UCLouvain
FUCaM MONS

LOUVAIN
School of Management

Louvain School of Management

Quel est le lien entre la « Grande Démission » et l'émergence croissante de l'intelligence artificielle et de l'automatisation dans le monde du travail ? : Analyse LinkedIn

Auteur : Venuti Antoine
Promoteur : Corentin Vande Kerckhove
Année académique : 2023- 2024
Master [120] : Ingénieur de Gestion à finalité spécialisée

Declaration Regarding AI Tool Usage in Master's Thesis

We recognize that AI tools might be valuable aids during the master's thesis work, but they are not infallible. Remember that transparency fosters trust, and acknowledging AI's role enhances the credibility of your work.

Therefore, when deciding to use such a tool, you need to adhere to the following principles of responsible use of AI.

1. **Critical Evaluation :**
    - We critically assessed the AI-generated output, ensuring its alignment with our research objectives.
    - Any modifications or corrections were made based on our expertise and domain knowledge.
2. **Transparency :**
    - We acknowledge the use of ChatGPT transparently, emphasizing that it contributed to our work but did not replace human judgment.
    - Our commitment to transparency ensures the integrity of this thesis.
3. **Ethical Considerations :**
    - We actively monitored for biases or unintended consequences introduced by the AI tool.
    - Our ethical responsibility guided our decisions throughout the research process.

Declaration

During the preparation of this master's thesis, the author(s) utilized ChatGPT 3.5 for the following purpose:

1. Help with coding: Using this AI to fix errors and improve the performance of some Python code.

After using ChatGPT, the author(s) diligently reviewed and edited the content produced by the tool. We take full responsibility for the final content presented in this thesis.

By signing this declaration, we affirm that the content of this master's thesis reflects our original work, augmented by the responsible use of AI.

28/05/2024 Venuti Antoine

Résumé

La pandémie de COVID-19 a transformé le monde du travail, entraînant la "Grande Démission" et une adoption accrue de l'IA pour faire face aux restrictions imposées. Cette étude a pour but d'utiliser plusieurs techniques de Natural Language Processing (NLP) afin d'examiner comment l'automatisation pourrait être à la fois une réponse à la démission massive des employés et une cause des changements radicaux de carrière. Dans un premier temps, nous extrayons les offres d'emploi liées à l'automatisation pour les années 2019 et 2023. Ensuite, nous procédons à une vectorisation des données en utilisant SBERT. Nous appliquons des techniques de réduction de dimensionnalité et de clustering, notamment K-means et HDBSCAN, sur les offres d'emploi liées à l'automatisation de 2019 et 2023 pour identifier les emplois menacés avant et après la Grande Démission. Les résultats montrent que la réduction de dimensionnalité par UMAP est la plus interprétable, tandis que l'évaluation de la qualité du clustering sur plusieurs métriques indique que HDBSCAN offre les meilleures performances. Les résultats du clustering nous indiquent donc quels étaient les emplois automatisés en 2019 et en 2023. Pour examiner les démissions des emplois identifiés dans les clusters survenues pendant la Grande Démission, nous utilisons une base de données LinkedIn contenant les expériences professionnelles des profils. Nous commençons par normaliser les titres d'emplois en utilisant JobBERT pour les associer aux emplois ESCO standardisés. Nous calculons ensuite la similarité au sein de chaque transition de jobs avec JobBERT et la similarité cosinus pour pouvoir déterminer quelles transitions sont des changements de carrière. Les résultats de ce calcul de similarité indiquent que cette technique présente de bons résultats, mais il serait nécessaire de l'évaluer plus en profondeur dans de futures recherches. Enfin, nous observons si les jobs automatisés en 2019 et en 2023 ont connu une augmentation de changements de carrière durant la Grande Démission. Les résultats montrent que les emplois menacés par l'automatisation avant la Grande Démission n'ont pas connu une augmentation des changements de carrière. En revanche, les demandes d'automatisation en 2023 concernent exclusivement des emplois ayant enregistré une hausse des démissions ou des changements de carrière. Cependant, cette étude n'établit pas de lien de corrélation direct, ce qui constitue une piste d'amélioration pour les recherches futures.

Remerciements

Je tiens à exprimer ma sincère gratitude au Professeur Corentin Vande Kerckhove pour son aide et ses précieux conseils tout au long de la réalisation de ce mémoire.

Je souhaite également particulièrement remercier ma famille et mes amis qui m'ont toujours soutenu durant la réalisation de ce mémoire.

Table des matières

Liste des figures

Liste des tableaux

Introduction

La pandémie de COVID-19 a transformé le monde du travail à de nombreux niveaux. Parmi les phénomènes notables causés par le coronavirus et les restrictions associées, la Grande Démission se distingue particulièrement. Celle-ci décrit le phénomène d'un grand nombre d'employés démissionnant volontairement depuis mars 2021 (Serenko,2023), lequel s'est renforcé au cours de la seconde moitié de l'année, peu de temps après le début de la reprise économique post-COVID-19, alors que de plus en plus d'opportunités d'emploi se sont présentées (Liu-Lastres et al. 2023). En parallèle de ce phénomène, les entreprises ont intensifié l'adoption de l'IA pour répondre aux défis posés par la pandémie. Ces deux phénomènes, bien que distincts, sont intimement liés et soulèvent des questions cruciales : l'automatisation est-elle une réponse au départ massif des employés, ou contribue-t-elle à exacerber ce phénomène? Alors que certains voient dans l'automatisation un moyen de combler les vides laissés par la démission des travailleurs, d'autres s'inquiètent des implications à long terme sur les emplois et les compétences nécessaires. Cette étude explore cette relation complexe et bidirectionnelle, offrant des perspectives essentielles pour comprendre l'impact de ces transformations simultanées sur le marché du travail.

La première partie de ce mémoire consiste en une mise en contexte où nous présenterons la « Grande Démission » ainsi que l'accélération de l'automatisation dans le monde du travail. L'objectif est de poser les bases et de mieux comprendre le cadre dans lequel s'inscrit ce mémoire.

La seconde partie de ce mémoire est dédiée à la revue de littérature. L'objectif est de recenser les travaux existants sur le thème de la Grande Démission et de présenter les concepts théoriques pertinents pour cette étude.

Dans la troisième partie de cette recherche, nous présenterons la méthodologie employée pour identifier les emplois menacés par l'automatisation avant et après la Grande Démission, ainsi que celle utilisée pour étudier les changements de carrière qui ont eu lieu à partir de ces emplois durant cette période.

Enfin, nous présenterons les résultats des différentes méthodes et les métriques utilisées pour les évaluer. Nous terminerons cette étude par une conclusion générale et par une présentation des limites et des possibilités d'amélioration de ce travail.

1. Mise en contexte

1.1. La Grande Démission

Partout dans le monde, les gens démissionnent de leur emploi. Plus de 24 millions d'Américains ont quitté leur emploi au cours des deuxième et troisième trimestres de 2021 (Sull et al. 2022). Une tendance similaire a été observée en Europe occidentale, en particulier au Royaume-Uni (YPulse, 2021). Une étude menée par Microsoft auprès de plus de 30 000 individus a révélé que plus de 40 % de la main-d'œuvre mondiale envisageait sérieusement de changer d'employeur en 2021 (Microsoft 2021). Une autre enquête récente de Workhuman menée auprès de 3 500 employés aux États-Unis, au Royaume-Uni, en Irlande et au Canada présente un chiffre similaire, soit deux fois plus élevé qu'en 2019, où seuls 20 % des travailleurs envisageaient de partir (Workhuman, 2021). Alors que ce phénomène se poursuit, les dirigeants d'entreprise tentent de saisir les raisons profondes derrière ce départ massif des employés et s'efforcent de trouver des stratégies efficaces pour retenir les employés talentueux (Sull et al. 2022).

La « Grande Démission », un terme récemment introduit par Anthony Klotz, professeur à l'University College London, lors de son interview avec Arianne Cohen (2021) pour Bloomberg Businessweek, transforme le marché de l'emploi de manière inattendue et à une rapidité sans précédent (Jaffe et al. 2022). De nombreux universitaires et commentateurs se sont saisis de cette expression accrocheuse pour affirmer que les employés « retrouvent » leur vie. Ils mettent en avant des taux significatifs d'épuisement professionnel, le souhait d'établir un meilleur équilibre entre vie professionnelle et personnelle, ainsi que la présence d'environnements de travail nocifs comme principales raisons de ce phénomène (Ng & Stanton, 2023). Ce phénomène de Grande Démission (également appelé Big Quit) constitue un problème économique majeur actuel.

La Grande Démission n'a pas impacté tous les secteurs de la même manière, l'industrie de l'hôtellerie demeurant parmi les plus durement touchées (Kundu et al. 2022). Selon un récent rapport du Bureau of Labor Statistics (US Bureau of Labor, 2022), le secteur de l'hébergement et de la restauration a enregistré le taux de démission le plus élevé en janvier 2022, atteignant 6,0 %. Ce chiffre est suivi par un taux de 4,6 % dans le commerce de détail et de 3,4 % dans les domaines des arts, du divertissement et des loisirs. Les recherches sur la Grande Démission ont également montré que les taux de démission diffèrent selon la génération et le genre. La génération Z a été en première ligne de ce mouvement : selon une enquête mondiale, près de 59 % des participants de cette génération ont exprimé leur mécontentement vis-à-vis de leur

travail actuel, tandis que plus de 50 % ont déclaré avoir l'intention de changer d'emploi dans l'année à venir. La génération Z a été la plus active dans les changements d'emploi, avec une augmentation de 80 %, suivie des millennials avec un taux de transition d'environ 50 %, de la génération X avec un taux d'environ 31 %, et enfin des baby-boomers avec un taux d'environ 5 % (Kundu et al. 2022). LinkedIn a examiné cette tendance et constaté qu'il y a eu une augmentation de 54 % entre 2021 et 2022 du nombre de membres de LinkedIn ayant mis à jour leur profil et trouvé un emploi. D'après l'analyse de plus de 9 millions de dossiers d'employés issus de plus de 4 000 entreprises, une augmentation de plus de 20 % du taux de démission chez les employés âgés de 30 à 45 ans entre 2020 et 2021 a été constatée (Liu-Lastres et al. 2023). En outre, une disparité importante dans les taux de démission entre les genres a été observée en août 2021 depuis le début de la pandémie, avec un rythme de démissions plus élevé chez les travailleuses par rapport à leurs homologues masculins (Pardue, 2021).

1.2. L'accélération de l'automatisation

En parallèle de ce phénomène de « Grande Démission », l'économie a subi une transformation structurelle, marquée par l'avancée de l'intelligence artificielle et de l'automatisation, générant une incertitude accrue quant à la stabilité professionnelle tout en demandant de nouvelles compétences. De nombreuses entreprises ont intensifié leurs investissements dans l'automatisation et l'intelligence artificielle (IA) pendant la pandémie de COVID-19 pour répondre aux exigences de distanciation sociale et limiter les interactions physiques, incitant de nombreux travailleurs à acquérir de nouvelles compétences ou à se reconvertir pour demeurer compétitifs dans un marché du travail en constante évolution (Ng & Stanton, 2023). L'automatisation pose de nombreux défis aux entreprises et aux employés concernant des tâches auparavant effectuées par des personnes, telles que la saisie de données, la comptabilité et le travail sur les chaînes de montage, qui sont désormais soutenues par les technologies numériques et l'intelligence artificielle. Cette tendance à l'automatisation est appelée à se poursuivre, entraînant une augmentation des défis. Par exemple, aux États-Unis, il est prévu que 47 % des emplois seront automatisés dans les décennies à venir (Lukauskas et al., 2023).

Cependant, la numérisation croissante pose un autre défi : la demande de nouvelles professions et compétences (Bacher & Tamesberger,2021). L'automatisation d'une grande partie du travail des employés crée un problème, car les entreprises manquent de personnel qualifié pour mettre en œuvre et maintenir les dernières technologies. Il existe une pénurie significative de travailleurs spécialisés en cybersécurité et en analyse de données. Avec la popularité et le

développement de l'intelligence artificielle, les entreprises intègrent de plus en plus cet outil dans leurs activités, entraînant la création de nouveaux emplois et la disparition d'autres. Les changements technologiques révèlent que certains travailleurs possèdent les compétences nécessaires et s'adaptent plus facilement au marché du travail, tandis que d'autres peinent à suivre cette évolution (Arntz et al.,2016). En effet, la montée en puissance de la technologie est déjà en train de transformer le marché du travail, conduisant à l'automatisation de certains emplois et tâches d'une part, et à l'émergence de nouvelles sortes d'emplois d'autre part (Skills, not job titles, are the new metric for the labour market, 2020). Par exemple, Sobeys, une entreprise alimentaire majeure, utilise des robots et l'IA pour gérer les commandes d'épicerie en ligne, réduisant ainsi les emplois nécessitant peu d'interactions humaines. Même l'industrie immobilière tire parti de la réalité virtuelle pour des visites à distance. Cependant, l'automatisation menace également les emplois à forte composante humaine. Les domaines créatifs comme la composition musicale peuvent être impactés par l'apprentissage automatique et l'IA. Bien que la pandémie de COVID-19 n'ait pas initié un changement fondamental dans nos tâches professionnelles, elle a considérablement accéléré l'adoption massive de l'automatisation et de l'IA. La rapidité de l'évolution technologique remet en question les compétences nécessaires, encourageant le développement de compétences fonctionnelles, de leadership et de soft skills difficiles à reproduire (Ng & Stanton, 2023 ; Serenko, 2024).

