

# **L'utilisation de l'intelligence artificielle pour l'analyse textuelle des accords d'entreprise sur le télétravail**

Laura Khoury et Arthur de Amorim, économistes, Université Paris Dauphine – PSL, LEDA (UMR CNRS-IRD)

Cet article présente un exemple concret d'utilisation de l'intelligence artificielle dans la recherche en sciences sociales et en économie du travail. Il expose une méthode permettant d'exploiter à grande échelle le contenu de milliers d'accords d'entreprise afin d'analyser plus finement les conditions de mise en œuvre du télétravail. Au-delà des résultats obtenus sur ce sujet particulier, la démarche illustre le potentiel de l'IA générative pour transformer les méthodes d'analyse documentaire et statistique appliquées aux relations professionnelles.

## **Un projet de recherche sur les effets du télétravail**

L'étude s'inscrit dans une recherche plus large consacrée à l'impact du télétravail sur les trajectoires professionnelles des salariés. Elle vise notamment à comprendre comment la mise en place du télétravail influence les rémunérations, les mobilités professionnelles, les changements d'entreprise ou encore les distances entre le domicile et le lieu de travail.

Pour cela, nous avons mobilisé une importante base de données regroupant les accords d'entreprise et d'établissement déposés en France. Parmi l'ensemble de ces accords, nous avons sélectionné ceux portant sur le télétravail afin d'observer les pratiques négociées avant la crise sanitaire de la Covid-19. Dans une première phase de nos travaux, nous utilisons essentiellement une information binaire : l'établissement avait-il signé ou non un accord sur le télétravail ? Cette approche permettait déjà de mesurer certains effets du dispositif. Les résultats indiquent notamment que la signature d'un accord favorise effectivement le recours au télétravail et s'accompagne d'une hausse du salaire horaire ainsi que d'un allongement de la distance moyenne domicile-travail des salariés concernés.

Toutefois, cette première approche ne permettait pas de tenir compte de la diversité des modalités prévues dans les accords. Or, tous les accords de télétravail ne se ressemblent pas : certains restreignent l'accès au dispositif, d'autres autorisent un nombre plus élevé de jours télétravaillés, imposent certaines conditions ou laissent davantage de souplesse aux salariés. C'est pour analyser cette diversité que l'intelligence artificielle a été mobilisée.

## **Le défi de l'exploitation de milliers d'accords**

La principale difficulté résidait dans le volume de documents à traiter. Sur la période 2014-2019, nous avons recensé 2 230 accords relatifs au télétravail. Analyser manuellement chacun de ces textes aurait nécessité un travail extrêmement long et coûteux.

La première étape a donc consisté à récupérer automatiquement les accords à partir de la plateforme [Legifrance](#). Cette collecte a été réalisée grâce à des techniques d'automatisation permettant de télécharger l'ensemble des documents de manière systématique. Cette phase a permis de constituer une base documentaire homogène et exploitable pour l'analyse.

Une fois les accords rassemblés, le véritable enjeu était d'extraire les informations pertinentes contenues dans les textes. Nous souhaitons notamment identifier des variables telles que le nombre maximal de jours de télétravail autorisés, les conditions d'éligibilité, l'ancienneté requise, les lieux autorisés pour le télétravail, ou le caractère fixe ou flexible des jours télétravaillés. Ces informations existent bien dans les documents,

mais elles sont formulées de manière très différente selon les entreprises. Leur extraction automatique nécessite donc une compréhension fine du langage juridique et des formulations utilisées.

## **Une méthode fondée sur l'IA générative**

Pour relever ce défi, nous avons développé une méthode combinant travail humain et intelligence artificielle.

Dans un premier temps, un échantillon de 100 accords a été sélectionné. Soixante-dix documents ont servi à élaborer et tester la méthode, tandis que trente autres ont été conservés pour évaluer de manière indépendante la qualité des résultats. Les accords ont été lus manuellement et annotés afin d'identifier précisément les informations recherchées. Cette base de référence a ensuite servi à construire un prompt détaillé destiné à un modèle d'intelligence artificielle générative.

Le principe consistait à transmettre à l'IA le texte intégral de chaque accord accompagné d'instructions très précises décrivant les variables à extraire et les règles à appliquer. Pour chaque document, le modèle produisait alors automatiquement une fiche structurée contenant les informations recherchées. Ces résultats étaient ensuite convertis sous forme de tableau statistique exploitable, et comparés aux informations issues de l'analyse manuelle pour évaluer la performance du modèle d'IA.

L'un des enseignements majeurs de cette expérience est que la qualité des résultats dépend fortement du travail de conception du prompt. Nous avons dû procéder à de nombreux allers-retours pour préciser les consignes et réduire les erreurs. Certaines variables se sont révélées particulièrement complexes à coder. Ainsi, déterminer si un accord imposait explicitement le domicile comme lieu exclusif de télétravail a nécessité un important travail d'affinement des instructions afin de distinguer les cas réellement explicites des formulations plus ambiguës.

Cette phase de calibration a permis d'obtenir une précision particulièrement élevée. Les résultats atteignent environ 95 % de concordance avec les codages réalisés manuellement, avec seulement une légère diminution des performances lors de la phase de validation sur des documents que le modèle n'avait jamais analysés.

## **Des gains considérables en temps et en coût**

L'utilisation de l'intelligence artificielle a rendu possible une opération qui aurait été extrêmement difficile à réaliser manuellement.

Une fois la méthode finalisée, l'analyse des milliers d'accords restants a été automatisée. L'ensemble du traitement a nécessité environ dix heures de calcul pour un coût global d'une quarantaine d'euros, soit quelques centimes par document. Ces chiffres illustrent le potentiel des outils d'IA pour traiter de très grands corpus documentaires à des coûts relativement faibles.

Il convient toutefois de souligner que ces gains reposent sur un investissement préalable important : la définition rigoureuse des variables à extraire, la constitution d'un échantillon de référence et les multiples tests nécessaires pour améliorer progressivement les performances du système. L'IA ne supprime donc pas le travail de recherche, mais elle en modifie profondément les méthodes.

## **Premiers enseignements sur les accords de télétravail**

La base ainsi constituée permet déjà de dégager plusieurs tendances.

Environ 40 % des accords précisent explicitement que le télétravail doit être réalisé au domicile du salarié. Dans de nombreux autres cas, cette question n'est pas clairement encadrée. Nous observons également que la majorité des accords laissent une certaine flexibilité dans le choix des jours télétravaillés, tandis que très peu imposent des jours fixes.

L'analyse montre aussi que les accords sont principalement présents dans les grandes entreprises. En revanche, lorsque les petites entreprises concluent de tels accords, elles tendent à accorder davantage de jours de télétravail que les structures de grande taille.

Autre résultat notable : avant la pandémie, près des deux tiers des accords autorisaient au maximum un jour de télétravail par semaine. Cette situation contraste fortement avec les pratiques observées aujourd'hui et témoigne de la transformation rapide des organisations du travail depuis la crise sanitaire.

## **Conclusion**

Cette expérience de recherche montre comment l'intelligence artificielle peut devenir un outil puissant d'analyse des relations professionnelles. En permettant l'extraction automatisée d'informations contenues dans des milliers d'accords collectifs, elle ouvre la voie à des recherches jusqu'alors difficiles à mener à grande échelle. Toutefois, l'expérience met également en évidence que la qualité des résultats dépend avant tout du travail humain de conception, de validation et d'interprétation. L'IA apparaît ainsi moins comme un substitut à l'expertise du chercheur que comme un levier permettant d'amplifier ses capacités d'analyse et d'explorer des corpus documentaires d'une ampleur inédite.