Thème: IA et éthique

Résumé de notes prises lors de la réunion du « Cercle d'échanges de l'AEFR » du 9 mai 2023.
A l'invitation de Sopra Steria.

Les débats entre les membres de l'AEFR, accueillis chez Sopra Steria par Pierre Lahbabi, Directeur des services financiers du Groupe, ont été introduits par Didier Valet, Directeur Général de l'Institut Louis Bachelier et président du Conseil d'orientation des cercles d'échanges de l'AEFR, qui a souligné l'actualité relancée du thème depuis l'apparition aux yeux du grand public de systèmes d'IA générative.

Bertrand Pailhès, Directeur des technologies et de l'innovation de la Commission Nationale de l'Informatique et des Libertés (CNIL), a ouvert les échanges en confirmant combien ces questions connaissent, depuis quelques mois, un très vif regain d'intérêt public, alors qu'elles ont été l'objet d'initiatives importantes avec le plan Obama lancé aux Etats-Unis en 2016 ou des actions entreprise en France suite au rapport Villani de 2018. Ainsi, la CNIL avait-elle été chargée en 2017 d'une mission sur les algorithmes, le HLEG (Hight-level expert group) constitué par la Commission européenne avait dégagé des principes qui reposent sur des concepts élaborés au cours des années 70, mais qui demeurent (robustesse, fairness, transparence) repris notamment dans le rapport Villani. L'enjeu est aujourd'hui de passer des principes à une mise en œuvre opérationnelle. En France, un Comité national pilote d'éthique du numérique vient d'être pérennisé. On note une activité très intense en particulier aux Etats-Unis, et une concurrence forte entre organisations, y-compris publiques (OCDE, Unesco, Conseil de l'Europe…) pour proposer un cadre éthique pour ces innovations. Il a donné, comme exemple, les questions éthiques soulevées par la reconnaissance faciale.

Yves Nicolas, Deputy Group CTO de Sopra Steria, a alors exposé que son groupe avait naturellement comme approche l'accompagnement de ses clients, et était notamment parmi les fondateurs d'un groupe "Confiance.AI" auquel participaient des entreprises industrielles. La réflexion part donc de cas d'usage, comme la génération précédente de systèmes informatiques, et tente de discerner les biais qui peuvent peser sur les résultats. L'un des axes est de travailler les algorithmes d'explicabilité des réponses fournies, et des bibliothèques de tels programmes se mettent progressivement en place.

Un débat nourri et actif s'engage, entre les participants, nourris de nombreux exemples, sur des sujets divers dont certains sont précisés ici.

- Des enjeux nouveaux sont soudainement apparus, renouvelant des sujets classiques.

L'approche du règlement européen sur l'IA en cours d'élaboration entend classer désormais les algorithmes en fonction des risques sur une échelle de 5 (interdits) à 1, le 4 étant les "hauts risques".

De telles analyses peuvent s'appliquer au credit scoring qui était, il y a dix ans, géré par des systèmes déterministes aujourd'hui enrichis, mais tout autant à la reconnaissance faciale, et surtout à la sécurité des produits de toutes natures et dans tous les domaines. Cette grille nouvelle est ainsi travaillée pour s'appliquer à la reconnaissance faciale dans l'espace public, interdite en Europe sans le consentement des personnes. De même le "crédit social" tel que mis en œuvre en Chine serait-il interdit en Europe.

- La question « Comment le régulateur pourra-t-il auditer ces modèles et attirer les compétences nécessaires pour ce faire? », revêt alors une position centrale.

On peut à cet égard cependant, dresser un parallèle entre le RGPD (que les entreprises mettent désormais en œuvre) ou l'autorisation de mise sur le marché (AMM) dans l'industrie pharmaceutique et le règlement sur l'IA, dans la logique de ce dernier d'un auto-contrôle indispensable à mettre en place par le producteur avant toute mise sur le marché, processus dont l'irrespect ou la défaillance fait encourir des sanctions lourdes, par exemple, en l'état actuel des législations nationales, en Italie et en France.

L'applicabilité d'une telle logique pour des produits dématérialisés est plus complexe ; mais il est probable que les clients, et notamment les plus grands, n'achèteront pas un produit s'ils n'ont pas la conviction que celui-ci est "conforme".

