Développer et entraîner un algorithme

05 avril 2022

Mettre en place les bonnes pratiques lors de cette phase cruciale.

Le fournisseur du système d'IA devra se confronter à une série de vérifications et précautions afin de garantir la qualité du système obtenu. Toutefois, les questions qui suivent sont également d'intérêt pour l'utilisateur du système (s'il n'est pas le fournisseur) dont la responsabilité pourra être engagée en cas de non-conformité au RGPD du traitement mis en œuvre. Il pourra ainsi vérifier le niveau de prise en compte par le fournisseur des questions de protection des données lors de la conception du système.

Concevoir et développer un algorithme fiable

Lors de la conception du traitement, le choix de l'algorithme, des outils et de l'infrastructure de développement doit faire l'objet d'une attention particulière afin de parvenir à un traitement fiable et robuste.

Les questions suivantes pourront aider le responsable de traitement à apprécier s'il maintient un équilibre entre la complexité des solutions choisies et la perte en explicabilité.

☐ Quel type d'algorithme est utilisé et quel est son principe de fonctionnement (apprentissage supervisé, non-supervisé, continu, fédéré, par renforcement, etc.) ?

Pour quelle raison cet algorithme a été choisi ?

Comment a-t-il été mis en œuvre (source du code, librairies utilisées, etc.) ?

☐ L'algorithme utilisé a-t-il été testé par des tiers indépendants ?

☐ Une revue de la littérature a-t-elle été effectuée ?

☐ Un comparatif aux systèmes analogues a-t-il été réalisé ?

Sur la base de quels critères ?

☐ Des outils tiers sont-ils utilisés ?

☐ Sont-ils réputés fiables et éprouvés ?

☐ Une recherche des possibles défauts dans la documentation des outils et parmi la communauté de développeurs a-t-elle montré des points d'intérêt ?

☐ Un suivi des mises à jour des outils est-il prévu ?

☐ L'algorithme d'IA est-il disponible en open source ?

☐ Les algorithmes ont-ils été suffisamment testés par des tiers ?

Si oui, par qui et par quels moyens ?

☐ Les algorithmes correspondent-ils à l'état de l'art ?

☐ Les retours de la communauté à son sujet sont-ils encouragés et étudiés ?

Mener un protocole d'entraînement méticuleux

En établissant un protocole pour l'entraînement du système d'IA, le fournisseur pourra remettre en question le choix des méthodes de validation de la performance et de l'équité de l'algorithme, tout en intégrant des tests de validation aux étapes les plus critiques. Les métriques d'entraînement et de validation, en ce qu'elles offrent la première fenêtre d'observation sur l'algorithme doivent être choisies avec attention : limiter le taux de faux négatifs dans un test de séropositivité pour une maladie contagieuse est plus important que celui de faux positifs par exemple.

Quelles sont les stratégies d'apprentissage mises en œuvre ?

Quelle répartition a été utilisée entre ensembles d'entraînement, de test, de validation ?

☐ Des stratégies telles que par exemple la validation croisée sont-elles utilisées ?

☐ La qualité des sorties du système d'IA est-elle jugée suffisante ?

Quelles métriques sont utilisées ?

☐ Permettent-elles de mesurer la performance du système de manière satisfaisante et en prenant en compte les conséquences pour les personnes concernées ?

☐ Les cas d'erreur ont-ils été étudiés ?

☐ Une corrélation a-t-elle été recherchée avec une des variables des données ?

☐ Une corrélation a-t-elle été recherchée avec une des variables des données ?

☐ Les situations limites où les sorties du système ne sont pas suffisamment fiables sont-elles clairement identifiées ?

☐ Une sécurité est-elle ajoutée pour prendre en charge ces cas (en rendant systématiquement la main à un opérateur humain par exemple) ?

☐ Cas de l'apprentissage en continu : des mesures sont-elles prises en amont pour éviter une dégradation de la performance, une dérive du modèle ou une attaque visant à orienter les résultats de l'algorithme (par exemple des *chatbots* en ligne devenus « racistes ») ?

Lesquelles ?

Vérifier la qualité du système en environnement contrôlé

Les tests sur l'algorithme et le système dans son ensemble doivent être réalisés dans des conditions représentatives et permettant une validation exhaustive du traitement.

- [ ] Le traitement a-t-il été validé par une expérimentation ?
- [ ] Le contexte dans lequel évolue le système d'IA (quantité de variables à considérer, difficulté à évaluer la représentativité des données, etc.) est-il particulièrement complexe ?
- [ ] L'algorithme en question a-t-il déjà été utilisé dans un contexte similaire ?
- [ ] Dans quel environnement a été menée l'expérimentation (contrôlé/non-contrôlé ? fermé/ouvert ? sur des cas d'usage simulés/réels) ?
- [ ] Les conditions se rapprochent-elles suffisamment des conditions réelles ?
- [ ] Les précautions appropriées ont-elles été prises, comme un contrôle systématique par un opérateur humain qui pourrait être ensuite réduit lors de la phase de production ?
- [ ] Les métriques utilisées pour valider l'expérimentation sont-elles adaptées et suffisantes ?
- [ ] Un bilan exhaustif et objectif de l'expérimentation a-t-il été réalisé ?

Imprimer cette check-list

< Fiche précédente : collecter et qualifier les données d'entraînement

Sommaire

Fiche suivante : utiliser un système d'IA en production >

Vous souhaitez contribuer ?

Écrivez à ia[@]cnil.fr

Haut de page