Le remplacement des emplois par les machines est une réalité croissante qui concerne même des professions autrefois hautement spécialisées. Plusieurs recherches ont mis en lumière la menace que représentent les dispositifs autonomes et intelligents pour de nombreux emplois. Selon une étude d'économistes du travail, l'introduction d'un robot supplémentaire pour mille travailleurs pourrait réduire le ratio emploi/population de 0,18 à 0,34 point de pourcentage, et les salaires de 0,25 à 0,5 %. Quant à la question de savoir si l'IA et la robotique créeront plus d'emplois qu'elles n'en détruiront, les avis sont partagés : une enquête du Pew Research Center et du Imagining the Internet Center de l'Université Elon en 2014 a révélé que 48 % des répondants imaginaient un avenir où plus d'emplois seraient perdus que créés, tandis que 52 % estimaient que plus d'emplois seraient créés que perdus (Rainie et al. 2017).

2. Revue de littérature

Après avoir présenté le contexte du marché du travail actuel, marqué par les phénomènes de la « Grande Démission » et de la digitalisation, nous entamons maintenant l'examen de la littérature. Pour rappel, l'objectif principal de ce mémoire est d'étudier la relation entre la « Grande Démission » et les changements induits par l'automatisation. Dans cette revue de littérature, nous examinerons les causes et conséquences répertoriées de la « Grande Démission » afin de mieux comprendre ce phénomène et de mettre en évidence les liens déjà établis avec l'automatisation. Ensuite, nous présenterons les sources à travers lesquelles les tendances du marché du travail peuvent être analysées et comment les techniques de « Natural Language Processing » (NLP) peuvent nous y aider. Enfin, nous parcourrons les travaux déjà réalisés sur ce sujet.

2.1. Causes de la Grande Démission

Alaql et al. (2023) ont identifié que l'autonomie accordée aux travailleurs en termes de temps, lieu et flexibilité est l'une des principales causes de démission. De manière similaire, une étude de Microsoft (2021) révèle un besoin croissant de flexibilité et de travail hybride, ce qui rejoint les résultats du Pew Research Center (2022) indiquant que 48 % des travailleurs ayant quitté leur emploi ont cité des problèmes de garde d'enfants. La génération Z est particulièrement affectée par les défis liés à la conciliation de la vie professionnelle et personnelle (Microsoft, 2021). En effet, 20 % des individus âgés de 18 à 24 ans ont démissionné en raison d'un manque d'équilibre entre ces deux aspects (YPulse, 2021).

Zagorsky (2022) révèle que les travailleurs plus jeunes de la génération Z ont vu leur attachement et leur fidélité envers leur entreprise diminuer à cause du manque de socialisation dû aux mesures de distanciation sociale pendant la pandémie de COVID-19, ce qui a freiné leur progression professionnelle et les a poussés à explorer d'autres opportunités de carrière. L'étude du Pew Research Center (2022) rejoint cette conclusion en mettant en évidence une tendance selon laquelle presque le double de la génération Z et des milléniaux, comparativement aux baby-boomers, envisageaient de changer d'emploi en 2022. En effet, 77 % de la génération Z, 63 % de la génération Y et 33 % des baby-boomers exprimaient cette intention de changer de poste (Kuzior et al. 2022) .

Une autre principale raison de démission est le stress engendré par le travail. Une charge de travail importante et un stress élevé ont entraîné de nombreux burn-out, incitant les employés en milieu de carrière à reconsidérer leurs priorités professionnelles et personnelles. Selon YPulse (2021), le bien-être mental est devenu une préoccupation majeure pour de nombreux jeunes professionnels. 24 % des individus âgés de 18 à 24 ans ont quitté leur emploi en raison de son impact négatif sur leur santé mentale. Bien que le stress soit courant en milieu professionnel, la pandémie a intensifié les exigences dans divers secteurs, augmentant significativement le niveau de stress. Les emplois dans le secteur des services sont devenus plus exigeants, avec des risques accrus et une atmosphère plus tendue, tandis que le télétravail a rendu le travail omniprésent, perturbant l'équilibre entre vie professionnelle et vie personnelle.

Ensuite, la rémunération a également été l'une des principales raisons de démission, selon Alaql et al. (2023). En effet, 63 % des travailleurs la citent comme raison principale (Pew Research Center, 2022) et YPulse (2021) indique que le désir d'une rémunération plus élevée est la principale raison pour laquelle de nombreux jeunes Européens quittent leur emploi.

Enfin, selon Sull et al. (2022), la principale raison est une culture toxique dans l'entreprise. Les entreprises renommées pour leur culture saine, comme Southwest Airlines, Johnson & Johnson, Enterprise Rent-A-Car et LinkedIn, ont maintenu des taux de rotation du personnel inférieurs à la moyenne pendant les six premiers mois de la Grande Démission. À l'inverse, des entreprises plus axées sur l'innovation, telles que SpaceX, Tesla, Nvidia et Netflix, ont enregistré des taux de démission plus élevés que leurs concurrents plus traditionnels. Cette situation n'est pas spécifique aux seules industries technologiques, car même des entreprises innovantes comme Goldman Sachs et Red Bull ont été touchées par des taux de rotation plus élevés. Selon les auteurs, une culture toxique est 10.4 fois plus susceptible de contribuer à la démission qu'un problème de rémunération. L'étude du Pew Research Center (2022) appuie cette affirmation en révélant que 57% des démissions sont liées au manque de respect au travail.

2.2. Conséquences de la Grande Démission

Cette « Grande Démission » n'est pas sans conséquence. Elle est en train d'accélérer la Quatrième Révolution Industrielle, également connue sous le nom d'Industrie 4.0. Ce mouvement pourrait offrir une solution potentielle à la pénurie de main-d'œuvre actuelle. Steve Reinharz, le fondateur et PDG de Robotic Assistance Devices, recommande une augmentation de l'utilisation de la robotique et de l'intelligence artificielle en soulignant que ces technologies sont capables de s'occuper d'une variété de tâches telles que la livraison, les soins aux personnes âgées, le nettoyage et le service client. La robotique à distance est devenue cruciale pour résoudre les pénuries de main-d'œuvre, car les méthodes traditionnelles ne suffisent plus. Selon Reinharz, environ 25% des emplois seront effectués par des robots à l'avenir, mais cela ne devrait pas automatiquement conduire à un chômage massif. Au lieu de cela, cela induira des transformations dans les emplois existants, exigeant une adaptabilité et une flexibilité de la part des travailleurs (Kuehner-Hebert,2021).

Ensuite, les compétences sont devenues la nouvelle monnaie sur le marché du travail. Elles indiquent la demande et l'offre à un niveau plus nuancé que les métiers, dont l'expertise requise évolue de plus en plus rapidement, et les diplômes, souvent obsolètes au moment de leur obtention. Le rythme actuel du changement nécessite donc de suivre la voie d'un marché du travail basé sur les compétences, plutôt que sur les diplômes, une variable bien plus dynamique (Skills, not job titles, are the new metric for the labour market, 2020). Cette prise de conscience est également intégrée par les travailleurs eux-mêmes : selon une enquête du Pew Research Center menée en 2016, 87 % des travailleurs estiment qu'il sera crucial pour eux d'acquérir une formation continue pour s'adapter aux évolutions sur leur lieu de travail (Rainie et al. 2017).

Jaffe et al. (2022) ont réalisé une étude concernant les tendances du marché du travail en se concentrant sur les compétences recherchées par les employeurs dans les secteurs du savoir et des services. En ce qui concerne les compétences techniques, ils identifient six catégories principales : langages de programmation, outils de base de données, frameworks Back End, frameworks Front End, systèmes d'exploitation et systèmes ERP. Ils identifient une demande croissante de compétences techniques dans le secteur du savoir, en particulier les compétences en base de données.

2.3. La Grande Démission à travers les médias sociaux et application du Natural Language Processing (NLP)

Les plateformes de réseaux sociaux en ligne n'ont cessé de croître en nombre, en qualité et en influence, jouant un rôle de plus en plus profond au sein de la société (Alaql et al., 2023). Les données numériques issues de ces plateformes ont le potentiel de capturer l'opinion publique, d'identifier des communautés et des organisations, de repérer des tendances, de formuler des prédictions dans des domaines disposant de volumes conséquents de données, et d'explorer des sujets d'intérêt. Les perspectives offertes par ces données sont vastes (Suma et al., 2020; Alotaibi et al., 2020). Avec 850 millions d'abonnés, LinkedIn se classe en tant que réseau social le plus prisé des professionnels (LinkedIn Corporation, 2022). Cette plateforme offre aux utilisateurs la possibilité de mettre en avant leurs expériences professionnelles tout en favorisant l'établissement de connexions professionnelles. Utilisé comme un outil clé pour le recrutement, la diffusion de contenus et la création d'identités de marque, LinkedIn joue un rôle central dans le monde professionnel. Sa vaste base de données de publications représente une mine d'informations précieuses permettant de saisir des problématiques et de déceler des tendances émergentes (Alaql et al. 2023). D'autres sites sont également spécialisés dans les offres d'emploi, l'un des plus connus étant Indeed.

L'analyse des exigences et des tendances professionnelles devient de plus en plus essentielle. Comprendre les évolutions et les transitions du marché du travail est crucial pour les chercheurs d'emploi, les employeurs et les décideurs politiques.

La taille des données provenant de LinkedIn et des autres sites d'emploi étant substantielle et les données étant de nature essentiellement textuelles, il est difficile pour un humain de découvrir les connaissances qu'elles renferment. Les avancées en *machine learning*, en *clustering* et en traitement du langage naturel (NLP) ont considérablement amélioré notre capacité à analyser les offres d'emploi et à identifier les tendances professionnelles. Les algorithmes d'apprentissage automatique peuvent être entraînés sur de vastes ensembles de données d'offres d'emploi pour identifier les exigences et leurs schémas. Les algorithmes de clustering peuvent regrouper des offres d'emploi similaires en fonction de leurs exigences, permettant ainsi une analyse plus détaillée des compétences et qualifications nécessaires dans un domaine spécifique. Le traitement du langage naturel (NLP) est une branche de l'intelligence artificielle qui développe des algorithmes pour permettre aux ordinateurs de comprendre,

interpréter et générer le langage humain. Les techniques de NLP peuvent extraire des informations à partir de textes non structurés, classifier et analyser efficacement les diverses exigences présentées dans les offres d'emploi. Comparées aux analyses manuelles, les techniques de NLP peuvent traiter rapidement de vastes volumes de données, les rendant idéales pour des applications comme l'analyse de sentiments et l'extraction d'informations. Cependant, le NLP rencontre des défis, notamment l'ambiguïté et la variabilité des langues naturelles, ce qui complique l'interprétation précise du sens et du contexte des mots et des phrases, et limite les capacités des algorithmes à extraire des informations significatives. En combinant ces méthodes, nous pouvons mieux comprendre le marché du travail et identifier de nouvelles tendances et opportunités (Chowdhary et al., 2020 ; Lukauskas et al., 2023 ; Mahany et al., 2022 ; Verma et al.,2022).

La similarité entre des documents peut être évaluée en fonction de la similarité des *embeddings* correspondant à leur contenu textuel. Les *embeddings*, qui capturent les informations lexicales et sémantiques des textes, peuvent être obtenus grâce à des encodeurs pré-entraînés ou via des approches de « Bag of words » utilisant les *embeddings* des mots constitutifs. Les *embeddings* de mots capturent des représentations sémantiquement significatives des mots en se basant sur leurs co-occurrences locales dans les phrases. Ils sont devenus presque omniprésents en raison de leur utilité pour calculer facilement la similarité sémantique entre deux mots, ainsi qu'à trouver des mots similaires à un mot donné. Les méthodes de génération d'*embeddings* représentent les mots sous forme de vecteurs continus dans un espace de dimensions réduites, capturant ainsi les propriétés lexicales et sémantiques des mots. Ces *embeddings* de mots permettent d'identifier la signification et la structure syntaxique d'un mot, créant ainsi une représentation vectorielle de ces informations. Les *embeddings* peuvent représenter le contexte, la sémantique et d'autres caractéristiques subtiles de la phrase. Cependant, il n'existe pas de consensus universel sur la manière dont les *embeddings* de phrases doivent être calculés (Agarwala et al 2021). La vectorisation du texte est une étape essentielle en traitement du langage naturel, permettant entre autres la classification, le regroupement et l'extraction d'informations. Elle convertit les données textuelles en un format numérique exploitable par les algorithmes d'apprentissage automatique.

2.4. Travaux déjà réalisés

Diverses études ont été menées sur le phénomène de la « Grande Démission » ou sur l'automatisation en se basant sur des analyses de bases de données provenant de LinkedIn et de sites d'offres d'emploi, mais aucune n'a exploré les deux phénomènes simultanément.

Plusieurs recherches se sont concentrées sur la recherche des compétences requises par les employeurs dans la période de la Grande Démission en utilisant des méthodes de clustering. L'étude la plus proche de la nôtre est celle de Lukauskas et al. (2023) qui se sont focalisés sur la compréhension des compétences requises sur le marché du travail lithuanien en utilisant des techniques NLP et de clustering. Ils sont cependant restés généraux et ne se sont pas focalisés sur les conséquences de l'automatisation. Dai et al. (2015) ont, eux, collecté des profils publics LinkedIn, puis ont appliqué des techniques NLP pour classifier les parcours éducatifs et regrouper les expériences professionnelles. De plus, dans une autre étude, Dai et al. (2018) ont utilisé des méthodes de clustering pour analyser les profils LinkedIn en fonction de leurs trajectoires professionnelles, explorant ainsi les tendances prédominantes dans le marché du travail.