- Le caractère singulier de l'IA générative est de fournir une réponse plausible, mais non pas certaine. A contrario, pour les solutions d'IA "classique"pour lesquelles l'utilisateur cherche une réponse (quasi)-certaine pour permettre un gain de temps, la problématique principale est celle de la validation des produits. Se pose alors la question de la désirabilité d'une "offre souveraine"(par exemple le projet « Mistral.AI », en France).
- Il est manifeste que l'éthique dépend de la géographie (Europe, US, Chine…), ce qui ouvre alors le débat sur les effets des régulations future sur l'évolution de la compétitivité de la zone, que la discussion applique particulièrement aux projets européens de réglementation stricte. On peut rappeler ici que le RGPD ouvrait le même type de questionnement et, qu'à l'expérience de ce dernier sur maintenant une dizaine d'années, il n'est pas évident qu'une règle stricte génère forcément un handicap de compétitivité. Ceci étant, il convient de souligner qu'une réglementation sévère avantage les grandes structures, et donc les grandes plateformes que celles-ci peuvent mettre en œuvre dans leur propre intérêt.
- Plus généralement, tout gain de productivité conduit à de nouveaux équilibres dont le résultat peut être faste, ou néfaste en termes d'activité et de production. S'il est évident que prétendre interdire l'exploitation de gains de productivité n'a pas de sens, la question des risques sur l'emploi reste majeure. La discussion l'applique en particulier aux métiers bancaires et plus largement financiers et juridiques.
- Tous s'accordent sur le fait que la vitesse de transformation est exceptionnelle. Et que la quantité sinon la qualité de commentaires et d'analyses a explosé.

Il apparait que tous les intervenants sont à la recherche d'éléments de cadrage aujourd'hui manquants, tant la vitesse de développements de l'IA générative bouleverse les référentiels classiques et ouvre des problématiques nouvelles ; des débats intenses ont ainsi lieu dans l'espace public, mais aussi au sein de la plupart des entreprises et structures.

Les positions sont souvent très tranchées et poussées au bout de la logique. Est ainsi donné l'exemple du combat mené par Google sur le droit à l'oubli français poussé jusqu'en CJCE, dont la conclusion par la Cour a mis en lumière que l'on ne pouvait pas intervenir sur le marché européen de la même manière que sur celui des Etats-Unis.

- A partir des développements perceptibles en Chine s'ouvre alors la question des bases d'apprentissage éligibles, illustrée par l'exemple frappant d certains systèmes experts en cardiologie développés en 2018 sur des données chinoises, et donc aujourd'hui non utilisables en Europe.

La problématique se translate ainsi sur la qualité et la licéité des données-sources. Et leur exploitation, avec ou sans IA.

C'est ainsi que sont comparés par un participant le "modèle Amazon" (l'IA pilote les humains) et le modèle où l'opérateur "garde la main", comme exemple la pratique de la MAIF avec son "conseiller augmenté", et qu'il est rappelé qu'en fait la question se posait également, quoique différemment, dans les générations de systèmes précédentes.

- Il reste que selon le niveau d'IA et le secteur considéré les résultats sont très inégalement explicables, du fait-même que les degrés de sécurité acceptés dans des disciplines différentes sont très divers. Sont ainsi comparés ce qui est accepté en matière d'expérimentations médicales par opposition aux règles du trafic aérien. Le portail LINC ouvert par la CNIL présente dans ce champ de réflexion un cadre utile.
- Est alors abordée la question de la gouvernance, en particulier dans les entreprises à mission. Les établissements bancaires ont beaucoup travaillé la question, et celle-ci se pose, au fond, indépendamment de l'IA, comme le démontrent les différences d'approche suivant les grands offreurs de service américains dans leurs engagements éthiques sur l'IA. Alors que Google, par exemple, s'engage à ce que l'usage de ses services reste conforme à son éthique, AWS en laisse toute la responsabilité à ses clients.

Le site "éthique et IA"de Microsoft, extrêmement fourni, est très illustratif de ces problématiques.

- Pervenche Berès, Présidente de l'AEFR, pose les dernières questions. « Comment les entreprises s'approprient-elles ces questions au-delà de leurs procédures de gouvernance ? » ; « La CNIL va-t-elle utiliser des IA pour détecter les IA défaillants ? » ; « Est-ce que le futur règlement sur l'IA pourrait devenir un référent international, à l'instar du RGPD, ou plutôt un handicap en termes de compétition pour l'UE ? ».

Il lui est répondu, notamment, que "le règlement IA a du sens parce qu'il pose des interdictions ". Est rappelée à cet égard, par contre-point, l'expérience du Comité National d'Ethique lors de la Covid: la CNIL n'a pas été

saisie… Car il y a eu alors peu d'IA mobilisée, et les modèles classiques ont été reproduits - il est vrai que l'on n'avait pas encore, alors, les données d'une pandémie, ce qui sera le cas lors de la pandémie suivante…
La conclusion logique désormais générale apparaît ainsi : "notre problème est de savoir poser les questions !"

… ce qui était bien le point d'accord entre tous les participants, repris lors des échanges informels qui ont pris alors corps et se sont prolongés, grâce à l'hospitalité de Sopra Steria.