolution de l'application.

4. Résultats

```
[09:42:22] 🚨 Agent is thinking...
[09:42:25] 🚨 Reasoning: I'll help you find all professors in Marrakech and save the results to a CSV file.
[09:42:25] ✅ Tool Call: filter_profiles_tool with filters: {'keyword': 'professor', 'location': 'marrakech'}
[09:42:25] ✓ Found 3 profiles.
[09:42:36] 🚨 Reasoning: Now I'll save these results to a CSV file.
[09:42:36] ✅ Tool Call: export_csv_tool with 3 profiles.
[09:42:36] 📁 CSV saved to exports/5312a377-b0ae-411d-9147-f8f40a9a2d9e_professors_marrakech.csv
[09:42:41] ✅ Answer: Perfect! I found **3 professors in Marrakech** and successfully saved them to a CSV file. Here's a summary of the profiles found: 1. **Beddar Hicham** (Male, 31) - Assistant Professor in Higher Education 2. **Abdelouahab El Boukhar** (Male, 32) - Assistant Professor | Innovation & IT Management at Cadi Ayyad University 3. **Hajar Mousannif** (Female, 43) - AI Strategist | Full Professor at UCA The data has been exported to **professors_marrakech.csv** and is ready for download. You can access it using the download link provided.
[09:42:41] ✓ Agent processed request successfully
```

Found 3 profiles via tool



beddar hicham
Assistant Professor in Higher Education
📍 marrakech ⏰ 31 years ♂ Male

Education:

- Cadi Ayyad University - Master's degree, Translation Technology and Specialized Translation (2015 - 2018)
- Ibn Zohr University - Doctor of Philosophy - PhD, English Language and Literature/Letters (N/A)



Abdelouahab El Boukhar
Assistant Professor | Innovation & IT Management | Cadi Ayyad University (UCA)

Figure 4.9: Exemple d'extraction du database avec agent.

5

Analyse des Défis Transverses et Perspectives

5.1 Analyse des défis liés au traitement d'image et à l'impression

Le développement du pipeline de traitement d'image a révélé des défis qui dépassent le cadre de la simple programmation. Ils touchent à la nature même de la donnée générée par l'IA et aux contraintes rigoureuses de l'industrie graphique.

5.1.1 Le défi de l'équilibre entre "Hallucination" et "Netteté"

Le principal défi lors de l'utilisation de modèles de Super-Resolution comme Real-ESRGAN est la gestion des artefacts.

- L'hallucination de détails : Plus on augmente le facteur d'échelle, plus l'IA a tendance à inventer des micro-textures (sur la peau ou les tissus) qui n'existaient pas dans l'original.
- La solution apportée : Nous avons dû brider le pipeline à un maximum de deux passes de traitement et introduire une étape de pré-nettoyage via OpenCV pour stabiliser les zones de couleurs unies, évitant ainsi que le bruit numérique ne soit interprété comme un détail à accentuer.

5. Analyse des Défis Transverses et Perspectives

5.1.2 La gestion critique de la mémoire RAM

Le passage de l'espace sRGB (très étendu et lumineux sur écran) au CMJN (plus restreint pour l'encre) entraîne inévitablement une perte de vivacité.

- Le défi marketing : Comment expliquer à un utilisateur que l'image néon générée par l'IA sera plus terne une fois imprimée ?
- L'apport du projet : Le module de Soft Proofing a permis de réduire ce gap de perception en offrant une prévisualisation honnête, évitant ainsi les erreurs coûteuses en phase de production réelle.

5.2 Perspectives d'évolution pour la partie Image

Bien que le pipeline actuel soit opérationnel et performant, plusieurs axes d'amélioration peuvent être envisagés pour le futur :

- **Vectorisation Hybride par Segments** : Plutôt que de vectoriser l'image entière (ce qui a échoué), une perspective serait d'utiliser des modèles de segmentation (comme SAM - Segment Anything Model) pour isoler uniquement le texte et les logos, les vectoriser séparément, puis les réinjecter sur le fond traité en haute définition (Raster).
- **Entraînement de modèles d'Upscaling spécifiques au Marketing** : Utiliser des datasets de publicités haute définition pour affiner (fine-tuning) le modèle Real-ESRGAN afin qu'il reconnaissse mieux les textures spécifiques aux produits commerciaux (verre, métal poli, typographie).
- **Génération de fichiers PSD multicouches par segmentation** : Une évolution majeure consisterait à briser la nature "aplatie" des images d'IA pour générer des fichiers PSD (Photoshop) éditables. En utilisant des modèles de segmentation avancés comme SAM (Segment Anything Model), le pipeline pourrait isoler automatiquement le sujet, l'arrière-plan et le texte sur des calques distincts. Cette approche permettrait aux graphistes de modifier ou de déplacer des éléments spécifiques sans altérer le reste de la composition, offrant une flexibilité totale pour les déclinaisons publicitaires.

5. Analyse des Défis Transverses et Perspectives

5.3 Perspectives pour la partie ai scrapping

Les travaux présentés constituent une base solide pour l'automatisation contrôlée du sourcing clients à l'aide de workflows IA orchestrés. Néanmoins, plusieurs perspectives d'évolution peuvent être envisagées afin d'enrichir le système, d'améliorer ses performances et d'élargir son champ d'application.