Alaql et al. (2023) ont utilisé une approche dite de « deep journalism » basée sur l'intelligence artificielle (IA) pour extraire et analyser automatiquement de grandes quantités de données de LinkedIn afin de fournir au public des informations sur les marchés du travail, les compétences et l'éducation des personnes, ainsi que sur les entreprises et les industries, en prenant en compte des perspectives multigénérationnelles.

D'autres analyses se sont, elles, focalisées sur la construction d'un modèle de prédiction de démission. Sull et al. (2022) ont analysé plus de 1,4 million d'avis Glassdoor à l'aide de l'outil d'analyse de langage de CultureX. Pour chaque entreprise du Culture 500, ils ont étudié la façon dont les employés abordaient 172 sujets et évalué leur niveau de positivité. Cette analyse leur a permis d'identifier les sujets les plus prédictifs du taux de démission ajusté à l'industrie d'une entreprise. Enfin, en collaboration avec Henning Seip de Candogram, le NYU Human Capital Analytics Lab a appliqué des méthodes de prédiction dans le contexte de la Grande Démission (Jaffe et al., 2022).

2.5. Question de recherche

Dans les sections précédentes, il a été montré que l'essor de l'intelligence artificielle (IA) et de l'automatisation pouvait être observé à la fois comme une conséquence et une cause de la « Grande Démission ». D'une part, l'évolution de l'IA et l'automatisation accélérée pendant la pandémie ont suscité une incertitude croissante quant à la sécurité de l'emploi, en exigeant simultanément l'acquisition de nouvelles compétences. Cette pression a contraint de nombreux travailleurs à acquérir de nouvelles aptitudes ou à se reconvertir pour demeurer compétitifs dans un marché du travail en constante mutation (Ng & Stanton, 2023). D'autre part, l'IA et l'automatisation sont devenues des solutions cruciales pour résoudre les pénuries de main-d'œuvre, car les méthodes traditionnelles ne sont plus suffisantes (Kuehner-Hebert, 2021).

De nombreuses recherches ont déjà été effectuées sur la Grande Démission, dont certaines en utilisant des données LinkedIn, mais aucune n'a encore permis d'observer dans les faits ces relations de cause à effet. La question de recherche étudiée sera donc celle-ci : « Quel est le lien entre la Grande Démission et l'émergence de l'intelligence artificielle et de l'automatisation ? ». Les deux hypothèses principales qui vont être testées sont celles-ci :

- H1 : L'automatisation provoque des changements de carrière radicaux.
- H2 : Les secteurs ayant connu une automatisation significative en 2023 sont également les secteurs ayant observé une augmentation de démissions et de changements de carrière pendant la Grande Démission.

Les deux hypothèses secondaires sont :

- H3 : La méthode de clustering permet d'identifier les emplois menacés par l'automatisation
- H4 : L'utilisation de JobBERT permet de calculer la similarité entre les emplois

L'objectif de cette recherche est d'observer si, dans le monde du travail, l'automatisation a véritablement influencé la « Grande Démission » (H1) ou si, au contraire, elle en est une conséquence (H2). Cette étude ne vise pas à répondre à ces deux hypothèses par le calcul de la corrélation, mais à proposer une méthodologie permettant de comprendre quels emplois sont menacés par l'automatisation et d'identifier les changements de carrière qui en résultent. L'étude va se concentrer sur le marché du travail des Etats-Unis. Cette recherche a le potentiel d'aider les employeurs à comprendre le contexte dans lequel ils évoluent et à prendre les meilleures décisions par rapport à l'automatisation.

3. Matériel et méthodes

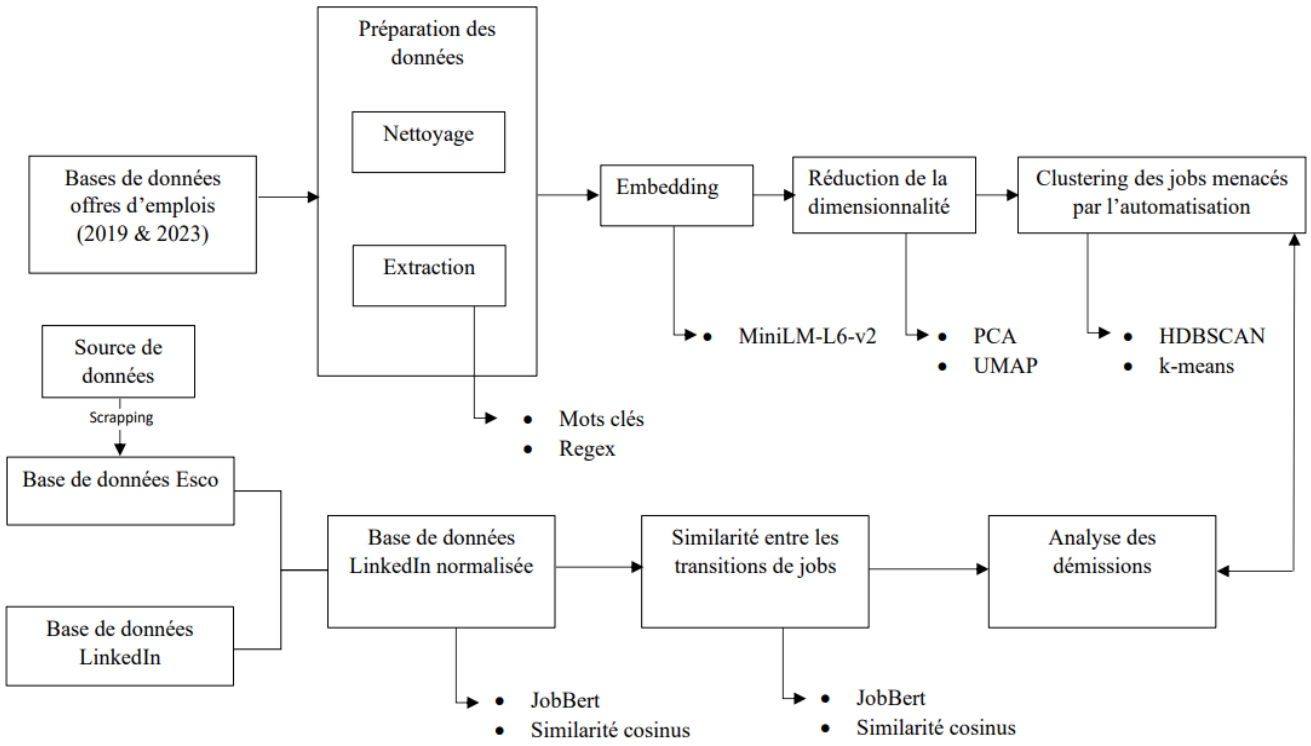
Cette section du travail présente les informations relatives aux données utilisées et à leur traitement ainsi que les différentes méthodes utilisées dans la recherche. Pour valider les hypothèses, deux composantes sont requises : comprendre quels sont les emplois menacés par l'automatisation et avoir une vue sur les transitions de jobs.

Pour tester la première hypothèse (l'automatisation conduit à des changements radicaux de jobs), les offres d'emploi de 2019, soit avant la Grande Démission, seront d'abord analysées afin de comprendre quels emplois étaient menacés par l'automatisation. Ensuite, la base de données LinkedIn sera examinée pour analyser les transitions de carrière autour de ces emplois pendant la Grande Démission (2021). Par exemple, si en 2019 il est découvert que l'automatisation vise principalement à remplacer le job X, il sera vérifié si, durant la Grande Démission, de nombreux changements de carrière ont eu lieu depuis ce job X.

Pour tester la seconde hypothèse (les secteurs les plus automatisés en 2023 sont ceux les plus touchés par les changements de carrière), le processus sera inversé. Les secteurs les plus touchés lors des premiers mois de la Grande Démission seront d'abord identifiés dans la base de données LinkedIn. Ensuite, ces informations seront comparées aux demandes d'automatisation des offres d'emploi de 2023, soit deux ans après le début de la Grande Démission. Le schéma suivant (Figure 1) reprend les principales étapes de l'étude pour aider à sa compréhension. Une description détaillée de chaque étape va être présentée dans le reste de cette section. Les principaux codes utilisés dans cette étude ont été rendus publics[1].

[1] https://github.com/AntoineVenuti/Job-analysis.git

Figure 1 : Schéma méthodologique



3.1. Analyse des emplois menacés par l'automatisation

3.1.1. Collecte, extraction et traitement des données

Cette sous-section explique de manière détaillée la méthodologie utilisée pour récupérer l'information sur les jobs menacés par l'automatisation dans les bases de données d'offres d'emploi de 2019 et 2023. La première base de données[2] contient 30 000 offres d'emploi postées sur Indeed aux USA entre le 01/08/2019 et le 31/10/2019. La seconde[3] contient des informations sur 23349 annonces de jobs postées sur LinkedIn aux USA entre le 22/07/2023 et le 23/07/2023. Les deux bases de données (BDD) sont disponibles publiquement sur Kaggle. Kaggle est une plateforme en ligne interactive qui met à disposition gratuitement des ensembles

[2] Source de la base de données 2019 : https://www.kaggle.com/datasets/promptcloud/indeed-job-posting-dataset

[3] Source de la base de données 2023 : https://www.kaggle.com/datasets/rajatraj0502/linkedin-job-2023

de données pour aider les scientifiques à mener à bien leurs projets d'apprentissage automatique. Les deux bases de données ont été utilisées suivant la même méthodologie.

Les deux BDD contiennent pour chaque offre d'emploi : le titre du job offert, le nom de la compagnie, le secteur et la description de l'offre. Nous utiliserons ici principalement cette dernière. La précision de la description de l'offre varie en fonction de la compagnie. En effet, certaines d'entre elles vont fournir par exemple, des informations sur la localisation, sur le salaire, sur l'équipe ou sur le type d'emploi. Pour rappel, nous essayons ici, de comprendre quels sont les jobs/compétences/processus que les entreprises cherchent à automatiser. Pour extraire cette information des descriptions de jobs, plusieurs techniques ont été explorées.

La première est la recherche de mots-clés. L'utilisation d'une liste de mots-clés liés à l'automatisation tels que « automation » ou « automate » pour filtrer les offres d'emploi intéressantes a été explorée. Le problème qui est apparu avec cette méthode est que toute la description du job contenant le mot-clé était extraite alors que souvent seule une partie de la description concerne l'automatisation.

Une autre méthode consiste à associer des règles à la recherche de mots-clés. Dans ce cas-ci, la règle était d'extraire uniquement la phrase contenant le mot « automation » ou les mots suivant le verbe « automate ».

Après avoir extrait les informations des offres d'emploi, beaucoup de mots inutiles subsistaient. Plusieurs méthodes de nettoyage des données ont été utilisées. L'utilisation des expressions régulières (*regex*) est une technique souvent utilisée dans le processus de nettoyage. Les expressions régulières sont à première vue de simples chaînes de caractères, mais une fois interprétées, elles décrivent alors un plus vaste ensemble de chaînes de caractères. Elles permettent de trouver ou de remplacer un motif précis. Dans ce cas-ci, elles ont été utilisées pour trouver et retirer tous les caractères spéciaux ainsi qu'une liste de mots présents dans la plupart des annonces et inutiles à notre recherche comme « job », « title » ou encore « summary ». Ensuite, les *stopwords* ont également été supprimés. Les *stopwords* sont des mots courants (comme "and", "of", "or") qui sont presque systématiquement filtrés dans les processus NLP car ils n'apportent généralement pas de valeur significative à l'analyse du texte.

La dernière méthode utilisée pour nettoyer les données est appelée « Named Entity Recognition » (NER). NER est une technique NLP qui vise à identifier et classifier les entités dans un texte non structuré. Elle permet d'extraire automatiquement des informations structurées, comme les noms de personnes, les organisations, les lieux, etc. Son objectif est

d'identifier et de catégoriser les entités mentionnées dans le texte selon des catégories prédéfinies. Dans notre cas, cette méthode a été utilisée pour identifier et supprimer toutes les localisations et les organisations. En effet, comme expliqué précédemment, beaucoup d'employeurs spécifient le nom de la ville ou de l'état et le nom de leur organisation dans l'offre d'emploi et ces noms constituent de l'information inutile pour notre recherche.

Après l'utilisation de ces différentes méthodes d'extraction et de nettoyage, un dataset composé uniquement d'offres d'emploi liées à l'automatisation et nettoyées de tous mots inutiles a été obtenu. Voici un exemple de phrases contenues dans le nouveau dataset : « reports dashboards allow business owners greater ability optimize performance key ». Cette offre est en lien avec une demande d'automatisation de dashboards.

3.1.2. Vectorisation des descriptions de jobs liées à l'automatisation

Après avoir nettoyé les données textuelles, il est essentiel de les vectoriser pour pouvoir les utiliser dans une analyse de clustering. Plusieurs méthodes de vectorisation existent telles que TF-IDF, Word2Vec, GloVe, Doc2Vec, FastText. Le choix de la méthode appropriée dépend principalement des tâches NLP, des caractéristiques des données et des ressources computationnelles (Lukauskas et al., 2023).

