

# L'Intelligence Artificielle en santé-environnement, au-delà de la boîte noire

## Résumés des interventions

21 Novembre 2024

# SESSION 1

## **Extraction de données structurées sur les services d'eau et d'assainissement à partir de PDFs**

Geneviève Fleury et Grégoire Etot, OFB

L'observatoire national SISPEA centralise les données sur le prix et la qualité des services publics d'eau et d'assainissement en France. Ces données sont toutefois incomplètes car certaines collectivités ne les saisissent pas dans le système.

Pour faciliter l'alimentation de la base de données SISPEA, nous travaillons au développement d'un outil d'IA permettant d'extraire les valeurs d'indicateurs codifiés à partir de rapports PDFs (dits RPQS) publiés en ligne chaque année.

Notre outil utilise des modèles open source et frugaux et vise à traiter des rapports PDFs aux formats variés, pouvant contenir à la fois du texte, des images et des tables.

---

## **SursynDC : traitement automatique du langage et technique de machine learning pour la surveillance syndromique des décès**

Johnny Platon et Anne Fouillet, Santé publique France

Les causes médicales de décès contenues dans les certificats de décès sont remplies en texte libre, puis elles sont codées en CIM-10 (classification internationale des maladies, 10ème révision) par l'Inserm-CépiDc dans un délai de plusieurs mois, incompatible avec une surveillance réactive de la mortalité.

Des techniques de traitement automatique des langues et machine learning (méthode SVM) ont été mises en œuvre sur les causes de décès exprimées en texte libre pour permettre leur classement automatique dans des regroupements syndromiques (regroupements de causes médicales ayant un intérêt pour la surveillance réactive).

Le projet a évalué leurs performances sur un échantillon de 1000 certificats électroniques de décès, après un entraînement des méthodes sur 3500 certificats. Ces 4500 certificats ont fait l'objet d'une annotation manuelle. De nouvelles perspectives entrent en jeu suite à l'arrivée de nouvelles techniques performantes sur les données textuelles (Transformers / LMM) et la mise en routine d'un premier algorithme d'intelligence artificielle.

# SESSION 1

## **CosIA : construire les cartes de couverture du sol par Intelligence Artificielle**

Eva Bookjans, IGN

**CoSIA, la Couverture de Sol par Intelligence Artificielle (IA), sont des cartes produites par l'IGN (Institut national de l'information géographique et forestière) de manière automatique grâce à un modèle d'IA. Le modèle s'appuie principalement sur les images aériennes de la BD ORTHO à 20 cm de résolution spatiale et composées de 4 canaux radiométriques (rouge, vert, bleu, infrarouge) pour prédire la couverture de sol de chaque pixel. Malgré les variations du paysage à travers le territoire français, le modèle est robuste et donne des bons résultats à l'échelle nationale.**

**Bien que CoSIA soit la sortie directe non corrigée du modèle d'IA, son niveau de détail géographique et ses 15 classes de couverture du sol en font une donnée à forte valeur ajoutée pour différents cas d'utilisation, tels que l'identification de la nature dans les zones urbaines, la cartographie des corridors écologiques, et l'étude des liens entre la santé des individus et leur environnement.**

---

## **Gérer les ressources et les risques du sol et du sous-sol : apport et potentiels de l'IA au BRGM**

Théophile Lohier, BRGM

**Le BRGM, service géologique national, est l'établissement public de référence dans les applications des sciences de la Terre pour gérer les ressources et les risques du sol et du sous-sol dans une perspective de développement durable. Son activité consiste notamment à comprendre les phénomènes géologiques et les risques associés, développer et mettre à disposition des méthodes et des outils, produire et diffuser des données et des connaissances. Depuis de nombreuses années, le BRGM développe son expertise dans la collecte, le traitement et la mise à disposition des données.**

**Aujourd'hui, les méthodes basées sur l'apprentissage statistique apportent de nouvelles solutions pour ces tâches. Dans cette présentation, nous proposons de montrer quelques exemples de travaux du BRGM utilisant des méthodes d'IA, en particulier dans le domaine des ressources minérales, de la ressource en eau, des risques naturels et de la pollution des sols.**

