



ÉCONOMIE NUMÉRIQUE

La difficile mesure de l'intelligence artificielle

Alexandre Bourgeois – Nicolas Prat
Division concepts, méthodes et évaluation des comptes nationaux
INSEE

Sommaire



01

Les actifs data et logiciels (dont IA)

3 types d'actifs et changements de périmètres

SNA2008/ESA2010 : l'IA fait partie des logiciels et bases de données (software and databases).

-SNA2025/ESA2030 : L'IA fait partie des logiciels, dont l'actif est séparé de la data (qui intègre les bases de données mais élargit aussi le périmètre à toutes les observations enregistrables ayant un potentiel de rentabilité économique.

→ Expérimentation Allemande sur data : → recommandation de calculer simultanément les 3 types d'actifs (data, logiciels et R&D) en même temps etc.

Les méthodes utilisées dans une grande partie de la littérature existante peuvent être classées, de manière générale, en **méthodes fondées sur les coûts (cost based)**, **sur les revenus (income based)** et **sur le marché (market based)**.

Risques liés à ces approches :

- Risque d'omettre la valeur économique (potentielle) plus large due à la non-rivalité ainsi que les externalités associées aux données.
- Risque de ne pas tenir nécessairement compte de la valeur du contenu

Méthodes qui reprennent la distinction, bien connue en économie, entre préférences révélées et préférences déclarées (ce que les individus font dans des situations économiques réelles et ce qu'ils affirment dans des scénarios hypothétiques)

Les méthodes par le marché (market-based methods)

Les méthodes de **tarification fondées sur le marché** s'appuient sur des données observables.

-C'est pourquoi, lorsqu'ils sont disponibles, **les prix du marché constituent toujours la méthode privilégiée** pour l'évaluation – même si ces prix ne représentent qu'une estimation partielle de la valeur sociale totale (Coyle et Diepeveen 2021).

-Cependant, il existe relativement peu de prix de marché observables pour les données. **La plupart des données sont actuellement créées et utilisées en interne** par des entreprises ou d'autres organisations plutôt que d'être vendues ou échangées (Coyle).

-Même les marchés de données bien établis tels que la notation de crédit ne publient généralement pas leurs prix (bien que les revenus provenant de la vente de notations de crédit soient déclarés dans les comptes des agences de notation, ce qui permet de tirer certaines conclusions sur les prix moyens)

-A noter que les prêteurs mettent en commun leurs propres données en les fournissant aux agences de notation de crédit, qui revendent ensuite des services aux prêteurs, ce qui démontre que la valeur des données agrégées est supérieure à la somme de la valeur de ses composantes.

-La littérature académique traitant des approches de l'évaluation des données fondées sur le marché peut être divisée en trois catégories :

- celles qui prônent **l'utilisation de marchés de données** et examinent les caractéristiques souhaitables de la tarification des données ;
- celles qui utilisent la **capitalisation boursière des entreprises** pour calculer la valeur ;
- celles qui utilisent les **flux mondiaux de données** pour établir une évaluation.

Les méthodes par les revenus (income-based methods)

Les approches fondées sur les revenus s'appuient sur les **flux de recettes futures attendus** (tels que la vente d'analyses marketing ou de services d'information) générés pour estimer la valeur des données sous-jacentes

- Cette méthode est évidemment plus facile à mettre en œuvre lorsqu'il existe des flux de recettes directement attribuables à l'utilisation des données.
- Une approche courante fondée sur les revenus est la **méthode dite de « l'exonération de redevances »**, souvent utilisée pour d'autres actifs incorporels tels que les marques et les brevets.
 - Il s'agit d'une mesure directe (révélée plutôt que issue d'enquêtes), car elle indique combien d'argent une entreprise peut économiser en créant/détenant les données plutôt qu'en versant des redevances pour en obtenir la licence.
- Ker et Mazzini (2020) utilisent également une **méthode fondée sur les revenus en calculant les revenus déclarés** liés à la compilation et à la vente de données, sur la base de données d'enquêtes auprès des entreprises.
 - Cependant, les nomenclatures actuelles des données sectorielles ne permettent pas d'identifier correctement ces cas.
 - De plus, en identifiant uniquement les cas où les données constituent le « produit », cette méthode exclut la valeur créée par les données dans les entreprises « enrichies par les données » qui ne vendent pas directement de données mais les utilisent pour emploi final propre afin d'améliorer leurs produits ou processus ou générer des revenus grâce à l'analyse de données (ex : entreprises mobilisant les données qu'elles collectent afin de mieux gérer leurs chaînes d'approvisionnement).
- Les approches fondées sur les revenus font davantage appel au jugement que les approches fondées sur les coûts.
 - Difficultés pour établir un contrefactuel car les flux de revenus peuvent inclure des flux de revenus améliorés par les informations issues des données, mais aussi d'autres facteurs.
 - Difficultés à prédire la valeur d'un flux de revenus futurs avant que les gains ne soient réalisés.