Une méthode importante est celle des "sentence transformers". Introduits par Reimers et Gurevych en 2019, ces transformers utilisent des modèles de transformateurs pré-entraînés tels que BERT pour générer des vecteurs denses représentant des phrases. Ces vecteurs capturent la signification sémantique du texte et peuvent être utilisés pour diverses tâches, telles que la classification de textes, la similarité sémantique et le clustering. Les sentence transformers peuvent saisir des structures linguistiques complexes et les relations textuelles, améliorant ainsi les performances sur diverses tâches de NLP. BERT, un modèle de transformateur pré-entraîné, utilise des mécanismes d'auto-attention pour capturer des informations contextuelles dans les deux directions (de gauche à droite et de droite à gauche). Ces modèles de transformateurs pré-entraînés permettent de capturer le contexte et l'ordre des mots au niveau des phrases ou des documents. Dans de nombreuses tâches de NLP, ils surpassent souvent les méthodes traditionnelles telles que TF-IDF et Word2Vec (Lukauskas et al., 2023).

Sentence-BERT (SBERT) est une modification de BERT qui utilise des structures de réseaux siamois et triplets pour obtenir des *embeddings* de phrases sémantiquement significatifs,

comparables à l'aide de la similarité cosinus. Cette approche réduit considérablement le temps et les ressources nécessaires par rapport à BERT, tout en conservant sa précision intrinsèque (Reimers & Gurevych, 2019). Reimers et Gurevych (2019) ont évalué SBERT sur des tâches courantes de STS et des tâches d'apprentissage par transfert, et ont constaté qu'il surpassait d'autres méthodes d'embeddings. Pour cette recherche, le modèle Sentence-BERT sera utilisé pour la vectorisation des données et plus précisément le modèle « all-MiniLM-L6-v2 ». Selon la documentation, le modèle all-mpnet-base-v2 offre une meilleure qualité, tandis que le modèle all-MiniLM-L6-v2 est 5 fois plus rapide tout en offrant une bonne qualité (Pretrained Models - Sentence-Transformers Documentation, s. d.).

3.1.3. Réduction de la dimensionnalité

La vectorisation des données peut parfois aboutir à de grandes dimensionnalités. Le modèle SBERT « all-MiniLM-L6-v2 » utilisé dans cette recherche génère des embeddings de dimension 384 (Pretrained Models - Sentence-Transformers Documentation, s. d.). Pour traiter ces données de manière adéquate, il est nécessaire de réduire leur dimensionnalité. En effet, une trop grande sparsité dans les espaces de haute dimension peut impacter négativement la qualité du clustering. La réduction de dimensionnalité consiste à transformer des données de haute dimension en une représentation significative de dimension réduite (Van Der Maaten et al., 2009).

Deux méthodes de réduction de dimensionnalité ont été explorées dans cette étude. La première est la méthode traditionnelle de l'analyse en composantes principales (PCA). PCA effectue cette réduction en incorporant les données dans un sous-espace linéaire de dimension inférieure et est de loin la technique linéaire de réduction de dimensionnalité la plus populaire (Van Der Maaten et al., 2009). Elle permet à la fois de maximiser la variance et de minimiser la perte d'informations (Lukauskas et al., 2023).

La deuxième méthode testée est la méthode de réduction de dimensionnalité UMAP (Uniform Manifold Approximation and Projection). L'algorithme UMAP est compétitif en termes de qualité de visualisation avec t-SNE et préserve davantage la structure globale avec des performances de temps d'exécution supérieures (McInnes et al.,2020). UMAP crée des représentations floues topologiques des données de haute dimension et les projette dans un espace de dimension inférieure. Cela signifie qu'il ne cherche pas à obtenir une projection précise des données, mais plutôt une représentation qui capture les relations importantes entre les points tout en permettant une certaine flexibilité. Les points proches dans l'espace de haute

dimension sont également proches les uns des autres dans la projection de basse dimension. Par rapport à PCA, UMAP est souvent plus rapide et offre la possibilité d'utiliser une variété de mesures de distance, incluant des métriques non euclidiennes qui permettent de saisir des relations de données complexes. La visualisation et l'interprétation des données sont facilitées par la préservation des structures locales et globales par UMAP (Lukauskas et al., 2023).

Les deux méthodes ont été testées dans ce travail et différentes métriques d'évaluation sont présentées dans la partie « résultats ».

3.1.4. Clustering des jobs menacés par l'automatisation

Pour mettre en évidence les groupes de jobs menacés par l'automatisation, un clustering a été effectué sur les descriptions d'offres d'emploi de 2019 et dans un second temps, de 2023. La méthode de clustering permet de regrouper des données similaires en ensembles distincts, afin de découvrir des structures intrinsèques ou des patterns dans les données et de faciliter leur interprétation ou leur utilisation ultérieure. Les éléments au sein d'un même groupe sont plus similaires entre eux qu'avec ceux des autres groupes. Deux méthodes de clustering ont été explorées et évaluées.

La première est le clustering K-means. L'algorithme K-Means est une méthode de clustering non supervisée, souvent utilisée en exploration de données et en reconnaissance de pattern. Il cherche à trouver K partitions qui satisfont un certain critère. Pour obtenir un résultat optimal, l'algorithme K-Means commence par sélectionner quelques points pour représenter les centres initiaux des clusters (souvent les K premiers points d'échantillon). Ensuite, les autres points d'échantillon sont regroupés autour de leurs centres respectifs selon le critère de distance minimale, ce qui donne une classification initiale. Si cette classification n'est pas satisfaisante, les centres des clusters sont recalculés et l'algorithme réitère ce processus jusqu'à obtenir une classification adéquate. L'algorithme K-Means se distingue par sa simplicité, son efficacité et sa rapidité. Cependant, le désavantage de cet algorithme est qu'il nécessite de spécifier le nombre de clusters à l'avance (Li & Wu, 2012).

La seconde technique de clustering utilisée est HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise). HDBSCAN est un algorithme de clustering basé sur la densité qui identifie les clusters en regroupant les points de données étroitement espacés. L'un des avantages par rapport à K-means est qu'il est capable de gérer efficacement les clusters de formes arbitraires et le bruit dans les données. HDBSCAN introduit également une composante hiérarchique dans le processus de clustering. Au lieu de se contenter de trouver des clusters à

une seule échelle de densité, HDBSCAN établit une hiérarchie de clusters en construisant un dendrogramme et en appliquant une méthode d'extraction basée sur la stabilité des clusters à travers différents niveaux de densité. HDBSCAN a 3 hyperparamètres : la taille minimale du cluster, le nombre minimum d'échantillons et un paramètre de sélections de clusters epsilon. Le premier détermine le nombre minimum de points nécessaires pour former un cluster dense. Le deuxième paramètre définit le nombre minimum de points dans le voisinage d'un point pour qu'il soit considéré comme un point « central » d'un cluster. Enfin, le rôle du troisième permet de gérer la tolérance de l'algorithme aux variations locales de densité lorsque les clusters sont extraits de la hiérarchie (Berba, 2021 ; Lukauskas et al., 2023).

Ces méthodes de clustering ont été appliquées et évaluées pour les deux bases de données. Les métriques d'évaluation seront présentées dans la partie « résultats ».

3.2. Démissions et transitions de jobs : Analyse LinkedIn

Cette sous-section décrit la méthodologie employée pour préparer les données de la BDD LinkedIn, afin d'analyser efficacement les mouvements professionnels qui s'effectuent à partir des types de postes menacés par l'automatisation, identifiés dans les clusters créés.

3.2.1. Préparation des données LinkedIn

La base de données[4] utilisée pour observer les démissions et les transitions contient les données LinkedIn de 10.000 profils américains choisis aléatoirement entre 1995 et 2022. Parmi celles-ci figurent le nom et le prénom, la localisation, l'éducation ainsi que les expériences professionnelles incluant pour chacune d'entre elles le titre du job, la date de début et de fin, le nom de l'entreprise et une description du job. Seules les données relatives aux expériences sont ici gardées.

Après un nettoyage de la base de données pour supprimer les éléments de texte résultant d'erreurs de scrapping, les lignes vides, les doublons et les profils sans expérience, une base de

[4] Source de la base de données : https://nubela.co/blog/sample-data-for-linkdb/

données contenant 7844 profils américains travaillant aux USA et possédant au moins une expérience professionnelle a été obtenue.

Tableau 1 : Exemple de valeurs de la base de données des profils LinkedIn

	Expérience la plus récente	Autres expériences	Expérience la plus ancienne
Profil 1	[2020, 2021, Manager, Microsoft]	…	[2015, 2017, Data analyst, Microsoft]

Note 1 : Dans l'ordre dans chaque liste : la date de début du job, la date de fin, le titre du job, l'entreprise.

A partir de là, un nouvel ensemble de données contenant l'historique entier des transitions tous profils confondus a pu être construit. Les colonnes « job1 » et « job2 » du tableau 2 représentent la transition professionnelle d'un poste initial (job1) vers un poste ultérieur (job2).

Tableau 2 : Exemple de valeurs de la table des transitions

Compagnie 1	Date de début job 1	Date de fin job 2	Job 1	Job 2	Date de début job 2	Date de fin job 2	Compagnie 2
Apple	2017	2019	Data analyst	Senior data analyst	2019	2021	Apple
Microsoft	2012	2017	Functional analyst	Data analyst	2017	2019	Apple
…	…	…	…	…	…	…	…

Dans ce travail, par souci de simplification, tous les changements d'entreprise seront considérés comme des démissions, sans prendre en compte les autres possibilités telles que les licenciements, les départs à la retraite ou les fermetures d'entreprise.

3.2.2. Normalisation des titres de jobs

A ce stade, plusieurs problèmes surviennent avant de pouvoir analyser les démissions. Premièrement, aucune donnée sur les compétences des jobs ni sur les secteurs auxquels ils appartiennent n'est disponible. Deuxièmement, les intitulés des postes sont encore très bruts et ne suivent aucune structure commune, car aucun format prédéfini n'est imposé lors de la saisie des intitulés de poste sur LinkedIn.

Pour résoudre ces problèmes, les titres de jobs ont été normalisés en utilisant la méthode présentée par Decorte et al. dans leur papier publié en 2021. Dans celui-ci, les auteurs présentent

une modification du modèle BERT appelée JobBERT, qui est un modèle pré-entrainé sur les offres d'emploi et concluent que JobBERT performe mieux que Sentence-BERT (SBERT) et les autres modèles non adaptés. Leur modèle a été entraîné pour pouvoir associer au mieux les intitulés de postes à des intitulés standardisés selon la classification ESCO (European Skills, Competences, Qualifications and Occupations). ESCO est une classification européenne qui standardise les intitulés de métiers, les compétences, et les qualifications pour faciliter la comparabilité et la reconnaissance des compétences à travers l'Europe. Dans notre cas, les intitulés de postes ESCO, ainsi que leurs compétences et secteurs associés, ont été directement extraits de la page web ESCO en utilisant le module Python BeautifulSoup.

Les intitulés des postes présents dans notre dataset de transitions et les intitulés standardisés ont ensuite été vectorisés à l'aide du modèle JobBERT. Chaque poste a été associé à l'intitulé standardisé le plus similaire en utilisant la similarité cosinus. Seules les associations ayant un score de similarité supérieur à un seuil spécifique ont été gardées. Ce seuil sera discuté dans la section « résultats ».

Étant donné que les créateurs de JobBERT ont déjà évalué et prouvé la capacité de leur modèle à associer correctement les intitulés de postes aux intitulés standardisés, la qualité de l'embedding de JobBERT ne sera pas à nouveau évaluée dans cette étude, puisque la même tâche de normalisation est effectuée. De plus, le seul ensemble de données d'évaluation trouvé est celui mis à disposition par les auteurs de JobBERT eux-mêmes.

3.2.3. Similarité entre les jobs

Pour tester la première hypothèse selon laquelle l'automatisation a entraîné davantage de changements de carrière, un moyen de calculer la similarité entre deux postes doit être défini. Cela permettra d'établir un seuil en dessous duquel une transition sera considérée comme un changement radical de carrière.

Pour effectuer cette tâche, le modèle JobBERT va de nouveau être utilisé pour vectoriser chaque job impliqué dans une transition. La similarité entre les embeddings peut ensuite être calculée en utilisant la similarité cosinus.

3.3. Analyse des démissions dans les secteurs automatisés

Pour analyser les changements de carrière dans les secteurs menacés par l'automatisation, la base de données des transitions a été filtrée successivement sur chaque cluster identifié, puis l'évolution des changements radicaux de carrière au fil du temps a été analysée.

Pour filtrer sur le cluster approprié, deux méthodes ont été utilisées. La première est la recherche de mots-clés. Pour chaque cluster une liste de mots-clés est construite. Par exemple, si le cluster concerne les jobs liés à l'analyse de données, la liste sera composée de mots comme « analysis», « analyst » ou encore « dashboards ». Ensuite, seules les transitions dont le titre ou les compétences du job quitté contiennent un de ces mots-clés sont gardées.

La deuxième méthode consiste à filtrer directement les transitions en se basant sur le secteur du poste quitté, dans le cas où le cluster analysé peut être associé à un secteur entier, par exemple « finance ».

4. Résultats

Dans cette section, les principaux résultats obtenus dans cette étude seront présentés.

La première sous-section portera sur l'analyse des bases de données d'offres d'emploi. Elle commencera par un examen des résultats de l'analyse exploratoire et des premières informations recueillies. Ensuite, les résultats des différentes méthodes d'extraction seront passés en revue, et la méthode finalement choisie sera présentée, tant pour la base de données de 2019 que celle de 2023. Les résultats des différentes techniques de réduction de dimensionnalité appliquées aux données de 2019 et 2023 seront ensuite exposés. Nous poursuivrons en présentant les résultats de l'évaluation de deux méthodes de clustering pour 2019 et 2023. Enfin, pour conclure cette partie, les résultats des clusters seront analysés et interprétés pour les deux années étudiées.

