**Proposition d’analyses - Semaine 1**

**Introduction**

Ce document présente les propositions d’analyses pour les différentes sources de données sur les politiques en Intelligence Artificielle qui sont les suivantes :

* OECD AI Policy Observatory
* IAPP – Global AI Law and Policy Tracker
* Stanford University – AI Index Report (HAI)

Premièrement, nous allons proposer des analyses individuelles pour chaque source de données puis, en second lieu, on présentera des analyses croisées entre ces différentes sources de données.

Ce document est la première version de la description des analyses que l’on va faire, je présente ci-dessous une liste non exhaustive et non définitive d’analyses intéressantes pour ces sources de données. En effet, il se peut que par la suite certaines analyses soient supprimées et/ou d’autres ajoutées.

**1. Analyses Descriptives par Source de Données**

Cette première partie est consacrée à l'analyse descriptive de chaque source de données prise individuellement. L'objectif est de résumer leurs caractéristiques principales et de comprendre le potentiel analytique intrinsèque de chacune. Ce travail constitue le socle fondamental sur lequel les analyses croisées plus complexes seront bâties.

1. OECD AI Policy Observatory

On commence par les analyses sur le fichier contenant les instruments politiques : oecd-ai-all-ai-policies.csv

**Chapitre 1 : Analyses descriptives**

**1.1** Axe Géographique : Cartographie de l'Activité Politique

* Question **:** Quels sont les pays et les régions les plus actifs en matière de publication de politiques sur l'IA ?
* Variable Clé : Country
* Visualisation Suggérée : Une carte du monde (choroplèthe) et un diagramme en barres horizontales du "Top 15" des pays.

**1.2** Axe Temporel : Évolution de la Production Politique

* Question : Comment le volume de politiques sur l'IA a-t-il évolué depuis les premières initiatives ? Y a-t-il eu une accélération notable ?
* Variable Clé : Start date
* Visualisation Suggérée : Un histogramme ou un graphique en courbes montrant le nombre de nouvelles politiques lancées par année.

**1.3** Axe Thématique : Nature des Instruments Politiques

* Question : Quels sont les types d'instruments politiques les plus utilisés par les gouvernements (stratégies, lois, financements...) ? Quels sont les domaines d'action (AI Policy Area(s)) les plus ciblés ?
* Variables Clés : Policy instrument type category, Theme area(s), AI Policy Area(s).
* Visualisation Suggérée : Des diagrammes en barres pour chaque variable, montrant la fréquence de chaque catégorie.

**1.4** Axe Budgétaire : Répartition des Moyens

* Question : Quelle est la distribution des fourchettes budgétaires déclarées pour ces politiques ?
* Variable Clé : Yearly budget range
* Visualisation Suggérée : Un diagramme en barres montrant le nombre de politiques dans chaque fourchette budgétaire (ex: <1M, 1M-5M, etc.).

### **Chapitre 2 : Analyses Croisées (Exploratoires)**

*Ce chapitre vise à explorer les relations entre les différentes variables pour découvrir des tendances plus complexes et générer des hypothèses.*

**2.1** Analyse : Spécialisation Thématique par Pays

* Question : Est-ce que certains pays se spécialisent dans des types de politiques spécifiques ? Par exemple, l'Union Européenne se concentre-t-elle davantage sur la "Gouvernance" que les États-Unis ?
* Variables Clés : Country + Policy instrument type category.
* Visualisation Suggérée : Un diagramme en barres empilées (ou 100% empilées) pour comparer la proportion de chaque type de politique pour une sélection de pays.

**2.2** Analyse : Évolution des Types de Politiques dans le Temps

* Question : La nature des politiques en IA a-t-elle changé au fil des ans ? Y avait-il plus de "Stratégies" au début, et observe-t-on plus de "Lois" contraignantes récemment ?
* Variables Clés : Start date + Policy instrument type category.
* Visualisation Suggérée : Un graphique en aires empilées, montrant l'évolution de la part de chaque type de politique année après année.

**2.3** Analyse : Caractéristiques et Financement

* Question : Les politiques considérées comme des "réformes structurelles" sont-elles plus souvent financées (ou non) par le secteur privé ?
* Variables Clés : Is a structural reform ? + Has funding from private sector ?.
* Visualisation Suggérée : Un tableau de contingence (heatmap) ou un diagramme en barres groupées pour visualiser la relation entre ces deux variables binaires.

