**Fiche descriptive : source 1. OECD AI Policy Observatory**

**Structure :**

Fichiers CSV (données tabulaires)

**Langue :** Anglais

**Source** : National AI policies & strategies from OECD.AI – Policy Observatory

Organisation de Coopération et de Développement Économiques (OCDE / OECD)

Lien de la source : <https://oecd.ai/en/dashboards/overview>

**Format** : Ces données sont directement téléchargeables en fichier CSV.

Cette source de données contient plusieurs type de données :

* Les instruments politiques en IA, (issues chacun d’une initiative politique) : fichier oecd-ai-all-ai-policies.csv contenant toutes les politiques en IA : 1884 données
* Live data, des données principalement quantitatif mis à jours constamment : fichier CSV;

Pour la partie “Live data” on a la possibilité de faire une API qui extrait les données en temps réel sachant qu'elle mentionne ses mises à jour récurrentes (*à voir avec le temps qu’on aura*).

**Volume approximatif :**

* 1 fichier CSV de 1864 data
* une vingtaine de fichiers avec des données numériques

**Variables disponibles**

Dans ce chapitre, pour plus de clarté, on va distinguer les deux sous-sources de données “All AI policies” et “Live data”.

1. **All AI policies**

oecd-ai-all-ai-policies.csv :

Cette partie présente les variables disponibles pour une initiative politique et celles pour un instrument politique qui est une sous-partie d’une initiative comme expliqué ci-après :

Le fichier contient plusieurs instruments politiques (un par ligne) qui sont les données sur lesquelles on va faire des analyses et des visualisations. Certains ont la même initiative politique (politique originale) et ces initiatives politiques sont classées par thèmes.

Pour cette section, j’ai décidé d’intégrer seulement les variables pour lesquelles des valeurs sont assez présentes dans le fichier CSV car il en existe avec une très faible couverture. Pour garantir la fiabilité de l'analyse, la sélection des variables présentées dans cette fiche s'est basée sur un critère objectif. Une règle a été instaurée : seules les variables ayant un taux de remplissage de 50 % ou plus ont été retenues. Cette méthode permet d'écarter les données trop éparses qui pourraient introduire des biais dans les résultats.

Bien que le fichier CSV soit très bien structuré avec les 1884 données, on distingue deux sous-parties de données (et donc de variables) **: l'initiative** (le "Pourquoi") vs **l'instrument** (le "Comment").

**L'initiative politique (niveau 1)** : C'est le programme ou la stratégie globale. C'est le "Pourquoi" et le "Quoi". Une initiative comme "Digital Luxembourg" est un grand projet national avec des objectifs généraux.

* Les variables à ce niveau sont uniques pour chaque initiative : le nom, le pays, la date de début, les objectifs généraux, le contexte.

**Les instruments politiques (niveau 2)** : ce sont les actions concrètes ou les outils mis en place pour réaliser les objectifs de l'initiative. C'est le "comment". Pour "Digital Luxembourg", les instruments peuvent être un fonds d'investissement, un nouveau règlement, une plateforme de collaboration, etc.

* Les variables à ce niveau décrivent l'action : le type d'instrument, l'organisation responsable, le budget alloué à cet instrument spécifique.

1.1 Tableau des variables pour les **Initiatives politiques**:

| **Nom de la Variable** | **Description** | **Nombre d'Apparitions (sur 1884)** |
| --- | --- | --- |
| Policy initiative ID | Identifiant unique pour chaque initiative politique dans la base de données de l'OCDE. | 1884 |
| Is a structural reform ? | Indicateur (Vrai/Faux) si la politique est considérée comme une réforme structurelle. | 1884 |
| Country | Le pays qui a initié la politique. | 1884 |
| Theme area(s) | La grande catégorie thématique de la politique (ex: 'National AI Policies', 'Regulation'). | 1884 |
| Theme(s) | Le sous-thème plus spécifique de la politique (ex: 'National AI policies', 'Data governance'). | 1884 |
| Platform URL | URL vers la page de l'initiative sur la plateforme OECD.ai. | 1884 |
| Is evaluated ? | Indicateur (Vrai/Faux) si l'initiative a fait l'objet d'une évaluation formelle. | 1884 |
| Yearly budget range | Fourchette budgétaire annuelle estimée pour l'initiative. | 1884 |
| Has funding from private sector ? | Indicateur (Vrai/Faux) si l'initiative a reçu un financement du secteur privé. | 1884 |
| English name | Le nom de l'initiative politique en anglais. | 1883 |
| Description | Une description textuelle détaillée de l'initiative politique. | 1854 |
| Objective(s) | Les objectifs déclarés de l'initiative. | 1831 |
| Start date | L'année de lancement de l'initiative (format AAAA.0). | 1830 |
| Target group(s) | Le public visé, de manière plus détaillée. | 1826 |
| Target group type(s) | Le type de public visé par la politique (ex: 'Firms', 'Research and education organisations'). | 1826 |
| Responsible organisation(s) | L'organisation ou le ministère responsable de la mise en œuvre. | 1793 |
| AI Principle(s) | Les principes éthiques de l'IA (selon la taxonomie de l'OCDE) que la politique adresse. | 1745 |
| AI Policy Area(s) | Le domaine d'action spécifique de la politique en matière d'IA (ex: 'AI R&D', 'Data governance'). | 1555 |
| Public access URL | URL vers un document ou une page publique décrivant l'initiative. | 1546 |
| Background | Contexte et informations générales sur la mise en place de l'initiative. | 1151 |
| Original name(s) | Le nom de l'initiative dans sa langue d'origine. | 985 |