Les méthodes part les coûts (cost-based methods)

Constituent actuellement l'une des approches les plus courantes pour évaluer les données dans la pratique.

- Cette logique s'inspire de la pratique standard du Système de comptabilité nationale (SCN) selon laquelle, si la valeur ne peut être directement observée par le biais d'une transaction de marché, une approche de la « somme des coûts » est recommandée (Ker et Mazzini 2020).

- On suppose que le chiffre obtenu donne une estimation raisonnable de la valeur minimale des données, en partant du principe qu'un actif a généralement une valeur attendue au moins égale à son coût, même en tenant compte de la dépréciation

- → hypothèse forte car certaines collectes de données sont imposées par la loi.

- Approche qui consiste à identifier les coûts liés à la production, à la collecte, au stockage et au remplacement d'un ensemble de données, ainsi que les coûts pour l'entreprise en cas de perte des données ou d'autres coûts (tels que l'indemnisation en cas de failles de sécurité).

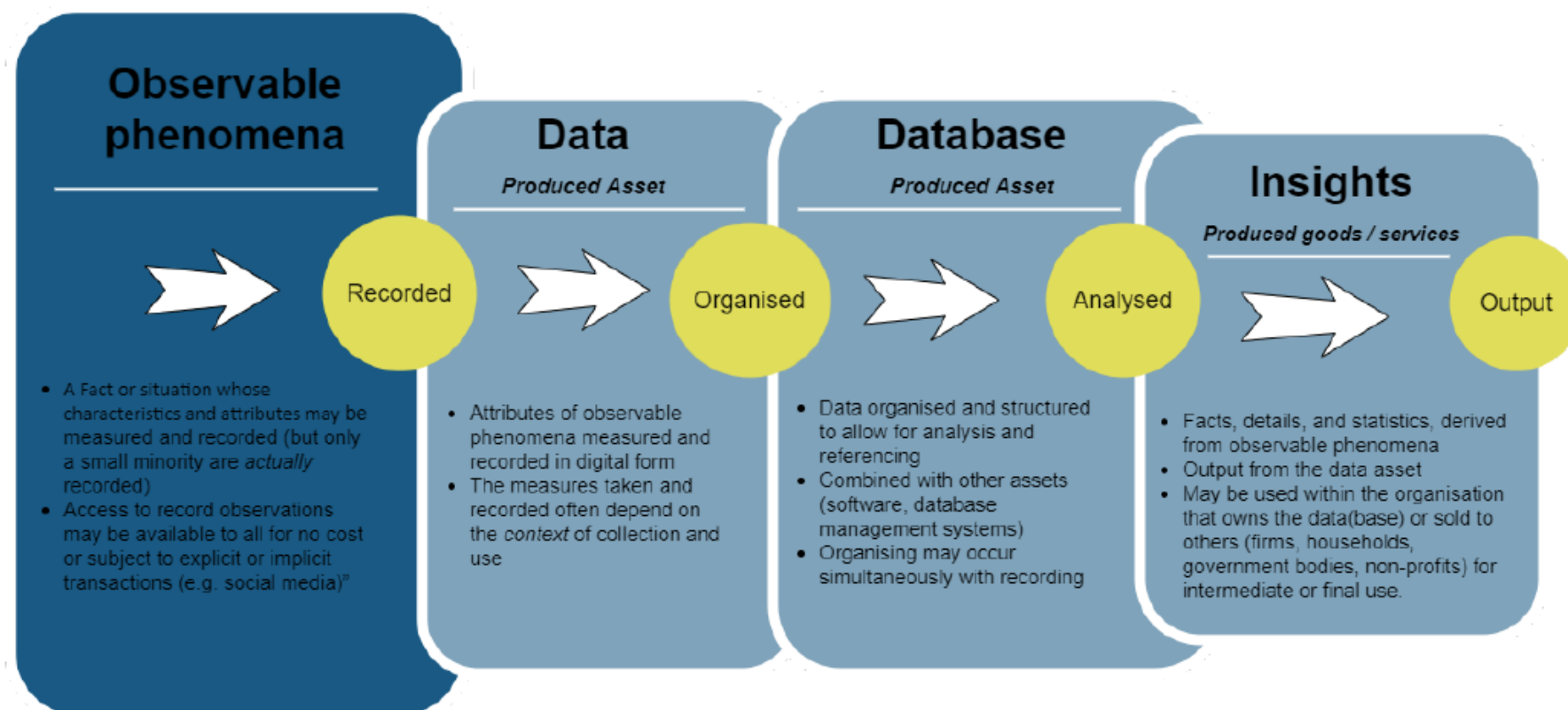
- → relative facilité de calcul et habitude des comptables nationaux de l'utilisation d'approches de somme des coûts pour d'autres actifs incorporels, cette **méthode a été privilégiée par le SNA / SEC.**

