

Synthèse par thèmes

Politique globale de la France en matière d'IA

Une compétition internationale féroce fait rage depuis plusieurs années pour attirer les cerveaux et les garder au sein des industries ou centres de recherche nationaux. Mais les français semblent être absents de cette course à l'IA. Toutefois, si les progrès accomplis restent insuffisants un an après la parution du rapport Villani, les projets entrepris ont été nombreux.

L'objectif défini par le plan national de recherche en IA était en effet que la France figure parmi les cinq premiers pays au monde en matière d'IA, et d'être la première puissance en Europe. Pour atteindre ce but, divers travaux ont été menés autour des trois piliers du plan national : talents, IA dans l'économie et la société, et dimension sociale et éthique. Concernant le volet recherche, la mesure la plus visible a été la mise en place des Instituts 3IA, pour répondre à l'ambition formulée par le Président de la République de développer des centres de recherche, d'éducation et de développement économique en IA à forte visibilité mondiale.

Alors même que le vol des données et l'espionnage font rage ces dernières années, la France n'est pas trop mal placée dans la compétition internationale dans ce domaine précis, et les technologies d'IA sur lesquelles se concentre l'attention aujourd'hui sont principalement les technologies de reconnaissance de formes ou dédiées à des applications bien spécifiques (domaine du jeu par exemple).

Par ailleurs, l'IA a vocation à être utilisée par les militaires pour gérer les systèmes d'armes de façon plus réactive et précise, dans un contexte marqué par une interconnexion croissante des systèmes militaires et une multiplication des données. La Ministre des Armées Madame Florence Parly l'a d'ailleurs définie comme une technologie stratégique indispensable pour garantir la supériorité opérationnelle. C'est pourquoi l'IA a été érigée au rang de priorité pour la défense nationale, une attention particulière étant portée à son application dans le champ de la conduite des opérations (aide à la décision), du renseignement, de la mutualisation collaborative et du cyberspace, entre autres.

Dimension européenne et comparaison internationale

La France et l'Europe bénéficient de plusieurs atouts sur lesquels s'appuyer dans la course à l'IA : une excellence scientifique historique, un environnement globalement favorable aux investisseurs, et une bonne association entre le public et le privé. Toutefois, ces atouts ne suffisent pas à surmonter les difficultés auxquelles l'Europe et la France font face, à savoir en premier lieu un manque d'interdisciplinarité et d'intégration du business, couplé à un manque d'intégration entre la BITD et la recherche. D'autre part, l'UE souffre de problématique d'échelles (échelles de marché, de culture, d'industrialisation), et d'un retard capitalistique conséquent en matière de digital. Il faudrait, pour pallier cela, établir en Europe une CNIL européenne et une union bancaire aboutie, afin de favoriser le véritable développement de l'IA au niveau supranational. Enfin, à ce problème capitalistique s'ajoute un problème culturel aux lourdes conséquences : la fuite des cerveaux à l'étranger. La réalité balkanique de l'écosystème du digital et des traitements algorithmiques européens incite par conséquent au pragmatisme et au réalisme, pour entamer la phase de rattrapage des géants du numérique, tout en acceptant que les américains et les chinois nous domineront encore quelque temps dans le domaine. Pour entamer ce rattrapage, la stratégie européenne repose actuellement sur un cadre commun de perspectives, et passe par l'initiative JEDI (*Joint European Disruptive Initiative*). JEDI se veut l'équivalent de la DARPA américaine, en créant un outil à mi-chemin entre la recherche fondamentale et l'utilisation pratique sur le marché et dans les forces. Son ambition est de financer les projets aujourd'hui trop risqués ou s'inscrivant dans un temps trop long que le secteur privé ne s'y engage. L'enjeu de

JEDI est central, car les technologies sont devenues le vecteur de nos valeurs, et *in fine*, la croissance, la prospérité, et la force des économies et sociétés européennes dépendent de ces technologies. Avec JEDI, c'est donc d'un véritable leadership technologique que l'on souhaite doter l'Europe.

