

Industrial Dynamics of Artificial Intelligence: Looking at Job Offers and Skills in Europe

François Maublanc



Contexte

- 1 Contexte
- 2 Base de données
- 3 Questions empiriques
- 4 Compétences & Savoir-être
- 5 Objectifs

Contexte I

- L'Intelligence Artificielle (IA) suscite une certaine fascination du grand public... mais également de nombreux fantasmes, notamment sur le travail.
- Les effets de l'intelligence artificielle restent relativement méconnus d'un point de vue économique. De très rares études ont été menées aux Etats-Unis, mais pour l'instant aucune au niveau européen.

Contexte II

- Pourtant, l'IA est considérée souvent comme une technologie disruptive, pouvant impliquer :
 - une transformation des activités de production et d'innovation,
 - des modifications dans le management et l'organisation des entreprises,
 - des créations d'emplois très qualifiés mais aussi des effets de polarisation (une recomposition de l'emploi autour de deux pôles, avec des écarts de plus en plus marqués de revenus et de conditions de travail),
 - une modification des besoins en compétences et des modes de recrutement des entreprises,
 - un impact profond sur le marché du travail...

Contexte III

- L'objectif de ce projet est de comprendre la dynamique du secteur de l'intelligence artificielle au niveau européen, et ses conséquences notamment sur le marché de l'emploi et l'innovation.

Base de données

- 1 Contexte
- 2 Base de données**
- 3 Questions empiriques
- 4 Compétences & Savoir-être
- 5 Objectifs

Base de données I

- Existe-t-il une base de données sur les entreprises présentes dans le domaine de l'intelligence artificielle ?
- ... à notre connaissance, non !
- Il va falloir constituer cette base de données.
- Deux options se présentent à nous :
 - étudier les personnes ayant déjà eu une expérience dans le domaine de l'intelligence artificielle.
 - étudier les offres d'emploi dans le domaine de l'intelligence artificielle.

Base de données II

- Pour étudier les personnes ayant déjà eu une expérience dans le domaine de l'intelligence artificielle => LinkedIn.
- Près de 70% des salariés en France ont un compte LinkedIn... et il est probable que dans un secteur comme l'IA, ce chiffre soit encore supérieur.
- Nous allons utiliser des outils du type *Phantombuster* pour récupérer des informations sur les profils LinkedIn :
 - 1 Obtenir une liste de comptes LinkedIn de personnes travaillant dans le domaine de l'intelligence artificielle.
 - 2 Récupérer leurs expériences passées (scraping du profil), notamment les entreprises, les dates de début et de fin de contrat...

Base de données III

- ③ Récupérer les informations des entreprises sur leurs pages LinkedIn.
- Ceci permettra de constituer un panel d'entreprises ayant embauché des personnes en IA au cours des dernières années, et ainsi d'analyser la dynamique du secteur !
- Limites de la méthode :
 - Nombre de profils que l'on peut scraper sur LinkedIn est limité (80/jour, soit 30000/an par compte gratuit)
 - Des questions de Droit sur l'utilisation des données...

Base de données IV

- Pour étudier les offres d'emploi dans le domaine de l'intelligence artificielle => Glassdoor, Indeed...
 - 1 Collecter ces offres en écrivant un code Python en utilisant la librairie *Beautiful Soup* pour récupérer automatiquement les annonces et les traiter dans un format *friendly* pour les équipes (fichier csv).
 - 2 Assurer une collecte régulière (une fois par mois) et traiter les doublons (annonce persistant plusieurs mois par exemple).
 - 3 Récupérer les informations des entreprises sur leurs pages Glassdoor ou Indeed.

Base de données V

- Quelle que soit la méthode utilisée, il est important d'avoir des annonces ou des profils pertinents.
- Plusieurs méthodes seront mises en place :
 - une recherche par mots clés : intelligence artificielle, deep learning, machine learning, réseau de neurones... ?
 - implémentation d'une intelligence artificielle capable à partir du contenu de l'annonce, d'indiquer si celle-ci se réfère à un emploi dans le domaine de l'intelligence artificielle ou non (algorithme de classification, apprentissage supervisé ou non supervisé ?,...).

Base de données VI

- Il faudra apparier les données sur les entreprises dans le secteur de l'IA avec des données financières sur les entreprises. => Nous utiliserons la base de données BVD Orpis IP (accès payant).
- Nous apparierons les données sur les entreprises dans le secteur de l'IA avec les données de dépôts de brevets et de publications scientifiques (collaborations universités/organismes de recherche avec les entreprises).

Questions empiriques

- 1 Contexte
- 2 Base de données
- 3 Questions empiriques**
- 4 Compétences & Savoir-être
- 5 Objectifs

Questions empiriques I

- Qu'est-ce qui explique qu'une entreprise publie des annonces (et cherche donc à recruter) dans le secteur de l'IA ?