02

Task force « *Data as an asset* »

Les données comme actif produit

- Le SCN 2025 remplace le concept de « bases de données » par un concept plus large de « données et bases de données »
- **Données** = enregistrement numérique de « **phénomènes observables** » (contenu informationnel), donnant un **bénéfice économique**



Les données comme actif produit

- Valorisation par la **Somme des Coûts (SoC)**
- **Quels termes dans la somme ?**
 - Rémunération du travail
 - Consommations Intermédiaires
 - Consommation de Capital Fixe
 - Taxes nettes des subventions
 - Rémunération du capital (*Return to Capital*)
- **Que doit recouvrir la valeur de la data ?**
 - Conception, mise en œuvre de la stratégie de production des données
 - Accès, enregistrement et stockage des données : y-compris les éventuels achats de données ou de droits d'accès
 - Nettoyage, traitement, organisation

« non-labour costs mark-up »

En résumé, quel processus ?

- Déterminer la « masse salariale en data »
 - Identifier les métiers concernés, et surtout les pondérer par la fraction de leur temps de travail passée à produire de la donnée → ***involvement rates***
- Multiplier par un *non-labour costs mark-up*
 - Idéalement, en distinguant :
 - les coûts de production propres de la donnée (identiques pour toutes les APE)
 - la rentabilité propre, et éventuellement la dépréciation propre à chaque APE
- Mettre en place la Méthode de l'Inventaire Permanent (MIP) pour calculer la dépréciation du capital
 - Durée de vie moyenne, profil de mortalité, lissage linéaire

Non-labour costs mark-up

Rappel : il s'agit de représenter le **passage des coûts salariaux à la production**