Au niveau international, ces dernières années ont été le théâtre d'une véritable course à l'IA à l'échelle planétaire. Dans cette compétition entre des acteurs de différentes envergures, les divergences structurantes qui apparaissent entre les pays concernent surtout les masses monétaires et les niveaux d'investissements mis en jeu (la Chine se démarquant de loin comme le pays investissant le plus en IA), les orientations déterminées dans les stratégies d'IA (servir la défense, l'industrie, la recherche), et le degré de soutien des citoyens et du gouvernement. Les acteurs majeurs en terme d'IA semblent ainsi être les Etats-Unis – qui ont placé l'IA au cœur si bien de la politique que de l'économie, le Canada qui a investi très tôt dans le domaine, la Chine avec son ambition affichée de reprendre une position de leader dans le domaine via des investissements massifs et un soutien de sa population, la France qui montre des résultats encourageants mais demeure en retard par rapports aux « grands », et l'Europe qui se distingue par l'objectif qu'elle affiche pour sa politique en matière d'IA : la mettre au service des citoyens pour en maximiser les bénéfices.

Différentes logiques s'opposent selon les pays en ce qui concerne les systèmes. En Chine, les systèmes sont parfaitement connectés mais s'investissent également énormément dans la traçabilité des individus, tandis qu'aux USA ils sont beaucoup plus libéraux et exécutent uniquement les tâches qu'on leur donne jusqu'à ce qu'on les arrête. Enfin, en Europe, les systèmes sont plutôt basés sur le discernement, plus complexes, avec un objectif d'être plus en phase avec l'individu que la machine centrale. L'approche française en matière d'IA demeure toutefois plus traditionaliste et dans la rigueur par rapport à ce qui est fait à l'étranger (qui a tendance à être plus pragmatique). Par rapport aux autres acteurs, par ailleurs, la France se trouve confrontée à un problème de « saupoudrage » des financements qui, parce qu'ils sont très dispersés, s'avèrent finalement peu efficaces. Pour devenir plus performant et concurrencer la Chine et les Etats-Unis, tous deux en tête du classement, il faudra développer en France et en Europe une véritable culture de l'IA, pour enfin industrialiser la production de l'innovation et en finir avec l'artisanat (en incluant les PME au processus).

IA en régions

Le rôle des régions en matière d'IA est essentiellement d'aider à faire croître cet écosystème, de permettre aux acteurs de se rencontrer et de donner aux entreprises implantées sur leur territoire les ressources pour accéder aux atouts de la recherche en IA. Elles sont l'échelle idéale pour mener à bien des politiques publiques agiles de développement de pôles d'excellence et d'écosystèmes innovants en IA, tout comme le vecteur privilégié pour faciliter l'appropriation des technologies d'IA par les PME et ETI. Leur objectif dans cette entreprise est double : développer l'écosystème d'innovation et d'IA en soutenant les projets à fort potentiel et en attirant les cerveaux du secteur sur leur territoire, et rendre accessible l'excellence scientifique et technologique aux entreprises implantées chez elles, afin qu'elles se modernisent et continuent à se développer.

L'Ile-de-France a un rôle central à jouer dans la phase d'industrialisation de l'IA dans la mesure où, en tant que première région d'Europe en termes de recherche et de développement, il s'agit de l'endroit où se jouera l'avenir de l'IA en Europe. Un plan (plan IA 2021) a d'ailleurs été mis en place pour permettre au tissu francilien d'entreprises d'accéder aux technologies d'IA et à ses innovations, se moderniser, prospérer et créer croissance et emplois. L'IDF a également conscience que favoriser l'accès à l'IA passe par un accès à des volumes de données suffisants et a par conséquent développé une plateforme de mutualisation de données industrielles en partenariat avec de grosses entreprises telles qu'ATOS ou la SNCF. Des efforts sont aussi menés en ce qui concerne la formation en IA, avec la création de nouveaux programmes de bourses doctorales et le

projet de lancement d'un lycée IA, pour former 200 profils par an. Mais l'accent est aussi mis sur la dimension internationale grâce à une accélération des coopérations et le lancement de challenges avec des partenaires issus du privé et de la recherche.