**Chapitre 3 : Analyses Croisées (Politiques vs. Indicateurs Quantitatifs)**

**Objectif :**

Cette section vise à explorer les relations entre les cadres politiques définis par les pays (issues de la base oecd-ai-all-ai-policies) et les indicateurs quantitatifs de performance de leur écosystème IA (issus des différentes sources Live Data). Ces deux sous-sources de données sont issues de la même source : “OECD AI Policy Observatory”. L'objectif est de générer des hypothèses sur l'impact potentiel des politiques publiques.

#### **Méthodologie Commune**

Toutes ces analyses reposent sur une méthodologie en trois étapes :

1. **Préparation des données politiques :** Agréger les données du fichier oecd-ai-all-ai-policies au niveau de chaque pays pour créer des indicateurs (ex: nombre total de politiques, type de politique dominant, année de la première stratégie, etc.).
2. **Préparation des données quantitatives :** Agréger les données des fichiers Live Data au niveau de chaque pays (ex: total des investissements sur 5 ans, nombre total de publications de haute qualité, etc.).
3. **Fusion (Enrichissement) :** Fusionner ces deux types de tables par pays pour permettre l'analyse croisée.

#### 

#### **Propositions d'Analyses Croisées**

**1. Analyse : Impact des Stratégies Nationales sur la Recherche**

* Question : Est-ce que les pays ayant une stratégie nationale d'IA ancienne et bien établie publient plus de recherches de haute qualité ?
* Données requises :
  + oecd-ai-all-ai-policies : Country, Start date (filtré sur les "Stratégies Nationales").
  + Live Data - AI Research : Données pour Publication type = Article et Quality ranking = High impact.
* Visualisation suggérée : Un nuage de points avec l'année de la première stratégie en Axe X et le total des publications de haute qualité en Axe Y.

**2. Analyse : Type de Cadre Politique et Attractivité des Investissements**

* Question : Les pays avec une approche politique axée sur le "Soutien Financier" attirent-ils plus d'investissements en capital-risque que les pays avec une approche axée sur la "Gouvernance" et la régulation ?
* Données requises :
  + oecd-ai-all-ai-policies : Country, Policy instrument type category.
  + Live Data - Investments in AI : Données sur les investissements VC.
* Visualisation suggérée : Un diagramme en barres comparant l'investissement moyen pour le groupe de pays "pro-financement" vs. le groupe "pro-régulation".

**3. Analyse : Réformes Structurelles et Dynamiques Économiques**

* Question : Les pays qui ont mis en place des "réformes structurelles" liées à l'IA sont-ils perçus comme plus attractifs par les investisseurs ?
* Données requises :
  + oecd-ai-all-ai-policies : Country, Is a structural reform ?.
  + Live Data - Investments in AI : Données sur les investissements VC.
* Visualisation suggérée : Un diagramme en boîtes (box plot) comparant la distribution des investissements pour le groupe de pays Is a structural reform ? = TRUE vs. le groupe FALSE.

B. IAPP – Global AI Law and Policy Tracker

#### **A. Analyse univariée (Descriptive)**

* Fréquence des politiques d'IA par pays (histogramme simple).
* Répartition entre les lois et les politiques (Law\_or\_Policy).
* Part des pays ayant une stratégie nationale d'IA (Is\_National\_Strategy).
* Distribution des années d’adoption (Date\_Adopted\_or\_Announced) pour créer une timeline.

#### **B. Analyse bivariée (Exploratoire)**

* Comparaison du type de document (loi, stratégie, etc.) par continent/région.
* Relation entre la présence d'une stratégie nationale et l'année d’adoption pour analyser la croissance dans le temps.
* Identification des autorités les plus citées dans les politiques (graphe de co-occurrence ou diagramme en barres).
* Lien entre les mentions de coopération internationale et le type de posture stratégique d'un pays.

#### **C. Analyse qualitative thématique**

* Nuage de mots sur la variable objectif\_strategique pour détecter les grandes ambitions.
* Carte du monde colorée selon la posture\_strategique (Pro-Innovation, Pro-Régulation, etc.).
* Matrice croisant les pays et le modèle\_inspiré pour identifier qui suit l'UE, l'OCDE, etc.
* Analyse de la présence d’investissements publics par région.