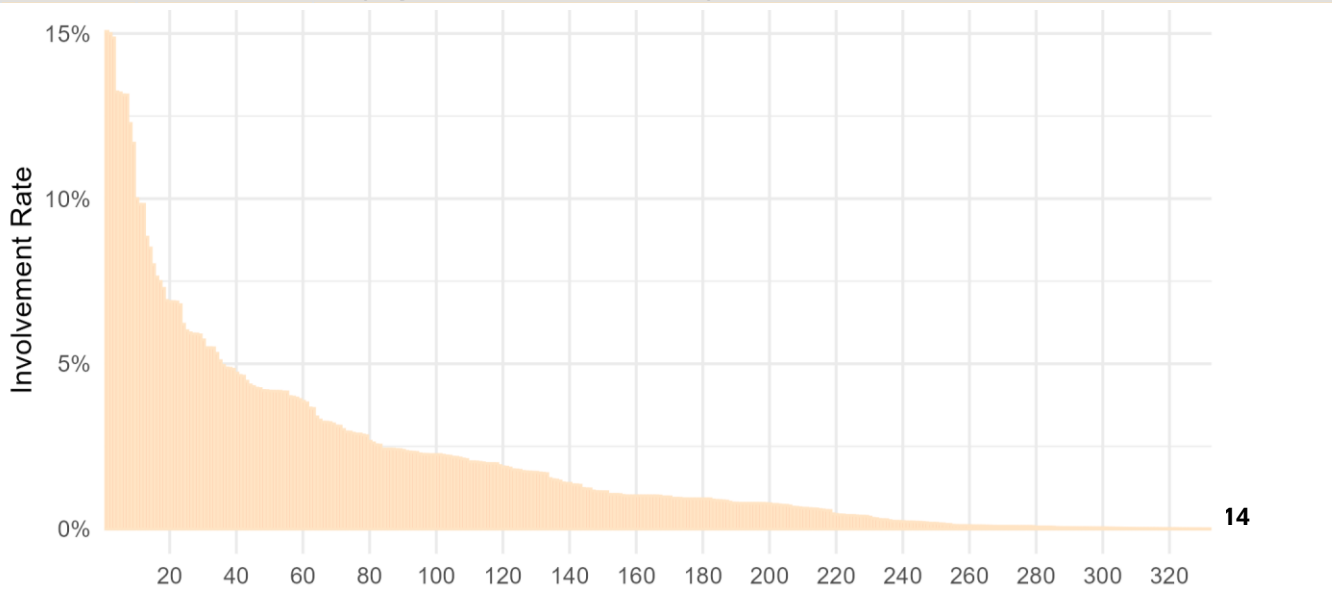
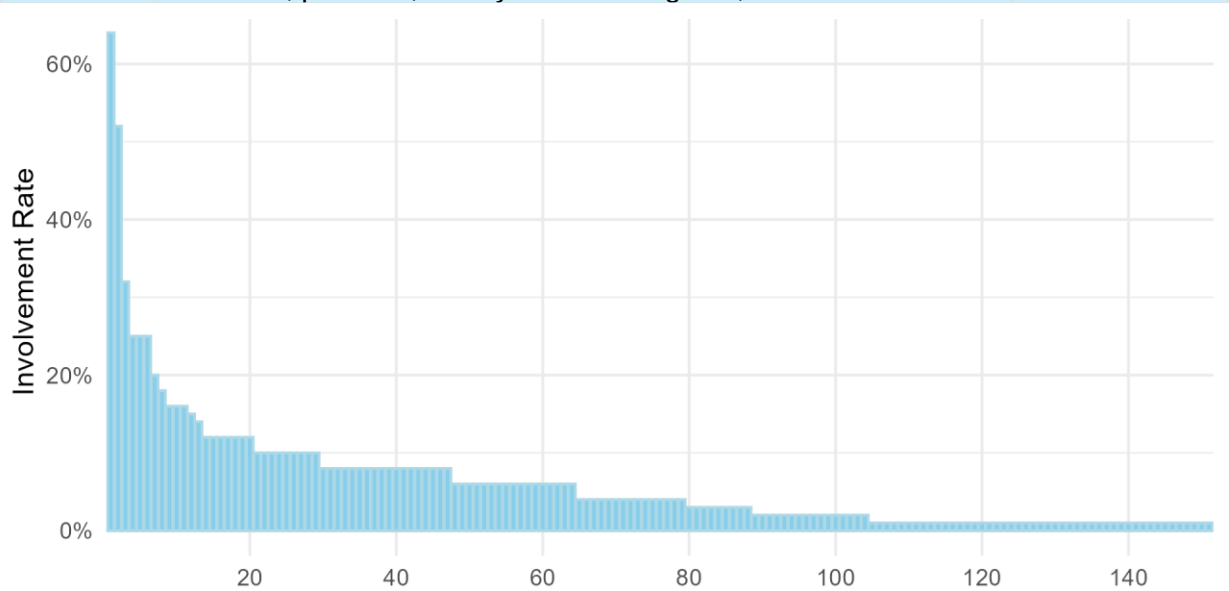
- Recommandation *standard* : utiliser le ratio P.1/D.1 (Production / Rémunérations) des secteurs d'activité (APE) 62+63
 - « *or similar available aggregate* »...
 - L'idée : trouver les entreprises qui ont une **production marchande** en données
 - 62 = Programmation, conseil et autres activités informatiques
 - 63 = Services d'information (*activités des portails de recherche sur Internet, de traitement et d'hébergement de données, ainsi que d'autres activités dont l'objectif essentiel est de fournir des informations*)
 - Avantage : ce sont des grandeurs standard de la compta nat (comptes des ENF)
- Recommandation *avancée* : utiliser des taux différenciés, pour refléter les différentes **rentabilités** et **dépréciations** selon les secteurs

Involvement rates

- Recommandation *standard* : utiliser les métiers et taux du *Handbook*
 - Avantage : On n'a besoin « que » de la BTS, et de relier les nomenclatures PCS et ISCO
 - Inconvénient : Ces coefficients « standard » ne reflètent pas spécifiquement l'économie française
- Recommandation *avancée* : utiliser nos propres métiers et coefficients, déterminés d'une manière « *objective et systématique* » (limiter le « dire d'expert »)
 - L'accent est mis sur le machine learning... *Mais attention, l'IA ce n'est pas magique !*
 - Les Japonais ont une enquête où les gens estiment leur propre pourcentage *mais ce sont les seuls*
 - Les Allemands ont développé une méthode « objective », *mais qui repose sur des hypothèses peut-être un peu fortes* (→ on en reparle plus loin !)