# PITCHS

## **Predict AI'r : des données de téléphonie à une meilleure qualité de l'air grâce à l'IA**

Giovanni De Nunzio, IFP Energies nouvelles

La planification urbaine et la gestion de la mobilité requièrent des outils permettant d'analyser et de prédire les effets d'actions et décisions à l'échelle des territoires. Cependant, les outils actuels utilisent des données statiques représentatives de l'instant auquel elles ont été récoltées, provenant d'enquêtes périodiques. Ces dernières sont insuffisantes pour anticiper et accompagner les décisions, en particulier lorsque les comportements varient rapidement comme cela a été le cas lors de la crise sanitaire du Covid-19.

Les pouvoirs publics peuvent donc se retrouver « dépassés » et n'avoir qu'une vision partielle ou biaisée pour établir et planifier des stratégies efficaces. L'objectif principal du projet predict AI'r est de créer un véritable observatoire et prédicteur dynamique de la mobilité et de ses impacts (sur l'air et le climat), qui soit fiable, représentatif à toute échelle territoriale et frugal en matière de duplication de la donnée personnelle.

En utilisant les données de bornage téléphonique déjà collectées par les opérateurs, ainsi que les outils de l'intelligence artificielle, le projet vise à fournir des services permettant de quantifier quotidiennement l'empreinte de la mobilité sur le climat et la qualité de l'air, afin d'accompagner les décisions des pouvoirs publics.

---

## **AMELIA : développer une cartographie dynamique air-bruit-mobilités**

Mathieu Vavrille, WaltR

Le bruit, la pollution de l'air et la mobilité sont à la fois très locaux et fluctuants : les collectivités manquent aujourd'hui d'informations cartographiques fines et d'outils d'aide à la décision.

Le projet AMELIA a réuni un consortium mixte afin de développer des outils de suivi de l'impact sanitaire et environnemental des pollutions liées au bruit et à l'air. Les cartes et outils d'AMELIA s'appuient sur des méthodes d'IA économes en temps de calcul et en besoin de stockage, dont l'apprentissage repose pour l'essentiel sur des mesures existantes.

# SESSION 2

## **Mobiliser le machine learning pour mesurer l'hétérogénéité de l'effet de la pollution de l'air sur le jeune enfant**

Milena Suarez Castillo, DREES

**La vulnérabilité à la pollution de l'air est vraisemblablement variable d'un enfant à l'autre, ce qu'occultent les estimations en moyenne. Dans le cadre d'une étude quasi-expérimentale s'appuyant sur la définition d'un groupe d'enfants « exposé » comparé à un groupe d'enfants « contrôle », l'hétérogénéité de l'effet de la pollution de l'air est estimée en fonction des caractéristiques des enfants, via une méthode de machine learning générique cherchant à définir des groupes d'impact de façon non supervisée.**

**L'effet de la pollution de l'air masquerait en réalité de fortes disparités : 10 % des enfants concentrent l'essentiel des effets détectables statistiquement lors d'une augmentation de l'exposition avant leur premier anniversaire. Ces enfants sont caractérisés par un certain nombre de facteurs de risques, comme la prématurité, mais aussi par un niveau de vie plus faible : parmi ces enfants les plus affectés, le dixième le plus modeste est 1,6 fois plus représenté que le dixième le plus aisé.**

---

## **Geospatial AI for One Health**

Jessica Abbate, Géomatys

**The increasing frequency of zoonotic disease emergence in recent decades has been linked to global change, particularly climate warming, habitat destruction and activities (e.g., farming, hunting) that reduce biodiversity, thus requiring a One Health approach to mitigate risk. Tracking the impacts of these changes requires massive amounts of data processing.**

**Earth observation data can not only help estimate human population distributions and behaviors but also enables near-real time observation of environmental changes, particularly when publication of official survey data may be sparse or slow. We use the example of Ebola virus disease emergence, linked specifically to host habitat fragmentation, to show how AI and machine learning (ML) methods play crucial roles in developing epidemic risk models and processing massive geospatial datasets. This combination of technology allows for timely interventions and better-informed public health strategies, ultimately advancing One Health goals by fostering a more holistic understanding of disease dynamics at the interface of humans, animals, and the environment.**

