

# HACKATHON CONSULT'IN France

**Thème** : L'IA, ou intelligence artificielle : une révolution en marche....

L'arrivée de l'Intelligence artificielle (IA), qui recouvre un large champ de technologies, va engendrer l'une des plus grandes transformations que le monde du travail ait connue : ses premières réalisations suscitent d'immenses espoirs, et pourraient contribuer à hauteur de 15 700 milliards de dollars à l'économie mondiale en 2030, selon les experts du cabinet d'audit PwC<sup>1</sup>.

Afin de donner les meilleures chances aux entreprises françaises de rester dans la course, le secrétaire d'État au Numérique, Mounir Mahjoubi, a confié le 3 septembre 2017 une mission sur l'intelligence artificielle au député mathématicien Cédric Villani.

Celui-ci a rendu public son rapport *"Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne"* le mercredi 28 mars 2018, avec notamment la proposition suivante :

- « Concentrer l'effort économique et industriel sur quatre domaines prioritaires : la santé, les transports (mobilité), l'environnement (écologie) et la défense (sécurité). »

Dans ce contexte en pleine (r)évolution, les entreprises vont faire face à un enjeu technologique et managérial de taille : que ce soit le marketing, la maintenance, la logistique, le contrôle, ou les ressources humaines... toutes les activités seront concernées.

Plus que jamais, les cabinets de conseil en stratégie et management se positionnent comme des « partenaires », dont l'objectif est d'accompagner au mieux les dirigeants et leurs collaborateurs, en facilitant notamment l'intégration des solutions d'intelligence artificielle dans les entreprises.

Face à cet enjeu, Consult'in France<sup>2</sup>, l'instance représentative de la filière du conseil en stratégie et en management, a choisi de placer le sujet de l'intelligence artificielle au cœur de ses réflexions, en organisant un Hackathon autour des 4 axes prioritaires identifiés par le rapport Villani : Santé, environnement, transport et défense.

L'objectif est de réunir des étudiants issus de diverses formations durant deux jours, et de les faire réfléchir aux enjeux de l'IA dans ces quatre domaines. Les équipes seront pour cela encadrées par des consultants issus des cabinets membres de Consult'in France, qui les « coacheront » tout au long de l'évènement.

Afin de leur soumettre des problématiques concrètes, nous avons recensé les différents jeux de données accessibles directement en open source, et identifier les partenaires qui pourraient nous proposer des thématiques et les jeux de données associées.

---

<sup>1</sup> <https://www.pwc.fr/fr/espace-presse/communiqués-de-presse/2017/juillet/intelligence-artificielle-un-potentiel-de-15700-milliards-de-dollars.html>

<sup>2</sup> Consult'in France rassemble 83 cabinets (soit 60 % du marché) très diversifiés, tant par leur taille que par leur domaine d'expertise, qui interviennent pour des entreprises du secteur privé et public.

Consult'in France fait partie de la Fédération SYNTEC qui rassemble et représente les métiers de l'« intelligence productive ».

## Thématique Transports

Le secteur des transports se trouve au cœur des promesses de l'IA et plus particulièrement dans les domaines du véhicule autonome, de l'optimisation des flux, de l'inter-mobilité ainsi que de la maintenance. Après avoir établi un diagnostic des problèmes rencontrés par les acteurs de la mobilité, les équipes proposeront des solutions pour y remédier en soulignant leur proposition de valeur, leur impact, leur stratégie de mise en œuvre, leur communication ainsi que ses leviers d'adoption. Pour cela, les équipes disposent du cadrage suivant :

### 1. Champs d'application

- L'exemple le plus frappant d'utilisation de l'intelligence artificielle dans le domaine des transports correspond au développement de la voiture autonome. Six degrés d'autonomie sont possibles et aujourd'hui certains véhicules tesla ou DS sont classés au niveau 3. Ils peuvent ainsi prendre en charge le changement de file sur l'autoroute ou les manœuvres de stationnement.
- La maintenance des équipements et des chaînes de production sera fortement impactée par le développement de l'IA : maintenance prédictive, gestion des risques, réduction des coûts, limitation des immobilisations...
- L'IA se prête particulièrement bien à la gestion des flux de circulation : optimisation de la circulation automobile, réduction du nombre d'incidents, développement des transports multimodaux, optimisation des transports en commun...