1.2 Tableau des variables pour les **Instrument politiques** (dépendent des Initiatives politiques et change en fonction des Instruments politiques :

| Nom de la Variable | Description | Nombre d'Apparitions (sur 1884) |
| --- | --- | --- |
| Policy instrument ID | Identifiant unique pour chaque instrument politique. Chaque initiative peut avoir plusieurs instruments. | 1884 |
| Policy instrument type category | La grande famille à laquelle appartient l'instrument (ex: 'Governance', 'Financial support'). | 1837 |
| Policy instrument type | La nature spécifique de l'instrument (ex: 'National strategies', 'Equity financing'). | 1837 |

*Note : En complément de ce jeu de données principal, plusieurs variables à faible taux de complétion ont été conservées. Elles font l'objet d'analyses spécifiques et qualitatives détaillées dans le rapport final pour enrichir l'interprétation des tendances générales. Elles offrent également la possibilité de générer de nouvelles variables dérivées, si cela s’avère pertinent pour l’analyse.*

1. **LIVE DATA**

La section "Live Data" de l'observatoire de l'OCDE sur l'IA (OECD.AI) est une plateforme interactive qui vise à suivre en temps réel ou quasi réel les tendances et les dynamiques de l'écosystème mondial de l'intelligence artificielle. Contrairement à la base de données statique sur les politiques publiques, cette source offre des indicateurs dynamiques couvrant plusieurs dimensions clés telles que la recherche scientifique, les investissements en capital-risque et le marché du travail.

Chaque onglet de cette plateforme est considéré comme une source de données distincte, voici quelques sources décrites individuellement ci-dessous.

**LIVE DATA**

Onglet AI Research : <https://oecd.ai/en/data?selectedArea=ai-research>

| **AI publications by country** | | | |
| --- | --- | --- | --- |
| Filtre/dimension | Type de visualisation | Potentiel analytique | Remarques |
| - Publication type : dataset, review, report, article..  - Quality ranking : How, Medium or Low impact  - Indicator : Publications or Citations | Graphique en courbes (évolution temporelle) | - Comparer la spécialisation de la recherche en IA entre les pays.  - Suivre l'évolution de cette spécialisation dans le temps (depuis 2000).  - Analyser en détail la performance par type de publication (ex: Articles vs Livres) et par niveau de qualité (ex: High vs Low impact). | - Source des données : Scopus (Elsevier).  - L'option "All" pour les filtres Publication type et Quality ranking fournit une vue agrégée. Pour une analyse détaillée, la collecte doit se faire filtre par filtre. |

Onglet “Investments in AI and data” <https://oecd.ai/en/data?selectedArea=investments-in-ai-and-data>

| **Worldwide VC investments in AI** | | | | |
| --- | --- | --- | --- | --- |
| Description | Filtre/dimension | Type de visualisation | Potentiel analytique | Remarques |
| Mesure les flux d'investissements en capital-risque (VC), une forme de financement privé, dirigés vers les startups du secteur de l'IA. | - Indicator : Sum of investments (USD millions dollars) or Number of investments | Graphique en courbes (évolution temporelle) | - Suivre la croissance de l'investissement privé en IA dans le temps.  - Comparer les niveaux d'investissement entre les pays.  - Analyser la maturité de l'écosystème d'un pays via la répartition des investissements (Seed vs Series A, B, C+). | - Source des données : Crunchbase.  - Permet d'évaluer la vitalité économique et l'attractivité du secteur de l'IA d'un pays pour les investisseurs privés. |

Via ces deux tableau “Worldwide VC investments in AI“ et “AI publications by country” on peut, par exemple, proposer l’analyse suivante:

### **Proposition d'Analyse Croisée : Corrélation entre Recherche de Haute Qualité et Investissement Privé**

#### **Titre de l'analyse**

Corrélation entre l'Excellence de la Recherche et l'Attractivité des Investissements Privés en IA.

#### **Objectif**

Vérifier l'hypothèse selon laquelle une forte production scientifique de haute qualité est un facteur d'attractivité pour le capital-risque (VC) dans le secteur de l'intelligence artificielle. En d'autres termes, est-ce que les investisseurs financent les startups dans les pays où la recherche est la plus pointue ?