# SESSION 2

## **Amélioration de scores de risque environnemental par machine learning informé et IA explicable**

Jean-Baptiste Guimbaud, LIRIS

Dès la conception, des facteurs environnementaux tels que la qualité de l'air ou les habitudes alimentaires peuvent significativement influencer le risque de développer diverses maladies chroniques. Dans la littérature épidémiologique, des indicateurs connus sous le nom de Scores de Risque Environnemental (Environmental Risk Score, ERS) sont utilisés non seulement pour identifier les individus à risque, mais aussi pour étudier les relations entre les facteurs environnementaux et la santé. Une limite de la plupart des ERSs est qu'ils sont exprimés sous forme de combinaisons linéaires d'un nombre limité de facteurs.

Nos travaux visent à développer des indicateurs ERSs capables d'investiguer des relations non linéaires et des interactions à travers un large éventail d'expositions tout en découvrant des facteurs actionnables pour guider des mesures et interventions préventives, tant chez les adultes que chez les enfants.

Pour atteindre cet objectif, nous exploitons les capacités prédictives des méthodes d'apprentissage automatique non paramétriques, combinées avec des outils récents d'IA explicable et des connaissances existantes du domaine.

---

## **L'IA en santé environnementale : cas d'étude avec la chaleur extrême**

Jérémie Boudreault, INRS & INSPQ (Québec, Canada)

Les chaleurs extrêmes ont des impacts sanitaires importants qui seront amplifiés par les changements climatiques. Ces impacts doivent être prédits en temps opportun afin de bien protéger la santé de la population. Traditionnellement, la modélisation de ces effets à la santé reposait sur des approches épidémiologiques et/ou statistiques. Or, ces dernières années ont vu l'émergence croissante de techniques basées sur l'intelligence artificielle (IA) à cette fin.

Dans cette présentation, un survol des différents travaux effectués au Québec (Canada) visant à mettre à profit les approches d'IA pour prédire les effets à la santé de la chaleur extrême sera présenté. Les résultats obtenus ouvrent la voie à une surveillance améliorée des effets sanitaires de la chaleur par les autorités de santé publique.



# SESSION 3

## **L'IA générative au service de la démocratisation de l'accès aux données**

David Thoumas, Opendatasoft

**Comment exploiter de manière utile la puissance des technologies de l'IA et les mettre au service des utilisateurs, de manière raisonnable et contrôlée ?**

Opendatasoft illustrera sur la base d'exemples concrets comment ont été intégrées au cœur de sa solution de portail de données des fonctionnalités avancées de recherche sémantique et de génération interactive de visualisations.

---

## **Constituer un écosystème et mutualiser les corpus d'expertise pour élaborer une IA publique de confiance sur la transition écologique**

Philippe Guillouzie, Ademe

**L'ADEME, le Cerema et le CGDD mutualisent leurs corpus d'expertises sur la transition écologique pour élaborer une IA publique de confiance sur la transition écologique.**

La solution SOFIA permet aux utilisateurs internes d'interroger via un simple prompt l'ensemble des ressources ou de choisir un corpus spécifique. SOFIA génère une proposition de réponse à partir des meilleures sources identifiées et accessibles directement par l'utilisateur. Amélioration des performances, de la qualité des réponses, élargissement des partenaires et développement de nouvelles fonctionnalités sont au programme des chantiers 2025.

---

## **Perception et acceptabilité de l'Intelligence Artificielle : l'apport des sciences comportementales**

Paul Grignon, Ecolab (CGDD/SRI)

**De nombreuses composantes influent sur la façon dont l'Intelligence artificielle est comprise et perçue. En effet, les enjeux de perception et l'acceptabilité sont clés à considérer pour appréhender les modalités selon lesquelles l'IA est appréhendée et utilisée. Cette présentation mettra en avant des axes de réflexion concernant la façon dont les connaissances en sciences comportementales peuvent permettre de mieux prendre en compte la cognition humaine dans les différents défis posés par l'utilisation de l'IA et par l'interprétation de résultats qui en sont issus.**