Estimation d'un **modèle probit** ou **logit** avec pour variable expliquée une dummy valant 1 si l'entreprise publie des offres en IA et 0 sinon, et comme variables expliquées :

- l'ancienneté de l'entreprise,
- la taille de l'entreprise,
- le secteur d'activité,
- la zone géographique,
- les caractéristiques financières de l'entreprise,
- ...

Il faudra aussi collecter des informations sur des entreprises en-dehors du secteur de l'IA !

Questions empiriques II

- Qu'est-ce qui explique l'expérience demandée par les entreprises dans les annonces de recrutement ?

Estimation d'un modèle **probit multinomial ordonné** en s'appuyant sur la méthode du maximum de vraisemblance, compte-tenu de la nature de la variable expliquée (échelle de Likert du type 0, 0-2, 2-7, 7+).

On comparera en particulier l'IA avec toutes les autres compétences en Sciences de l'Informatique : en IA, on s'attend à ce que l'expérience demandée soit bien moindre que dans les autres secteurs :

Questions empiriques III

- les compétences en IA sont entièrement nouvelles pour tous les secteurs, et ces compétences sont censées se trouver chez les jeunes récemment diplômés (avec peu d'expérience).
- L'IA permet le développement de tâches automatisées, y compris certaines qui ne sont pas routinières. => Les personnes en IA travailleraient davantage dans des organisations moins hiérarchiques et plus décentralisées => plutôt une population jeune diplômée avec peu d'expérience.

Questions empiriques IV

Nous allons avoir besoin d'autres secteurs de comparaison en Sciences de l'informatique : développement web, cyber-sécurité, visual computing, informatique embarquée, business intelligence & Analytics, data science... => Il va falloir scraper des annonces (ou profils LinkedIn) dans ces domaines !

Questions empiriques V

- Quels sont les facteurs expliquant que certaines entreprises en IA innovent et pas d'autres ?
 - Regarder les brevets déposés par les entreprises dans le domaine de l'IA, ainsi que les collaborations scientifiques entre entreprises et universités / organismes de recherche.
 - S'agit-il de chercheurs qui montent leur entreprise, ou au contraire d'entreprises qui se rapprochent du monde de la recherche académique ?

Estimation du nombre de brevets déposés par une entreprise au cours d'une période de temps par un modèle de **Poisson**, **Négatif Binomial** ou par des modèles **Zero inflated**.

Questions empiriques VI

- Regarder l'évolution de la dispersion spatiale des emplois dans le secteur de l'IA au cours du temps.
 - Construction d'une carte dynamique en Europe montrant la concentration spatiale des emplois (ou offres d'emplois) en IA.
 - Comparaison avec d'autres secteurs.

Compétences & Savoir-être

- 1 Contexte
- 2 Base de données
- 3 Questions empiriques
- 4 Compétences & Savoir-être**
- 5 Objectifs

Compétences I

- Compétences requises par le projet :
 - Web scraping avec Python
 - Intelligence artificielle
 - Régression linéaire
 - Statistiques inférentielles
 - Économétrie des variables qualitatives
 - Microéconomie
 - Gestion financière
 - Data visualisation
 - Anglais
 - Expression orale et écrite
 - ...

Compétences II

- Naturellement, vous ne réaliserez pas toutes les tâches du projet !
- Suivant vos parcours, vous réaliserez des tâches différentes.
- Il devra y avoir de la collaboration dans le groupe, parfois entre les groupes (les tâches ne sont pas toujours indépendantes !), et aussi de la compétition.
- Sur certaines compétences (estimation de modèles statistiques entre autres), des **cours personnalisés** seront donnés par les membres du projet.
- Suivi une fois par mois pour faire le point, plus régulièrement si vous avancez bien ou si vous avez des questions.

Savoir-Être

- Savoir travailler en groupe
- Etre autonome
- Etre force de proposition, savoir prendre des initiatives et les partager au sein du groupe et avec les membres des équipes vous encadrant.
- Ne pas cacher les problèmes ! Il y aura des difficultés, il faudra réfléchir pour les résoudre ou les contourner.

Objectifs

- 1 Contexte
- 2 Base de données
- 3 Questions empiriques
- 4 Compétences & Savoir-être
- 5 Objectifs**

Objectifs I

- CY Cergy Paris Université (CY-Tech) sera le moteur du projet sur la partie données !
- Base de données qui servira :
 - de point de départ à plusieurs publications scientifiques,
 - pour chacun d'entre-vous, à mieux connaître votre marché du travail pour trouver des stages ou votre premier emploi !
 - au Service Relation Entreprises de CY-Tech pour améliorer la recherche de stages, les invitations des entreprises au forum Entreprises de l'Ecole...
- Apporter des réponses au débat public :
 - L'IA entraîne-t-elle une concentration des emplois très qualifiés dans certaines zones géographiques (grandes métropoles) et en détruit-elle dans les autres zones ? => Accroissement de l'effet de polarisation.

Objectifs II

- Comment mieux designer les politiques publiques pour favoriser l'innovation en IA ?
- ...
- Contact

fmc@cy-tech.fr