La région Occitanie se distingue également grâce à la stratégie numérique régionale qu'elle est en train de développer. L'une de ses ambitions maitresses est d'attirer les femmes vers les carrières du numérique, pour ne plus trouver la présence féminine uniquement dans les représentations mais également parmi les acteurs. La stratégie mise en place en Occitanie repose par ailleurs sur deux piliers, à savoir le soutien de l'Institut 3IA de Toulouse et l'accompagnement des entreprises à travers l'agence de développement économique AdOcc.

IA et santé

En santé, si la question récurrente consiste à se demander si l'IA remplacera, à terme, le médecin, il semblerait que la réponse demeure négative : la médecine repose sur l'expérience humaine, que par définition les robots n'ont pas, et ces derniers ne pourront jamais remplacer la compassion, la capacité à rassurer ou à négocier l'observance d'un traitement que peut offrir un médecin.

Dans le domaine de la santé, une question d'éthique fondamentale est directrice : l'introduction de l'IA ne doit pas créer d'inégalités supplémentaires. Plusieurs éléments doivent par ailleurs être conjugués, à savoir la confiance en ce que font les algorithmes et la gestion des datas, l'acceptabilité et l'interopérabilité. Le bouleversement majeur de l'IA dans la médecine a été, en juin 2018, l'autorisation donnée par la *Food and Drug Administration* américaine à un robot pour faire le diagnostic de la rétinopathie diabétique, quand seulement un certain nombre de professionnels de santé a l'autorisation de faire du diagnostic en France. Ce bouleversement a eu lieu dans un contexte particulier qui est celui de la représentation croissante du nombre de patients alors que persiste une tendance à la réduction considérable des moyens.

Au niveau national, une innovation majeure dans le champ de la santé repose sur la création du Health Data Hub, initiative lancée à la suite du rapport Villani, qui conseillait de faire de la santé un des quatre piliers forts de la stratégie de la France en IA. Ce hub prend la forme d'une plateforme gérée par l'Etat français, accueillant et partageant les données des citoyens avec des acteurs privés ou publics. L'objectif est de permettre une exploitation massive et éthique de ces données pour des chercheurs, des startups et grands groupes. Le HDH apparait donc comme un guichet unique où chercher en cas de besoin de jeu de données à exploiter. Avoir une plateforme nationale de partage des données est alors un élément de garantie éthique pour permettre un déploiement facile et rapide de l'innovation en santé. Par ailleurs, l'opportunité du HDH est colossale, notamment grâce à plusieurs aspects clés que sont la confiance, l'organisation voire la régulation de l'écosystème, l'accessibilité des données, et la vocation à faire émerger un modèle économique propre. Il y a un intérêt commun à mutualiser les données au niveau local, régional et national, d'autant plus que la France possède les meilleures bases de données de santé au monde grâce à la Caisse nationale d'assurance maladie.

IA et transport

La mobilité-transport est le deuxième grand sujet évoqué dans le rapport Villani, et représente quelques centaines de milliards d'euros en terme d'impact économique sur l'Europe. Dans le transport, l'IA apporte une dynamique vertueuse car elle permet d'optimiser les déplacements (suppression des intermédiaires, moins de perte de temps, diminution du nombre d'accidents, réduction des émissions de CO₂), et de nombreuses technologies sont déjà opérationnelles dans le domaine. L'accélération des progrès de l'IA dans le champ des transports a été très importante, dépassant toutes les attentes. L'accent a été mis sur la promotion de

l'autonomie, tout en raisonnant selon un principe d'économie d'énergie pour les process de *machine learning*. D'ailleurs, la France est en avance sur ces logiques de consommation d'énergies finies et contraintes. Les opportunités de l'IA dans ce domaine se déclinent à trois niveaux : un niveau en lien avec les enjeux industriels (optimisation des capacités prédictives et aide à la décision pour la gestion des opérations en temps réel), un autre concernant l'interaction humain-machine et sur l'assistance aux agents et voyageurs. Enfin, le dernier niveau est en lien avec la cybersécurité (système ferroviaire caractérisé par le *security by design* avec des preuves de sécurité exigeantes). Une fiabilité de l'IA se met en place par rapport à l'intelligence humaine en ce qui concerne les transports et la mobilité en général, d'autant plus que l'IA apporte à l'utilisateur l'opportunité de paramétrer lui-même ses choix éthiques en la matière.