### 2. Données accessibles

Les données accessibles sont proposées par les acteurs suivants (les dates entre parenthèses indiquent la date de mise à jour) :

- Mairie de Paris (<https://opendata.paris.fr>) :
  - Disponibilité en temps-réel stations Vélib' (indisponible actuellement) / Localisation des stations Vélib' (04/01/2018)
  - Disponibilité en temps-réel des voitures et bornes des stations Autolib' / Localisation des stations Autolib' (21/03/2016)
  - Accidentologie (27/05/2014)
  - Données trafic issues des capteurs permanents (22/06/2015)
  - Structure du réseau (trottoirs, panneaux, feux, barrières, zones de stationnement, chantiers...)
- Ile-de-France mobilités (<https://opendata.stif.info>) :
  - API disponible pour faire des requêtes en temps-réel sur les horaires de prochain passage, soit pour un arrêt, soit pour tous les arrêts (nombre de requêtes journalières limité)
  - Structure du réseau (référentiel des lignes, gares, stations, arrêts de bus...)
  - Données de validation d'abonnements (nombre de validation par lieu et par jour, profil horaire des validations) pour les années 2015, 2016 et 2017.
- RATP (<https://data.ratp.fr>) :
  - Horaires en temps réel
  - Trafic annuel entrant par station (2013 à 2017)
  - Qualité de l'air mesurée en direct à Châtelet, Franklin Roosevelt et Auber
- Région Ile-de-France (<https://data.iledefrance.fr>) :
  - Déplacements domicile-lieu de travail en 2014
  - Déplacements domicile-lieu d'études en 2014

- Accidentologie Paris entre 2012 et 2013
- SNCF (<https://data.sncf.com/>) :
  - Structure du réseau (gares, passages à niveau, voies, signalisation, chantiers...)
  - Horaires des trains
  - Régularité mensuelle des trains depuis 2013
  - Comptage des voyageurs montant dans les Transiliens (comptage manuel tous les 3 ans)
  - Enquêtes en gare sur les motifs de voyage, la répartition socio-culturelle des voyageurs...
  - API similaire à celle du STIF avec requêtes limitées
  - Incidents de sécurité (mis à jour chaque semaine)
- Gouvernement (<https://data.gouv.fr/>) :
  - Trafic aérien
  - Accidents de la route
  - Trafic routier

## Thématique Sécurité

Le secteur de la sécurité se trouve au cœur des promesses de l'IA et plus particulièrement dans les domaines de la prédiction des crimes et délits, de la réduction des accidents de la route mais aussi de l'optimisation des ressources. Après avoir établi un diagnostic des problèmes rencontrés par les acteurs de la sécurité, les équipes proposeront des solutions pour y remédier en soulignant leur proposition de valeur, leur impact, leur stratégie de mise en œuvre, leur communication ainsi que ses leviers d'adoption. Pour cela, les équipes disposent du cadrage suivant :

### 1. Champs d'application

- La robotisation des outils de défense et sécurité va grandement profiter du développement de l'IA. En effet, la reconnaissance physique de lieux automatisée sera suivie d'une robotisation des métiers à risque comme le déminage puis des combattants.
- L'IA va permettre d'accélérer la collecte et le traitement d'informations. Le nombre de données disponibles est amené à croître exponentiellement et les algorithmes vont permettre d'automatiser leur étude : visionnage de vidéos, reconnaissance faciale et vocale, répartition d'appels d'urgence, localisation...
- Le développement de l'intelligence artificielle va enfin permettre de prédire et de prévenir les actions des individus : détection de formation de réseaux dangereux, protection d'attaques réelles ou virtuelles, prédiction de délits, de crimes et d'accidents...

### 2. Données accessibles

Les données accessibles sont proposées par les acteurs suivants (les dates entre parenthèses indiquent la date de mise à jour) :

- Ministère des Armées (<https://www.data.gouv.fr/fr/datasets/?sort=-created&organization=534fff92a3a7292c64a77f94>) :
  - Évolution du recrutement des militaires (2001-2016)
  - Effectifs, dépenses, patrimoine, budget (2014-2017)
  - Dépenses par mission
- Ministère de l'Intérieur (<https://www.interieur.gouv.fr/Publications/Statistiques>) :
  - Accidents corporels de la circulation (2005-2016)
  - Points d'accueil de la police et de la gendarmerie
  - Effectif par commune de la police municipale
  - Crimes et délits enregistrés depuis 2012
  - Interventions des services d'incendie et de secours (2010-2016)
  - Délivrances de visas et titres de séjours (jusqu'à 2013)

