Etiquetage

Background :

* Les données sont les choses les plus demandées sur le marché
* Les progrès de l’IA montrent que celle-ci va continuer à avoir une place de plus en plus importante sur le marché
* Le besoin primaire de l’IA est la labellisation de données pour pouvoir établir des modèles d’apprentissage supervisé
* Ces IA aujourd’hui souffrent d’un manque de données labelisées pour tout un tas de cas spécifique ou alors la labélisation n’est pas suffisamment précise et le résultat après entraînement n’est pas optimale
* Une intelligence ne nécessitant que quelques données pour apprendre est encore loin devant nous

Apprentissage supervisé

Avantages :

* Permet d’avoir des modèles très performants et d’apprendre de façon optimisée
* Va rester la méthode la plus efficace pendant encore de nombreuses années

Inconvénients :

* Nécessite une très grande quantité de données labellisées ce qui est dur à trouver sur des cas spécifiques

Apprentissage auto-supervisé

Avantages :

* Ne nécessite pas de données labellisées et peut s'entraîner sur des données brutes afin d’en tirer des paternes et de la classification basée dessus

Inconvénients :

* N’est pas aussi performant que les modèles basés sur de la donnée labellisée
* Est plus compliqué à rendre spécifique et ne peut pas répondre à des problèmes aussi larges qu’avec de l’apprentissage supervisé

Idée :

* Proposer une plateforme où chacun pourrait étiqueter des données fournies par des clients

Objectif : Avoir de la labélisation aussi efficace que possible, assurer par une communauté de confiance créer au fil du temps et répartie dans des catégories d’activités pour bénéficier du savoir-faire des gens (activité rémunérée par l’entreprise)

Compétition :

* [Amazon Mechanical truck](https://www.mturk.com/) : Propose à des HIT workers de créer un compte et ensuite d’être assigné à des tâches plus ou moins durs dépendamment du niveau d’études de la personne lorsqu’elle a apply pour le job (compagnie qui se rapproche le plus du produit)
  + Ils outsource le travail à des milliers de personnes qui sont qualifiées pour le travail de labélisation demandé
  + Rémunère les online workers 1$ - 6$ de l’heure
  + Est basé sur la mentalité du job sur le côté qui te permet de gagner plus d’argent
  + Sont une marketplace pour que la demande clients puisse arriver à être traiter par un exemple d’individus externes
  + Philosiophie complètement crowdsourcing
* [Innovatiana](https://www.innovatiana.com/nos-services/labellisation-computer-vision) : Equipe de Data Labelers qui s’occupe de l’étiquetages et labellisation des données pour les modèles de clients.
  + Points forts :
    - labellisation de qualité avec des outils optimisés : CVAT, YoloV7, Instance Label studio, Roboflow, Lable Studio, LabelBox, Prodigy etc.)
    - peuvent renvoyer ces données dans plusieurs types de format (JSON, XML etc.) via des canaux sécurisés
    - vont eux-mêmes collectés les données dites de qualité, ne veulent pas faire du crowdsourcing mais du outsourcing éthique
    - Nombreux domaines : Mode/Luxe, Immobilier, Banque, Santé, Tech
  + Points faibles :
    - Ont besoin d’une équipe en permanence pour accomplir les tâches
* [Data Labeler](https://www.datalabeler.com/) : Entreprise située à New York et qui s’occupe de fournir une labélisation de qualité et expéditive pour des modèles d’IA
  + Points forts :
    - Ils ont une équipe full-time de data-labelers qui crée aussi le dataset nécessaire pour répondre aux besoins de l’entreprise
    - Type d’industries : Medical, Agriculture, Consumer, Transportation, Insurance, Space, Robotics, Manufacturing
  + Points faibles :
    - Ne fait que du boxing : bounding boxes, polygons, points, text for image captioning, select, semantic segmentation
* [Lionbridge](https://www.lionbridge.com/content-creation-services/elearning-services/) : spécialisé dans le data labelling de données linguistiques pour de la traduction
  + Ils ont une base de données très poussée basé sur plus de 20 ans de recherche
  + Domaines : life science, games, technology, legal services, automotive, travel, retail, consumer packages, e-commerce
* [Appen](https://appen.com/what-we-do/#Cases) : ils font vraiment beaucoup de choses