Formation et recherche

La formation en IA en France reste encore balbutiante, et ceci bien que les besoins des entreprises en la matière soient conséquents. On observe effectivement un manque cruel d'étudiants (manque de ressource) et une absence regrettable des femmes dans ce domaine (manque de diversité). S'il faut saluer les évolutions considérables de la formation en IA française ces 20 dernières années, le problème majeur reste qu'elle ne se cantonne généralement qu'à un seul profil : le data scientist. Afin d'apporter une réponse au décalage entre la formation et les exigences professionnelles, développer les parcours interdisciplinaires, les nouveaux modes de formation et la réforme des grilles de postes semblent être des réponses adéquates et non négligeables. En effet, la formation demeure un enjeu majeur pour le futur de l'IA, afin d'être en mesure de répondre aux besoins des mondes professionnel et académique.

Enfin, en matière de recherche, pour éviter une déconnexion entre les enjeux sociétaux et industriels, ou entre les besoins et les solutions proposées, des chaires d'industrie et de recherche ont été lancées pour favoriser les coopérations entre industriels et chercheurs. Pour pallier le manque de chercheurs en IA, la loi Pacte a introduit la possibilité pour les chercheurs d'aller en industrie. Toutefois, si cette innovation a révélé bon nombre d'avantages (hybrider les connaissances des mondes académiques et universitaires, permettre une connaissance plus exacte du monde industriel par les professeurs pour mieux ancrer leurs cours dans la réalité, avoir davantage de données et de matériel), elle a aussi mis en lumière quelques effets pervers. Dès lors qu'un nombre important de professeurs en bénéficie, il y a moins d'effectifs pour enseigner aux étudiants. Par ailleurs, si la manœuvre ne permet que de décaler des forces de l'université vers l'industrie, on ne gagne pas de nouveaux chercheurs : elle perd alors tout son sens.

IA et cybersécurité

La nouveauté portée par l'IA en matière de cyber est qu'elle doit permettre de détecter et d'analyser les signaux faibles et d'augmenter la capacité des outils utilisés par les ingénieurs pour leur donner des aides à la décision. L'objectif est alors d'accroître les capacités de diagnostic et de remédiation. Mais s'il est possible de faire de la cybersécurité avec de l'IA, encore faut-il convaincre les acteurs concernés d'y avoir recours dans ce domaine où les conséquences d'une erreur sont lourdes et coûteuses.

Aspects juridiques

Lancé par les institutions européennes, le RGPD est une avancée louable en matière de « droit de l'IA » compte tenu de l'enjeu de souveraineté et d'anticipation que représente la possession de la data. Si l'on cédait auparavant nos données aux GAFAMS et BATX, le RGPD permet de se réapproprier ces informations. Mais même si les démarches légales ont été lancées pour encadrer les utilisations de l'IA et que les ébauches d'un cadre légal ont été posées, la jurisprudence en la matière manque toujours cruellement. D'autant plus que des techniques toujours plus poussées se développent pour toucher les systèmes, notamment la modification volontaire de matériel de manière imperceptible pour que les systèmes se trompent dans leur classification. La majorité des plaintes relatives à l'IA aujourd'hui concernent le consentement qui, la plupart du temps, reste trop théorique. Des sanctions ont donc été mises en places pour tenter de mieux l'encadrer. En France, une importante réglementation entoure déjà l'utilisation et la protection des données, ce qui confère aux français un avantage concurrentiel pour récupérer leurs données face aux géants chinois et américains. La réglementation trouve en effet son origine dans la loi informatique et libertés de 1978, renforcée par la modification du cadre légal entraînée par le RGDP.