## Thématique Santé

Le secteur de la santé se trouve au cœur des promesses de l'IA et plus particulièrement dans les domaines de la robotisation des interventions, de l'aide au diagnostic, de la médecine prédictive ainsi que de l'évaluation des politiques publiques. Après avoir établi un diagnostic des problèmes rencontrés par les acteurs de la santé, les équipes proposeront des solutions pour y remédier en soulignant leur proposition de valeur, leur impact, leur stratégie de mise en œuvre, leur communication ainsi que ses leviers d'adoption. Pour cela, les équipes disposent du cadrage suivant :

### 1. Champs d'application

- L'IA va révolutionner le diagnostic et la prise de décision médicale. Les algorithmes pourront interpréter des analyses médicales et détecter des pathologies. On pense ainsi à l'analyse des radiologies, la détection de pathologies médicales et psychologiques, l'aide à la décision des médecins, la détection d'effets secondaires ou encore de corrélations médicales.
- La médecine va devenir personnalisée grâce au développement des objets connectés qui permettront de suivre son état de santé en temps réel et d'avoir un avis médical automatisé. Les bracelets, balances, pacemakers et injecteurs connectés, couplés à des logiciels médicaux permettront une médecine préventive adaptée à chaque individu.
- Les robots, qui assistent aujourd'hui les opérations médicales, vont peu à peu remplacer les chirurgiens. De plus, les robots pourront intégrer le logement de certains patients qui ont besoin d'une assistance à domicile et assister les infirmières dans leurs tâches.
- Enfin, les politiques de santé publiques vont être améliorées grâce à l'analyse des données issues des services de santé : gestion des flux hospitaliers, lancement de campagnes de prévention ciblées, optimisation des dépenses de santé...

### 2. Données accessibles

Les données accessibles sont proposées par les acteurs suivants (les dates entre parenthèses indiquent la date de mise à jour) :

- Système National des Données de Santé (<https://www.snds.gouv.fr/SNDS/Open-Data>) :  
Le SNDS recense les données de plusieurs acteurs de la santé. Il est possible de faire une demande pour avoir accès à des données individuelles anonymisées. Néanmoins, cette demande prend du temps et doit être motivée.
- Direction de la recherche, des études, de l'évaluation et des statistiques (<http://www.data.drees.sante.gouv.fr/ReportFolders/reportFolders.aspx>) :
  - Répartition des séjours dans les établissements de soins de courte durée selon sexe, âge, pathologie et durée du séjour.
  - Indicateurs de natalité : effectif annuel des naissances par département, taux de mortalité, effectif annuel des naissances prématurées, taux annuels de poids des nouveau-nés, part des naissances multiples, classes d'âges des mères.
  - Diagnostic d'accès aux soins urgents (31/12/2015)
  - Statistiques annuelles des établissements de santé (position, capacité...)
  - Données sur les patients s'étant présentés aux urgences le 11/06/2013 : heure d'arrivée, âge, patientèle, raisons, parcours intra-hospitalier.
- Assurance Maladie (<http://open-data-assurance-maladie.ameli.fr/index.php>) :
  - Open Damir : ensemble des dépenses d'assurance maladie interrégimes (2009-2017) annuelles selon période, prestation, organisme de prise en charge, bénéficiaire, exécutant, prescripteur, montant et quantité.
  - Open Medic : dépenses annuelles de médicaments en officine de villes (2014-2017) en fonction du profil du client (tranche d'âge, sexe, région de résidence)

- Open PHMEV : prescriptions hospitalières annuelles de médicaments en fonction du profil du patient (2014-2017)
- Open Bio : dépenses annuelles de biologie médicale (2014-2017) classées par acte, tranche d'âge, sexe, région de résidence, prescripteur.
- Centre d'épidémiologie sur les causes médicales de décès (<http://www.cepidc.inserm.fr/inserm/html/index2.htm>) :
  - Effectifs de décès et taux de mortalité par département et grande ville depuis 1979 selon cause, sexe, tranche d'âge
- ScanSanté (<http://www.scansante.fr/>) :
  - Parts de marché consommation et production des soins
  - Statistiques des actes
  - Flux entre établissements
- Gouvernement (<https://data.gouv.fr>):
  - Base de données publiques des médicaments
  - Indicateurs avancés sanitaires quotidiens baclofene et dépendance alcoolique, incidence de la gale, incidence des poux, manifestations allergiques, gastro-enterite, syndrome grippal depuis 2009.