La seconde sous-section exposera les résultats de la préparation des données LinkedIn. Elle commencera par une brève analyse exploratoire, suivie d'un examen des résultats de la normalisation. Enfin, les résultats du calcul de similarité entre les emplois seront présentés.

Pour conclure cette section, les résultats combinés des deux sous-sections précédentes seront présentés avant d'être discutés.

4.1. Résultats de l'analyse des offres d'emploi

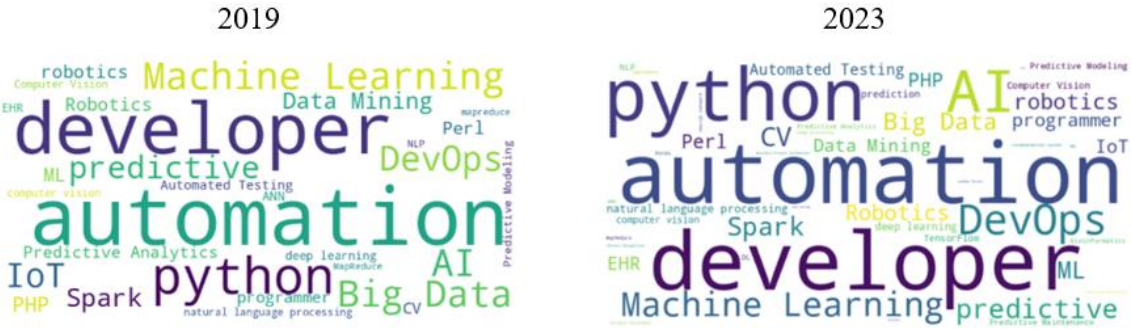
4.1.1. Analyse exploratoire des offres d'emploi

Cette section présente les différents résultats obtenus lors de l'analyse exploratoire des données d'offres d'emploi de 2019 et 2023. Elle offre un aperçu des tendances globales du marché de l'emploi au cours de ces deux années. Les nuages de mots suivants nous montrent les jobs les plus recherchés. Pour les deux années, les postes d'ingénieur et d'analyste sont très demandés, ce qui vient confirmer le manque de personnels qualifiés en analyse de données et en compétences informatiques.

Figure 2 : Nuages de mots des emplois offerts

Ensuite, les compétences en IA les plus demandées ont été identifiées. Pour les deux années examinées, la compétence en automatisation arrive en tête. Nous pouvons également constater que le langage Python a gagné en importance.

Figure 3 : Nuages de mots des compétences IA



4.1.2. Résultats de l'extraction des données

Pour rappel, les deux méthodes d'extraction des données testées consistaient soit à extraire toutes les offres contenant certains mots-clés, soit à extraire uniquement les mots de la phrase suivant des règles spécifiques. Après avoir comparé les résultats des deux techniques, la méthode basée sur les règles spécifiques a été sélectionnée. En effet, cette méthode nous a permis de sélectionner uniquement la partie de la description de job contenant les informations sur le but de l'automatisation.

4.1.3. Résultats de la réduction de la dimensionnalité

Après avoir vectorisé les descriptions extraites avec SBERT, la dimension des données était de 384. Dans cette sous-section, les résultats de l'évaluation des différentes méthodes utilisées pour réduire la dimensionnalité seront présentés. Pour évaluer ces méthodes, la métrique de « trustworthiness » présentée et utilisée par Lukauskas et al. (2023) dans leur papier sera employée. Cette métrique évalue dans quelle mesure les informations et les relations entre les observations dans les dimensions originales et réduites sont préservées. Dans notre étude cette métrique a été calculée en utilisant la fonction « trustworthiness » de la bibliothèque scikit-learn. Les deux méthodes ont été évaluées sur plusieurs valeurs d'hyper-paramètres.

Tableau 3 : Evaluation de la trustworthiness pour les données de 2019

COMPOSANTES	PCA	UMAP
2	0.699	0.870
3	0.744	0.921
4	0.793	0.931
5	0.832	0.941
6	0.855	0.949
7	0.876	0.947
8	0.896	0.949
9	0.909	0.950
10	0.924	0.949
15	0.946	0.951
20	0.962	0.949
25	0.973	**0.953**
30	0.978	0.950
35	0.985	0.949
40	0.987	0.951
50	**0.991**	0.949

Pour l'analyse de 2019, la PCA performe le mieux à 50 composantes tandis que UMAP performe mieux à 25 composantes. Les résultats de l'évaluation indiquent que la PCA est plus performante. Cependant, les résultats du clustering ont montré que l'interprétation était plus facile avec UMAP. Par conséquent, nous avons décidé de conserver UMAP avec 25 composantes.

Tableau 4 : Evaluation de la trustworthiness pour les données de 2023

COMPOSANTES	PCA	UMAP
2	0.712	0.869
3	0.761	0.909
4	0.808	0.922
5	0.836	0.932
6	0.859	0.935
7	0.881	0.937
8	0.887	0.939
9	0.904	0.942
10	0.912	0.938
15	0.939	**0.944**
20	0.961	0.939
25	0.970	0.939
30	0.976	0.938
35	0.984	0.940
40	0.986	0.943
50	**0.992**	0.941

Pour les données de 2023, la PCA offre les meilleures performances avec 50 composantes, tandis que UMAP excelle avec 15 composantes. Bien que la PCA soit globalement plus performante que UMAP, nous avons décidé de choisir UMAP pour les mêmes raisons d'interprétabilité mentionnées précédemment.

4.1.4. Résultats de l'évaluation du clustering

Dans cette sous-section, les résultats de l'évaluation des deux méthodes de clustering, à savoir K-means et HDBSCAN, seront présentés.

Dans les tâches supervisées, diverses méthodes peuvent être utilisées pour optimiser les caractéristiques d'entrée selon l'objectif de la tâche, comme la précision de la classification. En revanche, dans les problèmes non supervisés, ces outils sont rarement disponibles, en partie parce qu'il est difficile de mesurer la pertinence des caractéristiques sans étiquettes. Dans leur étude, McCrory et Thomas (2024) montrent que le Coefficient de Silhouette et le score de Davies-Bouldin sont les plus sensibles aux caractéristiques ajoutées et non pertinentes. Ainsi, le Coefficient de Silhouette et le score de Davies-Bouldin sont de bons candidats pour évaluer la sélection des caractéristiques dans les tâches de clustering non supervisées. Le Coefficient de Silhouette évalue les résultats du clustering en supposant que l'objectif souhaité est d'obtenir

des clusters denses et bien séparés. Il est défini en utilisant la distance moyenne entre un point et tous les points du même cluster, ainsi que la distance moyenne entre ce point et tous les autres points du cluster le plus proche. Une plus grande valeur de coefficient de silhouette indique une meilleure qualité de clustering. Il varie entre -1 et 1. L'indice de Davies-Bouldin compare la similarité de chaque cluster avec le cluster le plus similaire dans l'ensemble de données, en faisant la moyenne sur tous les clusters k. Il est calculé en utilisant la distance moyenne de tous les points au centroïde au sein de chaque cluster, ainsi que la distance entre les centroïdes des paires de clusters. Le score parfait est de 0 et signifie que les clusters sont bien séparés, denses et clairement définis et il n'a pas de limite supérieure (McCrory & Thomas, 2024).

Une troisième métrique d'évaluation a été utilisée : l'index de Calinski–Harabasz. Celui-ci évalue la qualité du clustering en comparant le rapport entre la dispersion inter-clusters et la dispersion intra-cluster. Une valeur élevée de l'indice CH indique un clustering de meilleure qualité, car elle reflète une plus grande séparation entre les clusters par rapport à la dispersion à l'intérieur de ceux-ci (Lukauskas et al., 2023).

Différents hyper-paramètres ont été testés pour chaque algorithme. Pour la méthode K-means, plusieurs tailles de clusters ont été explorées et pour HDBSCAN différentes variations des hyper-paramètres présentés plutôt ont été examinées. Pour le calcul, les trois métriques ont été importées depuis scikit-learn. Dans les tableaux ci-dessous, les meilleures combinaisons de paramètres par méthode sont reprises.

Tableau 5 : Evaluation du clustering des données de 2019

Méthode	Paramètres	Davies-Bouldin	Silhouette	Calinski-Harabasz
K-means	n_clusters : 5	1.18	0.33	498.16
	n_clusters : 10	1.06	0.35	479.61
	n_clusters : 20	1.08	0.34	393.10
HDBSCAN	{'min_cluster_size': 20, , 'min_samples': 20, 'cluster_selection_epsilon': 0.0}	0.61	**0.56**	643.49
	{'min_cluster_size': 30, 'min_samples': 30, 'cluster_selection_epsilon': 0.0}	0.67	0.55	**762.52**
	{'min_cluster_size': 50, 'min_samples': 5, 'cluster_selection_epsilon': 0.2}	**0.56**	0.45	487.07

Pour les données de 2019, HDBSCAN performe le mieux sur toutes les métriques. La combinaison de paramètres qui performe le mieux en moyenne sur les 3 métriques est HDBSCAN{'min_cluster_size': 20, 'min_samples': 20, 'cluster_selection_epsilon': 0.0}.

Tableau 6 : Evaluation du clustering des données de 2023

Méthode	Paramètres	Davies-Bouldin	Silhouette	Calinski-Harabasz
K-means	n_clusters : 5	1.24	0.31	314.03
	n_clusters : 10	1.18	0.29	243.97
	n_clusters : 20	1.15	0.28	200.78
HDBSCAN	{'min_cluster_size': 50, , 'min_samples': 50, 'cluster_selection_epsilon': 0.2}	**0.62**	0.57	371.01
	{'min_cluster_size': 30, 'min_samples': 30, 'cluster_selection_epsilon': 0.0}	0.70	**0.58**	**372.52**

Pour les données de 2023, HDBSCAN performe également le mieux sur toutes les métriques. La combinaison de paramètres qui performe le mieux en moyenne sur les 3 métriques est HDBSCAN{'min_cluster_size': 50, 'min_samples': 50, 'cluster_selection_epsilon': 0.2}.

4.1.5. Résultats du clustering

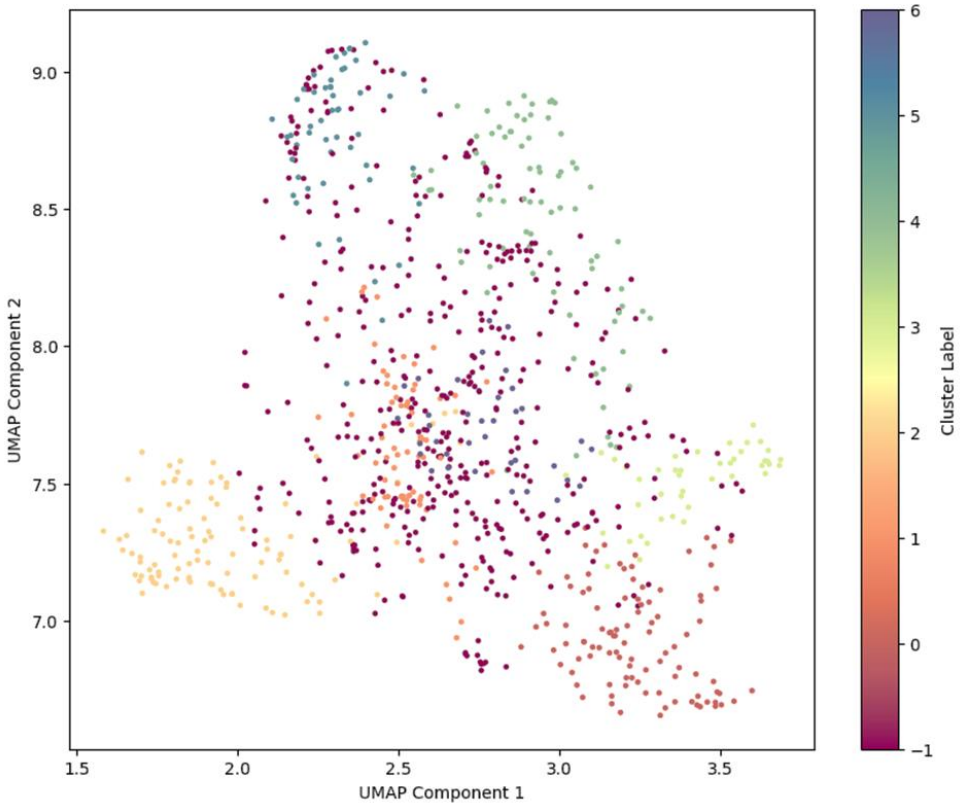
Dans cette section, les résultats de notre clustering visant à extraire les emplois menacés par l'automatisation à partir d'un corpus de documents d'offres d'emploi pour 2019 et 2023 seront présentés.

À la suite du clustering sur le corpus de 2019, 7 clusters ont été identifiés, ainsi qu'un cluster de bruit. La figure ci-dessous illustre les différents clusters, où le cluster -1 représente le cluster de bruit. En analysant les différentes offres d'emploi de chaque cluster, il a été possible de les étiqueter comme suit :

- Cluster 0 : Automatisation des tests.
- Cluster 1 : Automatisation de la finance et de la comptabilité.
- Cluster 2 : Automatisation de la conception et de la gestion des systèmes.
- Cluster 3 : Automatisation des services et de la gestion du cloud computing.
- Cluster 4 : Automatisation de l'analyse des données et des dashboards.
- Cluster 5 : Automatisation du marketing et de la gestion de la relation client.
- Cluster 6 : Automatisation de l'optimisation et l'intégration des flux de travail.