Enfin, si l'on se penche sur l'utilisation de l'IA dans le milieu juridique, la justice prédictive n'est pas encore arrivée en France, mais l'IA est tout de même utilisée dans ce domaine dont elle est susceptible de faire évoluer les pratiques et la culture, notamment via la justice prévisionnelle ou jurimétrie. L'IA pourra ainsi être utilisée pour rendre les recherches en jurisprudence des Cours européennes et internationales moins chronophages et artisanales, grâce aux algorithmes et à l'analyse de données (loi Lemaire du 7 octobre 2016). Cependant, il est primordial que les magistrats conservent leur liberté d'appréciation et leur indépendance dans l'exercice de leurs fonctions, l'utilisation des algorithmes étant basée sur la neutralité et la transparence.

IA, confiance éthique

S'il faut être conscients du fait qu'une perte de confiance en les systèmes d'IA et les algorithmes représentent un frein au développement et à l'innovation, à quel point pouvons-nous avoir confiance en des systèmes issus de l'apprentissage automatique, apprentissage pouvant faire les frais de biais du fait de leur programmation originelle par des humains influencés par leurs émotions ?

Dans ce contexte marqué par une certaine crainte de l'IA, l'évangélisation est nécessaire. Les études d'impact semblent alors être le meilleur moyen de rassurer les diverses parties prenantes et d'apporter des réponses à leurs préoccupations, qui se concentrent autour de la peur de voir les systèmes d'IA remplacer les humains dans certaines branches, de la responsabilité, de la sécurité, de la performance et de l'utilité de l'IA. Une action de démystification doit être entreprise pour rassurer les consommateurs affolés par les potentielles dérives de l'IA et la perspective qu'elle dominera, à terme, le monde. En ce sens, l'absence d'une culture statistique et le caractère peu accessible des mathématiques constituent des freins à l'explicabilité, qui est pourtant primordiale car sans cela, les systèmes d'IA dont la compréhension reste obscure pour les chefs militaires par exemple, ne seront pas utilisés. Pour adresser ce besoin d'explicabilité et favoriser la confiance en l'IA, plusieurs solutions sont actuellement proposées telles qu'apporter de la transparence au niveau global des algorithmes, ou privilégier la pédagogie pour mettre fin aux clichés. Mais *in fine*, les systèmes d'IA ne pourront être acceptés que s'ils répondent à trois critères : explicabilité, validité et responsabilité. Dans cette perspective d'explicabilité, l'intérêt de l'IA est de venir enrichir la compréhension de décisions et de raisonnements illisibles par un cerveau humain. Mais selon le projet mené, le rapport à l'explicabilité et à la confiance peut varier (on privilégiera parfois la performance et la précision à l'explicabilité).

En matière de confiance, l'IA se trouve confrontée à trois vulnérabilités majeures. Dans un premier temps, elle est critiquée pour sa similitude à une « boîte noire », dans la mesure où les technologies de *deep learning* ne

présentent pas nécessairement de processus de prise de décision explicables. Ensuite se pose la question de l'apprenabilité : les machines apprennent réellement ce que l'Homme veut qu'elles apprennent. Il peut donc y avoir des incohérences entre les attentes et les résultats, ou même une internalisation par les machines de biais introduits par l'humain lors de la programmation originelle. Enfin, nous parlons de systèmes pouvant être victimes de leurres : il faut être en mesure d'anticiper et de réagir au possible maquillage de malveillances.

Au vu de l'impact de l'IA, du numérique et des nouvelles technologies sur les sociétés et leur transformation, le développement d'une IA basée sur des principes éthiques a pris de l'ampleur ces dernières années. Parmi les principes définis peu à peu comme centraux, celui qui semble s'imposer et guider le développement de l'IA est la transparence couplée à l'explicabilité. De plus, à l'heure où la collecte massive des données devient la norme, les pratiques sont peu à peu remise en question une fois confrontées au principe de finalité et à la nécessité des objectifs préalables (affichés notamment dans le RGPD). Cependant, ce qui importera à l'avenir n'est pas tant d'avoir des principes mais de s'interroger à propos de leur mise en œuvre et des arbitrages entre eux.