

Traitement automatiques (NLP, IA générative) de texte d'accords d'entreprises portant sur l'environnement et la transition écologique

Sujet pour le Tremplin Recherche ESIEE Paris, 2025-26

Équipe ou projet dans le laboratoire

Ce stage se déroulera au sein de l'équipe de recherche du Conservatoire National des Arts et Métiers - Centre d'Etudes de l'Emploi et du Travail, qui mène un projet de recherche interdisciplinaire visant à explorer les pratiques de négociation collective au sein des entreprises sur plusieurs thèmes, dont l'environnement et la transition écologique. L'équipe développe des approches innovantes combinant intelligence artificielle et sciences sociales pour analyser à grande échelle les accords d'entreprise français.

Le tuteur du projet sera Silvia NAPOLITANO, Maitresse de conférences en science économiques à ESIEE-Pais et chercheuse à l'Erudite et au CNAM-Ceet .

Adresse électronique : silvia.napolitano@esiee.fr

Le projet est coordonné par Nathalie GREENAN, Professeure des Universités au CNAM et directrice des programmes de recherche Changements organisationnels, travail et emploi du CEET.

Filière visée

Datascience et intelligence artificielle

Présentation générale du sujet

Le stage consistera à développer un protocole de recherche et d'un cadre de codage pour exploiter la base publique de textes d'accords d'entreprise français portant sur l'environnement et la transition écologique (plusieurs milliers de documents). Le défi technique consiste à créer un système d'extraction d'information utilisant des modèles NLP avancés et des LLM pour identifier automatiquement les clauses environnementales, analyser les lignes d'action des entreprises en matière de protection de l'environnement et de transition écologique, et détecter les tendances émergentes. Le projet combine preprocessing de texte à grande échelle, fine-tuning de modèles, construction d'indicateurs quantitatifs, et validation de la robustesse des algorithmes développés.

Objectif du projet

L'objectif du projet est de mieux comprendre les pratiques de négociation collective en matière d'environnement et de transition écologique, et de développer des outils pour analyser et interpréter ces données. Le projet pourra également inclure l'élaboration de scénarios prospectifs sur les emplois et les compétences pour une industrie soutenable, en lien avec les résultats de l'analyse. Cette approche interdisciplinaire permettra de croiser les regards et les méthodes pour aborder un sujet d'actualité.

Le stagiaire aura donc comme objectif de :

- Comprendre les pratiques de négociation collective au sein des entreprises en matière de protection de l'environnement et de transition écologique, en réalisant une revue de la littérature.
- Contribuer à la définition d'un protocole de recherche et d'un cadre de codage pour exploiter la vaste base publique des accords collectifs d'entreprise sur la négociation portant sur l'environnement et la transition écologique.
- Sélectionner les outils pertinents pour le traitement automatique de textes (NLP, LLM, IA générative) pour coder les accords selon des axes prédéfinis théoriquement.
- Construire des indicateurs exploitables pour l'analyse quantitative.

- Vérifier la robustesse des outils mobilisés pour le codage, des indicateurs construits et de la méthode poursuivie.
- Analyser les lignes d'action des entreprises en matière de protection de l'environnement et de la transition écologique et leur dynamique en fonction des caractéristiques des entreprises et des tendances qu'elles identifient.

Bibliographie

Buyse, K., & Verbeke, A. (2003). Proactive environmental strategies: A stakeholder management perspective. *Strategic management journal*, 24(5), 453-470.

Do, S., Ollion, É., & Shen, R. (2024). The augmented social scientist: Using sequential transfer learning to annotate millions of texts with human-level accuracy. *Sociological Methods & Research*, 53(3), 1167-1200.

Grybauskas, A., & Cárdenas-Rubio, J. (2024). Unlocking employer insights: Using large language models to explore human-centric aspects in the context of industry 5.0. *Technological Forecasting and Social Change*, 208, 123719.