Ces différents clusters représentent les diverses catégories de tâches que les employeurs cherchaient à automatiser en 2019. Par conséquent, tous les emplois associés à ces clusters étaient potentiellement menacés par l'automatisation cette année-là et les suivantes. Un aperçu des valeurs de chaque cluster se trouve en annexe 7.1.

Figure 4 : Visualisation du clustering HDBSCAN de 2019

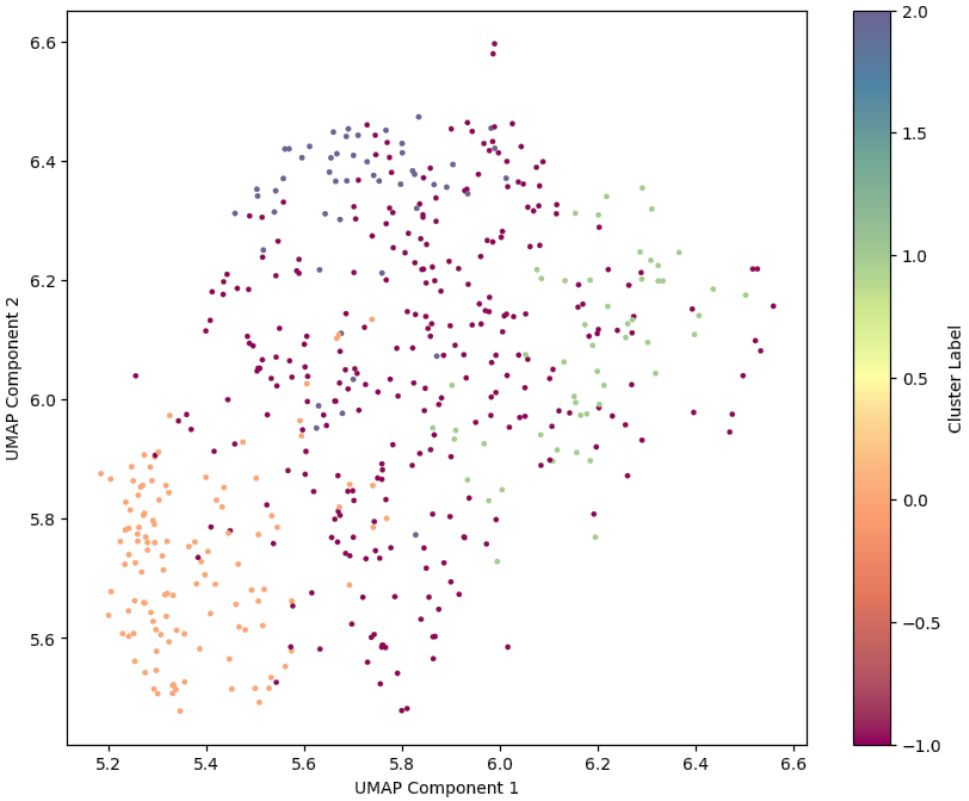


Pour le clustering sur le corpus de 2023, 3 clusters ont été identifiés, ainsi qu'un cluster de bruit. La figure ci-dessous illustre ces différents clusters, où le cluster -1 correspond au cluster de bruit. L'analyse des différentes offres d'emploi de chaque cluster a permis de les étiqueter comme suit.

- Cluster 0 : Automatisation des tests.
- Cluster 1 : Automatisation de la gestion financière et de l'analyse de données.
- Cluster 2 : Automatisation industrielle.

Ces différents clusters représentent les différentes catégories de tâches que les employeurs envisageaient d'automatiser en 2023. Cette demande d'automatisation visait potentiellement à aider les entreprises à faire face à la pénurie de travailleurs qualifiés pour ces tâches. Un exemple des valeurs de chaque cluster se trouve en annexe 7.2.

Figure 5 : Visualisation du clustering HDBSCAN de 2023

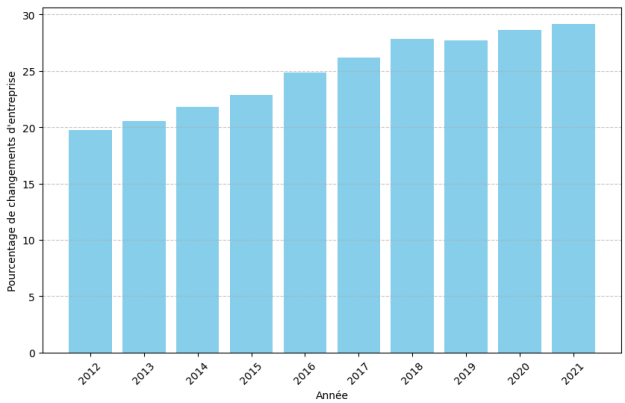


4.2. Résultats de la préparation des données LinkedIn

4.2.1. Analyse exploratoire des données LinkedIn

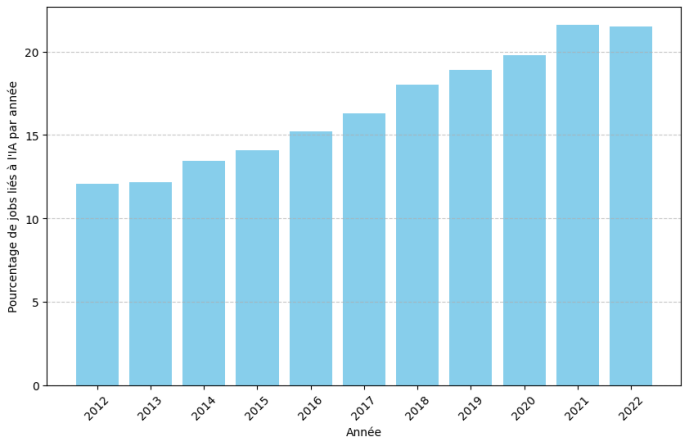
Cette première sous-section a pour but de présenter les tendances générales qui ont pu être analysées lors de l'analyse exploratoire de la base de données LinkedIn. Premièrement, une augmentation des démissions peut être observée au fil du temps (Figure 6).

Figure 6 : Pourcentage de démissions par année



Ensuite, en effectuant une recherche de mots-clés spécifiques dans les titres des emplois, le pourcentage d'emplois liés à l'IA par année a été obtenu (Figure 7). Il ressort que de plus en plus d'emplois sont liés de près ou de loin à l'IA.

Figure 7 : Pourcentage d'emplois liés à l'IA par année



4.2.2. Résultats de la normalisation des titres de jobs

Cette section va présenter les résultats obtenus en normalisant les titres de jobs avec la méthode JobBERT. Le tableau ci-dessous montre quelques exemples de normalisation.

Tableau 7 : Exemples d'emplois normalisés

Titre de job	Job standardisé (ESCO)	Similarité
TD Process integration engineer	integration engineer	0.834
Graduate research assistant	university research assistant	0.859
Store Cashier	cashier	0.831
Graduate Student Management Information Systems	IT auditor	0.543

Pour éviter de fausser les futures analyses, la décision a été prise de ne conserver que les données ayant un score de similarité supérieur à un certain seuil. En effet, il n'est pas toujours clair dans quelle mesure la valeur de similarité peut être interprétée et refléter une réelle similarité. Dans leur papier, Rekabsaz et al. (2018) proposent un seuil général, indépendant des termes des documents et du modèle d'embedding, pour évaluer la similarité (uniquement pour la fonction cosinus). En se basant sur les conclusions de leur étude, un seuil de similarité de 0.708 a été choisi. Ce seuil est défini comme le seuil général pour les embeddings de dimension 300, ce qui correspond à la dimension de sortie des embeddings de JobBERT.

4.2.3. Similarité entre les jobs

Cette section expose les résultats issus du calcul de la similarité entre les postes pour identifier les changements de carrière radicaux.

Pour évaluer la qualité de nos similarités, plusieurs tests ont été réalisés en partant du principe que deux emplois appartenant au même secteur devraient être plus similaires que deux emplois de secteurs différents. Lors de la normalisation, chaque job s'est vu attribuer 5 niveaux de classification allant du plus général au plus détaillé. Par exemple, pour un conseiller en banque d'entreprise, les 5 niveaux sont « Professionals », « Business and administration professionals », « Finance professionals », « Financial and investment advisers » et « corporate banking adviser ». Au plus deux jobs ont des niveaux en communs, au plus ils doivent être similaires et inversement. Le tableau suivant montre que ce principe a été respecté dans notre calcul de similarité. Le score moyen de similarité augmente avec le nombre de niveaux de classification que deux emplois ont en commun. Lorsqu'ils ne partagent que 2 niveaux, leur similarité moyenne est de 0.646. En revanche, lorsqu'ils partagent 4 niveaux, cette similarité moyenne s'élève à 0.76.

Tableau 8 : Similarité moyenne en fonction du nombre de classifications communes

Nombre de classifications communes	Similarité moyenne
0	0.591
1	0.608
2	0.646
3	0.686
4	0.760
5	0.980

De nouveau, un seuil a dû être choisi pour établir à partir de quand une transition était considérée comme radicale. Bien que le seuil de 0.708 ait pu être réutilisé, un seuil de 0.608 a été choisi pour refléter au mieux le principe présenté dans le tableau (8). Ce seuil représente la similarité moyenne des transitions partageant la même classification générale. En partant de ce principe, une transition est considérée comme brutale lorsque le nouveau poste et l'ancien ne partagent pas la même classification générale. Ce choix peut néanmoins être discuté et être envisagé comme une possibilité d'amélioration pour de futures recherches. Le tableau ci-dessous présente des exemples de transitions de postes et leur score de similarité. Chacun des trois exemples illustre un changement d'entreprise, mais dans des contextes différents. Dans la

première transition, le score de similarité est de 1, ce qui signifie que les postes sont parfaitement identiques. Dans le deuxième exemple, le score de 0.89 indique une transition vers un poste très similaire. Enfin, dans la troisième situation, le score de 0.47 montre que les deux postes ne sont pas similaires.

Tableau 9 : Exemples de similarité entre les emplois

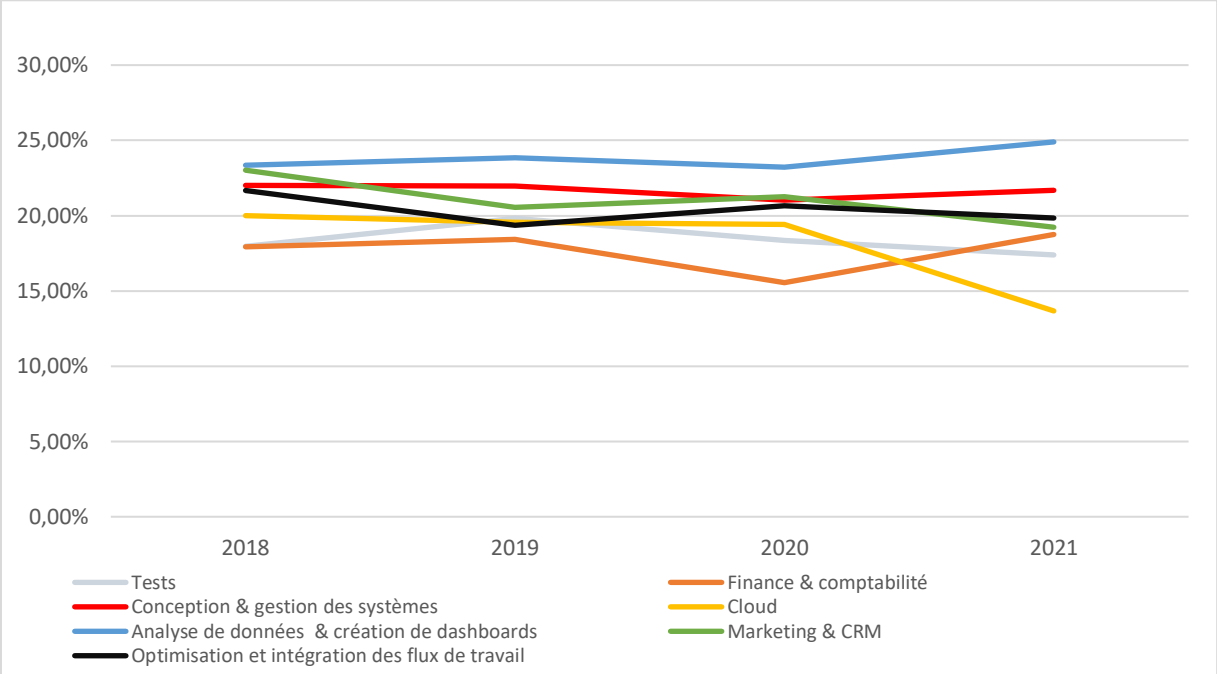
Job 1 (standardisé)	Compagnie 1	Job 2 (standardisé)	Compagnie 2	Similarité
product and services manager	Pebble Tech	product and services manager	Tesla	1.0
software architect	Partsearch Technologies	software developer	Munich Re (Group)	0.89
cashier	Whole Foods Market	secretary	Placer Title Company	0.47

4.3. Relation de cause à effet entre automatisation et démissions

Cette section va présenter les résultats de la synthèse des 2 principales analyses. En effet, les principaux jobs menacés par l'automatisation ayant été identifiés dans les clusters, ils ont pu être observés dans l'ensemble de données des démissions de LinkedIn.

La figure 8 montre le pourcentage de démissions au cours du temps pour chacun des clusters identifiés. Les emplois liés à l'analyse de données ainsi qu'à la finance et à la comptabilité ont connu une chute des démissions en 2020, suivie d'une forte remontée en 2021, dépassant le pourcentage observé avant la pandémie de COVID-19. Les emplois liés à l'optimisation des flux de travail ainsi qu'au marketing et CRM ont connu un pic de démissions en 2020, suivi d'une diminution en 2021, revenant à des niveaux proches de ceux d'avant 2020. Le pourcentage de démissions des travailleurs liés aux tests et au cloud n'a cessé de diminuer depuis 2019. Enfin, les jobs liés à la conception et à la gestion des systèmes restent constants à travers le temps malgré une petite chute en 2020.