## Thématique Environnement

L'environnement doit profiter de l'IA et plus particulièrement dans les domaines de l'agriculture, de l'énergie et du développement durable. Après avoir établi un diagnostic des problèmes rencontrés par les acteurs de la santé, les équipes proposeront des solutions pour y remédier en soulignant leur proposition de valeur, leur impact, leur stratégie de mise en œuvre, leur communication ainsi que ses leviers d'adoption. Pour cela, les équipes disposent du cadrage suivant :

### 1. Champs d'application

La protection de l'environnement et le développement durable sont des problématiques multi-sectorielles. Les champs d'application sont donc beaucoup plus larges que pour les trois autres thématiques du rapport Villani. Néanmoins, on peut identifier certains secteurs où l'IA est particulièrement pertinente :

- Dans l'agriculture, l'IA va permettre non seulement une robotisation des outils mais également une optimisation de l'utilisation des produits chimiques ou de l'utilisation des terres. De plus, l'IA va permettre l'émergence de nouveaux circuits de production plus courts qui tiendront compte des besoins et de la production en temps réel pour répartir les flux de consommation.
- L'IA aura un impact énorme dans les domaines de l'énergie et de sa distribution. Les réseaux vont devenir intelligents et permettront de produire l'énergie au plus proche du lieu de sa consommation, tout en tenant compte des difficultés de stockage pour les renouvelables afin de limiter les pertes.
- Les transports vont être révolutionnés par l'IA qui va permettre d'optimiser les flux de transport pour réduire la pollution. De plus, des moyens de transport automatisés et partagés vont émerger, ce qui permettra de désengorger les voies de circulation tout en étant plus pratiques pour les usagers.
- Cependant, l'émergence de l'IA et la collecte des données nécessaires à son fonctionnement pose une problématique en termes de consommation énergétique car ce sont des systèmes gourmands en énergie, ce qui pose des problèmes de développement durable.

### 2. Données accessibles

Les données accessibles sont proposées par les acteurs suivants (les dates entre parenthèses indiquent la date de mise à jour) :

- Entrepôt d'indicateurs et de données sur l'environnement (<http://www.stats.environnement.developpement-durable.gouv.fr/Eider/>) : dépassement des seuils de pollution (2000-2014), consommation d'énergie par secteur d'activité (1990-2009), parc éolien, production d'énergie (1990-2009), immatriculations de véhicules (1993-2014)
- Enedis (<https://data.enedis.fr/explore/?sort=modified>) : consommation et production électrique toutes les demi-heures, infrastructure du réseau
- Ministère de l'agriculture (<https://www.data.gouv.fr/fr/datasets/?sort=-created&organization=534fff8ca3a7292c64a77edf>) : liste des exploitations agricoles et agro-alimentaires, liste des routes de France
- Engie (<https://opendata-renewables.engie.com/explore/?sort=modified>) : production des turbines éoliennes du parc de la Haute Borne depuis 2013.
- GRDF (<https://opendata.grdf.fr/pages/accueil/>) : consommation de gaz, quantités acheminées à la France
- Les bases de données sur les transports peuvent être trouvées sur la note des transports.

- Le hackathon C3 lancé à l'occasion de la COP21 en 2015 présente un grand nombre de bases de données disponibles sur les thématiques de l'environnement (<http://c3challenge.com/donnees/>)



## Annexe

### 1. Licences d'utilisation des données

Plusieurs licences sont proposées et chaque jeu de données possède sa propre licence :

- Open Licence Etalab qui permet de reproduire, copier, transmettre, diffuser, adapter, modifier, extraire, transformer et exploiter à titre commercial sous réserve de mentionner la paternité des données (source et date de dernière mise à jour).
- ODbI permet de copier, distribuer, utiliser, créer, modifier, transformer et construire à partir des données sous réserve de mentionner la paternité, partager aux termes identiques et garder les données ouvertes en cas de redistribution sous version modifiée
- Creative Commons permet de copier, distribuer et communiquer les données en mentionnant la paternité, sans les modifier ni les utiliser de façon commerciale
- SNCF Open Data permet de reproduire, d'extraire, distribuer, communiquer, diffuser les données sous réserve d'intégrer une copie du texte de la licence, que les données diffusées respectent les mêmes termes.