#### **Sources de Données Requises**

1. Live Data - AI Research : Le fichier CSV contenant le nombre (ou la part) de publications pour la combinaison de filtres :
   * Publication type = Article
   * Quality ranking = High impact
2. Live Data - Investments in AI : Le fichier CSV principal contenant les montants des investissements en capital-risque, avec :
   * Country = All countries

La description et l'intégration des autres sources de "Live Data" (sur les investissements, les compétences, etc.) se feront de manière itérative. Cette approche permettra de sélectionner et de qualifier les nouvelles données en fonction des questions d'analyse croisée qui seront définies, assurant ainsi la pertinence de chaque ajout au périmètre de l'étude.

**Proposition d’analyses descriptives pour OECD : All ai policies**

### **1. Analyse de la Répartition Géographique**

Cette analyse répond à la question : "Où sont lancées les politiques en IA ?"

* Question exploratoire : Quels sont les pays les plus actifs ? Comment se répartissent les initiatives au niveau mondial ?
* Variables à utiliser : Country, Policy initiative ID (pour compter).
* Visualisations possibles :
  + Une carte du monde (choroplèthe) où l'intensité de la couleur représente le nombre d'initiatives par pays.
  + Un diagramme en barres horizontales montrant le top 15 des pays par nombre d'initiatives.
* Faisabilité : Automatisable. Il s'agit de compter le nombre d'initiatives par pays, ce qui est une opération de base.

**2. Analyse de l'Évolution Temporelle**

Cette analyse répond à la question : "Quand ces politiques sont-elles lancées ?"

* Question exploratoire : Comment le nombre de nouvelles politiques en IA a-t-il évolué au fil des ans ? Y a-t-il eu une année pic ?
* Variables à utiliser : Start date, Policy initiative ID.
* Visualisations possibles :
  + Un histogramme ou un diagramme en barres montrant le nombre de politiques lancées chaque année.
  + Un graphique en ligne pour visualiser la tendance sur le long terme.
* Faisabilité : Automatisable. Il suffit de regrouper les initiatives par année de départ et de les compter.

**3. Analyse Thématique (Contenu des politiques)**

Cette analyse répond à la question : "De quoi parlent ces politiques ?"

* Question exploratoire : Quels sont les grands thèmes (Theme area(s)), les sous-thèmes (Theme(s)) et les domaines d'action (AI Policy Area(s)) les plus fréquents ?
* Variables à utiliser : Theme area(s), Theme(s), AI Policy Area(s).
* Visualisations possibles :
  + Des diagrammes en barres pour chaque variable, montrant la fréquence de chaque catégorie.
  + Un diagramme circulaire (camembert) pour Theme area(s) qui a moins de catégories.
* Faisabilité : Automatisable. Un simple comptage de la fréquence de chaque catégorie suffit. *Note : Pour AI Policy Area(s), un traitement simple peut être nécessaire pour séparer les multiples valeurs contenues dans une même cellule.*

**4. Analyse par Caractéristiques**

Cette analyse répond à la question : "Quelle est la nature de ces politiques ?"

* Question exploratoire : Quelle est la proportion de réformes structurelles ? Quelle part bénéficie de financements privés ? Quelle part est évaluée ? Quel est le budget typique alloué ?
* Variables à utiliser : Is a structural reform ?, Has funding from private sector ?, Is evaluated ?, Yearly budget range.
* Visualisations possibles :
  + Des diagrammes circulaires pour les variables binaires (Oui/Non).
  + Un diagramme en barres pour la répartition des fourchettes budgétaires.
* Faisabilité : Automatisable.

**5. Analyses Croisées**

Cette analyse combine les axes précédents pour trouver des relations plus complexes.

* Question exploratoire : Est-ce que certains pays se concentrent sur des thèmes spécifiques ? Les réformes structurelles sont-elles plus récentes ? Les politiques financées par le privé sont-elles plus souvent évaluées ?
* Variables à utiliser : Combinaisons de deux variables (ex: Country et Theme(s)).
* Visualisations possibles :
  + Diagramme en barres empilées pour visualiser la répartition des thèmes par pays.
  + Tableau de contingence (heatmap) pour voir les corrélations entre variables catégorielles.
* Faisabilité : Automatisable. Cela demande des opérations de type "tableau croisé dynamique" (pivot table).

**6. Analyse de Contenu (Approche Avancée)**

Cette analyse vise à comprendre le discours et les détails fins des politiques.

* Question exploratoire : Quels sont les mots et concepts les plus fréquents dans les descriptions et objectifs des politiques ? Peut-on identifier automatiquement des groupes de politiques similaires en fonction de leur description ?
* Variables à utiliser : Description, Objective(s), Background.
* Visualisations possibles :
  + Nuage de mots (Word Cloud).
  + Graphiques de "Topic Modeling" montrant les sujets latents.
* Faisabilité : Traitement plus avancé (NLP - Traitement du Langage Naturel). Ce type d'analyse nécessite des techniques comme la tokenisation, la lemmatisation, et des modèles statistiques (ex: TF-IDF, LDA) ou encore Iramuteq pour extraire des informations pertinentes du texte brut.