#### **D. Exploration temporelle**

* Graphique de l'évolution du nombre de politiques d'IA publiées par année.
* Analyse de l'émergence des références à l'"EU AI Act" ou aux principes de l'OCDE dans les textes de politique au fil du temps.

C. Stanford University – AI Index Report (HAI)

Même type d’analyses pour cette source de données contenant des initiatives politiques tout en ajoutant en complément les données quantitatives dont dispose cette source de données, comme le “nombre de projets de loi relatifs à l'IA adoptés par pays, 2016-2024” ou autres que l’on pourra associer au nombre d’initiative politiques présente dans ce rapport. de l'OCDE afin de voir, par pays, le nombre de projets de loi qui ont été adoptés contrairement à ceux qui ont juste été présentés, etc.

**2. Analyses Croisées Multi-Sources**

Cette seconde partie représente le cœur analytique du projet, où la valeur ajoutée est créée en croisant les différentes sources de données. En fusionnant les données sur les politiques publiques avec les indicateurs quantitatifs (recherche, investissement, etc.), l'objectif est d'explorer des corrélations, de tester des hypothèses et de révéler des tendances qui ne sont pas visibles lorsque les sources sont analysées isolément.

1. “*Instruments politiques” de OCDE* avec “*Nombre de mentions de l'IA dans les procédures législatives dans 75 zones géographiques sélectionnées, 2016-2024*” de l’Index AI report de stanford.

* Analyse : Corrélation entre le Débat Législatif et la Production de Politiques en IA
* Objectif : Comparer le "bruit" législatif (le nombre de mentions de l'IA dans les débats parlementaires, issu du rapport de Stanford) avec l'action politique concrète (le nombre de politiques et d'initiatives effectivement publiées, issues de la base de données de l'OCDE). L'analyse vise à déterminer si les pays qui débattent le plus de l'IA sont aussi ceux qui produisent le plus de politiques.
* Méthode : Mettre en corrélation, par pays et par année, le nombre de mentions législatives avec le nombre de nouvelles politiques initiées. Une visualisation possible est un graphique combiné montrant l'évolution des deux indicateurs dans le temps pour un pays donné, afin de détecter d'éventuels décalages (par exemple, un pic de débats suivi un an plus tard par un pic de production de politiques).

1. Millions de dollars dépensé dans l’ia par pays par an de “Live data (OCDE observatory) par rapport au "Nombre de projets de loi relatifs à l'IA adoptés dans certaines zones géographiques, 2016-2024 (somme)”

* Analyse : Influence du Capital-Risque sur l'Activité Législative en IA
* Objectif : Analyser la corrélation entre le volume d'investissement en capital-risque et le nombre de lois sur l'IA adoptées par un pays. L'hypothèse à tester est qu'une forte activité d'investissement privé pourrait être un facteur d'accélération pour l'avancement législatif. Il s'agit de déterminer si les écosystèmes où les fonds d'investissement sont très actifs sont également ceux qui se dotent le plus rapidement d'un cadre réglementaire, suggérant un lien potentiel entre les intérêts du secteur privé et l'agenda politique.
* Données requises :
* Live Data - Investments in AI (OCDE) : La somme des investissements en capital-risque par pays sur la période 2016-2024.
* Stanford AI Index Report : Le nombre total de lois relatives à l'IA adoptées par pays sur la même période.
* Méthode et Visualisation : La visualisation principale reste un nuage de points (scatter plot) pour observer la nature de la relation entre les deux variables. Une corrélation positive supporterait l'hypothèse, tandis qu'une absence de corrélation suggérerait que les deux dynamiques sont largement indépendantes.

**Conclusion**

Il convient de considérer ce document comme une première feuille de route stratégique pour la phase d'analyse. La démarche adoptée se veut volontairement itérative ; les aperçus obtenus lors des premières explorations guideront l'ajout et la priorisation d'analyses plus fines. Ce processus agile garantira que l'analyse finale se concentre sur les angles les plus pertinents et à plus forte valeur ajoutée pour les objectifs globaux du projet.