On aimerait se diriger vers la recommandation avancée, mais c'est la partie la plus complexe.
Pas sûrs de véritablement disposer de l'information

ISCO code	Occupation	Involvement rate	PCS2003	IR.PCS2003	Libellé
4132	Data entry clerks	64%	388A	15,06%	Ingénieurs et cadres d'étude, recherche et développement en informatique
4227	Survey and market research interviewers	52%	388B	15,00%	Ingénieurs et cadres d'administration, maintenance, support et services aux utilisateurs
2521	Database designers and administrators	32%	388E	14,87%	Ingénieurs et cadres spécialistes des télécommunications
2120	Mathematicians, actuaries and statisticians	25%	478A	13,23%	Techniciens d'étude et de développement en informatique
3314	Statistical, mathematical and related associate professionals	25%	478B	13,20%	Techniciens de production, d'exploitation en informatique
3511	Information and communications technology operations technicians	25%	478C	13,14%	Techniciens d'installation, de maintenance, support et services aux utilisateurs en informatique
2511	Systems analysts	20%	478D	13,14%	Techniciens des télécommunications et de l'informatique des réseaux
4131	Typists and word processing operators	18%	388C	12,27%	Chefs de projets informatiques, responsables informatiques
2165	Cartographers and surveyors	16%	312E	11,68%	Ingénieurs conseils libéraux en études techniques
2529	Database and network professionals not elsewhere classified	16%	433D	10,00%	Préparateurs en pharmacie
3513	Computer network and systems technicians	16%	376A	9,83%	Cadres des marchés financiers
2512	Software developers	15%	376F	9,83%	Cadres des services techniques des organismes de sécurité sociale et assimilés
3312	Credit and loans officers	14%	433A	8,83%	Techniciens médicaux
2164	Town and traffic planners	12%	372A	8,50%	Cadres chargés d'études économiques, financières, commerciales
2514	Applications programmers	12%	226A	8,00%	Agents généraux et courtiers d'assurance indépendants, de 0 à 9 salariés
2519	Software and applications developers and analysts n.e.c.	12%	545D	7,63%	Employés des services techniques des organismes de sécurité sociale et assimilés
2522	Systems administrators	12%	376D	7,49%	Chefs d'établissements et responsables de l'exploitation bancaire
2523	Computer network professionals	12%	376B	7,28%	Cadres des opérations bancaires
3315	Valuers and loss assessors	12%	371A	6,92%	Cadres d'état-major administratifs, financiers, commerciaux des grandes entreprises
3433	Gallery, Museum and Library Technicians	12%	467B	6,88%	Techniciens des opérations bancaires
1330	Information and communications technology service managers	10%	467D	6,88%	Professions intermédiaires techniques des organismes de sécurité sociale
2160	Architects, planners, surveyors and designers, nos	10%	545C	6,87%	Employés des services techniques des assurances



Resultats (2021) – D.11

Total **D.11** en data (*champ des salariés*) :

18 018 M€

Total **D.11** sur le champ des salariés :

886 998 M€

D.11 data / D.11 total, ou « *intensité en données* » de l'économie dans son ensemble (*champ des salariés*) :

2 %

Caveat 1 :

Nous étudions ici D.11.

Le passage à D.1, puis à P.1, n'est pas forcément neutre.

Caveat 2 :

Nous n'étudions ici que le champ des salariés (avec la BTS)

Risques de double-compte

- Les actifs de propriété intellectuelle (IPP) : proches mais à distinguer
 - Données
 - Logiciels
 - R&D – *spécificité : elle est calculée via une enquête dédiée du MESRI sur la DIRDE*
 - IA – *aujourd'hui dans Logiciels, mais séparation future à prévoir...*

Vers un calcul conjoint, avec la même méthode

- Permet de s'assurer que la **somme des taux** pour un métier **ne dépasse pas 1 !**
- Pourrait faire partie des points à discuter au sein de la *task force* si celle-ci est prolongée

Calcul des *involvement rates*

- Ce que propose DESTATIS :

Utiliser les compétences « essentielles » et « optionnelles » du référentiel de métiers **ESCO**