5.3.1 Amélioration du raisonnement sémantique

Une première perspective concerne le renforcement des capacités de raisonnement du modèle de langage. Actuellement, le LLM intervient principalement pour l'enrichissement des données et l'interprétation des requêtes utilisateur. À l'avenir, il serait possible d'introduire :

- Une analyse sémantique plus fine des intitulés de postes, afin de mieux gérer les variations lexicales et les titres non standardisés ;
- Une pondération automatique de la pertinence des profils, produisant un score de matching entre la requête et chaque profil ;
- Une meilleure gestion des ambiguïtés linguistiques dans les requêtes complexes ou imprécises.

5.3.2 Extension vers une approche multi-plateformes

Le workflow présenté est actuellement centré sur LinkedIn. Une évolution naturelle consisterait à étendre l'architecture à d'autres plateformes professionnelles ou sources de données ouvertes (sites d'entreprises, annuaires professionnels, bases publiques).

Grâce à l'abstraction offerte par le MCP, cette extension pourrait être réalisée en :

- Déployant de nouveaux serveurs MCP spécifiques à chaque source ;
- Conservant un orchestrateur et un agent de raisonnement uniques ;
- Mutualisant les mécanismes de filtrage, d'enrichissement et d'export.

Cette approche permettrait de construire un système de sourcing unifié et multi-sources, tout en conservant les garanties de contrôle et de conformité.

5. Analyse des Défis Transverses et Perspectives

5.3.3 Automatisation adaptative sous supervision humaine

Une autre perspective intéressante consiste à introduire un niveau intermédiaire entre workflow strictement scripté et agent autonome. Il s'agirait d'un mode *semi-autonome*, dans lequel :

- L'agent propose des actions ou des stratégies alternatives ;
- L'utilisateur valide explicitement les décisions sensibles ;
- Les actions à risque restent bloquées sans confirmation humaine.

Ce paradigme *Human-in-the-Loop* renforcerait la sécurité et la confiance, tout en augmentant progressivement le degré d'automatisation.

5.3.4 Analyse avancée, auditabilité et conformité

Du point de vue industriel, le système pourrait être enrichi par :

- Des tableaux de bord analytiques (taux de profils pertinents, efficacité des campagnes, taux de réponse) ;
- Une exploitation avancée des logs pour l'audit, la conformité réglementaire et la détection d'anomalies ;
- Une traçabilité complète des décisions prises par l'agent, facilitant l'explicabilité (*Explainable AI*).

Ces éléments sont particulièrement importants dans des contextes professionnels soumis à des exigences réglementaires strictes.

5.3.5 Vers une industrialisation du workflow

Enfin, une perspective majeure réside dans l'industrialisation du système :

- Déploiement sous forme de microservices conteneurisés (Docker) ;
- Gestion multi-utilisateurs avec authentification et rôles ;
- Planification de campagnes et exécutions différées ;
- Scalabilité contrôlée via files de tâches (queues).

5. Analyse des Défis Transverses et Perspectives

Cette évolution permettrait de transformer le prototype académique présenté en une solution robuste, exploitable dans un environnement professionnel réel.

En résumé, les perspectives ouvertes par ce travail montrent que les **AI Workflows orchestrés** constituent une alternative crédible et durable aux agents autonomes classiques, en particulier dans des environnements contraints. Ils offrent un compromis équilibré entre intelligence, contrôle et conformité, ouvrant la voie à des systèmes d'automatisation responsables et évolutifs.

6

Conclusion Générale

Ce projet s'est articulé autour de l'ambition de transformer les capacités brutes de l'Intelligence Artificielle générative en outils opérationnels, fiables et adaptés aux exigences du marketing digital moderne. En adressant simultanément la problématique de la qualité visuelle pour l'impression et celle de l'efficacité de la prospection B2B, nous avons construit un pont entre l'innovation technologique et les réalités du terrain.

La première partie de ce travail a démontré que, si les images issues de l'IA ne sont pas nativement prêtes pour l'industrie physique, le développement d'un pipeline de traitement rigoureux permet de lever ces barrières. L'échec de la vectorisation nous a conduits vers une solution plus sophistiquée : un workflow combinant l'Upscaling par Deep Learning (Real-ESRGAN), la gestion de la mémoire par blocs (tiling) et une conversion colorimétrique CMJN conforme aux standards PDF/X-1a. Cette approche garantit que la créativité de l'IA peut désormais s'exprimer sur des supports de très grand format sans compromis sur la netteté.

La seconde partie, dédiée à l'automatisation intelligente sur LinkedIn, a prouvé la supériorité des architectures agentiques sur les méthodes de scraping traditionnelles. En intégrant des capacités de raisonnement sémantique, nous avons réussi à automatiser le sourcing de clients ciblés tout en maintenant une précision de filtrage proche de celle d'un expert humain.

Les défis relevés tout au long de ce projet — qu'ils soient liés à la gestion des ressources matérielles (RAM), à la fidélité chromatique ou à l'éthique de l'extraction de données — soulignent que la véritable valeur de l'IA réside dans sa médiation technique. L'interface graphique développée permet enfin de démocratiser ces

6. Conclusion Générale

processus complexes, les rendant accessibles aux professionnels du marketing sans nécessiter de compétences en programmation.

En perspective, l'évolution vers une segmentation intelligente des visuels pour générer des fichiers PSD multicouches ouvrira la voie à une personnalisation encore plus poussée. Ce projet pose ainsi les jalons d'un futur où l'IA n'est plus seulement un moteur de création, mais le cœur d'un écosystème de production marketing complet, agile et hautement qualitatif.