Figure 8 : Pourcentage de démissions par cluster et par année

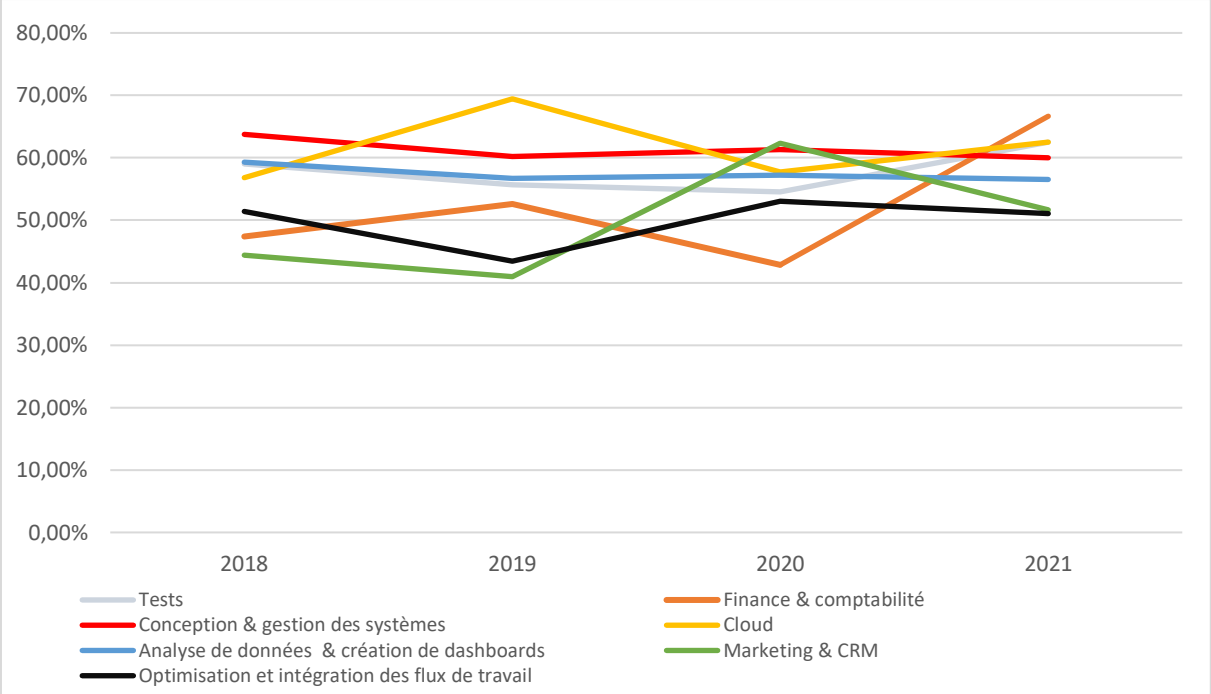


La figure suivante nous montre le pourcentage de démissions qui sont considérées comme radicales, à travers le temps. Les résultats montrent que le pourcentage de changements de carrières est stable pour les métiers liés à l'analyse de données et à la conception de systèmes. La première hypothèse selon laquelle l'automatisation a augmenté le nombre de changements radicaux de carrière pendant la Grande Démission peut donc déjà être réfutée pour les secteurs de la conception de systèmes et de l'analyse de données. Les données montrent que les pourcentages de changements de carrière dans ces secteurs sont restés relativement stables ou ont même diminué en 2021 par rapport à 2020, suggérant que l'automatisation n'a pas conduit à une augmentation notable des changements radicaux de carrière dans ces domaines pendant cette période. Pour l'optimisation des flux de travail, le marketing et le CRM, le pourcentage augmente en 2020 et rediminue en 2021 sans pour autant revenir à ses valeurs pré-pandémie. Les résultats de l'étude Nazareno et Schiff (2021) apportent peut-être un début d'hypothèse pour expliquer la chute des démissions dans certains secteurs fortement impactés par l'automatisation. Dans celle-ci, ils concluent que l'utilisation de l'IA peut réduire le stress des employés.

Il est intéressant de noter la forte augmentation des changements de carrière dans les domaines de la finance et de la comptabilité. En 2021, le pourcentage de changements radicaux y atteint 67 %, tandis que la moyenne des autres secteurs (hors clusters) est de 58 %. En conclusion, la

première hypothèse peut être rejetée pour tous les secteurs menacés par l'automatisation sauf pour le secteur financier et comptable.

Figure 9 : Pourcentage de changements de carrière par année



Pour conclure, les résultats présentés en figure 8 et 9 rejoignent les résultats du clustering de 2023. En effet, l'augmentation du nombre de démissions des analystes de données et l'augmentation du taux de changements de carrière dans le secteur de la finance et de la comptabilité peuvent être rattachées aux demandes d'automatisations identifiées dans le cluster «Automatisation de la gestion financière et de l'analyse de données » de 2023. D'autres résultats nous montrent d'ailleurs que la majorité des démissions de 2021 viennent de « Software and applications developers and analysts ». Parmi les demandes d'automatisation en 2023 figurait également l'automatisation des tests, qui peuvent être reliées à l'augmentation des changements de carrière dans ce type d'emplois. La deuxième hypothèse peut donc être validée.

5. Conclusion générale

5.1. Méthodologie et résultats de l'étude

L'objectif de ce mémoire était d'étudier la relation entre la « Grande Démission » et l'augmentation de l'automatisation. Les deux hypothèses centrales de cette étude étaient les suivantes : d'une part, l'automatisation a conduit à des changements de carrière, et d'autre part, les secteurs où la demande d'automatisation est actuellement élevée sont ceux qui ont connu d'importantes démissions et transformations de carrière durant la "Grande Démission".

Dans le but de répondre à ces hypothèses, nous avons commencé par analyser les bases de données d'offres d'emploi de 2019 et 2023 à la recherche des emplois menacés par l'automatisation. Pour ce faire, nous avons commencé par nettoyer les bases de données et en extraire uniquement les informations utiles des descriptions de jobs. Ensuite, celles-ci ont été vectorisées avec sentence-BERT et l'algorithme UMAP a été utilisé pour réduire la dimensionnalité de nos embeddings. Enfin, un clustering a été effectué pour chacune des deux années en utilisant HDBSCAN.

Le but de cette étude était d'analyser, à partir des données d'expériences LinkedIn, les démissions survenues dans ces clusters pendant la Grande Démission. Pour préparer au mieux nos données pour cette analyse, nous avons d'abord normalisé les intitulés de postes en utilisant JobBERT et la similarité cosinus pour associer chaque emploi au titre ESCO le plus pertinent. Ensuite, pour chaque paire d'emplois impliqués dans une transition, nous avons calculé un score de similarité, à nouveau avec JobBERT et la similarité cosinus. Cela nous a permis de caractériser chaque transition comme radicale ou non, en fonction de son score de similarité.

Enfin, nous avons observé l'évolution des démissions au début de la Grande Démission pour chacun des clusters identifiés en 2019, ce qui nous a permis d'évaluer notre première hypothèse. Ensuite, nous avons comparé les clusters de 2023 avec les démissions et les changements de carrière survenus pendant la Grande Démission pour évaluer notre deuxième hypothèse.

Les principaux résultats de notre analyse sont :

- Sentence-BERT, la réduction de dimensionalité UMAP et l'algorithme de clustering HDBSCAN ont été identifiés comme les plus performants et les plus aptes à permettre une interprétation simple des résultats du clustering des offres d'emploi. L'intérêt de ces résultats réside dans leur capacité à fournir une méthode efficace pour analyser les emplois menacés par l'automatisation.

- Les emplois liés aux tests de vérification, à la finance, à l'analyse de données et aux industries de manufactures sont ceux que les entreprises cherchent le plus à automatiser. Ces résultats sont utiles pour les employeurs qui cherchent à comprendre les tendances qui se manifestent actuellement dans le monde du travail et les employés qui veulent s'adapter au changement. De plus, ces résultats nous prouvent que le clustering est une méthode efficace pour identifier les emplois menacés par l'automatisation à partir de données d'offres d'emploi.

- JobBERT peut être utilisé pour calculer la similarité entre deux jobs, ce qui permet de détecter les changements de carrières. Les résultats de cette méthode peuvent être utiles pour de futures recherches sur le sujet.

- Parmi les emplois menacés par l'automatisation avant le début de la Grande Démission, seuls ceux du secteur de la finance et de la comptabilité ont vu une augmentation des changements de carrière durant celle-ci.

- Les emplois que les entreprises cherchent à automatiser en 2023 ont tous connu une augmentation des démissions et des changements de carrière pendant la Grande Démission.

5.2. Limites de l'étude et possibilités d'amélioration

Avant de conclure ce mémoire, nous allons discuter des différentes limitations qui ont émergé ainsi que des différentes pistes pour de futures recherches.

5.2.1. Résultats basés sur l'observation

Nos données et méthodes d'analyse sont observationnelles, rendant l'identification de la causalité difficile. Ainsi, lorsque nous analysons les changements de carrière pour les emplois menacés par l'automatisation, nous ne pouvons être certains que d'autres facteurs n'entrent pas en jeu. Cependant, notre étude propose une méthodologie pour identifier les emplois automatisés, normaliser les intitulés de postes et déterminer les changements de carrière. Nous utilisons ces méthodes pour esquisser un lien entre la Grande Démission et l'automatisation.

Des recherches futures pourront s'appuyer sur ces résultats pour calculer la corrélation entre l'automatisation et les changements de carrière.

5.2.2. Limites géographiques et temporelles de l'étude

Notre étude se concentre sur des données de profils des Etats-Unis. Une prochaine étude pourrait donc élargir l'exploration à d'autres pays. Une autre limite de ce travail a été l'accès à des bases de données de profils LinkedIn récentes. En effet, la base de données utilisée pour analyser les expériences des profils s'arrête en 2022. Une piste d'analyse future pourrait consister à examiner les tendances postérieures à 2022, en mettant particulièrement l'accent sur la période depuis laquelle ChatGPT a été introduit et adopté par de nombreuses entreprises. Avec l'avènement de ChatGPT et d'autres technologies d'intelligence artificielle, il est probable que le paysage du marché du travail ait continué à évoluer de manière significative.

5.2.3. Différenciation entre licenciements et démissions

La méthodologie utilisée ne permet pas de différencier une démission et un licenciement. En effet, les données de profils LinkedIn ne contiennent pas les raisons du changement de jobs. Or, dans cette étude, un changement d'entreprise a été comptabilisé comme une démission, ce qui peut conduire à une surestimation du nombre réel de démissions.

5.2.4. Choix du seuil de similarité

Enfin, le seuil de similarité choisi pour caractériser les transitions de changements de carrière radicales n'a pas été évalué de manière statistiquement fiable. Une amélioration possible serait de tester et d'évaluer sur un ensemble d'évaluation la capacité de cette méthode à réellement détecter les changements radicaux de carrière. Plusieurs techniques peuvent également être explorées afin de déterminer un seuil plus fiable.

6. Bibliographie

Agarwala, S., Anagawadi, A., & Reddy Guddeti, R. M. (2021). Detecting semantic similarity of documents using natural language processing. Procedia CIRP, 189, 128-135. https://doi.org/10.1016/j.procs.2021.05.076

Alaql, A. A., AlQurashi, F., & Mehmood, R. (2023). Data-driven deep journalism to discover age dynamics in multi-generational labour markets from LinkedIn media. Journalism and Media, 4(1), 120-145. https://doi.org/10.3390/journalmedia4010010

Alotaibi, S., Mehmood, R., Katib, I., Rana, O., & Albeshri, A. (2020). Sehaa: A big data analytics tool for healthcare symptoms and diseases detection using twitter, apache spark, and machine learning. Applied Sciences, 10(4), 1398. https://doi.org/10.3390/app10041398

Alruwaili, F., & Alahmadi, D. (2021). Enhanced clustering-based MOOC recommendations using LinkedIn profiles (MR-LI). International Journal of Advanced Computer Science & Applications, 12(8), 151-160. https://doi.org/10.14569/IJACSA.2021.0120818

Arntz, M., Gregory, T., Zierahn, U., & OECD / Directorate for Employment, Labour and Social Affairs. (2016). The risk of automation for jobs in OECD countries: A comparative analysis. OECD Publishing. https://doi.org/10.1787/5jlz9h56dvq7-en

Avitzur, O. (2021). The Great Resignation: The Workforce Exodus Hits Neurology Practice and Research. Neurology Today, 21(23), 1-26.

Bacher, J.; Tamesberger, D. The Corona Generation: (Not) Finding Employment during the Pandemic. CESifo Forum 2021, 22, 3–7. 6.

