Journée ARISTOTE du 15 novembre 2018

Proposition de questions de la table ronde sur l'explicabilité de l'IA…

Bruno Teboul (Ph.D)
Gfi Informatique
Global Practice Director – Artificial Intelligence & Blockchain

---

D'un point de vue des besoins
1/ Comment définir l'explicabilité de l'IA à travers d'applications concrètes ?

Les besoins pour les entreprises qui intègrent des plateformes embarquant de l'IA sont de plusieurs natures.

Si l'on accepte de déléguer des tâches à un système « intelligent » ont doit être capable à tout moment de l'auditer, de le questionner sur l'origine des résultats ou sur les choix préconisés.

Un système d'IA auditable (idéal) est en mesure de répondre aux questions clés de l'explicabilité :

- Pourquoi fais-tu cela ?
- Pourquoi ne pas appliquer une autre méthode ?
- Quand ta méthode fonctionne-t-elle ?
- Quand échoue-t-elle ?
- Quand puis-je te faire confiance ?
- Comment corriger une erreur ?

Ce n'est pas encore le cas pour la majorité des solutions embarquant de l'IA. Cela dépend en fait fortement des solutions algorithmiques mises en œuvre : moteurs de règles, systèmes experts revisités, apprentissage supervisé ou non, simple régression, Foret d'arbres décisionnels (Random Forest) qui permettent de l'auditabilité à un bon niveau. Tout dépend de la technologie se cachant derrière le terme générique IA.

Concrètement, tous les domaines d'application sont concernés par ce besoin d'explicabilité de l'IA :

Le transport (semi-autonome, autonome)

La sécurité des données : traçabilité, intégrité, certification

La médecine : choix d'un process thérapeutique plutôt qu'un autre, validation à postériori par une expertise humaine.

La finance : en particulier le trading haute et moyenne fréquence > nécessité de traçabilité et d'explicabilité dans le choix d'une stratégie de trading, qu'elle soit positive ou négative pour celui qui la met en place.

Le juridique : dans l'aide à la décision et l'analyse automatique des grands volumes documentaires juridiques.

le secteur militaire – défense : robotique militaire et aide à la décision. Autonomie des systèmes armés : comment déléguer l'ordre au système sans être capable de répondre aux 6 questions primordiales ?

D'un point de vue des enjeux (aspects réglementaires, éthiques, …)
2/ Quelles sont les enjeux industriels et sociétaux autour de l'explicabilité de l'IA ?

Les succès de l'apprentissage machine automatique, de l'apprentissage profond… liés à l'essor de l'approche connexionniste en IA (vs IA Symbolique) nous conduit au développement, d'applications en IA de plus en plus rapide, robustes et autonomes. Ces systèmes complexes sont capables de traiter une multitude d'informations visuelles, sonores, lexicales, sémantiques et établir des inventaires de correspondances congruents entre les éléments et situations réels et ceux représentés au sein des machines comme chez les humains. Dans ce contexte, les machines sont capables de très rapidement et avec un taux d'erreur très faible de décider et le cas échéant d'agir dans des délais très rapides. Ces dépassements des simples capacités cognitives humaines engendrent la création de machines autonomes, dont le contrôle ne peut être fait qu'à posteriori, dans des temps compatibles avec ceux de la lenteur de l'appareillage informatique et/ou cognitif humain. Cela pose alors le problème du blocage de l'efficacité des systèmes par l'impossibilité que l'homme a de faire confiance à la machine numérique ("black box society"). On connait ce type de situation dans le domaine du véhicule autonome, du nanotrading et plus encore dans celui des SALA (systèmes d'armes létales autonomes).
Par ailleurs, l'Europe s'étant dotée du RGPD dont les obligations sont parfois en opposition avec le mode de fonctionnement des solutions actuelles de Machine Learning et de Deep Learning. Il est urgent d'établir des normes réglementaires, des règles d'éthique, pour rendre possible la traçabilité des process et de la chaine d'information au sein de l'entreprise (pas de boite noire) qui créé de l'incertitude et du doute raisonnable ou non.

D'un point de vue des solutions
3/ Travaillez-vous cette question actuellement et si oui, comment, avec qui, dans quel écosystème ?
Quelles sont les pistes scientifiques et techniques pour apporter de l'explicabilité à l'IA ?

"L'explicabilité" du traitement algorithmique et du fonctionnement de l'IA deviennent tout l'enjeu de la confiance, de l'éthique et de la sécurité numérique. La solution est une forme d'apprentissage collaboratif, partenarial, de l'homme et de la machine, dans laquelle la machine numérique est capable d'expliquer à l'homme le sens de ses décisions et les limites de son autonomie ; ce que ne savent pas faire les systèmes actuels. De tels systèmes symbiotiques (HAT pour Human-Autonomy Teaming) nécessitent une Intelligence Artificielle capable de s'expliquer. Les machines décisionnelles, de contrôle, autonomes, devront alors s'inscrire dans un projet global d'intégration bidirectionnelle hommes-machines, de type XAI (Explainable Artificial Intelligence). GFI Informatique met en place une stratégie IA et Blockchain très innovante en la matière. Nous allons nouer des partenariats académiques de premier plan pour soutenir l'enseignement et la recherche en IA. Nous allons associer notre FabLab et notre Direction R&D qui comprend déjà une quinzaine de scientifiques de haut niveau : Doctorants, Ph.D, post-docs et des startups, pour développer une approche « explicable-auditable » de l'IA dans le cadre de projets de co-innovation avec nos clients.

D'un point de vue de la perspective
4/ Quels sont les manques ?
Que seraient le réseau d'excellence et l'écosystème innovant sur cette thématique ?

Les manques : les plateformes actuelles de deep Learning ou de machine learning sont souvent incapables de répondre simplement aux six questions primordiales de l'IA. Des programmes sont lancés aux USA (par exemple le programme XAI de la DARPA) mais cela reste globalement insuffisant.

C’est la recherche qui va apporter des solutions opérationnelles à l’explicabilité (R&D). Il faut donc encourager et soutenir financièrement la recherche en ce sens !

Au niveau national, le rapport Villani a mis l’accent sur les exigences d’explicabilité de l’IA. Ces exigences sont répétées tout au long du rapport. Il existe également des initiatives comme le Hub FranceIA qui existe depuis un an avec l’objectif de fédérer l’écosystème française de l’IA. Le sujet de l’explicabilité fait partie des thématiques adressées par les différents groupes de travail du Hub et par les projets lancés (AMI Bercy sur la mutualisation des datasets).

Y-a-t-il d’autres points que vous souhaiteriez aborder ?

Nécessité de former des spécialistes de l’IA et de son explicabilité. Encourager et financer la recherche en IA auprès des grands groupes (conventions de mécénat, chaires, laboratoires communs…). Il existe encore trop peu de formation au niveau Master en France. Le Master MVA de l’ENS, Le Master Data scientist de l’X d’Erwan Lepennec, le Master Artificial Intelligence & Visual Computing de l,X de Marie-Laure Cani, ou encore le Master MASH de Dauphine dirigé par Marc Hoffmann sont des Masters d’excellence, mais ce n’est pas suffisant à l’échelle nationale. Il faut ouvrir des formations de ce type dans chaque région française, avec le soutien des grands groupes et des startups.