D'abord, **étiqueter** manuellement de façon binaire les compétences : data ou pas

$$IR_i = \frac{2 \times \text{data essential}_i + \text{data optional}_i}{2 \times \text{all essential}_i + \text{all optional}_i}$$

- « objectivité » (= éviter le jugement d'expert)
- mais le coefficient **2** pourrait peut-être être amélioré

Partir des compétences semble être une bonne idée, et on comprend l'intuition que certaines compétences valent plus que d'autres

- Changement de concept derrière l'*IR* : « fraction du temps » → « fraction du *contenu* du poste »
- Pour une pondération plus « poussée » et *data-driven*, une idée : le wage premium (inspiration D2E)

03

Défis posés par l'IA

Une méthode de la somme des coûts pour évaluer un general purpose technology (GPT) ?

Task Force CMFB sur le commerce numérique (Digital Trade) : l'utilisation des données de cartes de crédit n'est pas très appropriée pour l'IA

→ Il est difficile de distinguer les services d'IA, car ceux-ci sont généralement associés à des services cloud fournis par des multinationales qui exercent de nombreuses autres activités commerciales.

- Définition d'une technologie à usage général (**GPT**) : « nouvelle méthode de production et d'invention suffisamment importante pour avoir un impact global durable ». Comme l'électricité.

- 95 % de l'utilisation de l'IA est aujourd'hui gratuite (D. Coyle)

→ Les services gratuits dans les comptes nationaux constituent déjà un sujet difficile à traiter...

Méthode par les coûts pour valoriser l'actif IA au sein des logiciels : s'appuie sur les masses salariales associées : s'il y en a très peu, et de grosses masses de capital => **Forte volatilité liée à un immense markup et une base de salariés difficiles à identifier.**

IA : Quel talon d'Achille des comptes nationaux ?

L'IA est (et sera) dans les comptes, **mais où ?**

-Fragilité vis à vis des **clés fixes et des hypothèses** mobilisées dans la construction des comptes :

-Ex 1 : Les clés de la table de passage BdP pour les imports de services.

-Ex 2 : L'hypothèse de stabilité des coefficients techniques pour la projection des effets lignes. L'IA devrait entraîner des changements structurels dans certaines relations commerciales, notamment en internalisant des services qui étaient auparavant classés comme consommation intermédiaire.

-Si biais de ventilation => **déformation du partage volume prix** => potentiel mismeasurement de la croissance du PIB en volume.

-Comment **lisser nos hypothèses** jusqu'à l'obtention de sources ?

-Pour les CI, comment identifier les secteurs les plus touchés par ces phénomènes, et comment combler le fossé entre la situation actuelle (où les informations sur la consommation intermédiaire sont limitées) et la facturation électronique, qui, nous l'espérons, fournira des informations plus précises et plus à jour sur ces consommations intermédiaires.

-Un défi majeur pour la comptabilité nationale : notamment pour le **suivi de la productivité**, car les statistiques de l'emploi rendent compte de manière fiable du phénomène IA, mais les comptes nationaux arriveront-ils à **fiabiliser la VA par branche en volume** du numérateur ?

Entrelacement des produits de propriété intellectuelle (IPP) : le cas Pokémon GO

-Lancé en 2016, Pokémon Go est rapidement devenu un véritable phénomène en Europe et dans le monde entier, transformant les rues de Bruxelles, Paris et Rome en terrains de jeu en réalité augmentée. → Plus de 100 millions de joueurs en 2024.

-Selon le MIT Technology Review, Niantic Spatial a utilisé des images recueillies pendant le jeu pour aider à entraîner ses systèmes, conçus pour créer des cartes 3D d'environnements réels.

→ afin de développer des systèmes de navigation visuelle pour les robots de livraison.

-Une application gratuite qui génère des données → utilisées pour entraîner un modèle d'IA (cartes 3D) → utilisées pour les robots de livraison → utilisées pour collecter d'autres données au final ?...

→ Au final, quel est le coût total du robot ?

→ Question des liens entre les actifs IPP : chaîne de valeur data-IA



Merci

Alexandre Bourgeois
Chef de la division concepts
méthodes évaluation (CME)
Département des comptes
nationaux (DCN)
+33 1 87 69 58 65
alexandre.bourgeois@insee.fr

Insee^{nl}
Mesurer pour comprendre

Nicolas Prat
Méthodologue
Division Concepts, Méthodes et
Évaluation des Comptes
Nationaux (CME)
+33 1 87 69 58 85
nicolas.prat@insee.fr

insee.fr      [blog](#)