Baukal, C. E. (2023, June), Training New Employees during the Great Resignation Paper presented at 2023 ASEE Annual Conference & Exposition, Baltimore , Maryland. 10.18260/1-2—44515

Berba, P. (2021, 13 décembre). Understanding HDBSCAN and Density-Based Clustering – Towards Data Science. Medium. https://towardsdatascience.com/understanding-hdbscan-and-density-based-clustering-121dbee1320e

Chowdhary, K. R. (2020). Fundamentals of artificial intelligence (1st 2020. ed.). Springer India. https://doi.org/10.1007/978-81-322-3972-7

Dai, K., Nespereira, C. G., Vilas, A. F., & Redondo, R. P. D. (2015). Scraping and clustering techniques for the characterization of linkedin profiles. https://doi.org/10.48550/arxiv.1505.00989

Dai, K., Vilas, A. F., & Redondo, R. P. D. (2018). The Workforce Analyzer : Group discovery among LinkedIn public profiles. Journal of Ambient Intelligence and Humanized Computing, 9(6), 2025-2034. https://doi.org/10.1007/s12652-017-0484-6

Decorte, J., Van Hautte, J., Demeester, T., & Develder, C. (2021). JobBERT: Understanding job titles through skills. https://doi.org/10.48550/arxiv.2109.09605

Jaffe, E., Mo, F., Jaramillo, C., Squires, P., & Togelius, J. Understanding the Great Resignation.

Kuehner-Hebert, K. (2021). How the 'Great Resignation' is accelerating the use of robotics and AI-based tools. *BenefitsPRO,* https://www.proquest.com/trade-journals/how-great-resignation-is-accelerating-use/docview/2605723812/se-2

Kundu, M. S., Das, M. S., & Nag, M. S. (2022). The Great Resignation: A Quantitative Analysis of the Factors Leading to the Phenomenon. *PROCEEDINGS BOOK*, 534.

Kuzior, A., Kettler, K., & Rabą, Ł. (2022). Great Resignation—Ethical, cultural, relational, and personal dimensions of generation Y and Z employees' engagement. Sustainability, 14(11), 6764. https://doi.org/10.3390/su14116764

Li, Y., & Wu, H. (2012). A clustering method based on K-means algorithm. Physics Procedia, 25, 1104-1109. https://doi.org/10.1016/j.phpro.2012.03.206

LinkedIn Corporation. 2022. LinkedIn HomePage. Available online: www.linkedin.com (accessed on 02 January 2023).

Liu-Lastres, B., Wen, H., & Huang, W. (2023). A reflection on the great resignation in the hospitality and tourism industry. International Journal of Contemporary Hospitality Management, 35(1), 235-249. https://doi.org/10.1108/IJCHM-05-2022-0551

Lukauskas, M., Šarkauskaitė, V., Pilinkienė, V., Stundžienė, A., Grybauskas, A., & Bruneckienė, J. (2023). Enhancing skills demand understanding through job ad segmentation using NLP and clustering techniques. Applied Sciences, 13(10), 6119. https://doi.org/10.3390/app13106119

Mahany, A., Khaled, H., Elmitwally, N. S., Aljohani, N., & Ghoniemy, S. (2022). Negation and speculation in NLP: A survey, corpora, methods, and applications. Applied Sciences, 12(10), 5209. https://doi.org/10.3390/app12105209

McCrory, M., & Thomas, S. A. (2024). Cluster metric sensitivity to irrelevant features. https://doi.org/10.48550/arxiv.2402.12008

McInnes, L., Healy, J., & Melville, J. (2020). UMAP: Uniform manifold approximation and projection for dimension reduction. https://doi.org/10.48550/arxiv.1802.03426

Martin, M. (2021). How the great resignation will shape HR and the future of work. Recognition and Engagement Excellence Essentials.

Microsoft. 2021. The Next Great Disruption Is Hybrid Work—Are We Ready? Available online: https://www.microsoft.com/en-us/ worklab/work-trend-index/hybrid-work (accessed on 30 December 2023).

Nazareno, L., & Schiff, D. S. (2021). The impact of automation and artificial intelligence on worker well-being. Technology in Society, 67, 101679. https://doi.org/10.1016/j.techsoc.2021.101679

Ng, E., & Stanton, P. (2023). Editorial: The great resignation: Managing people in a post COVID-19 pandemic world. Personnel Review, 52(2), 401-407. https://doi.org/10.1108/PR-03-2023-914

Pardue, L. (2021), "A Real-Time look at the'great resignation", available at: https://gusto.com/companynews/a-real-time-look-at-the-great-resignation (accessed on 30 December 2023).

Pretrained Models — Sentence-Transformers documentation. (s. d.). https://www.sbert.net/docs/pretrained_models.html

Rainie, L. and Anderson, J. (2017), "*The future of jobs and jobs training*", Pew Research Center, available at: https://www.pewresearch.org/internet/2017/05/03/the-future-of-jobs-and-jobs-training/

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. https://doi.org/10.48550/arxiv.1908.10084

Rekabsaz, N., Lupu, M., & Hanbury, A. (2018). Uncertainty in neural network word embedding : Exploration of threshold for similarity. https://doi.org/10.48550/arxiv.1606.06086

Serenko, A. (2023). The great resignation: The great knowledge exodus or the onset of the great knowledge revolution? Journal of Knowledge Management, 27(4), 1042-1055. https://doi.org/10.1108/JKM-12-2021-0920

Serenko, A. (2024). The human capital management perspective on quiet quitting: Recommendations for employees, managers, and national policymakers. Journal of Knowledge Management, 28(1), 27-43. https://doi.org/10.1108/JKM-10-2022-0792

Skills, not job titles, are the new metric for the labour market. (2020, 8 février). World Economic Forum. https://www.weforum.org/agenda/2019/07/skills-not-job-titles-are-the-new-metric-for-the-labour-market/

Squicciarini, M., & Nachtigall, H. (2021). Demand for AI skills in jobs: Evidence from online job postings. (No. 2021). Paris: Organisation for Economic Cooperation and Development (OECD). https://doi.org/10.1787/3ed32d94-en

Statista (2022), "Median age of the global labor force from 1990 to 2025", available at: https://www.statista.com/statistics/996530/median-age-global-labor-force-years/

Sull, D., Sull, C., & Zweig, B. (2022). Toxic culture is driving the great resignation. MIT Sloan Management Review, 63(2), 1-9.

Suma, S., Mehmood, R., & Albeshri, A. (2020). Automatic detection and validation of smart City events using HPC and Apache Spark platforms. Dans EAI/Springer Innovations in Communication and Computing (p. 55-78). https://doi.org/10.1007/978-3-030-13705-2_3

US Bureau of Labor (2022a), "Quit levels and rates by industry and region, seasonality adjusted", available at: www.bls.gov

Van Dam, A. (2021,). The latest twist in the 'great resignation': Retiring but delaying social security. The Washington Post

Van Der Maaten, L., Postma, E., & Van den Herik, J. (2009). Dimensionality reduction: a comparative. J Mach Learn Res, 10(66-71).

Verma, A., Lamsal, K., & Verma, P. (2022). An investigation of skill requirements in artificial intelligence and machine learning job advertisements. Industry & Higher Education, 36(1), 63-73. https://doi.org/10.1177/0950422221990990

YPulse (2021), “Why the great resignation is happening in Western Europe too”, available at: www.ypulse.com/article/2021/11/02/why-the-great-resignation-is-happening-in-western-europe-too

Zagorsky, J. L. (2022). The great resignation: Historical data and a deeper analysis show it’s not as great as screaming headlines suggest. The Conversation U.S

7. Annexes

7.1. Exemples de valeurs du clustering des offres d'emploi de 2019

Cluster 0 : Tests

testing debugging deploymentparticipate productive member agile team engineers produce high quality
testing tooling solutions primary focus automated functional regression testing projects
test engineers identify manual regression tests candidates automationmaintain comprehensive suite
testing frameworkscripting incl business rules script includes ui policies ui
solutions real world testing problems ll opportunity significant impact success
product test apt customer success 6 months build application knowledge
testing work define execute right level performance testing based release
test script development executing service test scripts reporting tracking defects
test suitesidentify opportunities recommend solutions improving services efficiency effectiveness abreast

Cluster 1: Finance et comptabilité

general ledger purchasing systems determine suitable cost effective purchasing methods
labor scheduling monthly status report quarterly business review cash management
solutions ability automate manual processes experiences include limited billing collection
accounts payable systemshyland onbase peoplesoft ap experience advantageous h2 class
use judgment analyzing cost benefits alternatives selection best combinations accurately
improve add new functionality existing financial accounting systems participate design
spending expense reporting process enabling companies individuals control greater convenience
subscription order revenue operations real time business companies industry launch
financial systems pc software excellent organizational skills high level acceptance
invoicing sales analytics customer data inventory details help owners quick

Cluster 2 : Conception et Gestion de l'Équipement et des Systèmes (supply chain)

test measurement equipment responsibilities include design document prototype oversee production
control systems design assistance sustainable maintenance service retro commissioning energy
playback attention quality control daily maintenance projection equipment responsible aspect
lighting window treatments designer enhance design division designer work internal
equipment systems standard optometry industryworking conditions position performed traditional retail
equipment systems standard optometry industry physical demands position requires ability
assembly test equipment ability read electrical mechanical pneumatic prints good
technology software services optimize supply chain dematic employs 7 000
test equipment characterize image sensors developing evaluation boards suing schematic

Cluster 3: Services et Gestion du Cloud Computing

configuration deployment familiarity cloud deployment workflows cloud iaas providers preferably aws
complexities orchestrating running cloud application services sre s dislike things
services support continuous security compliance aws cost optimization assist client
provisioning infrastructure applications cloud ready application architectures ll provide prescriptive
workloadsexperience production deployment monitoring operational support enterprise class applicationsexperience performance
cost effective resilient secure cloud based applications migrations existing workloads
transforming way build deliver consume empower customers radically simplifying virtualization
manner generate actionable intelligence cloud based open platform integrates ecosystem
orchestration consistent monitoring logging refine manage deployment orchestration use scm
deployments strong focus design build deployment security system hardening securing

Cluster 4: Analyse de données (dashboard)

streamline provide greater efficiencies accuracy current finance procedures maximize contribution

interactive dashboards forecasting models advertising businesspartner data technology manage requirements

everyday workload management processes big data analytics digital marketing line

reports dashboards allow business owners greater ability optimize performance key

reports increase efficiency provide key operational strategic support market investment

processes streamline reporting data distribution bachelor s degree math science

user journey identify growth opportunities contract lifecycleusing data analytics proactively

dashboards reporting ensure successcollaborate program strategy design lead performance models

improve scalable processes tools dashboards drive operational business efficiencyanalyze data

reporting tools processes utilizing current systems resources streamline sales margin

Cluster 5: Marketing et Gestion de la Relation Client

revenue growth founded 2009 recurly uses open platform approach easily

client onboarding improved personalized communications anytime content access tailored clients

customer identification program infrastructure vital component new sellers ready business

scale programs emphasis integrated marketing duties assignedknowledge skills abilities required

marketing programs supports development execution marketing programs critical projects support

grower user journeys engender trust loyalty measured shared yoy grower

sales environment accurate entry management lead data crm system requiredability

social management wordpress multisite known cpg brand s b2c b2b

point sale system ability drive motivate lead retail staff detail

processes improve efficiencies need bachelor s degree marketing graphic design

Cluster 6: Optimisation et Intégration des flux de travail

business processescleaning raw data sync multiple 3rd party apps systemsassisting
process integration essential functions participate aspects consulting systems analysis detailed
workflows beekeeper integrates existing operational systems makes accessible central portal
manual repetitive workflows accelerating optimizing business processes simpligov empowers public
business processes illuminate financial opportunities great enjoy working startup environment
eliminate applying business process improvements possible business intelligence manager evaluates
document based business processes improving accuracy increasing productivity reducing costs
business process discovery testing extensive customer list includes fortune 500
growth processesability embrace ambiguity define solutions obvious clearcommitted dedicated willing
administration tasks dba processes consult customers sql server including best

7.2. Exemples de valeurs du clustering des offres d'emploi de 2023

Cluster 0: Automatisation des tests

testing experienceqa testing java requiredexperience windows software qa testing experience
deployment shall echo achievements security shall guardian angel weave best
incident response capabilities azure sentinel automate develop monitors scripts python
testing frameworks data validation quality frameworks data lineage frameworksmetadata definition
tooling like jenkins azure devops xcode toolchainexperience working team ios
test casespreferred qualificationsminimum 5 7 years experience working agile environmentminimum
testing tools cypress blaze meter etc experience developing test automation
b testing optimize website performance user experience ensure team s
qa testing scripts thoroughly test application ci cd pipeline environment
build platforms continuous integration experience knowledge jenkins hudson maven gradle

Cluster 1 : Automatisation de la gestion financière et de l'analyse de données

decision platforms programs supervisory relationships reports directly svp chief lending

reports iteratively build prototype dashboards provide insights scale solving business

future finance businesses thrive hundreds thousands businesses trust bill solutions

modernize systems conducting meetings presentations share ideas findings performing requirements

ap processesmaintain policies set forth companyrequirements cpa required preferably public

possible regardless destination won t settle current state change agent

detailed financial reports dashboards presentations management finance team requirements bachelor s

accounts payable transactions research provide support documentation audits monitor manage

manage data modeling visualizations analysis existing new reports develop data

labor scheduling system ensure store workload accounted funded appropriately stores

Cluster 2 : Automatisation industrielle

rod pump control systems pump controllers integrated variable frequency drives

manual processes skills experience required 6 years experience systems engineering

manufacturing equipment transition processes production key elements role include responsible

machinery solutions position understanding design sales process involved large scale

use latest modern management technology manufacture wide range commercial residential

strong interpersonal communication skills effectively interact business stakeholders technical teams

fast paced manufacturing environment familiarity autocad draft sight previous experience

machine solutions deliver equipment global manufacturing locations day day responsibilities

machinery consider opportunity minimal travel 10 15 overnight travel minimal

dispensing processes including use automated dispensing units connect rx overwrap

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management
Chaussée de Binche 151, 7000 Mons, Belgique | **www.uclouvain.be/lsm**